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], C. Jessica E. Metcalf[1,16]

1. Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA
2. Madagascar Health and Environmental Research (MAHERY), 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 (Université de Paris), Institut Recherche et Développement, Paris, France
7. Centre for Vaccines and Immunology (CVI), National Institute for Communicable Diseases (NICD) a division of the National Health Laboratory Service (NHLS), 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, Madagascar
10. Teaching Hospital of Care and Public Health Analakely, Antananarivo, Madagascar
11. Department of Life Science, University of Fianarantsoa, Madagascar
12. Department of Mathematics, University of Fianarantsoa, Madagascar
13. Department of Integrative Biology, University of Wisconsin-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, NJ, USA

Abstract

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 (e.g., warmer environments [1], younger populations [2-4]) have yet to be framed within a comprehensive analysis. We synthesize factors hypothesized to drive the pace and burden of this pandemic in SSA during the period from February 25 to December 20, 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 sub-Saharan Africa. 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 Sub-Saharan Africa (SSA) remains uncertain. To date, reported case counts and mortality in SSA have lagged behind other geographic 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 (**Table S1**) - totals far lower than observed in Asia, Europe, and the Americas (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 suggest 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 (e.g., variation in testing rates among countries) and correlation between surveillance and infection reporting [9] (**Extended Data Figure 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 (i.e., less than 10%) but this varies more than ten-fold between SSA countries and, critically, is sensitive to assumptions about the death reporting rate (**Figure 1A**) and infection fatality ratio (IFR, **Figure 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 infection rates could be higher in some settings, but only the latter was designed as a representative sample and serology-based estimates remain sparse in SSA.

Given limitations in inference from direct measures of infection and death rates, experience from locations in which the pandemic has progressed more rapidly provides a valuable basis of knowledge to assess the relative risk of populations in SSA and identify those remaining at greatest risk. For example, individuals in lower socio-economic 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 available in Table 1.

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 (**Figure 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 non-communicable 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 (e.g., 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 (**Table S2**), 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 prior infection by widespread circulating coronaviruses remains 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. Here, we synthesize 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 novel 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 WHO, World Bank, UNPOP, DHS, GBD, and WorldPop, and newly generated data sets (see **Extended Data Figure 2**, **Table S3** for details). We organize our assessment around two aspects that will shape national outcomes and response priorities in the event of widespread outbreaks: i) the burden, or expected severity of the outcome of an infection, which emerges from age, comorbidities, and health systems functioning, and ii) the rate of spread within a geographic area, or pace of the pandemic.

We group factors that might drive the relative rates of these two features (mortality burden and pace of the outbreak) along six dimensions of risk: (A) demographic and socio-economic parameters related to transmission and burden, (B) comorbidities relevant to burden, (C) climatic variables that may impact the magnitude and seasonality of transmission, (D) prevention measures deployed to reduce transmission, (E) accessibility and coverage of existing healthcare systems to reduce burden, and (F) patterns of human mobility relevant to transmission (**Table S2**).

National scale variability in SSA among these dimensions of risk often exceeds ranges observed across the globe (**Figure 3A-D**, **Extended Data Figure 3**). For example, estimates of

access to basic handwashing (i.e., clean water and soap [28]) among urban households in Mali, Madagascar, Tanzania, and Namibia (62-70%) exceed the global average (58%), but are less than 10% for Liberia, Lesotho, Congo DRC, and Guinea-Bissau (**Figure 3D**). Conversely, the range in the number of physicians is low in SSA, with all countries other than Mauritius below the global average (168.78 per 100,000 population) (**Figure 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 elderly in determining exposure and clinical outcomes (e.g., for variation in household size see **Figure 3E-F**). Relative ranking across variables is also uneven among countries (**Extended Data Figure 4**) with the result that this diversity cannot be easily reduced (e.g., the first two principal components explain only 32.6%, and 13.1% of the total variance as shown in **Extended Data Figure 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 consider 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 (**Figure 4A**), where density and thus transmission are likely higher [29].

Comorbidities and access to clinical care also vary across SSA (e.g., for diabetes prevalence and hospital bed capacity see **Figure 4B**). By comparison to settings where previous SARS-CoV-2 infection fatality ratio (IFR) estimates have been reported, mortality due to noncommunicable diseases in SSA increases more rapidly with age (**Extended Data Figure 6**) suggesting risk for an elevated IFR in some 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 (e.g., 33 of 48 SSA countries have reported fewer than 200 total deaths as of December 2020) and incomplete associated age data. Consequently, we use global IFR by age estimates and explore the potential effect on mortality of deviations from the expected baseline IFR in diverse SSA settings.

Small shifts (e.g., of 2-10 years of age) in the IFR profile result 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 towards higher mortality in middle-aged groups (e.g., 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 (**Figure 4C**, see 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 (**Figure 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 (**Figure 4D**) and maps poorly to indicators of comorbidity (**Figure 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 (**Figure 5A**), stratified by the SARS-CoV-2 status at the departure location on the day of travel (**Figure 5B**). A small subset of SSA countries received a disproportionately large percentage (e.g., South Africa, Ethiopia, Kenya, Nigeria together contribute 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 four countries together contributing 74.3% of all reported cases in SSA as of December 20, 2020 [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 (**Figure 3D**), population contact patterns including mobility and urban crowding [29] (e.g., **Figure**

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 (e.g., as indicated by travel time to the nearest city with a population over 50,000 **Figure 5C**). As expected, in stochastic simulations using estimates of viral transmission parameters and mobility (e.g., **Extended Data Figure 7**), a smaller cumulative proportion of the population is infected at a given time in countries with larger populations in less connected subregions (**Figure 5D**, **Extended Data Figure 8**), and including non-pharmaceutical interventions (NPIs) reduces this proportion still further. At the national level, susceptibility declines more slowly and more unevenly in such settings (e.g., Ethiopia, South Sudan, Tanzania) due to a lower probability of introductions and re-introductions of the virus locally, an effect amplified by lockdowns (**Extended Data Figure 9**). It remains unclear whether the more prolonged, asynchronous epidemics expected in these countries or the overlapping, concurrent epidemics expected in countries with higher connectivity (e.g. Malawi, Kenya, Burundi) will be a greater stress to health systems. Outbreak control efforts are 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 (e.g., malaria).

Despite extreme variation among cities in SSA (**Figure 5E**), large epidemic peaks are expected in all cities (**Figure 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 US, but robust to parameter value choice; see 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 (e.g., Windhoek and Maseru) and later in more humid, coastal cities (e.g., Bissau, Lomé, and Lagos). Reductions in transmission due to control efforts, as expected, prolong the time to epidemic peak (**Figure 5F, Extended Data Figure 10**). Apart from these slight shifts in timing, the large proportion of the population that is susceptible overwhelm the effects of climate [25], and 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 drivers of the pandemic in SSA (**Figure 3**), indicating continued variation in the burden (**Figure 4**) and pace (**Figure 5**) is to be expected even across low income settings. As small perturbations in the age profile of mortality could drastically change the national level burden in SSA (**Figure 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 (e.g., for HIV, TB, Malaria, and non-communicable 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 (e.g., USA), effective governance and management prior to reaching large case counts will likely yield the largest rewards. Generalizing across SSA is difficult as the time course and estimates of the effect of intervention policies have varied greatly (**Extended Data Figure 9**), but 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 sub-Saharan Africa. 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 Figure 7**). As the

outbreak continues to unfold, critically evaluating this mapping (e.g., **Extended Data Figure 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 United Kingdom 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 possibility of a SARS-CoV-2 vaccine, 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.

Acknowledgements

REB is supported by the Cooperative Institute for Modeling the Earth System (CIMES). AA acknowledges support from the NIH Medical Scientist Training Program 1T32GM136577. AJT is funded by the BMGF (OPP1182425, OPP1134076 and INV-002697). MB is funded by NWO Rubicon grant 019.192EN.017. We thank the Center for Community Health and Wellbeing, Princeton University for support.

