Variation in SARS-CoV-2 outbreaks across sub-Saharan Africa

Benjamin L. Rice [1,2 ✉], Akshaya Annapragada [3], Rachel E. Baker [1,4], Marjolein Bruijning[1], Winfred Dotse-Gborgbortsi [5], Keitly Mensah [6], Ian F. Miller [1], Nkengafac Villyen Motaze[7,8], Antso Raherinandrasana[9,10], Malavika Rajeev[1], Julio Rakotonirina[9,10], Tanjona Ramiadantsoa[11,12,13], Fidisoa Rasambainarivo[1,14], Weiyu Yu [15], Bryan T. Grenfell[1,16], Andrew J. Tatem [5] and C. Jessica E. Metcalf [1,16]

A surprising feature of the SARS-CoV-2 pandemic to date is the low burdens reported in sub-Saharan Africa (SSA) countries relative to other global regions. Potential explanations (for example, warmer environments[1], younger populations[2-4]) have yet to be framed within a comprehensive analysis. We synthesized factors hypothesized to drive the pace and burden of this pandemic in SSA during the period from 25 February to 20 December 2020, encompassing demographic, comorbidity, climatic, healthcare capacity, intervention efforts and human mobility dimensions. Large diversity in the probable drivers indicates a need for caution in interpreting analyses that aggregate data across low- and middle-income settings. Our simulation shows that climatic variation between SSA population centers has little effect on early outbreak trajectories; however, heterogeneity in connectivity, although rarely considered, is likely an important contributor to variance in the pace of viral spread across SSA. Our synthesis points to the potential benefits of context-specific adaptation of surveillance systems during the ongoing pandemic. In particular, characterizing patterns of severity over age will be a priority in settings with high comorbidity burdens and poor access to care. Understanding the spatial extent of outbreaks warrants emphasis in settings where low connectivity could drive prolonged, asynchronous outbreaks resulting in extended stress to health systems.

The trajectory of the SARS-CoV-2 pandemic in SSA is uncertain. To date, reported case counts and mortality in SSA have lagged behind other geographical regions. All SSA countries, with the exception of South Africa and Ethiopia, reported fewer than 100,000 total cases and fewer than 1,800 deaths as of December 2020 (Supplementary Table 1)—totals far lower than observed in Asia, Europe and the Americas (Africa Centres for Disease Control and Prevention (Africa CDC) COVID-19 daily updates https://africacdc.org/covid-19/, Johns Hopkins Coronavirus Resource Center https://coronavirus.jhu.edu/data/mortality). However, variation in reporting between countries and some seroprevalence surveys that suggested high rates of local infection[5-7] make it unclear if the relatively few reported cases and deaths to date indicate a generally reduced epidemic potential in SSA[8].

Comparisons across SSA populations based on reported infection rates are obscured by heterogeneity in surveillance capacity (for example, variation in testing rates among countries) and correlation between surveillance and infection reporting[9] (Extended Data Fig. 1). Combining reported death counts with assumptions about the probability of mortality given infection[2] yields generally low estimates of the percentage of the population expected to have been infected (that is, <10%) but this varies more than tenfold between SSA countries and, critically, is sensitive to assumptions about the death reporting rate (Fig. 1a) and infection fatality ratio (IFR; Fig. 1b). Serology provides an alternative and more direct measure of the percentage infected. Initial serological studies of blood banks in Kenya (5–10%)[5], health care workers in urban Malawi (9–16%)[7] or from Niger State in Nigeria (20–30%)[6] indicate that infection rates could be higher in some settings, but only the latter was designed as a representative sample and serology-based estimates are sparse in SSA.

Given limitations in inference from direct measures of infection and death rates, experience from locations where the pandemic has progressed more rapidly provides a valuable basis of knowledge to assess the relative risk of populations in SSA and identify those at the greatest risk. For example, individuals in lower socioeconomic settings have been disproportionately affected in high latitude countries[10,11], indicating poverty as a determinant of risk of increased severity of disease. Widespread disruptions to routine health services have been reported[12-14] and are likely to contribute to the burden of the pandemic in SSA[15]. The role of other factors, from demography[2-4] to health system context[16] and intervention timing[17,18], is also increasingly well-characterized. A summary of the main findings and limitations of the study is shown in Table 1.

[1]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA. [2]Madagascar Health and Environmental Research, Maroantsetra, Madagascar. [3]Johns Hopkins University School of Medicine, Baltimore, MD, USA. [4]Princeton Environmental Institute, Princeton University, Princeton, NJ, USA. [5]WorldPop, School of Geography and Environmental Science, University of Southampton, Southampton, UK. [6]Centre Population et Développement (CEPED), Institut de Recherche pour le Développement (IRD) and Université de Paris, Inserm ERL 1244, Paris, France. [7]Centre for Vaccines and Immunology, National Institute for Commnunicable Diseases, National Health Laboratory Service, Johannesburg, South Africa. [8]Department of Global Health, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa. [9]Faculty of Medicine, University of Antananarivo, Antananarivo, Madagascar. [10]Teaching Hospital of Care and Public Health Analakely, Antananarivo, Madagascar. [11]Department of Life Science, University of Fianarantsoa, Fianarantsoa, Madagascar. [12]Department of Mathematics, University of Fianarantsoa, Fianarantsoa, Madagascar. [13]Department of Integrative Biology, University of Wisconsin-Madison, Madison, WI, USA. [14]Mahaliana Labs SARL, Antananarivo, Madagascar. [15]School of Geography and Environmental Science, University of Southampton, Southampton, UK. [16]Princeton School of Public and International Affairs, Princeton University, Princeton, NJ, USA. ✉e-mail: b.rice@princeton.edu
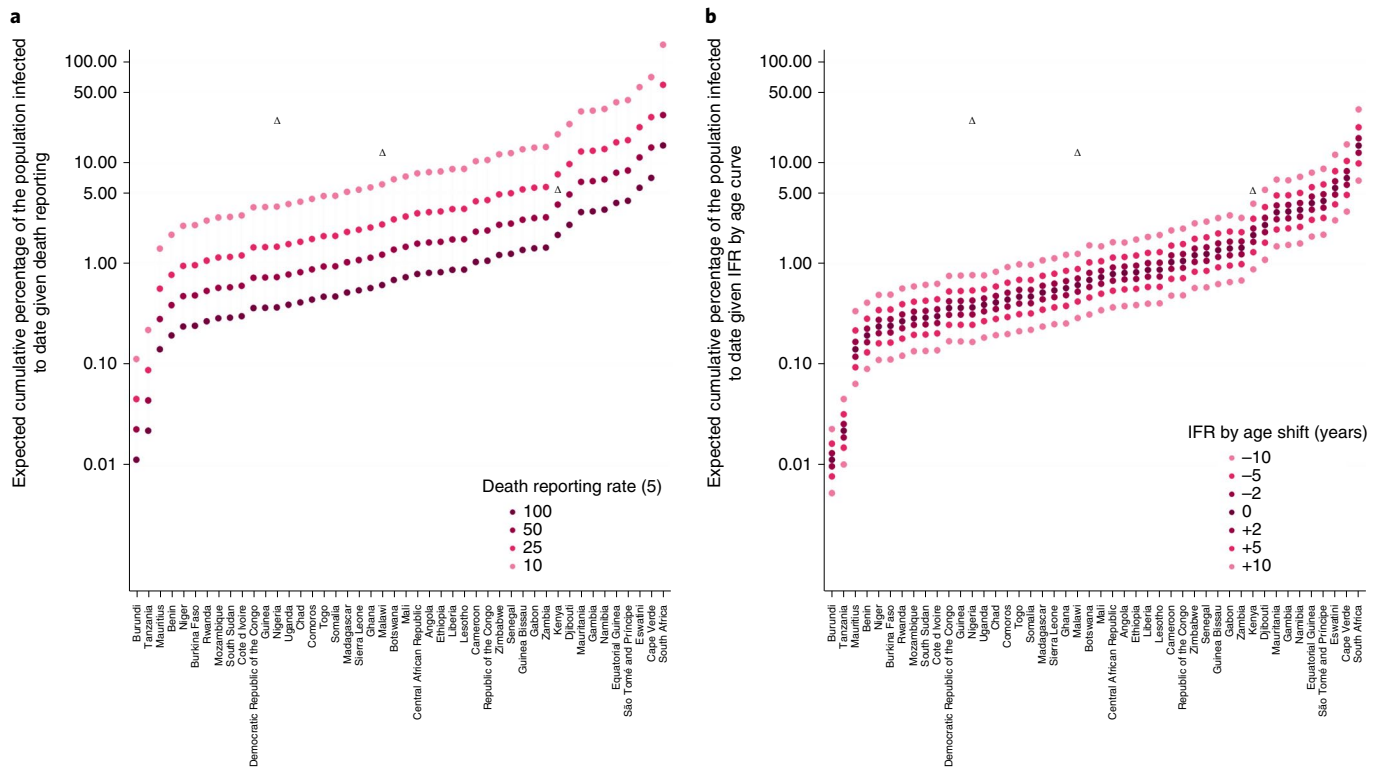
Fig. 1 | Variation in the cumulative percentage of the population infected in SSA countries as expected from reported mortality totals. a,b, The expected percentage of a country's population infected given the number of reported deaths to date, country-specific age structure and a range of death reporting completeness scenarios (**a**), or a range of IFR scenarios (**b**). The global IFR age curves were fitted to existing age-stratified IFR estimates (Methods and Supplementary Table 4) and shifted toward younger or older ages by the specified number of years to simulate higher or lower IFRs, respectively (**b**). Conservatively, we assumed no variation in infection rates by age. (Infections skewed toward older age groups would result in a higher average IFR and thus a lower expected percentage of the population infected for a given number of deaths.) Reported case and death counts are current as of December 2020 (sourced from the Africa CDC; Supplementary Table 1). Data from Eritrea and the Seychelles are not shown as they have zero reported deaths as of December 2020. Comparisons to serological surveys (unfilled triangles) available from blood banks in Kenya[5], health care workers in urban Malawi[7] and a subnational cluster-stratified random sample from Niger State in Nigeria[6] are shown.

Characterizing and anticipating the trajectory of ongoing outbreaks in SSA requires considering variability in known drivers and how they might interact to increase or decrease risk across populations in SSA and relative to non-SSA settings (Fig. 2). For example, while most countries in SSA have a relatively young population age structure, suggesting a decreased burden (since SARS-CoV-2 morbidity and mortality increase with age[2–4]), prevalent infectious and noncommunicable comorbidities could counterbalance this apparent demographic advantage[16,19–21]. Similarly, SSA countries have health systems that vary greatly in their infrastructure and dense, resource-limited urban populations could have fewer options for social distancing[22]. Yet, decentralized, community-based health systems that benefit from past experience with epidemic response (for example, to Ebola[23,24]) can be mobilized. Climate is frequently invoked as a potential mitigating factor for warmer and wetter settings[1], including SSA, but climate varies greatly between population centers in SSA and the reality of the existence of large susceptible populations could counteract any climate forcing during initial phases of the epidemic[25]. Connectivity, at international and subnational scales, also varies greatly[26,27] and the time interval between viral introductions and the onset of interventions, such as lockdowns, will modulate the trajectory[9]. Finally, burdens of malnutrition, infectious diseases and many other underlying health conditions are higher in SSA than in other regions (Supplementary Table 2) and their interactions with SARS-CoV-2 are, as of yet, poorly understood; conversely cross-protection from either SARS-CoV-2 infection or disease as a result of previous infection by widespread circulating coronaviruses is a possibility.

The highly variable social and health contexts of countries in SSA will drive location-specific variation in the magnitude of the burden, the time course of the outbreak and options for mitigation. In this study, we synthesized the range of factors hypothesized to modulate the potential outcomes of SARS-CoV-2 outbreaks in SSA settings by leveraging existing data sources and integrating new SARS-CoV-2-relevant mobility and climate transmission models. Data on direct measures and indirect indicators of risk factors were sourced from publicly available databases including from the World Health Organization (WHO), World Bank, United Nations Population Division, Demographic and Health Surveys (DHS-USAID), Global Burden of Diseases and WorldPop, and newly generated datasets (Extended Data Fig. 2 and Supplementary Table 3). We organized our assessment around two aspects that will shape national outcomes and response priorities in the event of widespread outbreaks: (1) the burden, or expected severity of the outcome of an infection, which emerges from age, comorbidities and health systems functioning; and (2) the rate of spread within a geographical area or pace of the pandemic.

We grouped factors that might drive the relative rates of these two features (mortality burden and pace of the outbreak) along six dimensions of risk: (1) demographic and socioeconomic parameters related to transmission and burden; (2) comorbidities relevant to burden; (3) climatic variables that may impact the magnitude and seasonality of transmission; (4) prevention measures deployed

Table 1 | Policy summary

| Category	Description
Background	As the SARS-CoV-2 pandemic expanded globally, reported incidence and mortality remained low in SSA. Yet, a general conclusion that SSA may avoid the high burdens seen elsewhere neglects considerable national and subnational variability in likely drivers of the pandemic's impacts, from its burden to its pace, and does not address variable surveillance and registration systems.
Main findings and limitations	Synthesizing data on the likely drivers of the pace and burden of SARS-CoV-2 in SSA reveals extensive variability in factors that can define the burden once individuals have become infected. Pairing this with simulations of the trajectory of the outbreak indicates little effect of climate but potentially prolonged outbreaks in many settings due to heterogeneities in connectivity, an effect that could be amplified by control efforts. However, although we provide a qualitative overview of the continued potential impact of the pandemic in SSA, quantitative projections remain intractable given a lack of information on the quantitative impact of important risk factors (from how comorbidities might shape the IFR to how remoteness will reduce spread). Additionally, uncertainties associated with existing surveillance and mortality registration data impede direct comparison of expectations with national data.
Policy implications	Strengthening surveillance and registration is necessary to narrow the range of expectations for country trajectories in SSA for incidence and mortality. Additional tools for surveillance, such as serology, or approaches that quantify excess mortality, will provide important complementary measures. Our national and subnational analyses point to where returns on investment in strengthening surveillance could yield the greatest returns. Countries with high comorbidity risk may have most to gain from understanding determinants of mortality; low connectivity countries will benefit from investments in delineating the spatial extent of outbreaks; all countries will benefit from evaluating the intersection between epidemic pace and health system disruptions.
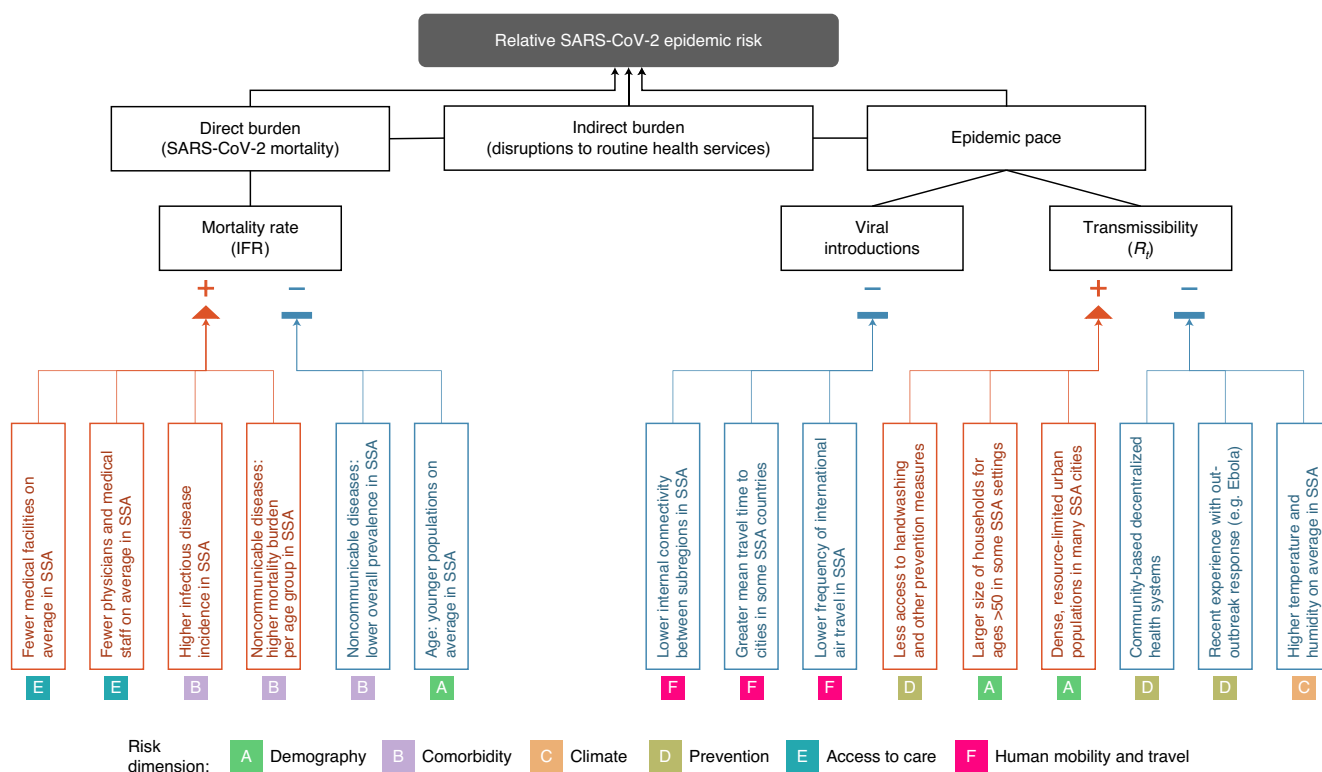
Fig. 2 | Hypothesized modulators of relative SARS-CoV-2 epidemic risk in SSA. The factors (A–F) hypothesized to increase (red) or decrease (blue) mortality burden or epidemic pace within SSA, relative to global averages, are grouped into six categories or dimensions of risk. In this framework, epidemic pace is determined by person-to-person transmissibility, which can be defined as the time-varying effective reproductive number, R_t, and introduction and geographical spread of the virus via human mobility. SARS-CoV-2 mortality (determined by the IFR) is modulated by demography, comorbidities (for example, noncommunicable diseases) and access to care. Overall burden is a function of direct burden and indirect effects due to, for example, socioeconomic disruptions and disruptions in health services, such as vaccination and infectious disease control. Supplementary Table 2 contains the details and references used as a basis to draw the hypothesized modulating pathways.

to reduce transmission; (5) accessibility and coverage of existing healthcare systems to reduce burden; and (6) patterns of human mobility relevant to transmission (Supplementary Table 2).