Author contributions

BLR, AA, REB, MB, WWD, KM, IFM, NVM, AR, MR, JR, TR, FR, WY, BTG, CJT, and CJEM conceived the analysis. BLR, MR, MB, WWD, and WY were responsible for data curation. BLR, AA, MB, MR, REB, and CJEM performed the analyses. BLR, MR, MB, REB, CJEM, and BTG were responsible for methodology. BLR, MR, MB, REB, and WY were responsible for software and the Shiny app online tool. BLR, MR, MB, REB, and WY led data visualization. BLR and CJEM wrote the initial draft. BLR, AA, REB, MB, WWD, KM, IFM, NVM, AR, MR, JR, TR, FR, WY, BTG, CJT, and CJEM contributed to reviewing and editing the manuscript.

Competing interests

The authors declare no competing interests.

References

1. Mecenas, P., Bastos, R., Vallinoto, A. & Normando, D. Effects of temperature and humidity on the spread of COVID-19: A systematic review. doi:10.1101/2020.04.14.20064923.
2. Salje, H. *et al.* Estimating the burden of SARS-CoV-2 in France. *Science* (2020) doi:10.1126/science.abc3517.
3. Rinaldi, G. & Paradisi, M. An empirical estimate of the infection fatality rate of COVID-19 from the first Italian outbreak. doi:10.1101/2020.04.18.20070912.
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* (2021) doi:10.1126/science.abe1916.
6. Majiya, H. *et al.* Seroprevalence of COVID-19 in Niger State. doi:10.1101/2020.08.04.20168112.
7. Chibwana, M. G. *et al.* High SARS-CoV-2 seroprevalence in Health Care Workers but relatively low numbers of deaths in urban Malawi. *medRxiv* (2020) doi:10.1101/2020.07.30.20164970.
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. doi:10.1101/2020.04.01.20050203.
10. Deng, X. *et al.* Case fatality risk of novel coronavirus diseases 2019 in China. doi:10.1101/2020.03.04.20031005.
11. Collaborative, T. O. *et al.* OpenSAFELY: factors associated with COVID-19-related hospital death in the linked electronic health records of 17 million adult NHS patients. doi:10.1101/2020.05.06.20092999.
12. COVID-19 significantly impacts health services for noncommunicable diseases. 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. https://www.who.int/publications/i/item/covid-19-operational-guidance-for-maintaining-essential-health-services-during-an-outbreak.
14. Santoli, J. M. Effects of the COVID-19 Pandemic on Routine Pediatric Vaccine Ordering and Administration — United States, 2020. *MMWR Morb. Mortal. Wkly. Rep.* **69**, (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* (2020) doi:10.1126/science.abc0035.
17. Pei, S., Kandula, S. & Shaman, J. Differential Effects of Intervention Timing on COVID-19 Spread in the United States. *medRxiv* (2020) doi:10.1101/2020.05.15.20103655.
18. Lai, S. *et al.* Assessing the effect of global travel and contact reductions to mitigate the COVID-19 pandemic and resurgence. doi:10.1101/2020.06.17.20133843.
19. Nepomuceno, M. R. *et al.* Besides population age structure, health and other demographic factors can contribute to understanding the COVID-19 burden. *Proceedings of the National Academy of Sciences of the United States of America* vol. 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* (2020) doi:10.1016/S2214-109X(20)30264-3.
21. Ghisolfi, S. *et al.* Predicted COVID-19 fatality rates based on age, sex, comorbidities, and health system capacity. doi:10.1101/2020.06.05.20123489.
22. [No title]. 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. *International journal of infectious diseases: IJID: official publication of the International Society for Infectious Diseases* vol. 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.* (2020) doi:10.1093/cid/ciaa695.
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* (2020) doi:10.1126/science.abc2535.
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. *Scientific Data* vol. 4 (2017).
28. [No title]. https://washdata.org/sites/default/files/documents/reports/2019-05/JMP-2018-core-questions-for-household-surveys.pdf.
29. Korevaar, H. M. *et al.* Quantifying the impact of US state non-pharmaceutical interventions on COVID-19 transmission. doi:10.1101/2020.06.30.20142877.
30. O'Driscoll, M. *et al.* Age-specific mortality and immunity patterns of SARS-CoV-2 infection in 45 countries. *medRxiv* 2020.08.24.20180851 (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* vol. 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* (2020) doi:10.1001/jama.2020.3972.
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. medRxiv 2020.12.21.20248640v1 (2020).
39. Mina, M. J. *et al.* A Global lmmunological Observatory to meet a time of pandemics. *Elife* **9**, (2020).

Figure legends:

Figure 1

Variation in the cumulative percentage of the population infected in sub-Saharan Africa countries as expected from reported mortality totals

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 (panel **A**), or a range of infection fatality ratio (IFR) scenarios (panel **B**). The global IFR-age curves were fit to existing age-stratified IFR estimates (see methods, **Table S4**) and, for panel B, shifted towards younger or older ages by the specified number of years to simulate higher or lower IFRs, respectively. Conservatively, we assume no variation in infection rates by age (infections skewed towards 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 current as of December 2020 (sourced from Africa CDC, see **Table S1**); Data from Eritrea and the Seychelles 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.

Figure 2

Hypothesized modulators of relative SARS-CoV-2 epidemic risk in sub-Saharan Africa

Factors hypothesized to increase (red) or decrease (blue) mortality burden or epidemic pace within sub-Saharan Africa, relative to global averages, are grouped in six categories or dimensions of risk (A-F). 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 geographic spread of the virus via human mobility. SARS-CoV-2 mortality (determined by the infection fatality ratio, IFR) is modulated by demography, comorbidities (e.g., non-communicable diseases (NCDs)), and access to care. Overall burden is a function of direct burden and indirect effects due to, for example, socio-economic disruptions and disruptions in health services such as vaccination and infectious disease control. **Table S2** contains details and the references used as a basis to draw the hypothesized modulating pathways.

Figure 3

Variation among sub-Saharan African countries in select determinants of SARS-CoV-2 risk

A-D: At right, SSA countries are ranked from least to greatest for each indicator; bar color shows population age structure (% of the population above age 50). Solid horizontal lines show the global mean value; dotted lines show the mean among SSA countries. At left, boxplots show median, inner bounds corresponding to the interquartile range (IQR, 25th to 75th percentiles), and outer bounds corresponding to 1.5 * IQR, grouped by geographic region. Per WHO: sub-Saharan Africa (SSA); Americas Region (AMR); Eastern Mediterranean Region (EMR); Europe Region (EUR); Southeast Asia Region (SEA); Western Pacific Region (WPR) (*n* = 206, 172, 106, and 92 countries with available data for A-D, respectively). **E-F:** Dot size shows mean household (HH) size for HHs with individuals over age 50; dashed lines show median value among SSA countries; quadrants of greatest risk are outlined in red (e.g., fewer physicians and greater age standardized Chronic Obstructive Pulmonary Disease (COPD) mortality). See **Table S3** and **Extended Data Figure 3** for a full description and link to visualization of all variables.

Figure 4

Variation in expected burden for SARS-CoV-2 outbreaks in sub-Saharan Africa

A: Expected mortality in a scenario where cumulative infection reaches 20% across age groups and the infection fatality ratio (IFR) curve is fit to existing age-stratified IFR estimates (see methods, **Table S4**). **B**: National level variation in comorbidity and access to care variables, for e.g., diabetes prevalence among adults and the number of hospital beds per 100,000 population for sub-Saharan African countries. **C**: The range in mortality per 100,000 population expected in scenarios where 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. **D-E**: Select national level indicators; estimates of reduced access to care (e.g., fewer hospitals) or increased comorbidity burden (e.g., higher prevalence of raised blood pressure) shown with darker red for higher risk quartiles (see **Extended Data Figure 4** for all indicators). Countries missing data for an indicator (NA) are shown in gray. For comparison between countries, estimates are age-standardized where applicable (see **Table S3** for details). High resolution maps for each variable and scenario available at the SSA-SARS-CoV-2-tool (https://labmetcalf.shinyapps.io/covid19-burden-africa/).

Figure 5

Variation in connectivity and climate in sub-Saharan Africa and expected effects on SARS-CoV-2

A: International travellers to sub-Saharan Africa (SSA) from January to April 2020, as inferred from the number of passenger seats on arriving aircraft. **B**: For the four countries with the most arrivals, the proportion of arrivals by month coming from countries with 0, 1-100, 101-1000, and 1000+ reported SARS-CoV-2 infections at the time of travel (see **Table S5** for all others). **C**: Connectivity within SSA countries as inferred from average population weighted mean travel time to the nearest urban area greater with a population greater than 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 (see methods). **E**: Climate variation across SSA as shown by seasonal range in specific humidity, *q* (g/kg) (max average *q* - min average *q*). **F:** The effect of local seasonality and control efforts (R_0 decreases by 0%, i.e., unmitigated, 10%, or 20%) on the timing of epidemic peaks (max *I/N*) in SSA cities (with three exemplar cities highlighted in pink, see methods).

480

481 **Table 1: Policy Summary**

Background	As the SARS-CoV-2 pandemic expanded globally, reported incidence and mortality remained low in Sub-Saharan Africa (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 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 which 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 infection fatality ratio 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	To narrow the range of expectations for country trajectories in SSA for incidence and mortality, strengthening surveillance and registration is necessary. Additional tools for surveillance such as serology, or approaches that quantify excess mortality, will provide important complementary measures. Our national and sub-national 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.

482

483

Methods

Reported SARS-CoV-2 case counts, mortality, and testing in sub-Saharan Africa as of December 2020

Variables and data sources for testing data

The numbers of reported cases, deaths, and tests for the 48 studied sub-Saharan Africa (SSA) countries (**Table S1**) were sourced from the Africa Centers for Disease Control (CDC) dashboard on December 20, 2020 (and previously on September 23 and June 30, 2020) (https://africacdc.org/covid-19/). 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 recency of data within the centralized Africa CDC repository, but the 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 (not including Sudan). From the Northern Africa region, Mauritania is included in SSA.

For comparison to non-SSA countries, the number of reported cases in other geographic regions were obtained from the Johns Hopkins University Coronavirus Resource Center on September 23, 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/). As reported confirmed cases are likely to be a significant underestimate of the true number of infections, CFRs may be a poor proxy for the infection fatality ratio (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 June 30 and remain highly variable as of September 23 and December 20, 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 Congo (DRC) to 18,321.83 in Mauritius in September 2020; and from 100.9 in Congo (DRC) to 23695.0 in Mauritius in December 2020 (**Extended Data Figure 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 (i.e., 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 Figure 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 one reported death, 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 Figure 1C**). Limitations in the ascertainment of infection rates and the rarity of reported deaths (e.g., 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 were used for the subsequent analyses in this study.

Synthesizing factors hypothesized to increase or decrease SARS-CoV-2 epidemic risk in SSA

Variable selection and data sources for variables hypothesized to associate 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 (**Table S2**). 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 respiratory effects of pollution [43]. Specific to Africa, vulnerability scores based on these hypothesized associations or combinations of risks factors have been developed (e.g., [44,45]).

Many possible sources of bias complicate interpretation of these associations [46], and while they provide a useful baseline, inference is also likely to change as the pandemic advances. To reflect this, our analysis combines a number of high level variables likely to broadly encompass these putative risk factors (e.g., non-communicable disease (NCD) related mortality and health life expectancy) with more specific measures encompassed in evidence to date (e.g., prevalence of diabetes, obesity, and respiratory illness such as chronic obstructive pulmonary disease (COPD)). We also include 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, barring with 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 (e.g., for malaria [49]), treatment [50], or prevention programs [51].

Data on the identified indicators were sourced in May 2020 from the World Health Organization (WHO) Global Health Observatory (GHO) database (https://www.who.int/data/gho), World Bank (https://data.worldbank.org/), and other sources detailed in **Table S3**. National level demographic data (population size and age structure) was sourced from United Nations World Population Prospects (UNPOP) (https://population.un.org/wpp/Download/Standard/Population/) 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 one person aged > 50 years, taken from the most recently available demographic health survey (DHS) 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 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 geographic areas we compared the year of most recent data for variables (**Extended Data Figure 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 uncertainty associated in the estimates available for a single variable, we also include multiple variables per category (e.g., demographic and socio-economic factors, comorbidities, access to care) to avoid reliance on a single metric. This allows 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 are likely to be correlated (see PCA methods below for further discussion) would bias inference of cumulative risk in a statistical framework, we do not attempt to quantitatively combine risk across variables for a country, nor project risk based on the variables included here. Rather, we characterize the magnitude of variation among countries for these variables (see **Figure 3** in the main text for a subset of the variables; **Figure 4B** for bivariate risk maps following [52]) and then explore the range of outcomes that would be expected under scenarios where IFR increases with age at different rates (see **Figure 4** in the main text).

Variable selection and data sources for variables hypothesized to modulate the rate of viral spread

In addition to characterizing variation among factors likely to modulate burden, we also synthesize 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 geographic spread include climatic factors (e.g., specific humidity), access to prevention measures (e.g., handwashing), and human mobility (e.g., international and domestic travel). **Table S2** outlines

the dimensions of risk selected and references the previous studies relevant to the selection of these factors.

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

International flight data was obtained from a custom report from OAG Aviation Worldwide (UK) and included the departure location, airport of arrival, 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 sub-Saharan Africa. To obtain a set of measures that broadly represent connectivity within different countries in the region, friction surfaces from [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 in the subnational connectivity section.

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

Principal component analysis of variables considered

Selection of data and variables

The 29 national level variables from **Table S3** were selected for principal component analysis (PCA). We conducted further PCA on the subset of eight indicators related to access to healthcare (Category E) and the 14 national indicators variables related to comorbidities (Category B).

We excluded disaggregated sub-national spatial variation data (variables A2, C1, E2, and category F), disaggregated or redundant variables derived from already included variables (variables A4 and D2), and disaggregated age-specific disease data from IHME global burden of disease study (variables B2, B4, and B13) from PCA analysis. COVID-19 tests per 100,000 population (variable D4, **Table S1**), per capita gross domestic product (GDP) (variable A8), and the GINI index of wealth inequality (variable A9) were used to visualize patterns among sub-Saharan Africa 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 as zero values (i.e., outside the range of typical values seen in the data) inflate the total variance in the data set and thus, if anything, deflate the percent of the variance explained by PCA. Therefore, this approach avoids mistakenly attributing predictive value to principal components due to incomplete data. See **Table S3** for data sources for each variable.

Principal Component Analysis

The PCA was conducted on each of the three subsets described above, using the scikitlearn library [70]. In order 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 prior to the analysis. Briefly, PCA applies a linear transformation to a set of *n* features to output a set of *n* orthogonal principal components which are uncorrelated and each explain a percentage of the total variance in the dataset [71]. A link to the code for this analysis is available online at the https://labmetcalf.shinyapps.io/covid19-burden-africa/.

The principal components were then analyzed for the percentage of variance explained, and compared to: (i) the number of COVID-19 tests per 100,000 population as of the end of June, 2020 (**Table S1**), (ii) the per capita GDP, and (iii) the GINI index of wealth inequality. For the GINI index, estimates from 2008-2018 were available for 45 of the 48 countries (no GINI index data were available for Eritrea, Equatorial Guinea, and Somalia) (see **Data File 1** for the year for each country for each metric).