National scale variability in SSA among these dimensions of risk often exceeds ranges observed across the globe (Fig. 3a–d and Extended Data Fig. 3). For example, estimates of access to

basic handwashing (that is, clean water and soap[28]) among urban households in Mali, Madagascar, Tanzania and Namibia (62–70%) exceed the global average (58%) but are <10% for Liberia, Lesotho, Democratic Republic of the Congo and Republic of Guinea-Bissau (Fig. 3d). Conversely, the range in the number of physicians is low in SSA, with all countries other than Mauritius below the global
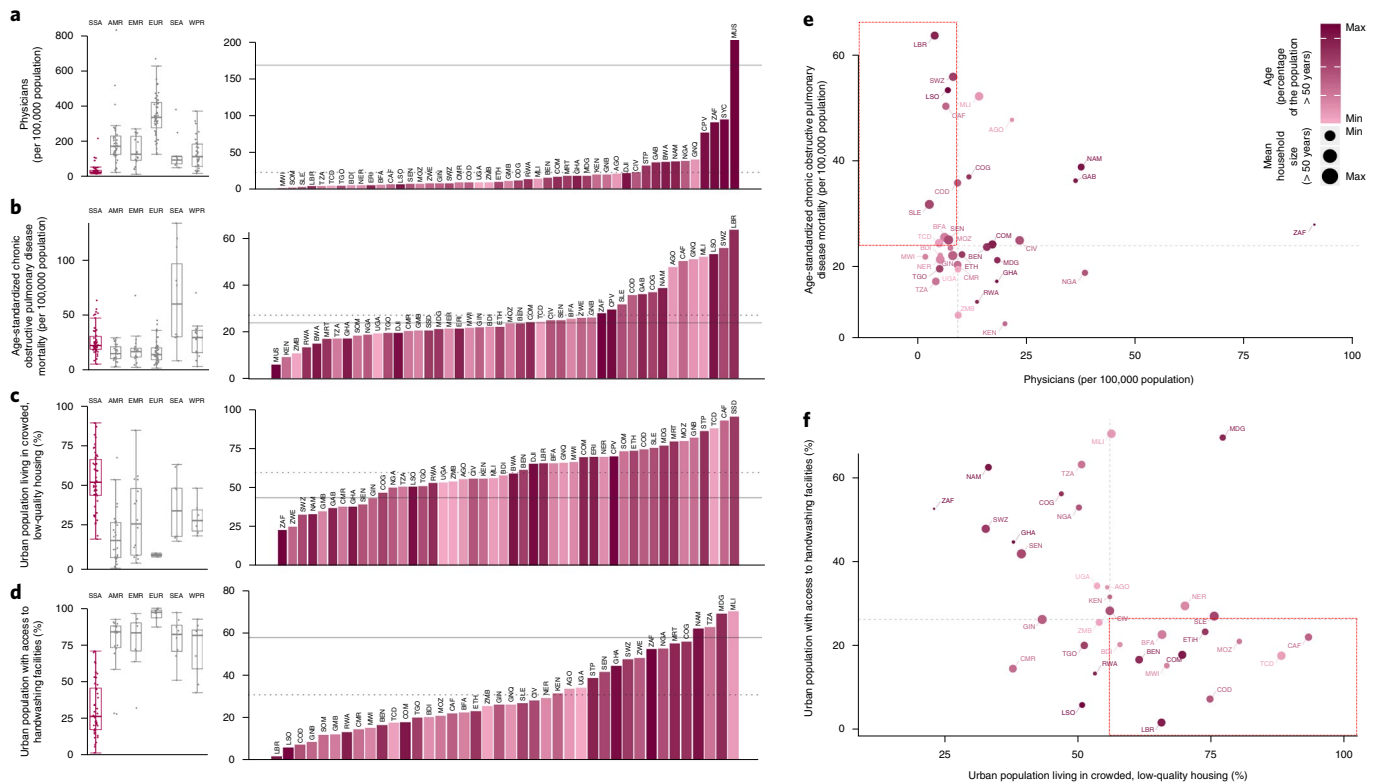
Fig. 3 | Variation among SSA countries in select determinants of SARS-CoV-2 risk. a–d, Right, SSA countries were ranked from least to greatest for each indicator; the bar color shows the population age structure (percentage of the population above the age of 50). The solid horizontal lines show the global mean value and the dotted lines show the mean among SSA countries. Left, The boxplots show the median, the inner bounds correspond to the interquartile range (IQR, 25th to 75th percentiles) and the outer bounds correspond to the 1.5 × IQR, grouped by WHO-defined geographic regions. SSA, Sub-Saharan Africa; AMR, Americas Region; EMR, Eastern Mediterranean Region; EUR, Europe Region; SEA, Southeast Asia Region; WPR, Western Pacific Region (*n* = 206, 172, 106 and 92 countries with available data for **a–d**, respectively). **e,f**, Bivariate comparisons of the variables shown in **a,b** and **c,d**, respectively. The dot size shows the mean household size for households with individuals aged over 50, the dashed lines show the median value among SSA countries and the quadrants with the greatest risk are outlined in red (for example, fewer physicians and greater age-standardized chronic obstructive pulmonary disease mortality). See Supplementary Table 3 and Extended Data Fig. 3 for a full description and link to visualization of all variables.

average (168.78 per 100,000 population) (Fig. 3a). Yet, estimates are still heterogeneous within SSA, with, for example, Gabon estimated to have more than 4 times the physicians of neighboring Cameroon (36.11 and 8.98 per 100,000 population, respectively). This disparity is likely to interact with social contact rates among the aged in determining exposure and clinical outcomes (for variation in household size, see Fig. 3e,f). Relative ranking across variables is also uneven among countries (Extended Data Fig. 4) with the result that this diversity cannot be easily reduced (for example, the first two principal components explain only 32.6 and 13.1% of the total variance as shown in Extended Data Fig. 5), indicating that approaches reliant on a small subset of variables will fail to capture the observed variation among SSA countries.

To first evaluate variation in the burden emerging from the severity of infection outcome, we considered how demography, comorbidity and access to care might modulate the age profile of SARS-CoV-2 morbidity and mortality[2–4]. Subnational variation in the distribution of high-risk age groups indicates considerable variability, with higher burden expected in urban settings in SSA (Fig. 4a), where density and thus transmission are likely higher[29].

Comorbidities and access to clinical care also vary across SSA (for example, for diabetes prevalence and hospital bed capacity, see Fig. 4b). In comparison to settings where previous SARS-CoV-2 IFR estimates have been reported, mortality due to noncommunicable diseases in SSA increases more rapidly with age (Extended Data Fig. 6) suggesting risk for an elevated IFR in some

SSA settings. Conversely, an analysis of the reported age-specific death data available from Kenya and South Africa suggested low IFRs in comparison to non-SSA countries[30]. Comparison of empirical age profiles of mortality more broadly across SSA is currently limited by the small number of total deaths reported to date for many countries (for example, 33 of 48 SSA countries have reported fewer than 200 total deaths as of December 2020) and incomplete associated age data. Consequently, we used the global IFR by age estimates and explored the potential effect on mortality of deviations from the expected baseline IFR in diverse SSA settings.

Small shifts (for example, of 2–10 years of age) in the IFR profile resulted in large effects on expected mortality for a given level of infection. For example, Chad, Burkina Faso and the Central African Republic, while among the youngest SSA countries, have a high prevalence of diabetes and low density of hospital beds. Given the age structure of these countries, a slight shift in the IFR by age profile toward higher mortality in middle-aged groups (for example, ages 50–60 years) would result in mortality increasing to a rate that would exceed a majority of the other, relatively older SSA countries at the unshifted baseline (Fig. 4c and Methods). Generally, minor shifts in the IFR lead to differences larger than the magnitude of the difference expected from differing age structures for countries in SSA.

Although there is greater access to care in older populations by some metrics (Fig. 3a; correlation between age and the number of physicians per capita, *r* = 0.896, *P* < 0.001), access to clinical care is highly variable overall (Fig. 4d) and maps poorly to indicators of
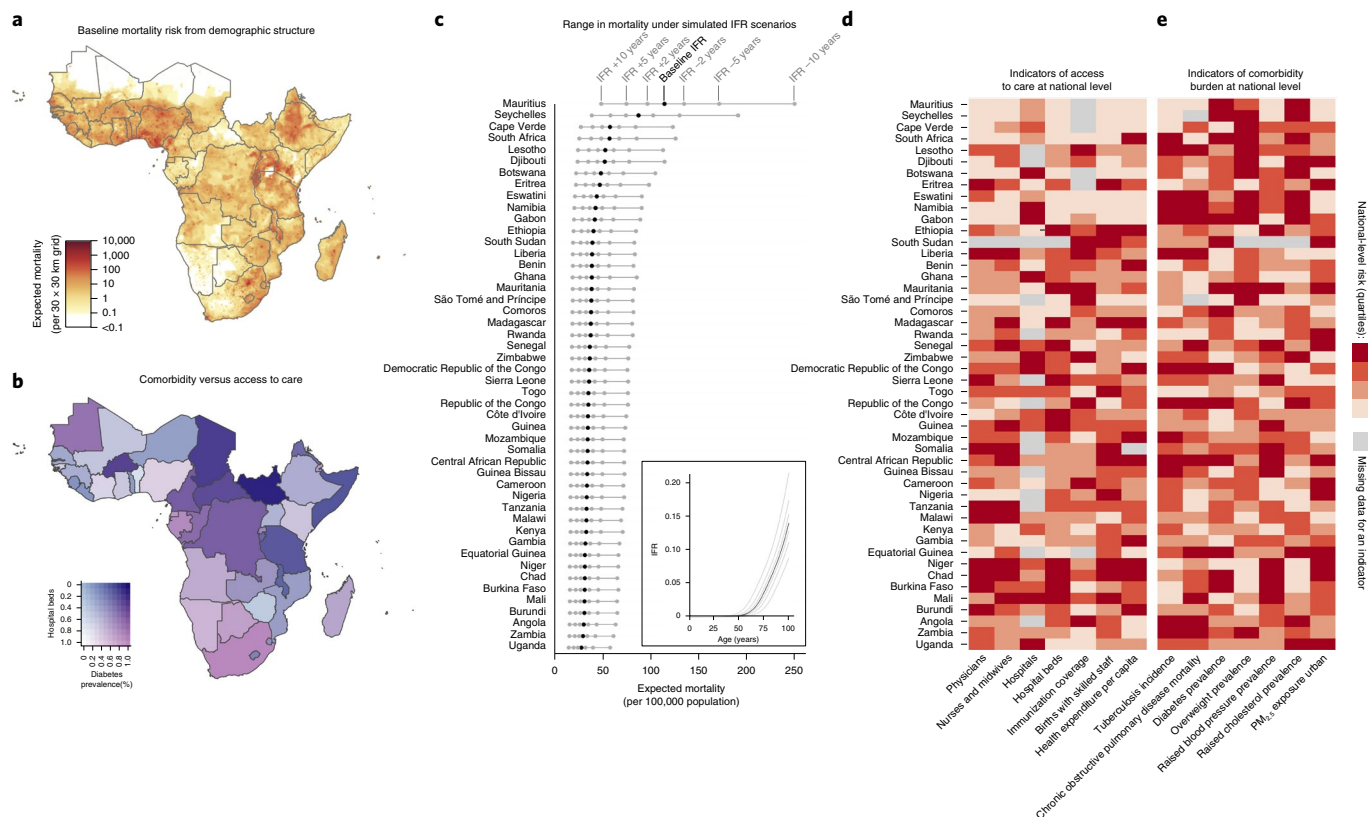
Fig. 4 | Variation in expected burden for SARS-CoV-2 outbreaks in SSA. a, Expected mortality in a scenario where cumulative infection reaches 20% across age groups and the IFR curve is fitted to existing age-stratified IFR estimates (Methods and Supplementary Table 4). **b**, National-level variation in comorbidity and access to care variables, for example, diabetes prevalence among adults and the number of hospital beds per 100,000 population for SSA countries. **c**, Range in mortality per 100,000 population expected in scenarios where the cumulative infection rate is 20% and IFR per age is the baseline (black) or shifted ±2, 5 or 10 years (gray). Inset, The IFR by age curves for each scenario are shown. **d,e**, Selected national-level indicators; estimates of reduced access to care (for example, fewer hospitals) (**d**) or increased comorbidity burden (for example, higher prevalence of raised blood pressure) (**e**) shown with darker red for higher-risk quartiles (see Extended Data Fig. 4 for all indicators). Countries missing data for an indicator are shown in gray. For comparison between countries, estimates are age-standardized where applicable (Supplementary Table 3). High-resolution maps for each variable and scenario are available at the SSA-SARS-CoV-2-tool for estimating the burden of SARS-CoV-2 in SSA (https://labmetcalf.shinyapps.io/covid19-burden-africa/).

comorbidity (Fig. 4e). Empirical data are urgently needed to assess the extent to which the IFR-age-comorbidity associations observed elsewhere are applicable to SSA settings with reduced access to advanced care. Yet both surveillance and mortality registration[31] are frequently under-resourced in SSA, complicating both evaluating and anticipating the burden of the pandemic and underscoring the urgency of strengthening existing systems[24].

The frequency of viral introduction to each country, likely governed by international air travel in SSA[32], determines both the timing of the first infections and the number of initial infection clusters that seed subsequent outbreaks. The relative importation risk among SSA cities and countries was assessed by compiling data from 108,894 flights arriving at 113 international airports in SSA from January to April 2020 (Fig. 5a), stratified by the SARS-CoV-2 status at the departure location on the day of travel (Fig. 5b). A small subset of SSA countries received a disproportionately large percentage (for example, South Africa, Ethiopia, Kenya and Nigeria together contributed 47.9%) of the total travel from countries with confirmed SARS-CoV-2 infections, which likely contributed to variation in the pace of the pandemic across settings and is consistent with those 4 countries together contributing 74.3% of all reported cases in SSA as of 20 December 2020 (refs. [32,33]).

Once local chains of infection are established, the rate of spread within countries will be shaped by efforts to reduce spread, such

as handwashing and other non-pharmaceutical interventions (Fig. 3d), population contact patterns including mobility and urban crowding[29] (Fig. 3c) and potentially the effect of climatic variation[1]. Where countries fall across this spectrum of pace will shape interactions with lockdowns and determine the length and severity of disruptions to routine health system functioning.