The first two 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 (Pearson correlation coefficient, r = 0.67, p = 1.1e-7, **Extended Data Figure 5A**). Similarly, high GDP countries seem to associate with an increase in principal component 1 (Pearson correlation coefficient, r = 0.80, p = 6.02e-12), **Extended Data Figure 5B**). In contrast, countries with greater wealth inequality (as measured by the GINI index) are associated with a decrease in principal component 2 (Pearson correlation coefficient, r = -0.42, p = 0.0042, **Extended Data Figure 5C**). Despite these correlations, a relatively low percentage of variance is explained by each principal component: for the 29 variables, 13 of the 29 principal components are required to explain 90% of the variance (**Extended Data Figure 5D**). When only the access to care subset of variables is considered, the first two principal components explain 50.7% and 19.1% of the variance, respectively, and five of eight 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 nine of 14 principal components are required to explain 90% of the variance (**Extended Data Figure 5D**).

These data suggest that inter-country variation in this dataset is not easily explained by a small number of variables. Moreover, though 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 as, 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 (indeed $r = 0.60$, $p = 5.7e\text{-}6$ 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 infection fatality ratio (IFR) that account for asymptomatic cases, underreporting, and delays in reporting are few, however, it is evident that IFR increases substantially with age [72]. We use age-stratified estimates of IFR from three studies (two published [2,4], one preprint [3]) that accounted for these factors in their estimation (**Table S4**).

To apply these estimates to other age-stratified data with different bin ranges and generate continuous predictions of IFR with age, we fit the relationship between the midpoint of the age bracket and the IFR estimate using a generalized additive model (GAM) using the 'mgcv` package [73] in R version 4.0.2 [74]. We use a beta distribution as the link function for IFR estimates (data distributed on [0, 1]). For the upper age bracket (80+ years), we take the upper range to be 100 years and the midpoint to be 90.

We assume a given level of cumulative infection (here 20% in each age class, i.e., a constant rate of infection among age classes) and then apply IFR*s* by age to the population structure of each country to generate estimates of burden. Age structure estimates were taken from the UNPOP (see **Table S3**) country level estimates of population in 1 year age groups (0 - 100 years of age) to generate estimates of burden.

Comorbidities over age from 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 sub-Saharan Africa will be 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 (i.e., IFR); to assess the relative risk of comorbidities across age in SSA,

743 estimates of comorbidity severity by age (in terms of annual deaths attributable) were obtained
744 from the Institute for Health Metrics and Evaluation (IHME) Global Burden of Disease (GBD)
745 study in 2017 [75]. Data were accessed through the GBD results tool for cardiovascular disease,
746 chronic respiratory disease (not including asthma), and diabetes, reflecting three categories of
747 comorbidity with demonstrated associations with risk (**Table S2**). We make the assumption that
748 higher mortality rates due to these NCDs, especially among younger age groups, is indicative of
749 increased severity and lesser access to sufficient care for these diseases - suggesting an
750 elevated risk for their interaction with SARS-CoV-2 as comorbidities. While there are significant
751 uncertainties in these data, they provide the best estimates of age specific risks and have been
752 used previously to estimate populations at risk [20].
753
754 The comorbidity by age curves for SSA countries were compared to those for the three
755 countries from which SARS-CoV-2 IFR by age estimates were sourced. Attributable mortality
756 due to all three NCD categories is higher at age 50 in all 48 SSA countries when compared to
757 estimates from France and Italy and for 42 of 48 SSA countries when compared to China
758 (**Extended Data Figure 6**).
759
760 Given the potential for populations in SSA to experience a differing burden of SARS-CoV-2 due
761 to their increased severity of comorbidities in younger age groups, we explore the effects of
762 shifting IFRs estimated by the GAM of IFR estimates from France, Italy, and China younger by
763 2, 5, and 10 years (**Figure 3**).
764

765 ## International air travel to SSA

766
767 The number of passenger seats on flights arriving to international airports were grouped by
768 country and month for January 2020 to April 2020 (**Table S5**) - the months when the
769 introduction of SARS-CoV-2 to SSA countries was likely to have first occurred. The first
770 confirmed case reported from a SSA country, per the Johns Hopkins Coronavirus Research
771 Center was in Nigeria on February 28, 2020. By March 31, 2020, 43 of 48 SSA countries had
772 reported SARS-CoV-2 infections and international travel was largely restricted by April. Lesotho
773 was the last SSA country to report a confirmed SARS-CoV-2 infection (on May 13, 2020);
774 however, given difficulties in surveillance, the first reported detections were likely delayed
775 relative to the first importations of the virus.
776
777 The probability of importation of the virus is defined by the number of travelers from each source
778 location each date and the probability that a traveler from that source location on that date was
779 infectious. Due to limitations in surveillance, especially early in the SARS-CoV-2 pandemic,
780 empirical data on infection rates among travelers is largely lacking. To account for differences in
781 the status of the SARS-CoV-2 pandemic across source locations, and thus differences in the
782 importation risk for travelers from those locations, we coarsely stratified travelers arriving each
783 day into four categories based on the status of their source countries:
784

i. Travelers from countries with zero reported cases (i.e., although undetected transmission was possibly occurring, SARS-CoV-2 had not yet been confirmed in the source country by that date)
ii. Those traveling from countries with more than one reported case (i.e., SARS-CoV-2 had been confirmed to be present in that source country by that date),
iii. Those traveling from countries with more than 100 reported cases (indicating community transmission was likely beginning), and
iv. Those traveling from countries with more than 1000 reported cases (indicating widespread transmission)

For determining reported case counts at source locations for travelers, no cases were reported outside of China until January 13, 2020 (the date of the first reported case in Thailand). Over January 13 to January 21, 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 [76] beginning January 22 (https://coronavirus.jhu.edu/map.html); we use that data from January 22 onwards and the WHO reports prior to January 22.

The number of travelers within each category arriving per month is shown in **Table S5**. 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 (i.e., travelers are not more likely to be exposed than non-travelers in that source location) and does not account for complex travel itineraries (i.e., 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, as the relative risk for viral importation will still scale with the number of travelers, comparisons among SSA countries can be informative (e.g., 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 sub-Saharan Africa

Indicators of subnational connectivity

To allow comparison of the relative connectivity across countries, we use 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 (e.g., transport from a city to the rural periphery), we use as an indicator the population weighted mean travel time to the nearest urban center (i.e., population density > 1,500 per square kilometer or a density of built-up areas > 50% coincident with population > 50,000) within administrative-2 units [69]. 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 will be governed by subnational mobility: the pathogen is more likely to be introduced into a location where individuals arrive more frequently than one where incoming travellers are less frequent. Large-scale consistent measures of mobility remain 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 as 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 use 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 in R to the friction surfaces supplied in [26]). With this, we develop a metapopulation model for each country to develop an overview of the possible range of trajectories of unchecked spread of SARS-CoV-2.

We assume that the pathogen first arrives into each country in the administrative 2 level unit with the largest population (e.g., 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 track 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 ~ 1 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:

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

where β captures the magnitude of transmission over the course of one discrete time-step, and 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 [77], and $\iota_{j,t}$ 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 Bank (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, as this is assumed to be the location of introduction; for this spatial unit, we set $I_{j,1} = 1$.

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

$$\iota_{j,t} \sim Bern(1 - exp(-\sum_{1}^{L} c_{i,j} I_{i,t}/N_i))$$

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 (e.g., a country with parameters that broadly reflect those available for Malawi, **Extended Data Figure 7**), while in others (e.g., 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 5 years and the time to the first local extinction of the pathogen (**Extended Data Figure 7**, right top). In countries with lower connectivity (e.g., that might resemble Madagascar), local outbreaks can go extinct rapidly before travelling very far; in other countries (e.g., that might resemble 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, 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 5 years are also those where within administrative unit travel is low (**Extended Data Figure 7**, right bottom).

These simulations provide a window onto 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, further complicated by 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 here. Variability in case reporting complicates clarifying this (**Extended Data Figure 7**, lower) but we highlight countries with less connectivity (i.e., less synchronous outbreaks expected) relative to the

908 median among SSA countries and with older populations (i.e., a greater proportion in higher risk
909 age groups) (**Extended Data Figure 8**).
910
911 The University of Oxford Blavatnik School of Government generated composite scores of
912 government response, interventions for containment, and economic support provided, with each
913 scored from 0-100 (Coronavirus Government Response Tracker
914 https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-
915 tracker). These data were compared with the day on which 10 cases were exceeded in a
916 country per the Johns Hopkins dashboard data (Johns Hopkins Coronavirus Resource Center
917 https://coronavirus.jhu.edu/map.html.).
918
919 While faster waning of immunity will act to increase the rate of spread of the infection, resulting
920 in a higher proportion infected after one year, control efforts will generally act to slow the rate of
921 spread of the infection (**Extended Data Figure 9**). As different countries are likely to have
922 differently effective control efforts (**Extended Data Figure 9**), this precludes making country
923 specific predictions as to the relative impact of control efforts on delay.
924

925 Modeling epidemic trajectories in scenarios where transmission rate 926 depends on climate

927
928 *Climate data sourcing: Variation in humidity in SSA*
929
930 Specific humidity data for selected urban centers comes from ERA5 using an average
931 climatology (1981-2017) [53]; we do not consider year-to-year climate variations. Selected cities (*n*
932 = 56) were chosen to represent the major urban areas in SSA. The largest city in each SSA
933 country was included as well as any additional cities that were among the 25 largest cities or
934 busiest airports in SSA.
935
936 *Methods for climate driven modelling of SARS-CoV-2*
937
938 We use a climate-driven SIRS (Susceptible-Infected-Recovered-Susceptible) model to estimate
939 epidemic trajectories (i.e., the time of peak incidence) in different cities in 2020, assuming no
940 control measures are in place or a 10% or 20% reduction in R_0 beginning 2 weeks after the total
941 reported cases for a country exceed 10 cases [25,81]. The model is given by:
942

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

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

943
944 where *S* is the susceptible population, *I* is the infected population and *N* is the total population.
945 *D* is the mean infectious period, set at 5 days following [25,55].
946

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 use parameters from the most climate-dependent scenario in 2^5, based on the endemic betacoronavirus HKU1 in the USA. In this scenario *L*, the duration of immunity, is found to be 66.25 weeks (i.e., greater than 1 year and such that waning immunity does not affect timing of the epidemic peak). We initially select a range where R_0 declines from R_{0max} = 2.5 to R_{0min} = 1.5 (i.e., transmission declines 40% at high humidity) as this exceeds the range observed for influenza and other coronaviruses for which there is data available (from the USA). R_{0max} = 2.5 is chosen as 2.5 is oft cited as the approximate R_0 for SARS-CoV-2. Thus, we initially assume that the climate dependence of SARS-CoV-2 in SSA will not greatly exceed that of other known coronaviruses from the US context. Then we explore the effects of different degrees of climate dependency (i.e., wider ranges between R_{0max} = 2.5 to R_{0min} = 1.5 and scenarios where R_{0min} approaches 1) (**Extended Data Figure 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 the climate and is related to specific humidity according to the equation:

$$R_0 = exp(a * q(t) + log(R_{0max} - R_{0min})) + R_{0min}$$

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

Sensitivity analysis

Selecting an R_{0min} value of 1, such that epidemic growth stops at high humidities is likely implausible as simulations indicate no outbreaks would occur in cities such as Antananarivo (countered by the observation that SARS-CoV-2 outbreaks did in fact occur) (**Extended Data Figure 10B**) (see **Table S1** for reported case counts at the country level). Expanding the range between R_{0min} and R_{0max} by increasing R_{0max} results in epidemic peaks being reached earlier after outbreak onset, but does not increase the difference in timing between cities with different climates (**Extended Data Figure 10C**; e.g., the difference in timing between peaks in Windhoek and Lome is similar in panels **A** and **C**). Finally, we explored scenarios where R_{0min} was between 1.0 and 1.5. When R_{0min} > 1.1, epidemic peaks are seen in each SSA city with the difference in timing of the peak growing larger when smaller values of R_{0min} are selected (**Extended Data Figure 10D**). However, the difference in timing, even when small values of R_{0min} are selected, is a maximum of 25 weeks (i.e., a shorter time period than the interval between present time (October 2020) and the beginning of outbreaks (approx. March 2020)) and rapidly reduces to only a few weeks as R_{0min} approaches 1.5.

Data Availability

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

Code Availability

All code has been deposited into the publicly available GitHub repository (same as above):
Link to GitHub repository containing code: https://github.com/labmetcalf/SSA-SARS-CoV-2

Methods Only References

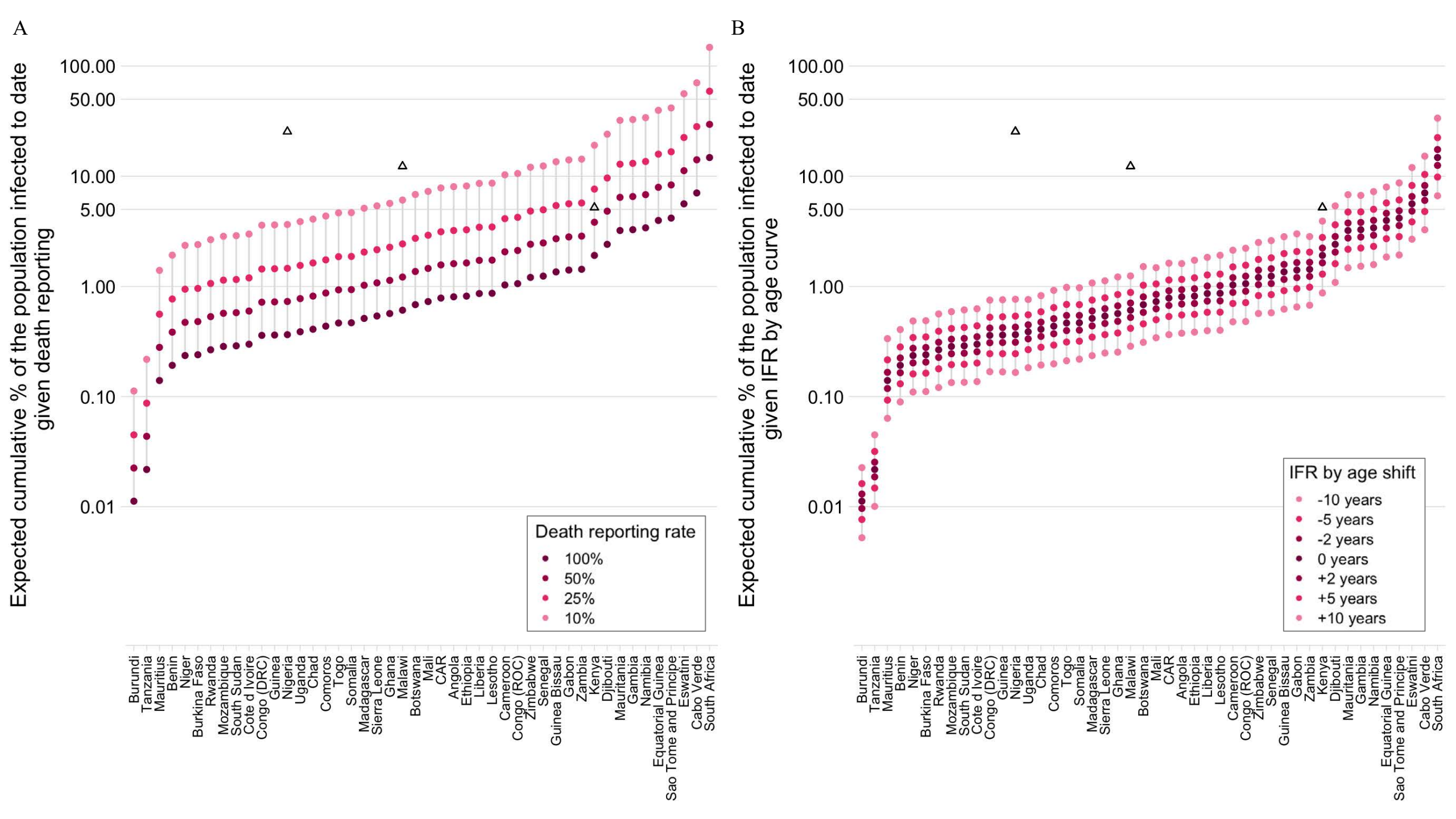
40. The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. The Epidemiological Characteristics of an Outbreak of 2019 Novel Coronavirus Diseases (COVID-19) — China, 2020. *CCDCW* **2**, 113–122 (2020).
41. Team, C. C.-19 R. *et al.* Severe Outcomes Among Patients with Coronavirus Disease 2019 (COVID-19) — United States, February 12–March 16, 2020. *MMWR. Morbidity and Mortality Weekly Report* vol. 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* (2020) doi:10.1002/oby.22831.
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 COVID-19 Risk Factors – Africa Center for Strategic Studies. https://africacenter.org/spotlight/mapping-risk-factors-spread-covid-19-africa/.
45. Bringing Greater Precision to the COVID-19 Response. https://precisionforcovid.org/africa.
46. Griffith, G. *et al.* Collider bias undermines our understanding of COVID-19 disease risk and severity. doi:10.1101/2020.05.04.20090506.
47. Shankar, A. H. Nutritional Modulation of Malaria Morbidity and Mortality. *The Journal of Infectious Diseases* vol. 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 Medicine* vol. 21 567–577 (2020).
49. Chanda-Kapata, P., Kapata, N. & Zumla, A. COVID-19 and malaria: A symptom screening