Subnational connectivity varies greatly across SSA, both between subregions of a country and between cities and their rural periphery (for example, as indicated by travel time to the nearest city with a population over 50,000; Fig. 5c). As expected, in stochastic simulations using estimates of viral transmission parameters and mobility (Extended Data Fig. 7), a smaller cumulative proportion of the population is infected at a given time in countries with larger populations in less connected subregions (Fig. 5d and Extended Data Fig. 8); including non-pharmaceutical interventions reduces this proportion still further. At the national level, susceptibility declines more slowly and more unevenly in such settings (for example, Ethiopia, South Sudan, Tanzania) due to a lower probability of introductions and reintroductions of the virus locally, an effect amplified by lockdowns (Extended Data Fig. 9). It is unclear whether the more prolonged, asynchronous epidemics expected in these countries or the overlapping, concurrent epidemics expected in countries with higher connectivity (for example, Malawi, Kenya, Burundi) will be a greater stress to health systems. Outbreak control efforts are
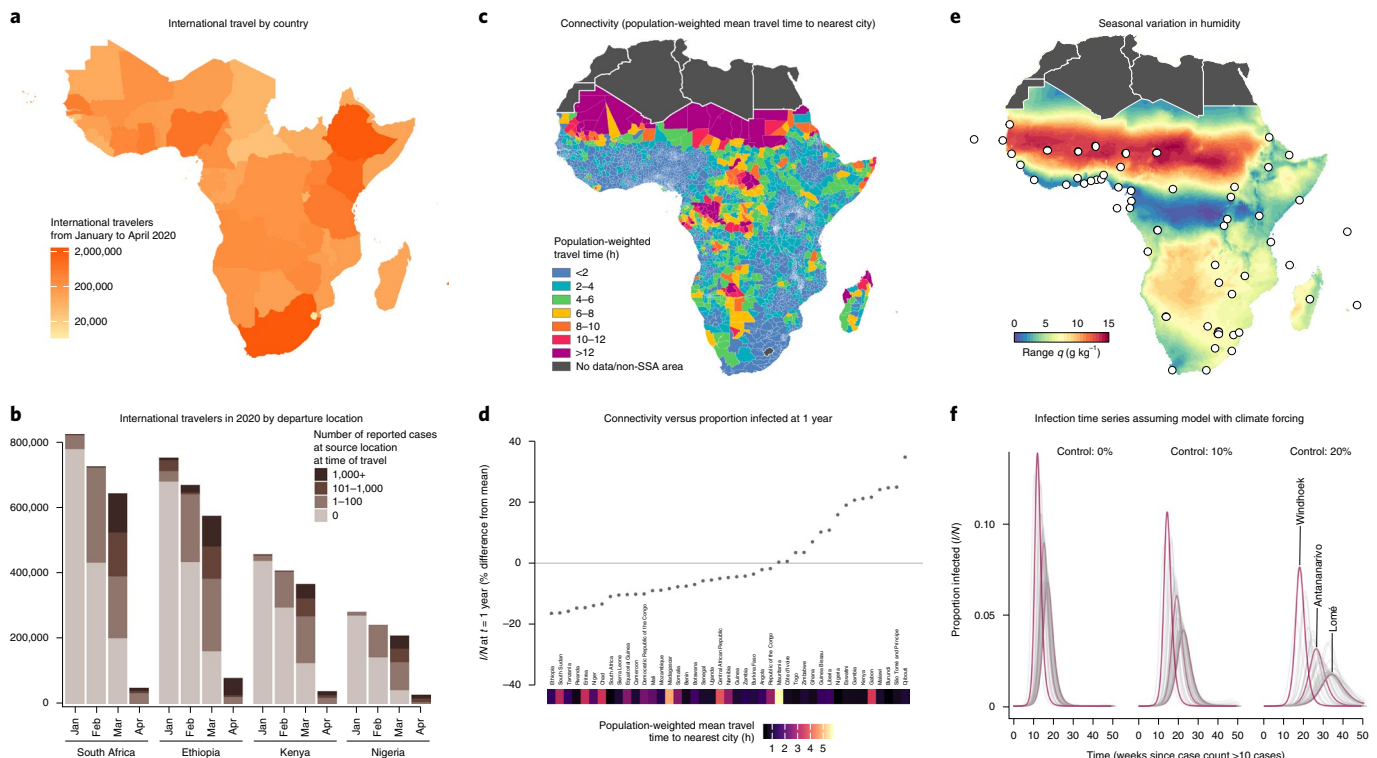
Fig. 5 | Variation in connectivity and climate in SSA and expected effects on SARS-CoV-2. a, International travelers to SSA from January to April 2020, as inferred from the number of passenger seats on arriving aircrafts. **b**, For the 4 countries with the most arrivals, the proportion of arrivals by month coming from countries with 0, 1–100, 101–1,000 and 1000+ reported SARS-CoV-2 infections at the time of travel (see Supplementary Table 5 for all others) is shown. **c**, Connectivity within SSA countries as inferred from average population-weighted mean travel time to the nearest urban area with a population >50,000. **d**, Mean travel time at the national level and variation in the fraction of the population expected to be infected (I/N) in the first year from stochastic simulations (Methods). **e**, Climate variation across SSA as shown by seasonal range in specific humidity, q (g kg^{-1}) (max average q − min average q). **f**, The effect of local seasonality and control efforts (R_0 decreases by 0%, that is, unmitigated, 10 or 20%) on the timing of epidemic peaks (max I/N) in SSA cities (with three exemplar cities highlighted in pink; Methods).

likely to be further complicated during prolonged epidemics if they intersect with seasonal events such as temporal patterns in human mobility[29] or other infections (for example, malaria).

Despite extreme variation among cities in SSA (Fig. 5e), large epidemic peaks are expected in all cities (Fig. 5f), even from our models incorporating interventions and transmission rates that decline in response to warmer, more humid local climates (climate-dependent variation in transmission rate for coronaviruses inferred from endemic circulation in the United States, but robust to parameter value choice; Methods). After accounting for differences in the date of introductions, simulated climate forcing generates a maximum of only 6–7 weeks' variation in the time to epidemic peaks, with peaks generally expected earlier in more southerly, colder, drier cities (for example, Windhoek and Maseru) and later in more humid, coastal cities (for example, Bissau, Lomé and Lagos). Reductions in transmission due to control efforts, as expected, prolong the time to epidemic peak (Fig. 5f and Extended Data Fig. 10). Apart from these slight shifts in timing, the large proportion of the population that is susceptible overwhelm the effects of climate[25]; earlier suggestions that Africa's generally more tropical environment alone may provide a protective effect[1] are not supported by evidence.

Our synthesis emphasizes striking country-to-country variation in the drivers of the pandemic in SSA (Fig. 3), indicating that continued variation in the burden (Fig. 4) and pace (Fig. 5) is to be expected even across low-income settings. Since small perturbations in the age profile of mortality could drastically change the national-level burden in SSA (Fig. 4), building expectations for the risk for each country requires monitoring for deviations in the

pattern of morbidity and mortality over age. Transparent and timely communication of these context-specific risk patterns could aid community engagement in efforts to reduce transmission, help motivate population behavioral changes and guide existing networks of community case management.

Because the largest impacts of SARS-CoV-2 outbreaks may be through indirect effects on routine health provisioning, understanding how existing programs may be disrupted differently by acute versus longer outbreaks is crucial to planning resource allocation. For example, population immunity will decline proportionally with the length of disruptions to routine vaccination programs[34], resulting in more severe consequences in areas with prolonged epidemic time courses.

Others have suggested that this crisis presents an opportunity to unify and mobilize across existing health programs (for example, for human immunodeficiency virus (HIV), tuberculosis, malaria and noncommunicable diseases)[24]. Although this might be a powerful strategy in the context of acute, temporally confined crises, long-term distraction and diversion of resources[35] could be harmful in settings with extended, asynchronous epidemics. A higher risk of infection among healthcare workers during epidemics[36,37] could also amplify this risk.

As evidenced by failures in locations where the epidemic progressed rapidly (for example, United States), effective governance and management before reaching large case counts will likely yield the largest rewards. Generalizing across SSA is difficult because the time course and estimates of the effect of intervention policies have varied greatly (Extended Data Fig. 9); however, Mauritius and

Rwanda, for example, have reported extremely low incidence thanks in part to a well-managed early response.

The burden and time course of SARS-CoV-2 is expected to be highly variable across SSA. Simulations show that variation in international and subnational connectivity are expected to be important determinants of pace, but variability in reporting regimes makes it difficult to compare observations to date with expectations (Extended Data Fig. 7). As the outbreak continues to unfold, critically evaluating this mapping (Extended Data Fig. 8) can focus surveillance efforts to areas expected to have prolonged epidemic trajectories and high mortality burdens. The emergence and rapid spread in southern Africa of lineage B.1.351, with multiple spike protein mutations including the N501Y mutation associated with increased transmission rate in the UK lineage B.1.1.7, indicates the importance of genomic surveillance of transmission foci in SSA[38]. Additional immunological surveys and country-specific analyses of the age profile of mortality are urgently needed in SSA and will likely be a powerful lens for understanding the current landscape of population risk[39]. When considering hopeful futures with the distribution of SARS-CoV-2 vaccines, it is imperative that vaccine distribution be equitable and in proportion with need. Understanding factors that both drive spatial variation in vulnerable populations and temporal variation in pandemic progression could help approach these goals in SSA.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41591-021-01234-8.

References

1. Mecenas, P., Bastos, R. T. D. M., Vallinoto, A. C. R. & Normando, D. Effects of temperature and humidity on the spread of COVID-19: a systematic review. *PLoS ONE* **15**, e0238339 (2020).
2. Salje, H. Estimating the burden of SARS-CoV-2 in France. *Science* **369**, 208–211 (2020).
3. Rinaldi, G. & Paradisi, M. An empirical estimate of the infection fatality rate of COVID-19 from the first Italian outbreak. Preprint at *medRxiv* https://doi.org/10.1101/2020.04.18.20070912 (2020).
4. Verity, R. et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect. Dis.* **20**, 669–677 (2020).
5. Uyoga, S. et al. Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Kenyan blood donors. *Science* **371**, 79–82 (2021).
6. Majiya, H. et al. Seroprevalence of COVID-19 in Niger State. Preprint at *medRxiv* https://doi.org/10.1101/2020.08.04.20168112 (2020).
7. Chibwana, M. G. et al. High SARS-CoV-2 seroprevalence in health care workers but relatively low numbers of deaths in urban Malawi. *Wellcome Open Res.* **5**, 199 (2020).
8. Mbow, M. et al. COVID-19 in Africa: dampening the storm? *Science* **369**, 624–626 (2020).
9. Skrip, L. A. et al. Seeding COVID-19 across sub-Saharan Africa: an analysis of reported importation events across 40 countries. Preprint at *medRxiv* https://doi.org/10.1101/2020.04.01.20050203 (2020).
10. Deng, X. et al. Case fatality risk of the first pandemic wave of coronavirus disease 2019 (COVID-19) in China. *Clin. Inf. Dis.* ciaa578 (2020).
11. Williamson, E. et al. Factors associated with COVID-19-related death using OpenSAFELY. *Nature* **584**, 430–436 (2020).
12. *COVID-19 Significantly Impacts Health Services for Noncommunicable Diseases* (World Health Organization, 2020); https://www.who.int/news-room/detail/01-06-2020-covid-19-significantly-impacts-health-services-for-noncommunicable-diseases
13. *Maintaining Essential Health Services: Operational Guidance for the COVID-19 Context* (World Health Organization, 2020); https://www.who.int/publications/i/item/covid-19-operational-guidance-for-maintaining-essential-health-services-during-an-outbreak (2020).
14. Santoli, J. M. et al. Effects of the COVID-19 pandemic on routine pediatric vaccine ordering and administration: United States, 2020. *MMWR Morb. Mortal. Wkly. Rep.* **69**, 591–593, (2020).
15. Roberton, T. et al. Early estimates of the indirect effects of the COVID-19 pandemic on maternal and child mortality in low-income and middle-income countries: a modelling study. *Lancet Glob. Health* **8**, e901–e908 (2020).
16. Walker, P. G. T. et al. The impact of COVID-19 and strategies for mitigation and suppression in low- and middle-income countries. *Science* **369**, 413–422 (2020).
17. Pei, S., Kandula, S. & Shaman, J. Differential effects of intervention timing on COVID-19 spread in the United States. *Sci. Adv.* **6**, eabd6370 (2020).
18. Lai, S. et al. Assessing the effect of global travel and contact reductions to mitigate the COVID-19 pandemic and resurgence. Preprint at *medRxiv* https://doi.org/10.1101/2020.06.17.20133843 (2020).
19. Nepomuceno, M. R. et al. Besides population age structure, health and other demographic factors can contribute to understanding the COVID-19 burden. *Proc. Natl Acad. Sci. USA* **117**, 13881–13883 (2020).
20. Clark, A. et al. Global, regional, and national estimates of the population at increased risk of severe COVID-19 due to underlying health conditions in 2020: a modelling study. *Lancet Glob. Health* **8**, e1003–e1017 (2020).
21. Ghisolfi, S. et al. Predicted COVID-19 fatality rates based on age, sex, comorbidities, and health system capacity. *BMJ Glob. Health* **5**, e003094 (2020).
22. Winskill P. et al. *Report 22: Equity in Response to the COVID-19 Pandemic: an Assessment of the Direct and Indirect Impacts on Dsadvantaged and Vulnerable Populations in Low- and Lower Middle-Income Countries* (Imperial College London, 2020); https://www.imperial.ac.uk/media/imperial-college/medicine/mrc-gida/2020-05-12-COVID19-Report-22.pdf
23. Kapata, N. et al. Is Africa prepared for tackling the COVID-19 (SARS-CoV-2) epidemic. Lessons from past outbreaks, ongoing pan-African public health efforts, and implications for the future. *Int. J. Infect. Dis.* **93**, 233–236 (2020).
24. Nachega, J. B. et al. From easing lockdowns to scaling-up community-based COVID-19 screening, testing, and contact tracing in Africa: shared approaches, innovations, and challenges to minimize morbidity and mortality. *Clin. Infect. Dis.* ciaa695 (2020).
25. Baker, R. E., Yang, W., Vecchi, G. A., Metcalf, C. J. E. & Grenfell, B. T. Susceptible supply limits the role of climate in the early SARS-CoV-2 pandemic. *Science* **369**, 315–319 (2020).
26. Weiss, D. J. et al. A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature* **553**, 333–336 (2018).
27. Tatem, A. J. WorldPop, open data for spatial demography. *Sci. Data* **4**, 170004 (2017).
28. World Health Organization and United Nations Children's Fund Joint Monitoring Programme for Water Supply, Sanitation and Hygiene. *Core Questions on Water, Sanitation and Hygiene for Household Surveys: 2018 Update.* https://washdata.org/sites/default/files/documents/reports/2019-05/JMP-2018-core-questions-for-household-surveys.pdf (UNICEF and WHO, 2018).
29. Korevaar, H. M. et al. Quantifying the impact of US state non-pharmaceutical interventions on COVID-19 transmission. Preprint at at *medRxiv* https://doi.org/10.1101/2020.06.30.20142877 (2020).
30. O'Driscoll, M. et al. Age-specific mortality and immunity patterns of SARS-CoV-2 infection in 45 countries. *Nature* https://doi.org/10.1038/s41586-020-2918-0 (2020).
31. Mikkelsen, L. et al. A global assessment of civil registration and vital statistics systems: monitoring data quality and progress. *Lancet* **386**, 1395–1406 (2015).
32. Gilbert, M. et al. Preparedness and vulnerability of African countries against importations of COVID-19: a modelling study. *Lancet* **395**, 871–877 (2020).
33. Haider, N. et al. Passengers' destinations from China: low risk of novel coronavirus (2019-nCoV) transmission into Africa and South America. *Epidemiol. Infect.* **148**, e41 (2020).
34. Takahashi, S. et al. Reduced vaccination and the risk of measles and other childhood infections post-Ebola. *Science* **347**, 1240–1242 (2015).
35. Cash, R. & Patel, V. Has COVID-19 subverted global health? *Lancet* **395**, 1687–1688 (2020).
36. Adams, J. G. & Walls, R. M. Supporting the health care workforce during the COVID-19 global epidemic. *JAMA* **323**, 1439–1440 (2020).
37. Kilmarx, P. H. et al. Ebola virus disease in health care workers: Sierra Leone, 2014. *MMWR Morb. Mortal. Wkly. Rep.* **63**, 1168–1171 (2014).
38. Tegally, H. et al. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. Preprint at *medRxiv* https://doi.org/10.1101/2020.12.21.20248640 (2020).
39. Mina, M. J. et al. A global immunological observatory to meet a time of pandemics. *eLife* **9**, e58989 (2020).