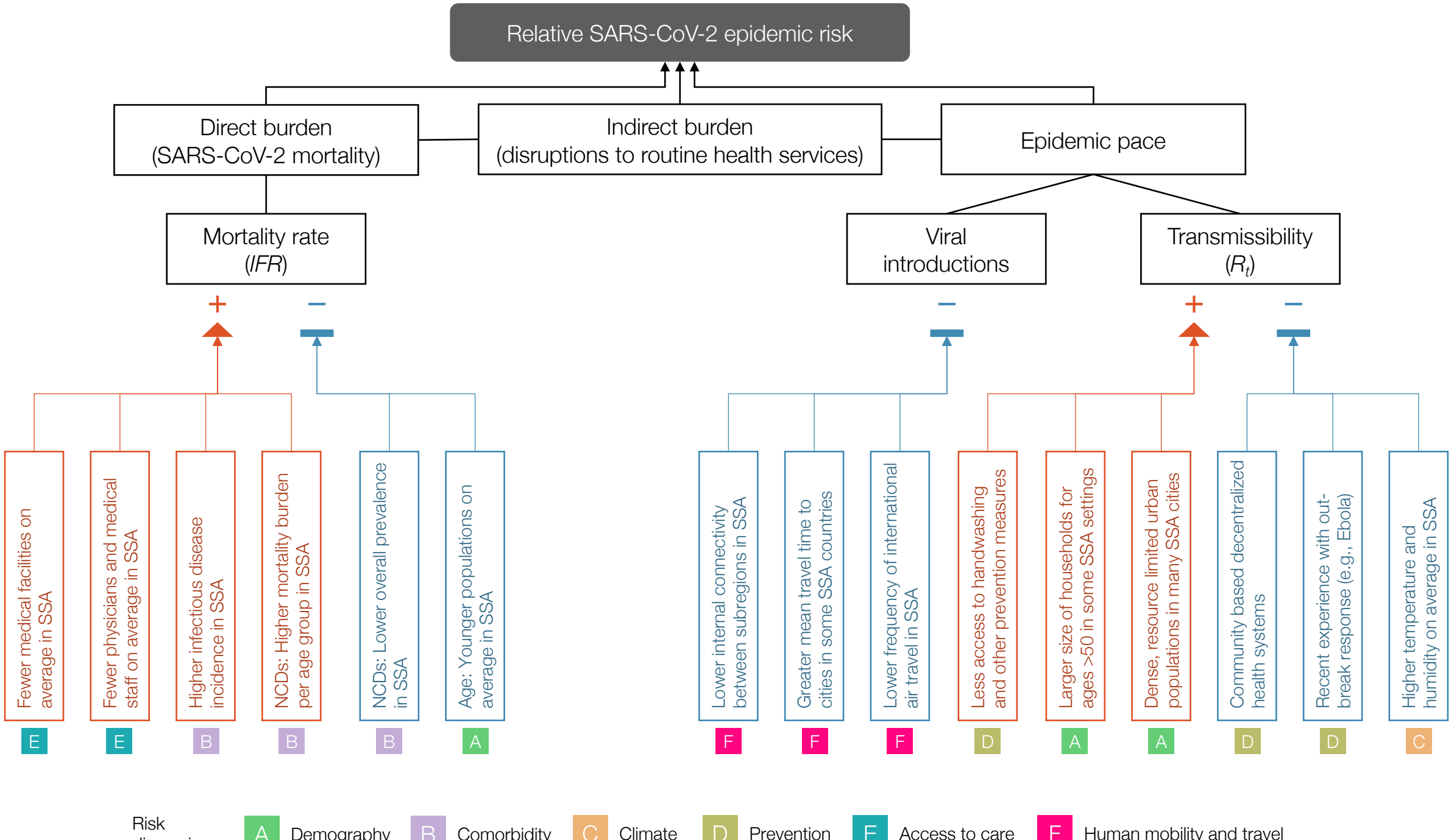
challenge for malaria endemic countries. *Int. J. Infect. Dis.* **94**, 151–153 (2020).