Methods

Reported SARS-CoV-2 case counts, mortality and testing in SSA as of December 2020. *Variables and data sources for reporting data.* The numbers of reported cases, deaths and tests for the 48 SSA countries studied (Supplementary Table 1) were sourced from the Africa CDC dashboard on 20 December 2020 (and previously on 23 September and 30 June 2020). The Africa CDC obtains data from the official Africa CDC Regional Collaborating Centre and member state reports. Differences in the timing of reporting by member states results in some variation in the recency of data within the centralized Africa CDC repository, but data should broadly reflect the relative scale of testing and reporting efforts across countries. For Mauritius (https://covid19.mu/) and Rwanda (https://covid19.who.int/region/afro/country/rw), reporting to the Africa CDC was confirmed by comparison to country-specific dashboards.

The countries or member states within SSA in this study follow the United Nations and Africa CDC-listed regions of Southern, Western, Central and Eastern Africa (excluding Sudan). From the Northern Africa region, Mauritania is included in SSA.

For comparison to non-SSA countries, the number of reported cases in other geographical regions were obtained from the Johns Hopkins University Coronavirus Resource Center on 23 September 2020 (https://coronavirus.jhu.edu/map.html).

Case fatality ratios (CFRs) were calculated by dividing the number of reported deaths by the number of reported cases and expressed as a percentage. Positivity was calculated by dividing the number of reported cases by the number of reported tests. Testing and case rates were calculated per 100,000 population using population size estimates for 2020 from the United Nations Population Division (https://population.un.org/wpp/Download/Standard/Population/). Since reported confirmed cases are likely to be an underestimate of the true number of infections, CFRs may be a poor proxy for the IFR, defined as the proportion of infections that result in mortality[4].

Variation in testing and mortality rates. Testing rates among SSA countries varied by multiple orders of magnitude as of 30 June and remain highly variable as of 23 September and 20 December 2020. The number of tests completed per 100,000 population ranged from 19.84 in Burundi to 13,508.13 in Mauritius in June 2020; from 65.98 in the Democratic Republic of the Congo to 18,321.83 in Mauritius in September 2020; and from 100.9 in the Democratic Republic of the Congo to 23695.0 in Mauritius in December 2020 (Extended Data Fig. 1a). Tanzania (6.50 tests per 100,000 population) has not reported new tests, cases or deaths to the Africa CDC since April 2020. The number of reported infections (that is, positive tests) was strongly correlated with the number of tests completed in June 2020 (Pearson's correlation coefficient, $r = 0.9667$, $P < 0.001$), September 2020 ($r = 0.9689$, $P < 0.001$) and December 2020 ($r = 0.9750$, $P < 0.001$) (Extended Data Fig. 1b). As of June 2020, no deaths due to SARS-CoV-2 were reported to the Africa CDC for five SSA countries (Eritrea, Lesotho, Namibia, Seychelles, Uganda). As of December 2020, still no deaths due to SARS-CoV-2 were reported to the Africa CDC for two of those countries (Eritrea and Seychelles). Among countries with at least 1 reported death, the CFR varied from 0.22% in Rwanda to 8.54% in Chad in June 2020; from 0.21% in Burundi to 6.96% in Chad in September 2020; and from 0.26% in Burundi to 5.40% in Chad in December 2020 (Extended Data Fig. 1c). Limitations in the ascertainment of infection rates and the rarity of reported deaths (for example, the median number of reported deaths per SSA country was 25.5 as of June 2020, 71.0 as of September 2020 and 101.0 as of December 2020), indicate that the data are insufficient to determine country-specific IFRs and IFR by age profiles for most countries. As a result, global IFR by age estimates was used for the subsequent analyses in this study.

Synthesizing factors that increase or decrease SARS-CoV-2 epidemic risk in SSA. *Variable selection and data sources for variables associated with an increased probability of severe clinical outcomes for an infection.* To characterize epidemic risk, defined as potential SARS-CoV-2 related morbidity and mortality, we first synthesized factors hypothesized to influence risk in SSA settings (Supplementary Table 2). Early during the pandemic, evidence suggested that age was an important risk factor associated with morbidity and mortality associated with SARS-CoV-2 infection[40], a pattern subsequently confirmed across settings[2,11,41]. Associations between SARS-CoV-2 mortality and comorbidities including hypertension, diabetes and cardiovascular disease emerged early[40] and have been observed across settings, with further growing evidence for associations with obesity[11,42], severe asthma[11] and the respiratory effects of pollution[43]. Specific to Africa, vulnerability scores based on these hypothesized associations or combinations of risks factors have been developed (for example, refs. [44,45]).

Many possible sources of bias complicate interpretation of these associations[46]; while they provide a useful baseline, inference is also likely to change as the pandemic advances. To reflect this, our analysis combined a number of high-level variables likely to broadly encompass these putative risk factors (for example, noncommunicable disease-related mortality and healthy life expectancy) with more specific measures encompassed in evidence to date (for example, prevalence of diabetes, obesity and respiratory illness, such as chronic obstructive pulmonary disease). We also included measures relating to infectious diseases, undernourishment and anemia given their interaction and effects in determining

health status in these settings[47]. Although interactions with such infectious diseases have been suggested, evidence is limited to date, except for HIV, where effects have been suggested to be minor[48]. We also note that the key concern raised around such infections to date is associated with disruption to routine screening (for example, for malaria[49]), treatment[50] or prevention programs[51].

Data on the identified indicators were sourced in May 2020 from the WHO Global Health Observatory database (https://www.who.int/data/gho), World Bank (https://data.worldbank.org/) and other sources detailed in Supplementary Table 3. National-level demographic data (population size and age structure) was sourced from the United Nations World Population Prospects and data on subnational variation in demography was sourced from WorldPop[27]. Household size data was defined by the mean number of individuals in a household with at least 1 person aged >50 years, taken from the most recently available Demographic and Health Surveys data (https://dhsprogram.com). All country-level data for all indicators can be found online at the SSA-SARS-CoV-2 tool (https://labmetcalf.shinyapps.io/covid19-burden-africa/).

Comparisons of national-level estimates sourced from the WHO and other sources are affected by variation within countries and variation in the uncertainty around estimates from different geographical areas. To assess potential differences in data quality between geographical areas, we compared the year of the most recent data for the variables (Extended Data Fig. 2). The mean (range varied from 2014.624 to 2014.928 by region) and median year (2016 for all regions) of the most recent data varied little between regions. To account for the uncertainty associated with the estimates available for a single variable, we also included multiple variables per category (for example, demographic and socioeconomic factors, comorbidities, access to care) to avoid reliance on a single metric. This allowed exploring variation between countries across a broad suite of variables likely to be indicative of the different dimensions of risk.

Although including multiple variables that were likely to be correlated (see the principal component analysis (PCA) methods below for further discussion) would bias inference of cumulative risk in a statistical framework, we did not attempt to quantitatively combine risk across variables for a country, nor project risk based on the variables included in this study. Rather, we characterized the magnitude of variation among countries for these variables (see Fig. 3 for a subset of the variables and Fig. 4b for the bivariate risk maps following Chin et al.[52]) and then explored the range of outcomes that would be expected under scenarios where the IFR increases with age at different rates (Fig. 4).

Variable selection and data sources for variables modulating the rate of viral spread. In addition to characterizing variation among factors likely to modulate burden, we also synthesized data sources relevant to the rate of viral spread, or pace, for the SARS-CoV-2 pandemic in SSA. Factors hypothesized to modulate viral transmission and geographical spread include climatic factors (for example, specific humidity), access to prevention measures (for example, handwashing) and human mobility (for example, international and domestic travel). Supplementary Table 2 outlines the dimensions of risk selected and references the previous studies relevant to the selection of these factors.

Climate data were sourced from the global, gridded ERA5 dataset[53] where model data were combined with global observation data (see Methods for climate-driven modeling of SARS-CoV-2 section for further details).

International flight data were obtained from a custom report from OAG Aviation Worldwide (UK) and included the departure location, arrival airport, date of travel and number of passenger seats for flights arriving to 113 international airports in SSA (see International air travel to SSA section).

As an estimate of connectivity within subregions of countries, the population-weighted mean travel time to the nearest city with a population greater than 50,000 was determined; details are provided in the section on Subnational connectivity among countries in SSA. To obtain a set of measures that broadly represent connectivity within different countries in the region, friction surfaces from Weiss et al.[26] were used to obtain estimates of the connectivity between different administrative level 2 units within each country. Details of this, alongside the metapopulation model framework used to simulate viral spread with variation in connectivity are found in the Subnational connectivity section.

Figure 3 shows variation among SSA countries for four of the variables and Extended Data Fig. 3 links to visualizations of variation for all variables. Figure 4 shows variation for a subset of the comorbidity and access to care indicators as a heatmap and Extended Data Fig. 4 shows variation for all the variables (also available at https://labmetcalf.shinyapps.io/covid19-burden-africa/).

PCA of variables considered. *Selection of data and variables.* The 29 national-level variables from Supplementary Table 3 were selected for the PCA. We conducted further PCA on the subset of 8 indicators related to access to healthcare (category E) and the 14 national indicators variables related to comorbidities (category B).

We excluded disaggregated subnational spatial variation data (variables A2, C1, E2 and category F), disaggregated or redundant variables derived from variables already included (variables A4 and D2) and disaggregated age-specific disease data from the Institute for Health Metrics and Evaluation (IHME) global burden of disease study (variables B2, B4 and B13) from the PCA analysis. COVID-19 tests per 100,000 population (variable D4; Supplementary Table 1), per capita gross

domestic product (GDP) (variable A8) and the Gini index of wealth inequality (variable A9) were used to visualize patterns among SSA countries.

In some cases, data were missing for a country for an indicator; in these cases, missing data were replaced with a zero value. This is a conservative approach since zero values (that is, outside the range of typical values seen in the data) inflate the total variance in the dataset and thus, if anything, deflate the percentage of the variance explained by the PCA. Therefore, this approach avoids mistakenly attributing predictive value to principal components due to incomplete data. See Supplementary Table 3 for data sources for each variable.

PCA. The PCA was conducted on each of the three subsets described above using the scikit-learn library[54]. To avoid biasing the PCA due to large differences in magnitude and scale, each feature was centered around the mean and scaled to unit variance before the analysis. Briefly, PCA applies a linear transformation to a set of n features to output a set of n orthogonal principal components that are uncorrelated and each explain a percentage of the total variance in the dataset[55]. A link to the code for this analysis is available at https://labmetcalf.shinyapps.io/covid19-burden-africa/.

The principal components were then analyzed for the percentage of variance explained and compared to: (1) the number of COVID-19 tests per 100,000 population as of the end of June 2020 (Supplementary Table 1); (2) the per capita GDP; and (3) the Gini index of wealth inequality. For the Gini index, estimates from 2008 to 2018 were available for 45 of the 48 countries (no Gini index data were available for Eritrea, Equatorial Guinea and Somalia).

The first 2 principal components from the analysis of 29 variables explain 32.6 and 13.1% the total variance, respectively, in the dataset. Countries with higher numbers of completed SARS-CoV-2 tests reported tended to associate with an increase in principal component 1 ($r = 0.67$, $P = 1.1 \times 10^{-7}$; Extended Data Fig. 5a). Similarly, countries with a high GDP seemed to associate with an increase in principal component 1 ($r = 0.80$, $P = 6.02 \times 10^{-12}$; Extended Data Fig. 5b). In contrast, countries with greater wealth inequality (as measured by the Gini index) were associated with a decrease in principal component 2 ($r = -0.42$, $P = 0.0042$; Extended Data Fig. 5c). Despite these correlations, a relatively low percentage of variance was explained by each principal component: for the 29 variables, 13 of the 29 principal components were required to explain 90% of the variance (Extended Data Fig. 5d). When only the access to care subset of variables is considered, the first 2 principal components explain 50.7 and 19.1% of the variance, respectively, and 5 of 8 principal components are required to explain 90% of the variance. When only the comorbidities subset is considered, the first two principal components explain 27.9 and 17.8% of the variance, respectively, and 9 of 14 principal components are required to explain 90% of the variance (Extended Data Fig. 5d).

These data suggest that intercountry variation in this dataset is not easily explained by a small number of variables. Moreover, although correlations exist between principal components and high-level explanatory variables (testing capacity, wealth), their magnitude is modest. These results highlight that dimensionality reduction is unlikely to be an effective analysis strategy for the variables considered in this study. Despite this overall finding, the PCA on the access to care subset of variables highlights that the variance in these variables is more easily explained by a small number of principal components and hence may be more amenable to dimensionality reduction. This finding is unsurprising since, for example, the number of hospital beds per 100,000 population is likely to be directly related to the number of hospitals per 100,000 population ($r = 0.60$, $P = 5.7 \times 10^{-6}$ for SSA). In contrast, for comorbidities, the relationship between different variables is less clear. Given the low percentages of variation captured by each principal component, and the high variability between different types of variables, these results motivate a holistic approach to using these data for assessing relative SARS-CoV-2 risk across SSA.

Evaluating the burden emerging from the severity of infection outcome.
Data sourcing: empirical estimates of IFR. Estimates of the IFR that account for asymptomatic cases, underreporting and delays in reporting are few; however, it is evident that the IFR increases substantially with age[56]. We used age-stratified estimates of IFR from three studies (two published[2,4] and one preprint[3]) that accounted for these factors in their estimation (Supplementary Table 4).

To apply these estimates to other age-stratified data with different bin ranges and generate continuous predictions of IFR with age, we fitted the relationship between the midpoint of the age bracket and the IFR estimate using a generalized additive model using the mgcv package v1.8-33 (ref. [57]) in R v.4.0.2 (ref. [58]). We used a beta distribution as the link function for the IFR estimates (data distributed on [0, 1]). For the upper age bracket (80+ years), we took the upper range to be 100 years and the midpoint to be 90.

We assumed a given level of cumulative infection (20% in each age class, that is, a constant rate of infection among age classes) and then applied IFRs by age to the population structure of each country to generate estimates of burden. Age structure estimates were taken from the United Nations World Population Prospects (Supplementary Table 3) country-level estimates of population in 1-year age groups (0–100 years of age) to generate estimates of burden.

Comorbidities over age from the IHME. Applying these IFR estimates to the demographic structure of SSA countries provides a baseline expectation for

mortality but depends on the assumption that mortality patterns in SSA are similar to those from where the IFR estimates were sourced (France, China and Italy). Comorbidities have been shown to be an important determinant of the severity of infection outcomes (that is, IFR). To assess the relative risk of comorbidities across age in SSA, estimates of comorbidity severity by age (in terms of annual deaths attributable) were obtained from the IHME Global Burden of Disease (GBD) study in 2017 (ref. [59]). Data were accessed through the GBD results tool for cardiovascular disease, chronic respiratory disease (not including asthma) and diabetes, reflecting three categories of comorbidity with demonstrated associations with risk (Supplementary Table 2). We assumed that higher mortality rates due to these noncommunicable diseases, especially among younger age groups, is indicative of increased severity and lesser access to sufficient care for these diseases, suggesting an elevated risk for their interaction with SARS-CoV-2 as comorbidities. While there are uncertainties in these data, they provide the best estimates of age-specific risks and have been used previously to estimate populations at risk[20].

The comorbidity by age curves for SSA countries were compared to those for the three countries from which SARS-CoV-2 IFR by age estimates were sourced. Attributable mortality due to all three noncommunicable disease categories was higher at age 50 in all 48 SSA countries when compared to estimates from France and Italy and for 42 of 48 SSA countries when compared to China (Extended Data Fig. 6).

Given the potential for populations in SSA to experience a differing burden of SARS-CoV-2 due to their increased severity of comorbidities in younger age groups, we explored the effects of shifting IFRs estimated by the generalized additive model of IFR estimates from France, Italy and China younger by 2, 5 and 10 years (Fig. 3).