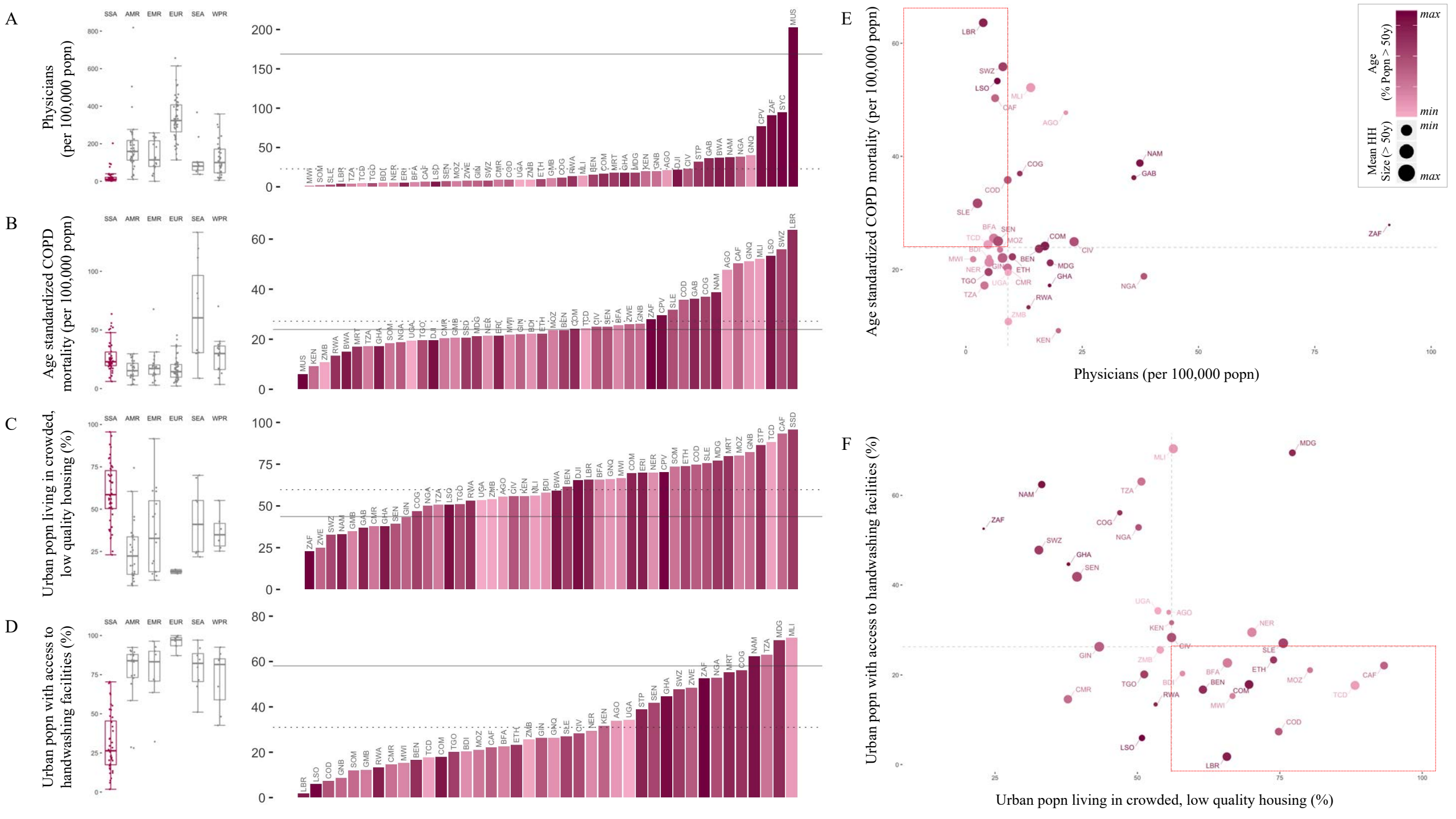
50. Lesosky, M. & Myer, L. Modelling the impact of COVID-19 on HIV. *The lancet. HIV* vol. 7 e596–e598 (2020).
51. Sherrard-Smith, E. *et al.* The potential public health consequences of COVID-19 on malaria in Africa. *Nat. Med.* **26**, 1411–1416 (2020).
52. Chin, T. *et al.* U.S. county-level characteristics to inform equitable COVID-19 response. *medRxiv* (2020) doi:10.1101/2020.04.08.20058248.
53. Hersbach, H. *et al.* The ERA5 global reanalysis. *Q.J.R. Meteorol. Soc.* **64**, 29 (2020).
54. Report 23 - State-level tracking of COVID-19 in the United States. *Imperial College London* http://www.imperial.ac.uk/medicine/departments/school-public-health/infectious-disease-epidemiology/mrc-global-infectious-disease-analysis/covid-19/report-23-united-states/.
55. Kissler, S. M., Tedijanto, C., Goldstein, E., Grad, Y. H. & Lipsitch, M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science* **368**, 860–868 (2020).
56. Jiwani, S. S. & Antiporta, D. A. Inequalities in access to water and soap matter for the COVID-19 response in sub-Saharan Africa. *Int. J. Equity Health* **19**, 82 (2020).
57. Stoler, J., Jepson, W. E. & Wutich, A. Beyond handwashing: Water insecurity undermines COVID-19 response in developing areas. *J. Glob. Health* **10**, 010355 (2020).
58. Makoni, M. Keeping COVID-19 at bay in Africa. *Lancet Respir Med* **8**, 553–554 (2020).
59. Tatem, A. J. *et al.* Ranking of elimination feasibility between malaria-endemic countries. *Lancet* **376**, 1579–1591 (2010).
60. Metcalf, C. J. E. *et al.* Transport networks and inequities in vaccination: remoteness shapes measles vaccine coverage and prospects for elimination across Africa. *Epidemiol. Infect.* **143**, 1457–1466 (2015).
61. Adepoju, P. Africa's struggle with inadequate COVID-19 testing. *The Lancet Microbe* vol. 1 e12 (2020).

62. Pindolia, D. K. *et al.* The demographics of human and malaria movement and migration patterns in East Africa. *Malar. J.* **12**, 397 (2013).
63. Qiu, Y., Chen, X. & Shi, W. Impacts of social and economic factors on the transmission of coronavirus disease (COVID-19) in China. doi:10.1101/2020.03.13.20035238.
64. Li, H. *et al.* Age-Dependent Risks of Incidence and Mortality of COVID-19 in Hubei Province and Other Parts of China. *Front. Med.* **7**, 190 (2020).
65. Global Burden of Disease Study 2017 (GBD 2017) Population Estimates 1950-2017 | GHDx. http://ghdx.healthdata.org/record/ihme-data/gbd-2017-population-estimates-1950-2017.
66. Brauer, M., Zhao, J. T., Bennitt, F. B. & Stanaway, J. D. Global access to handwashing: implications for COVID-19 control in low-income countries. doi:10.1101/2020.04.07.20057117.
67. Alegana, V. A. *et al.* National and sub-national variation in patterns of febrile case management in sub-Saharan Africa. *Nat. Commun.* **9**, 4994 (2018).
68. Murthy, S., Leligdowicz, A. & Adhikari, N. K. J. Intensive care unit capacity in low-income countries: a systematic review. *PLoS One* **10**, e0116949 (2015).
69. 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).
70. [No title]. https://dl.acm.org/doi/10.5555/1953048.2078195.
71. Jolliffe, I. T. & Cadima, J. Principal component analysis: a review and recent developments. *Philos. Trans. A Math. Phys. Eng. Sci.* **374**, 20150202 (2016).
72. Meyerowitz-Katz, G. & Merone, L. A systematic review and meta-analysis of published research data on COVID-19 infection-fatality rates. doi:10.1101/2020.05.03.20089854.
73. Wood, S. N. *Generalized Additive Models: An Introduction with R, Second Edition*. (CRC Press, 2017).
74. R: The R Project for Statistical Computing. https://www.R-project.org/.

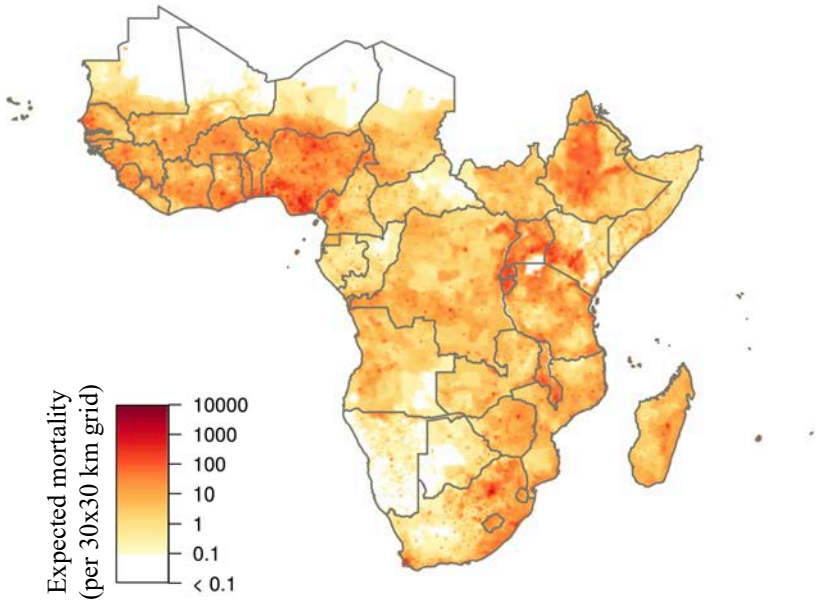
75. Global Burden of Disease Study 2017 (GBD 2017) Population Estimates 1950-2017 | GHDx. http://ghdx.healthdata.org/record/ihme-data/gbd-2017-population-estimates-1950-2017.
76. Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **20**, 533–534 (2020).
77. Bjornstad, O. N., Finkenstadt, B. F. & Grenfell, B. T. Dynamics of Measles Epidemics: Estimating Scaling of Transmission Rates Using a Time Series SIR Model. *Ecological Monographs* vol. 72 169 (2002).
78. Google. COVID-19 Community Mobility Reports.
79. WHO | Measles and Rubella Surveillance Data. (2020).
80. Nimpa, M. M. *et al.* Measles outbreak in 2018-2019, Madagascar: epidemiology and public health implications. *Pan Afr. Med. J.* **35**, 84 (2020).
81. 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).