International air travel to SSA. The number of passenger seats on flights arriving to international airports were grouped by country and month for January to April 2020 (Supplementary Table 5), the months when the introduction of SARS-CoV-2 to SSA countries was likely to have first occurred. The first confirmed case reported from an SSA country, according to the Johns Hopkins Coronavirus Research Center was in Nigeria on 28 February 2020. By 31 March 2020, 43 of 48 SSA countries had reported SARS-CoV-2 infections and international travel was largely restricted by April. Lesotho was the last SSA country to report a confirmed SARS-CoV-2 infection (on 13 May 2020); however, given the difficulties in surveillance, the first reported detections were likely delayed relative to the first importations of the virus. The probability of importation of the virus is defined by the number of travelers from each source location, each date and the probability that a traveler from that source location on that date was infectious. Due to limitations in surveillance, especially early in the SARS-CoV-2 pandemic, empirical data on infection rates among travelers were largely lacking. To account for differences in the status of the SARS-CoV-2 pandemic across source locations and thus differences in the importation risk for travelers from those locations, we coarsely stratified travelers arriving each day into 4 categories based on the status of their source countries: (1) travelers from countries with zero reported cases (that is, although undetected transmission was possibly occurring, SARS-CoV-2 had not yet been confirmed in the source country by that date); (2) those traveling from countries with more than 1 reported case (that is, SARS-CoV-2 had been confirmed to be present in that source country by that date); (3) those traveling from countries with more than 100 reported cases (indicating community transmission was likely beginning); and (4) those traveling from countries with more than 1,000 reported cases (indicating widespread transmission).

To determine reported case counts at source locations for travelers, no cases were reported outside China until 13 January 2020 (the date of the first reported case in Thailand). Over 13 January to 21 January, cases were then reported in Japan, South Korea, Taiwan, Hong Kong and the United States (https://covid19.who.int/). Subsequently, counts per country were tabulated daily by the Johns Hopkins Coronavirus Resource Center[60] beginning on 22 January (https://coronavirus.jhu.edu/map.html); we used the data from 22 January onwards and the WHO reports before 22 January.

The number of travelers within each category arriving per month is shown in Supplementary Table 5. This approach makes the conservative assumption that the probability a traveler is infected reflects the general countrywide infection rate of the source country at the time of travel (that is, travelers are not more likely to be exposed than non-travelers in that source location) and does not account for complex travel itineraries (that is, a traveler from a high-risk source location transiting through a low-risk source location would be grouped with other travelers from the low-risk source location). Consequently, the risk for viral importation is likely systematically underestimated. However, since the relative risk for viral importation will still scale with the number of travelers, comparisons among SSA countries can be informative (for example, SSA countries with more travelers from countries with confirmed SARS-CoV-2 transmission are at higher risk for viral importation).

Subnational connectivity among countries in SSA. *Indicators of subnational connectivity.* To allow comparison of the relative connectivity across countries, we used the friction surface estimates provided by Weiss et al.[26] as a relative

measure of the rate of human movement between subregions of a country. For connectivity within subregions of a country (for example, transport from a city to the rural periphery), we used as an indicator the population-weighted mean travel time to the nearest urban center (that is, population density >1,500 per square kilometer or a density of built-up areas >50% coincident with population >50,000) within administrative 2 units[61]. For some countries, estimates at administrative 2 units were unavailable (Comoros, Cape Verde, Lesotho, Mauritius, Mayotte and Seychelles); estimates at the administrative-1 unit level were used for these cases (these were all island nations, with the exception of Lesotho).

Metapopulation model methods. Once SARS-CoV-2 has been introduced into a country, the degree of spread of the infection within the country is governed by subnational mobility: the pathogen is more likely to be introduced into a location where individuals arrive more frequently than one where incoming travelers are less frequent. Large-scale consistent measures of mobility are rare. However, recently, estimates of accessibility have been produced at a global scale[26]. Although this is unlikely to perfectly reflect mobility within countries, especially since interventions and travel restrictions are put in place, it provides a starting point for evaluating the role of human mobility in shaping the outbreak pace across SSA. We used the inverse of a measure of the cost of travel between the centroids of administrative level 2 spatial units to describe mobility between locations (estimated by applying the costDistance function in the gdistance package v1.3-6 in R to the friction surfaces supplied in Weiss et al.[26]). With this, we developed a metapopulation model for each country to develop an overview of the possible range of trajectories of unchecked spread of SARS-CoV-2.

We assumed that the pathogen first arrives in each country in the administrative 2 level unit with the largest population (for example, the largest city) and the population in each administrative 2 level (of size N_j) is entirely susceptible at the time of arrival. We then tracked the spread within and between each of the administrative 2 level units of each country. Within each administrative 2 level unit, dynamics are governed by a discrete time susceptible (S), infected (I) and recovered (R) model with a time step of approximately one week, which is broadly consistent with the serial interval of SARS-CoV-2. Within the spatial unit indexed j, with total size N_j, the number of infected individuals in the next time step is defined by:

$$I_{j,t+1} = \beta I_{j,t}^{\alpha} S_{j,t}/N_j + \iota_{j,t}$$

where β captures the magnitude of transmission over the course of one discrete time step; since the discrete time step chosen is set to approximate the serial interval of the virus, this will reflect the R_0 of SARS-CoV-2, and is thus set to 2.5; the exponent $\alpha = 0.97$ is used to capture the effects of discretization[62] and $I_{j,t}$ captures the introduction of new infections into site j at time t. Susceptible and recovered individuals are updated according to:

$$S_{j,t+1} = S_{j,t} + wR_{j,t} - I_{j,t+1} + b$$
$$R_{j,t+1} = (1 - w)R_{j,t} + I_{j,t}$$

where b reflects the introduction of new susceptible individuals resulting from the birth rate, set to reflect the most recent estimates for that country from the World Bank Data (https://data.worldbank.org/indicator/SP.DYN.CBRT.IN), and w reflects the rate of waning of immunity. The population is initiated with $S_{j,1} = N_j R_{j,1} = 0$, and $I_{j,1} = 0$ except for the spatial unit corresponding to the largest population size N_j for each country since this is assumed to be the location of introduction; for this spatial unit, we set $I_{j,1} = 1$.

We made the simplifying assumption that mobility linking locations i and j, denoted as $c_{i,j}$, scales with the inverse of the cost of travel between sites i and j evaluated according to the friction surface provided in Weiss et al.[26]. The introduction of an infected individual into location j is then defined by a draw from a Bernouilli distribution following:

$$\iota_{j,t} \approx \text{Bernouilli}\left(1 - \exp\left(-\sum_{1}^{L} c_{i,j} I_{i,t}/N_i\right)\right)$$

where L is the total number of administrative 2 units in that country and the rate of introduction is the product of connectivity between the focal location and each other location multiplied by the proportion of population in each other location that is infected.

Some countries show rapid spread between administrative units within the country (for example, a country with parameters that broadly reflect those available for Malawi; Extended Data Fig. 7), while in others (for example, reflecting Madagascar), connectivity may be so low that the outbreak may be over in the administrative unit of the largest size (where it was introduced) before introductions successfully reach other poorly connected administrative units. Where duration of immunity is sufficiently long, the result may be a hump-shaped relationship between the proportion of the population that is infected after five years and the time to the first local extinction of the pathogen (Extended Data Fig. 7, top right). In countries with lower connectivity (for example, resembling Madagascar), local outbreaks can go extinct rapidly before traveling very far; in other countries (for example, resembling Gabon), the pathogen goes extinct rapidly because it travels rapidly and rapidly depletes susceptible individuals everywhere.

The U-shaped pattern diminishes as the rate of waning of immunity increases and is replaced by a monotonic negative relationship. With sufficiently rapid waning of immunity, local extinction ceases to occur in the absence of control efforts.

The impact of the pattern of travel between centroids is echoed by the pattern of travel within administrative districts: countries where the pathogen does not reach a large fraction of the administrative 2 units within the country in five years are also those where within-administrative-unit travel is low (Extended Data Fig. 7, right).

These simulations provide a window into qualitative patterns expected for subnational spread of the pandemic virus but there is no clear way of calibrating the absolute rate of travel between regions of relevance for SARS-CoV-2; this is further complicated by the remaining uncertainties around rates of waning of immunity. Thus, the time scales of these simulations should be considered in relative, rather than absolute terms. Variation in lockdown effectiveness, or other changes in mobility for a given country, may also compromise relative comparisons as might large volumes of land border crossings in some settings, which we have not accounted for in this study. Variability in testing and case reporting complicates clarifying this (Extended Data Fig. 7, bottom left and bottom right, respectively) but we have highlighted countries with less connectivity (that is, less synchronous outbreaks expected) relative to the median among SSA countries and with older populations (that is, a greater proportion in higher-risk age groups) (Extended Data Fig. 8).

The University of Oxford's Blavatnik School of Government generated composite scores of government response, interventions for containment and economic support provided, with each scored from 0 to 100 (Coronavirus Government Response Tracker; https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker). These data were compared with the day on which ten cases were exceeded in a country according to the Johns Hopkins dashboard data (Johns Hopkins Coronavirus Resource Center; https://coronavirus.jhu.edu/map.html).

While faster waning of immunity will act to increase the rate of spread of the infection, resulting in a higher proportion infected after one year, control efforts will generally act to slow the rate of spread of the infection (Extended Data Fig. 9). Since different countries are likely to have differently effective control efforts (Extended Data Fig. 9), this precludes making country-specific predictions as to the relative impact of control efforts on delay.

Modeling epidemic trajectories in scenarios where transmission rate depends on climate. *Climate data sourcing: variation in humidity in SSA.* Specific humidity data for selected urban centers comes from the ERA5 using an average climatology (1981–2017)[53]; we did not consider year-to-year climate variations. Selected cities ($n = 56$) were chosen to represent the major urban areas in SSA. The largest city in each SSA country was included as well as any additional cities that were among the 25 largest cities or busiest airports in SSA.

Methods for climate-driven modeling of SARS-CoV-2. We used a climate-driven susceptible-infected-recovered-susceptible model to estimate epidemic trajectories (that is, the time of peak incidence) in different cities in 2020, assuming no control measures were in place or a 10 or 20% reduction in R_0 beginning 2 weeks after the total reported cases for a country exceeded 10 cases[25,63]. The model is given by:

$$\frac{dS}{dt} = \frac{N - S - L}{L} - \frac{\beta(t)IS}{N}$$

$$\frac{dI}{dt} = \frac{\beta(t)IS}{N} - \frac{I}{D}$$

where S is the susceptible population, I is the infected population and N is the total population. D is the mean infectious period, set at 5 d following ref. [25].

To investigate the effects on epidemic trajectories of a climate dependency of SARS-CoV-2 on cities with the climate patterns of the selected cities in SSA, we used parameters from the most climate-dependent scenario in ref. [25], based on the endemic betacoronavirus HKU1 in the United States. In this scenario L, the duration of immunity, was 66.25 weeks (that is, >1 year and such that waning immunity did not affect the timing of the epidemic peak). We initially selected a range where R_0 declined from $R_{0\text{max}} = 2.5$ to $R_{0\text{min}} = 1.5$ (that is, transmission declined 40% at high humidity) since this exceeds the range observed for influenza and other coronaviruses for which data are available (from the United States). $R_{0\text{max}} = 2.5$ was chosen because 2.5 is often cited as the approximate R_0 for SARS-CoV-2. Thus, we initially assumed that the climate dependence of SARS-CoV-2 in SSA would not greatly exceed that of other known coronaviruses from the US context. Then, we explored the effects of different degrees of climate dependency (that is, wider ranges between $R_{0\text{max}} = 2.5$ to $R_{0\text{min}} = 1.5$ and scenarios where $R_{0\text{min}}$ approached 1) (Extended Data Fig. 10).

Transmission is governed by $\beta(t)$, which is related to the basic reproduction number R_0 by $R_0(t) = \beta(t)D$. The basic reproduction number varies based on climate and is related to specific humidity according to the equation:

$$R_0 = \exp[a \times q(t) + \log(R_{0\text{max}} - R_{0\text{min}})] + R_{0\text{min}}$$

where $q(t)$ is specific humidity[53] and a is set at -227.5 based on estimated HKU1 parameters[25]. We assumed the time of introduction for cities to be the date at which the total reported cases for a country exceeded 10 cases.

Sensitivity analysis. Selecting an R_{0min} value of 1, such that epidemic growth stops at high humidities, is likely implausible since simulations indicated no outbreaks would occur in cities such as Antananarivo (countered by the observation that SARS-CoV-2 outbreaks did in fact occur) (Extended Data Fig. 10b; see Supplementary Table 1 for the reported case counts at the country level). Expanding the range between R_{0min} and R_{0max} by increasing R_{0max} resulted in epidemic peaks being reached earlier after outbreak onset but did not increase the difference in timing between cities with different climates (Extended Data Fig. 10c; for example, the difference in timing between peaks in Windhoek and Lomé is similar in 10a and 10c). Finally, we explored scenarios where the R_{0min} was between 1.0 and 1.5. When $R_{0min} > 1.1$, epidemic peaks were seen in each SSA city with the difference in timing of the peak growing larger when smaller values of R_{0min} were selected (Extended Data Fig. 10d). However, the difference in timing, even when small values of R_{0min} were selected, was a maximum of 25 weeks and rapidly reduced to only a few weeks when R_{0min} approached 1.5.

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

Data availability
All data have been deposited into a publicly available GitHub repository at https://github.com/labmetcalf/SSA-SARS-CoV-2.

Code availability
All code has been deposited into the publicly available GitHub repository at https://github.com/labmetcalf/SSA-SARS-CoV-2.

References
40. Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19): China, 2020. *China CDC Wkly.* **2**, 113–122 (2020).
41. Bialek, S. et al. Severe outcomes among patients with coronavirus disease 2019 (COVID-19): United States, February 12–March 16, 2020. *MMWR Morb. Mortal. Wkly. Rep.* **69**, 343–346 (2020).
42. Simonnet, A. et al. High prevalence of obesity in severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) requiring invasive mechanical ventilation. *Obesity (Silver Spring)* **28**, 1195–1199 (2020).
43. Conticini, E., Frediani, B. & Caro, D. Can atmospheric pollution be considered a co-factor in extremely high level of SARS-CoV-2 lethality in Northern Italy? *Environ. Pollut.* **261**, 114465 (2020).
44. *Mapping Risk Factors for the Spread of COVID-19 in Africa* https://africacenter.org/spotlight/mapping-risk-factors-spread-covid-19-africa/ (Africa Center for Strategic Studies, 2020).
45. *The Africa COVID Community Vulnerability Index (CCVI). An Index that Assesses Health, Economic, and Social Impacts of COVID-19 in Africa.* https://precisionforcovid.org/africa (Precision for COVID, 2020).
46. Griffith, G. J. et al. Collider bias undermines our understanding of COVID-19 disease risk and severity. *Nat. Commun.* **11**, 5749 (2020).
47. Shankar, A. H. Nutritional modulation of malaria morbidity and mortality. *J. Infect. Dis.* **182**, S37–S53 (2000).
48. Cooper, T. J., Woodward, B. L., Alom, S. & Harky, A. Coronavirus disease 2019 (COVID-19) outcomes in HIV/AIDS patients: a systematic review. *HIV Med.* **21**, 567–577 (2020).
49. Chanda-Kapata, P., Kapata, N. & Zumla, A. COVID-19 and malaria: a symptom screening challenge for malaria endemic countries. *Int. J. Infect. Dis.* **94**, 151–153 (2020).
50. Lesosky, M. & Myer, L. Modelling the impact of COVID-19 on HIV. *Lancet HIV* **7**, e596–e598 (2020).
51. Sherrard-Smith, E. et al. The potential public health consequences of COVID-19 on malaria in Africa. *Nat. Med.* **26**, 1411–1416 (2020).
52. Chin, T. et al. US-county level variation in intersecting individual, household and community characteristics relevant to COVID-19 and planning an equitable response: a cross-sectional analysis. *BMJ Open* **10**, e039886 (2020).
53. Hersbach, H. et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **146**, 1999–2049 (2020).
54. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
55. Jolliffe, I. T. & Cadima, J. Principal component analysis: a review and recent developments. *Philos. Trans. A Math. Phys. Eng. Sci.* **374**, 20150202 (2016).
56. Meyerowitz-Katz, G. & Merone, L. A systematic review and meta-analysis of published research data on COVID-19 infection fatality rates. *Int. J. Infect. Dis.* **101**, 138–148 (2020).
57. Wood, S. N. *Generalized Additive Models: An Introduction with R* (CRC Press, 2017).
58. R Core Team. *R: A Language and Environment for Statistical Computing.* https://www.R-project.org/ (R Foundation for Statistical Computing, 2020).
59. *Global Burden of Disease Study 2017 (GBD 2017) Population Estimates 1950–2017* (Institute for Health Metrics and Evaluation, 2019); http://ghdx.healthdata.org/record/ihme-data/gbd-2017-population-estimates-1950-2017
60. Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **20**, 533–534 (2020).
61. Linard, C., Gilbert, M., Snow, R. W., Noor, A. M. & Tatem, A. J. Population distribution, settlement patterns and accessibility across Africa in 2010. *PLoS ONE* **7**, e31743 (2012).
62. Bjørnstad, O. N., Finkenstäd, B. F. & Grenfell, B. T. Dynamics of measles epidemics: estimating scaling of transmission rates using a time series SIR model. *Ecol. Monogr.* **72**, 169–184 (2002).
63. Shaman, J., Pitzer, V. E., Viboud, C., Grenfell, B. T. & Lipsitch, M. Absolute humidity and the seasonal onset of influenza in the continental United States. *PLoS Biol.* **8**, e1000316 (2010).
64. *COVID-19 Community Mobility Reports.* https://www.google.com/covid19/mobility/ (Google, 2020).
65. *Measles and Rubella Surveillance Data.* http://158.232.12.119/immunization/monitoring_surveillance/burden/vpd/surveillance_type/active/measles_monthlydata/en/ (World Health Organization, 2020).
66. Nimpa, M. M. et al. Measles outbreak in 2018–2019, Madagascar: epidemiology and public health implications. *Pan Afr. Med. J.* **35**, 84 (2020).

Author contributions
B.L.R., A.A., R.E.B., M.B., W.D-G, K.M., I.F.M., N.V.M., A.R., M.R., J.R., T.R., F.R., W.Y., B.T.G., A.J.T. and C.J.E.M. conceived the analysis. B.L.R., M.R., M.B., W.D.-G. and W.Y. are responsible for data curation. B.L.R., A.A., M.B., M.R., R.E.B. and C.J.E.M. performed the analyses. B.L.R., M.R., M.B., R.E.B., C.J.E.M. and B.T.G. are responsible for the methodology. B.L.R., M.R., M.B., R.E.B. and W.Y. are responsible for the software and the Shiny app online tool. B.L.R., M.R., M.B., R.E.B. and W.Y. led the visualization of data. B.L.R. and C.J.E.M. wrote the initial draft of the manuscript. B.L.R., A.A., R.E.B., M.B., W.D.-G., K.M., I.F.M., N.V.M., A.R., M.R., J.R., T.R., F.R., W.Y., B.T.G., A.J.T. and C.J.E.M. contributed to reviewing and editing the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41591-021-01234-8.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41591-021-01234-8.

Correspondence and requests for materials should be addressed to B.L.R.

Peer review information Jennifer Sargent was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Reprints and permissions information is available at www.nature.com/reprints.
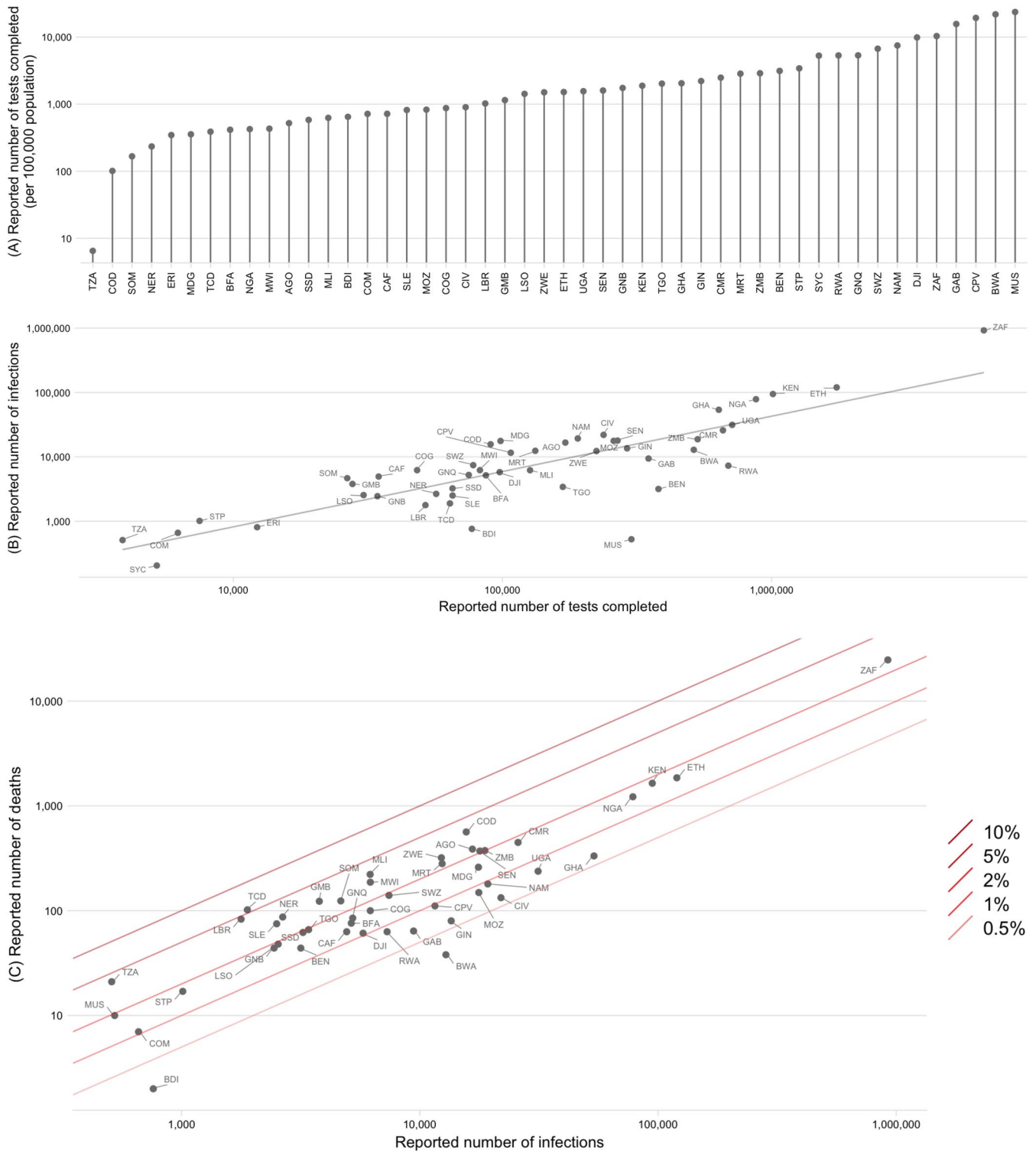
Extended Data Fig. 1 | Variation between SSA countries in testing and reporting rates. a, Reported number of tests completed per country as of December 20, 2020. **b,** Number of infections (*I*) per reported number of tests (*T*); line shows linear least squares regression: $I = 1.422 \times 10^{-1} \times T - 1.912 \times 10^4$ ($df = 46$, adjusted $R^2 = 0.9496$, Pearson's correlation coefficient, $r = 0.9750$, $p < 0.001$). **c,** Reported infections and deaths for sub-Saharan African countries with case fatality ratios (CFRs) shown as diagonal lines.
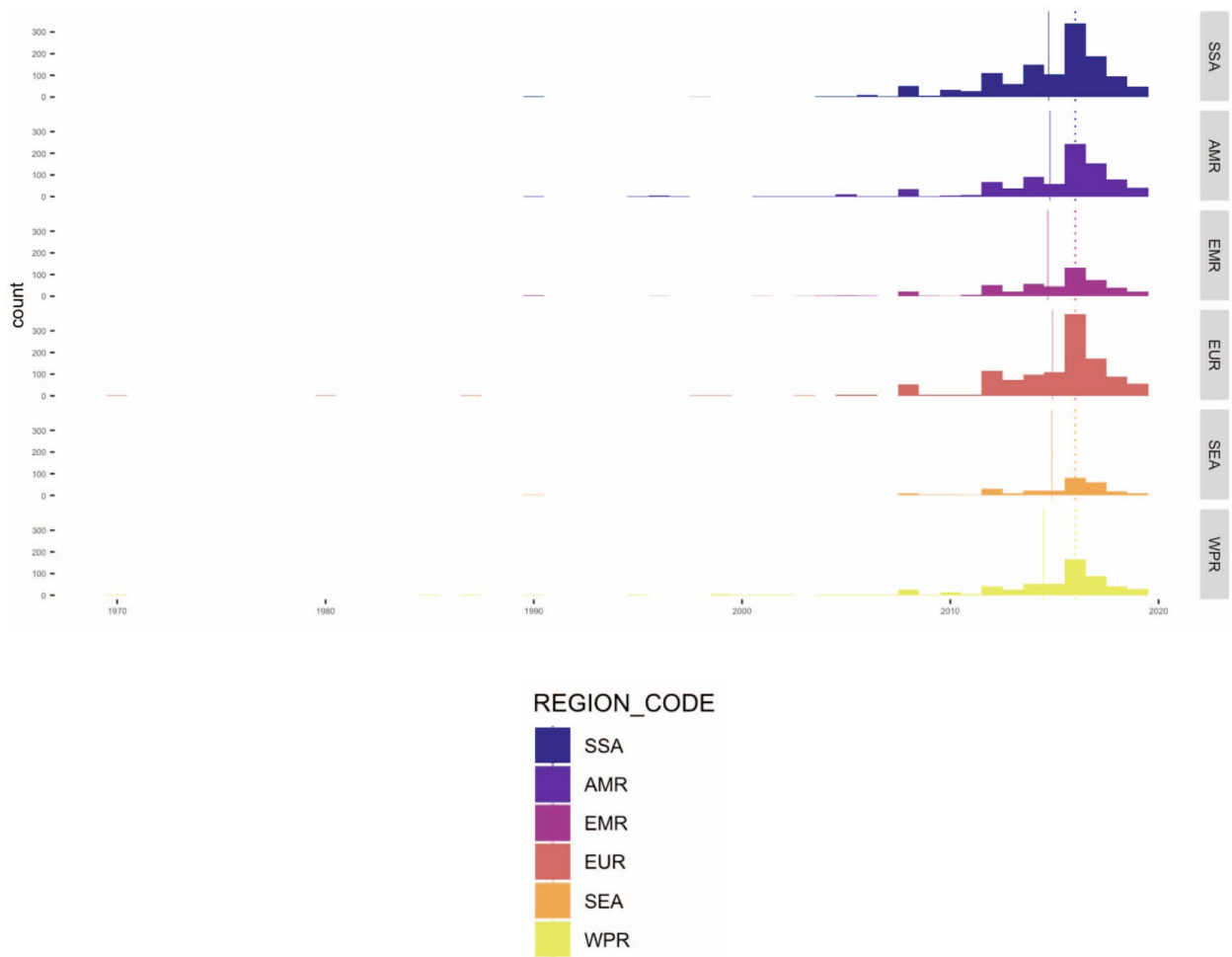
Extended Data Fig. 2 | Year of most recent data available for variables compared between global regions. Dotted vertical line shows regional median; solid vertical line shows regional mean. Note that most data comes from 2015-2019 (median = 2016, mean = 2014.62-2014.93).
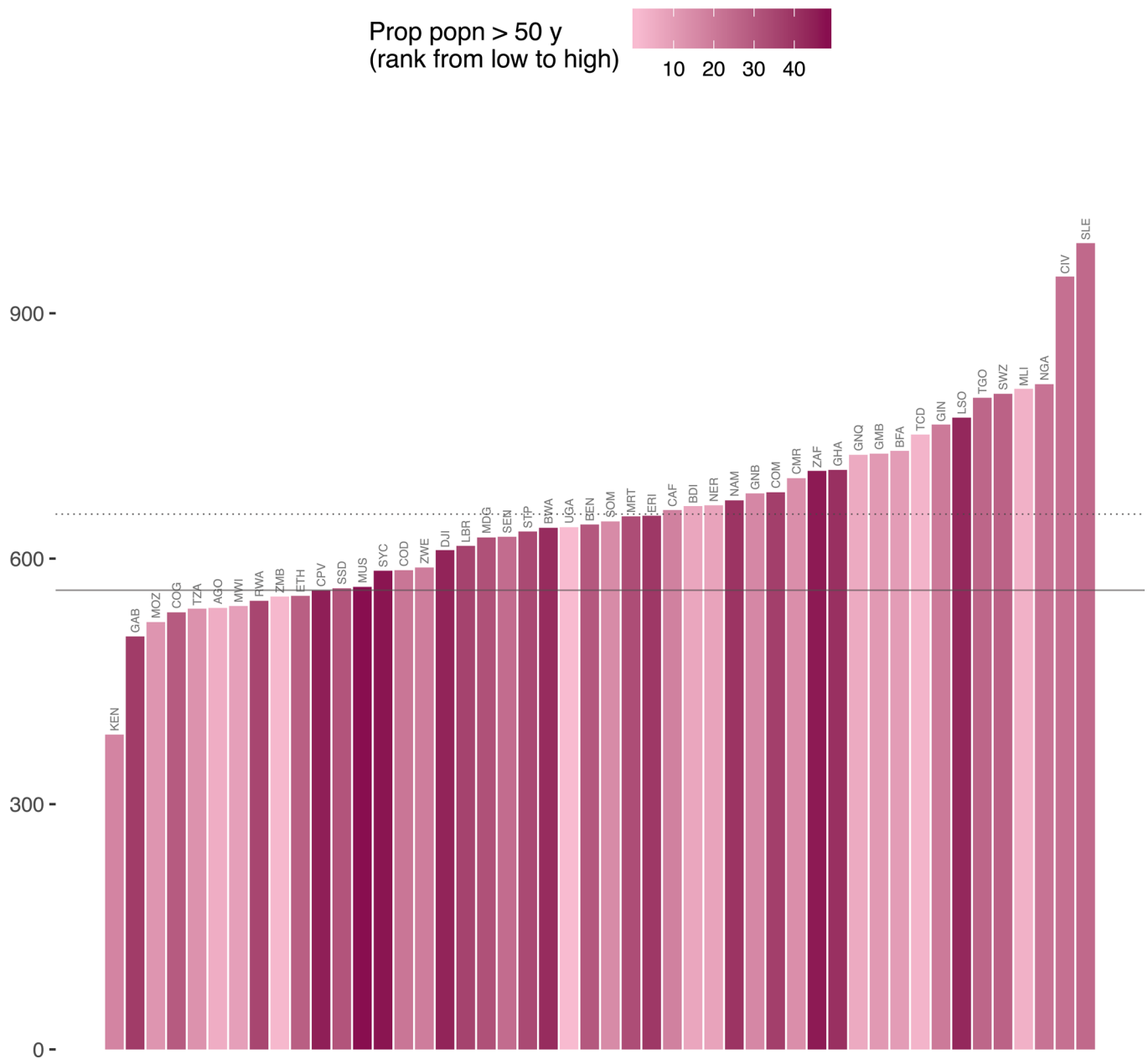