A
B
Expected cumulative % of the population infected to date given death reporting
Expected cumulative % of the population infected to date given IFR by age curve
100.00
50.00
10.00
5.00
1.00
0.10
0.01
Death reporting rate
100%
50%
25%
10%
IFR by age shift
-10 years
-5 years
-2 years
0 years
+2 years
+5 years
+10 years
Burundi
Tanzania
Mauritius
Benin
Niger
Burkina Faso
Rwanda
Mozambique
South Sudan
Cote d Ivoire
Congo (DRC)
Guinea
Nigeria
Uganda
Chad
Comoros
Togo
Somalia
Madagascar
Sierra Leone
Ghana
Malawi
Botswana
Mali
CAR
Angola
Ethiopia
Liberia
Lesotho
Cameroon
Congo (ROC)
Zimbabwe
Senegal
Guinea Bissau
Gabon
Zambia
Kenya
Djibouti
Mauritania
Gambia
Namibia
Equatorial Guinea
Sao Tome and Principe
Eswatini
Cabo Verde
South Africa

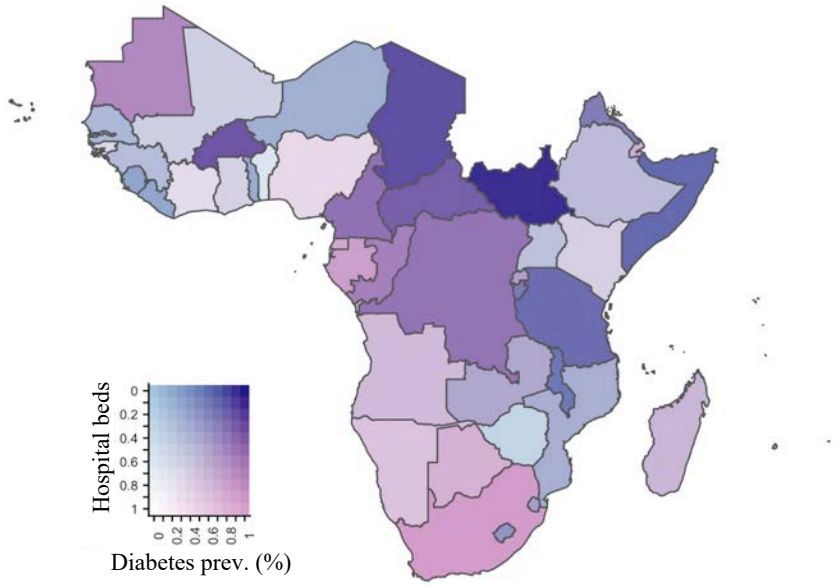

Relative SARS-CoV-2 epidemic risk
Direct burden (SARS-CoV-2 mortality)
Indirect burden (disruptions to routine health services)
Epidemic pace
Mortality rate (IFR)
Viral introductions
Transmissibility (R_t)
+
−
−
+
−
Fewer medical facilities on average in SSA
Fewer physicians and medical staff on average in SSA
Higher infectious disease incidence in SSA
NCDs: Higher mortality burden per age group in SSA
NCDs: Lower overall prevalence in SSA
Age: Younger populations on average in SSA
Lower internal connectivity between subregions in SSA
Greater mean travel time to cities in some SSA countries
Lower frequency of international air travel in SSA
Less access to handwashing and other prevention measures
Larger size of households for ages >50 in some SSA settings
Dense, resource limited urban populations in many SSA cities
Community based decentralized health systems
Recent experience with out-break response (e.g., Ebola)
Higher temperature and humidity on average in SSA
E
E
B
B
B
A
F
F
F
D
A
A
D
D
C
Risk dimension:
A Demography
B Comorbidity
C Climate
D Prevention
E Access to care
F Human mobility and travel


A
Physicians (per 100,000 popn)
SSA AMR EMR EUR SEA WPR
B
Age standardized COPD mortality (per 100,000 popn)
C
Urban popn living in crowded, low quality housing (%)
D
Urban popn with access to handwashing facilities (%)
E
Age standardized COPD mortality (per 100,000 popn)
Physicians (per 100,000 popn)
Age (% Popn > 50y)
max
min
Mean HH Size (> 50y)
F
Urban popn with access to handwashing facilities (%)
Urban popn living in crowded, low quality housing (%)

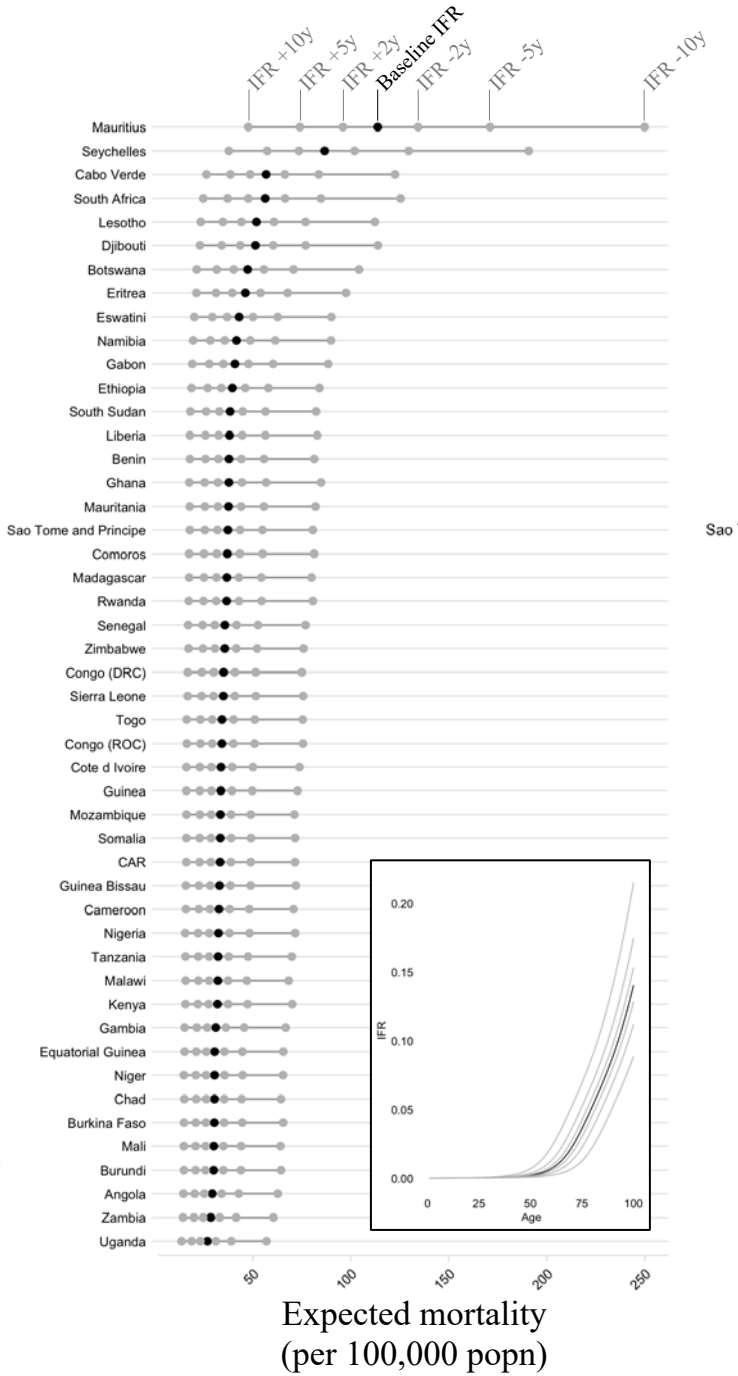
A | Baseline mortality risk from demographic structure



B | Comorbidity vs access to care