Extended Data Fig. 3 | Variation among sub-Saharan African countries in determinants of SARS-CoV-2 risk by variable. A subset of variables is shown in Fig. 3a–d in the main text. Non-communicable disease (NCD) overall mortality per 100,000 population (age standardized) is shown here as an exemplar. The remaining variables are shown online: SSA-SARS-CoV-2-tool (https://labmetcalf.shinyapps.io/covid19-burden-africa/).
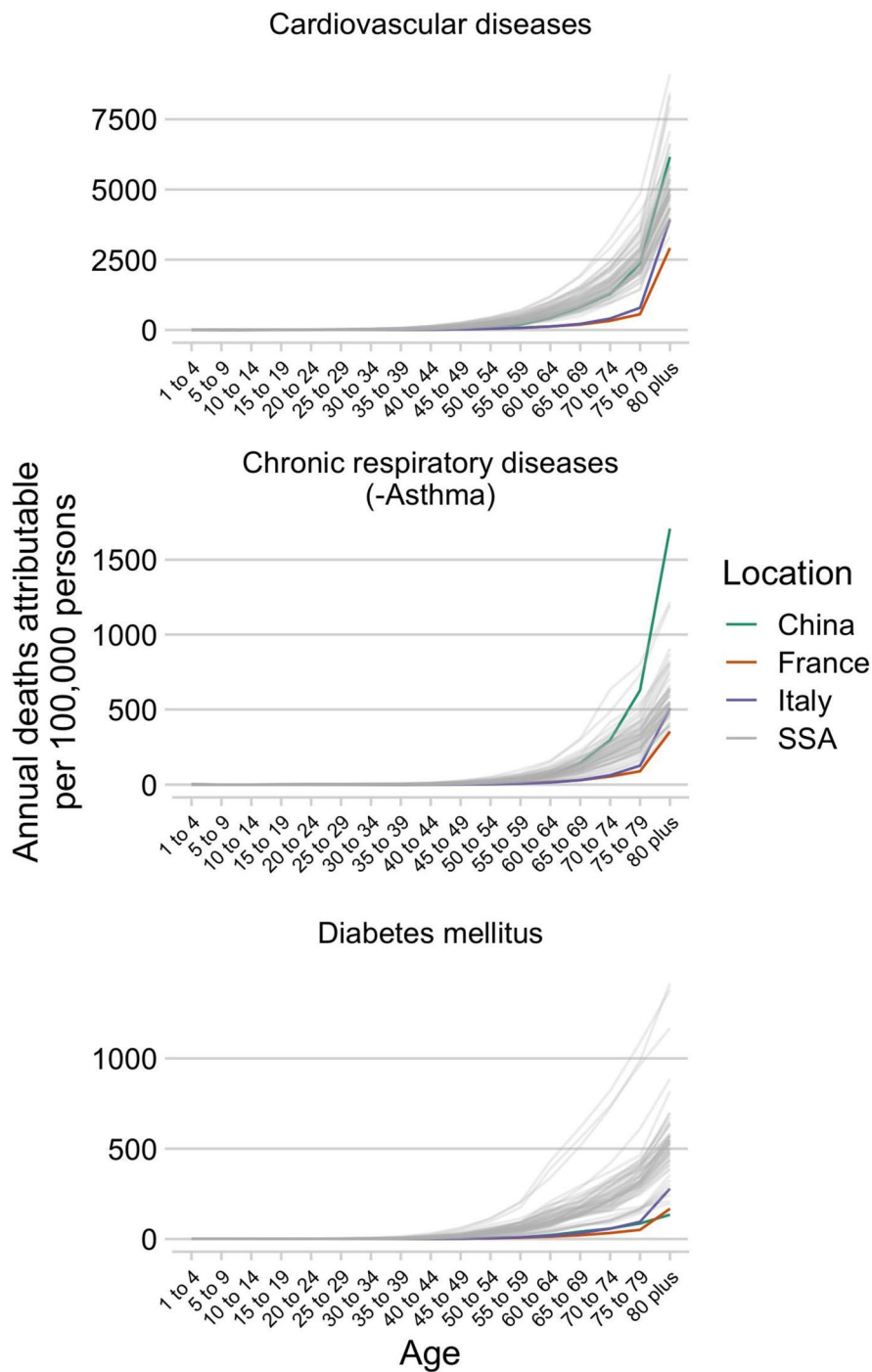
A. Comorbidities
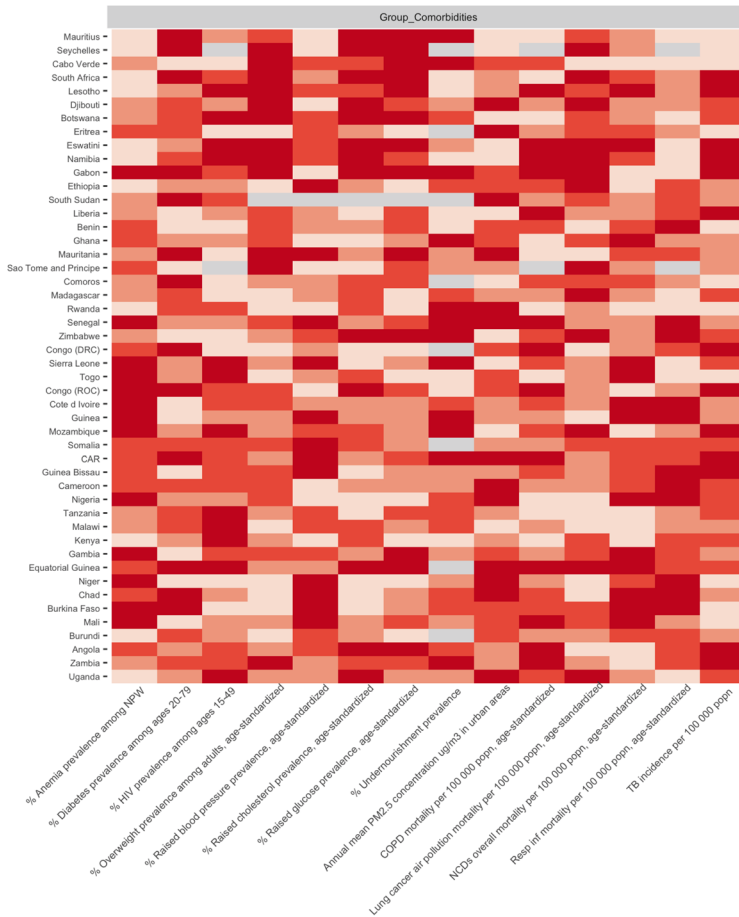
B. Access to care
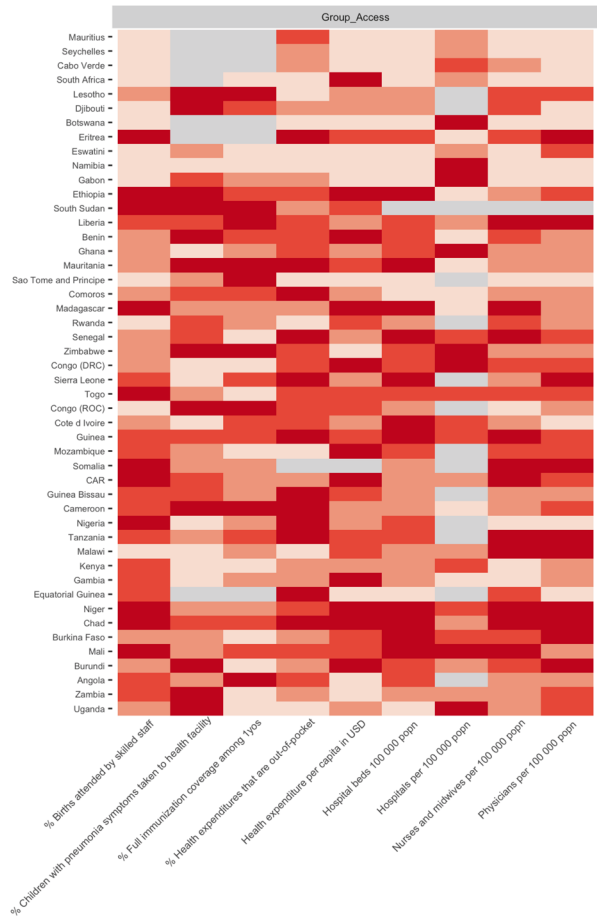
Extended Data Fig. 4 | Variation among sub-Saharan African countries in determinants of SARS-CoV-2 mortality risk by category. A subset of variables is shown in Fig. 4d,e in the main text. The remaining variables are shown online: SSA-SARS-CoV-2-tool (https://labmetcalf.shinyapps.io/covid19-burden-africa/). **a**, Select national level indicators; estimates of increased comorbidity burden (for example, higher prevalence of raised blood pressure) shown with darker red for higher risk quartiles. **b**, Select national level indicators; estimates of reduced access to care (for example, fewer hospitals) shown with darker red for higher risk quartiles. Countries missing data for an indicator (NA) are shown in gray. For comparison between countries, estimates are age-standardized where applicable (see Supplementary Table 3 for details).
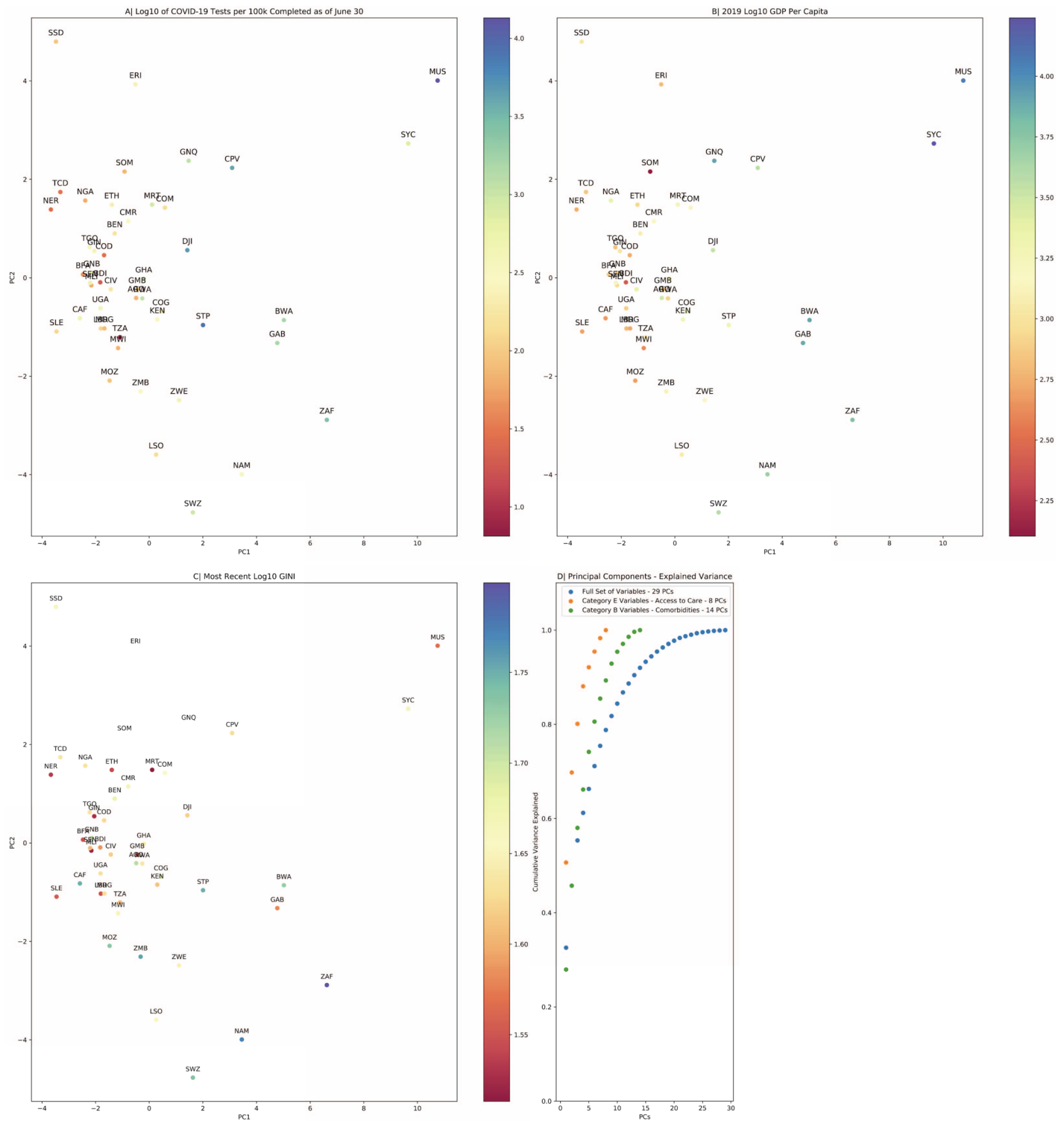
Extended Data Fig. 5 | Principal Component Analysis of all variables and category specific subsets of variables. a, Principal Component 1 and 2, countries colored by Log10 scaled tests per 100,000 population (as of June 30, 2020). **b**, Principal Component 1 and 2, countries colored by Log10 scaled GDP per capita. **c**, Principal Component 1 and 2, countries colored by the GINI index (a measure of wealth disparity). **d**, Scree plot showing the cumulative proportion of variance explained by principal component for analysis done using all variables (blue, 29 variables), comorbidity indicators (green, 14 variables, Section B in Supplementary Table 3)), and access to care indicators (orange, 8 variables, Section E in Supplementary Table 3).

Extended Data Fig. 6 | Comorbidity burden by age in sub-Saharan Africa. Estimated mortality per age group for sub-Saharan African countries (gray lines) compared to China, France, and Italy (the countries from which estimates of SARS-CoV-2 infection fatality ratios (IFRs) by age are available) for three NCD categories (cardiovascular diseases, chronic respiratory diseases excluding asthma, and diabetes).
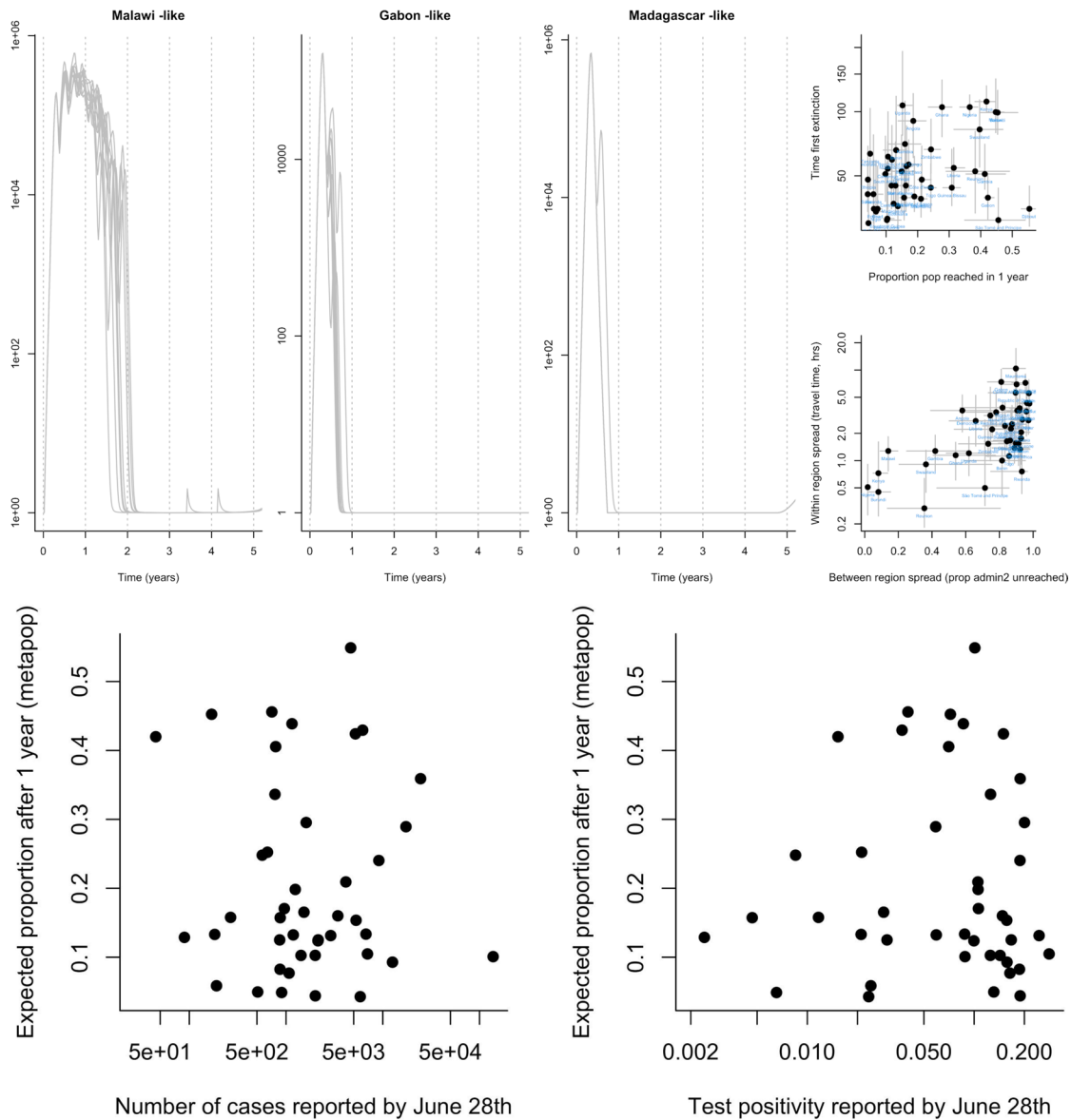
Extended Data Fig. 7 | Pace of the outbreak and cases and testing vs. the pace of the outbreak. Top: Each grey line on the left-hand panels indicates the total infected across all administrative units in a metapopulation simulation with parameters reflecting the country indicated by the plot title, assuming interventions are constant, and that immunity does not wane. Simulations with parameters reflecting three representative countries are shown, ranked from higher connectivity (Malawi-like) to lower connectivity (Madagascar-like). The top right-hand plot shows where more rapid disappearances of the outbreak locally are expected (y axis shows time to first extinction) and where a higher proportion of the countries' population is reached during simulation (x axis shows proportion of population infected by 1 year); grey horizontal bars indicate quartiles across 100 simulations. We note that a shorter duration of immunity will reduce the probability of extinction within an admin-2 (simulations shown do not include waning). The lower right-hand panel shows the fraction of administrative units unreached against the travel time in hours to the nearest city of 50,000 or more people; grey horizontal bars again reflect quartiles across 100 simulations. **Bottom**: The total number of confirmed cases reported by country (x axis, left, as reported for June 28th by Africa CDC) and the test positivity (x axis, right, defined as the total number of confirmed cases divided by the number of tests run, as reported by Africa CDC) compared with the proportion of the population estimated to be infected after one year using the metapopulation simulation described in the methods, assuming no waning of immunity (Pearson's correlation coefficients, respectively, $r = -0.04$, $p > 0.5$, $df = 41$; $r = 0.02$, $p > 0.5$, $df = 41$).
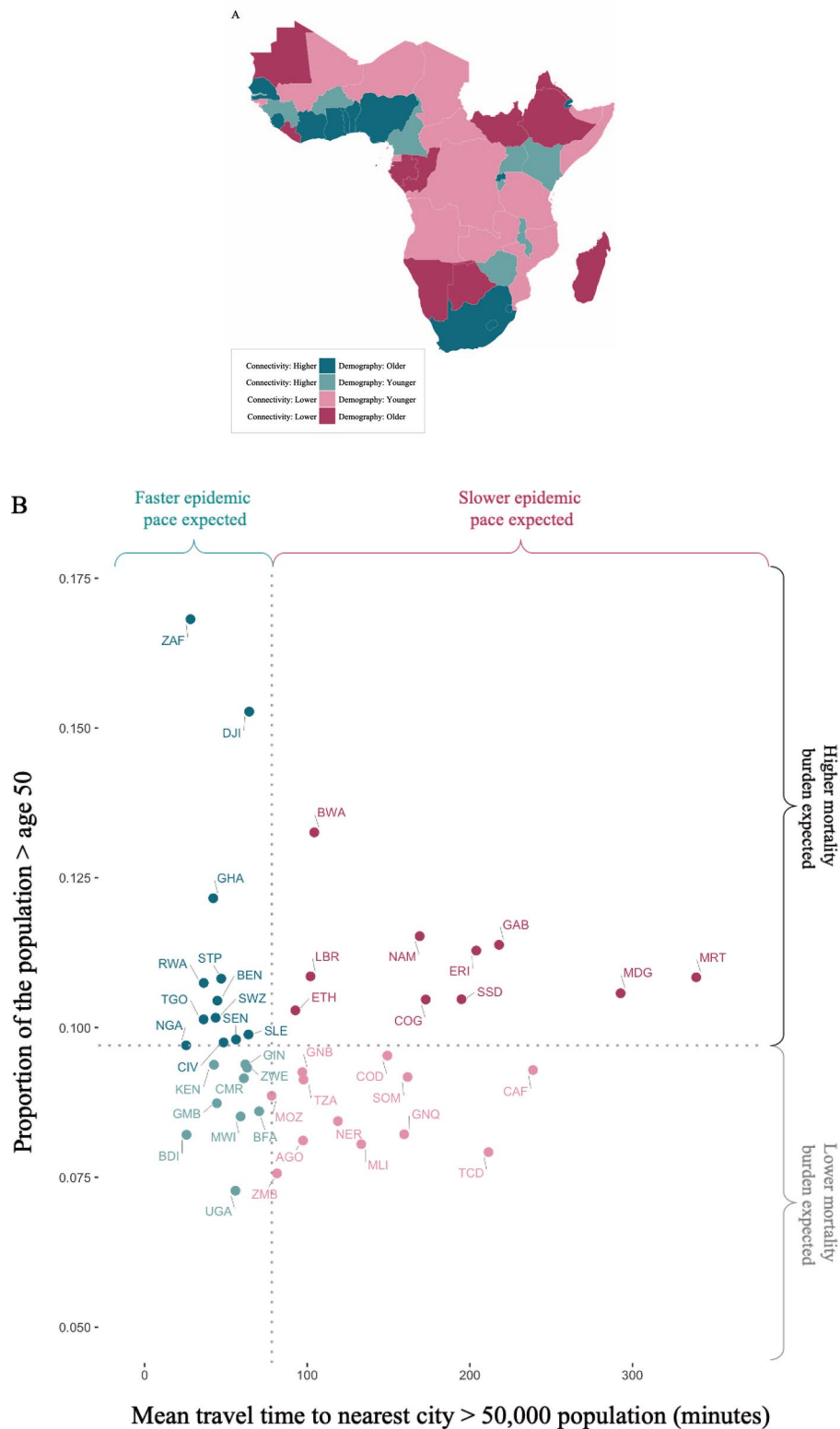
Extended Data Fig. 8 | Bivariate example of expected pace versus expected burden at the national level in SARS-CoV-2 outbreaks in sub-Saharan Africa. Countries are colored by with respect to indicators of their expected epidemic pace (using as an example subnational connectivity in terms of travel time to nearest city) and potential burden (using as an example the proportion of the population over age 50). **a**, In pink, countries with less connectivity (that is, less synchronous outbreaks) relative to the median among SSA countries; in blue, countries with more connectivity; darker colors show countries with older populations (that is, a greater proportion in higher risk age groups). **b**, Dotted lines show the median; in the upper right, in dark pink, countries are highlighted due to their increased potential risk for an outbreak to be prolonged (see metapopulation model methods) and high burden (see burden estimation methods).
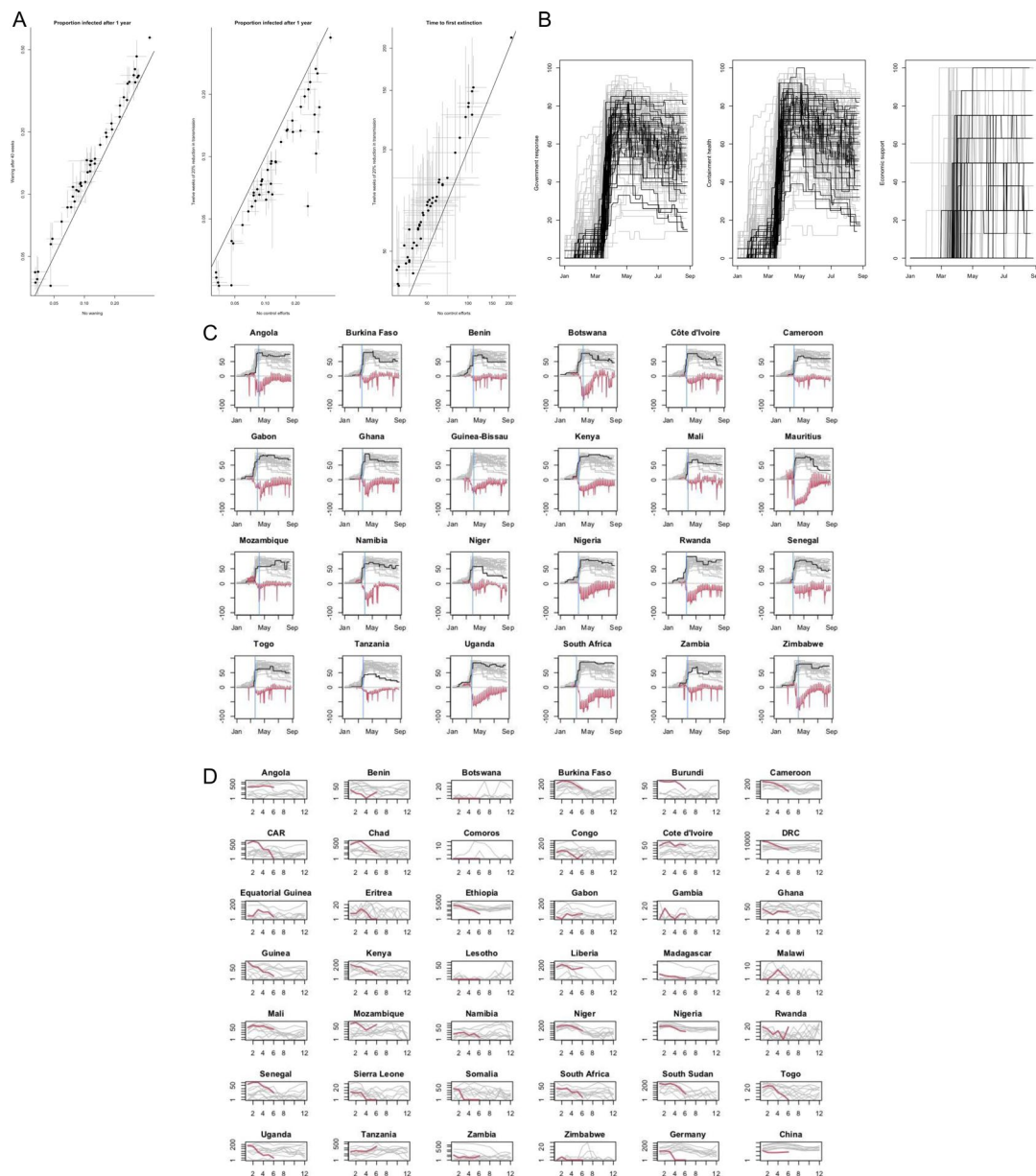
Extended Data Fig. 9 | Impact of waning of immunity and the introduction of control efforts on spatial spread. a, Impact of waning of immunity and the introduction of control efforts on spatial spread. The left panel indicates the proportion of the population infected after one year in the absence (x-axis) or presence (y-axis) of waning of immunity (duration of immunity taken to be ~40 weeks, that is, w=1/40, reflecting estimates for other coronaviruses HCoV-OC43 and HCoV-HKU1) across countries in SSA; grey horizontal lines indicate quartiles across 100 simulations. All points above the 0,1 line indicate that waning of immunity accelerates spatial spread. The central panel indicates the proportion of the population infected after one year in the absence (x axis) or presence (y axis) of control efforts with 12 weeks of a 20% reduction in transmission as an exemplar. All points below the 0,1 line indicate a lower proportion infected as a result of control efforts. All points above the 0,1 line in the right panel indicate more weeks until the first extinction in the presence of NPIs. Note that a duration of immunity of less than 40 weeks yields no local extinction. **b**, Time course of the range of policies deployed across different countries. A composite score of government response (left), interventions for containment (middle) and economic support provided (right) each scored from 0-100, provided by the University of Oxford Blavatnik School of Government; showing SSA countries (black lines) relative to other countries (grey lines). **c**, Comparison of policies implemented in SSA and google derived measures of mobility. The black line indicates a score of the magnitude of policies directed towards health containment for each country (plot title) on a scale from 0-100 with other SSA countries for which data on mobility was available (n = 24 of 48) shown for comparison in grey; the red line indicates the percent reductions in mobility to work relative to baseline[64] for that country (similar patterns seen for other mobility measures). The vertical blue line shows the day on which 10 cases were exceeded based on the Johns Hopkins dashboard data. **d**, Comparison of reductions in transmission with another directly transmitted infection. Monthly measles incidence (y-axis) between 2011 and 2019[65] is shown in gray, and the first 6 months of 2020 (months on the x-axis) shown in red for countries for which data is available in SSA (n = 34 of 48). China and Germany (which have been relatively successful in controlling the virus) shown for comparison at the bottom right. Although multi-annual features might drive measles incidence (for example, dynamics in Madagascar are largely dominated by a honeymoon outbreak that occurred in 2018-2019[66]) for countries that slowed the SARS-CoV-2 pandemic, signatures of reduction in measles can be identified (for example, Germany and China; similar patterns are seen in Viet Nam).
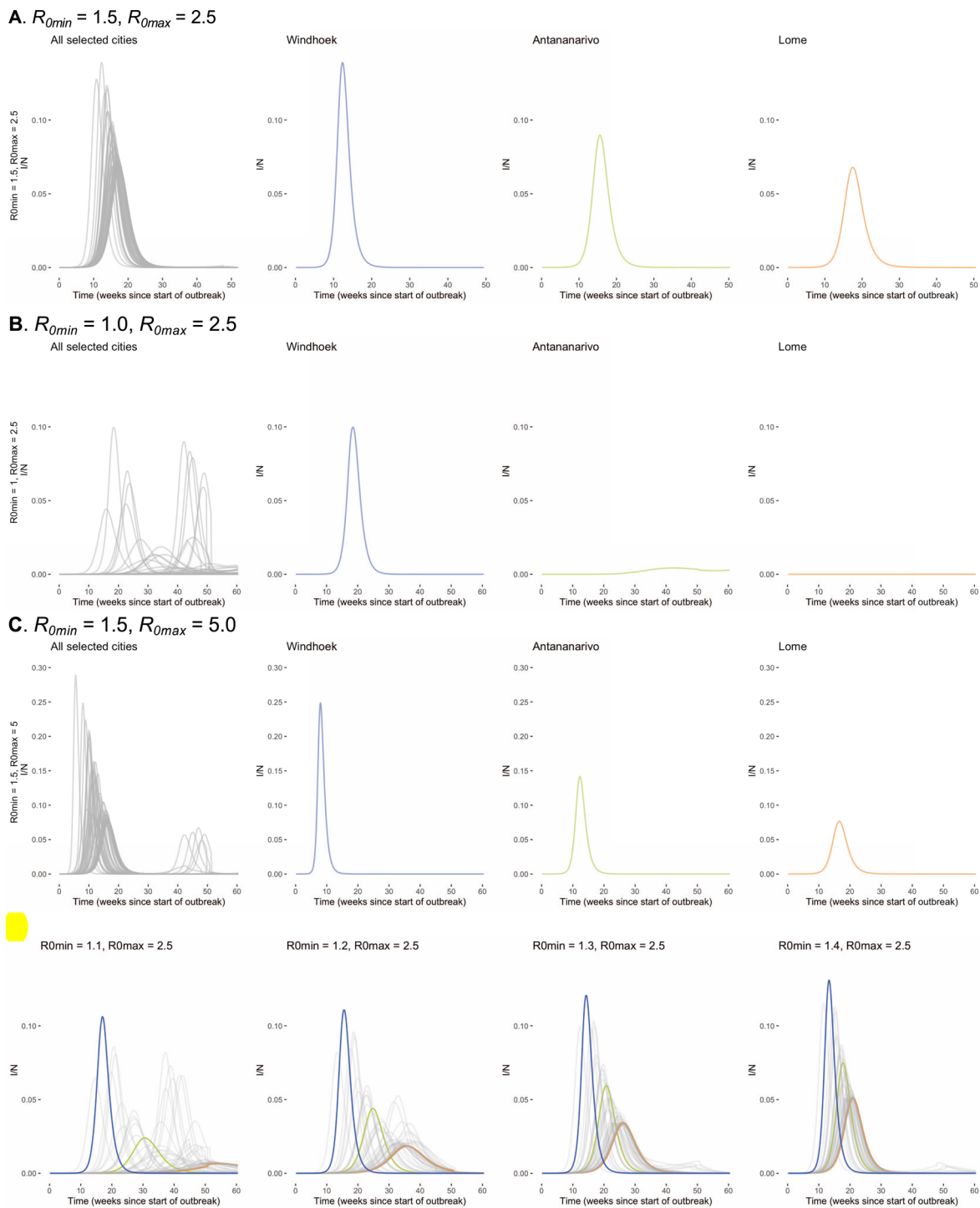
Extended Data Fig. 10 | Transmission climate-dependency and sensitivity to R_{0max} and R_{0min} value selection. Transmission (R_0) declines with increasing specific humidity from R_{0max} to R_{0min}. Three exemplar cities with low, intermediate, and high average specific humidity are shown across rows (Windhoek, Antananarivo, and Lome, respectively). **a-c**, Proportion of the population infected (I/N) over time for the specified R_{0min} and R_{0max} values. **d**, Variation in peak size and timing when $1.0 < R_{0min} < 1.5$.

nature research

Corresponding author(s): Benjamin L. Rice

Last updated by author(s): Jul 28, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	R version 4.0.2 (2020-06-22)
Data analysis	R version 4.0.2 (2020-06-22)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data have been deposited into a publicly available GitHub repository: https://github.com/labmetcalf/SSA-SARS-CoV-2

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	In this study, the target sample was the 48 countries within the geographic region of sub-Saharan Africa.
Data exclusions	No data were excluded from these analyses
Replication	Simulations and data processing code were re-run to verify the reproducibility of the findings. All attempts successfully reproduced the findings
Randomization	Randomization is not relevant to this study as all 48 countries within the geographic region of sub-Saharan Africa were considered; there were no treatment or control groups
Blinding	Blinding is not relevant to this study; all 48 countries within the geographic region of sub-Saharan Africa were considered; there were no treatment or control groups

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging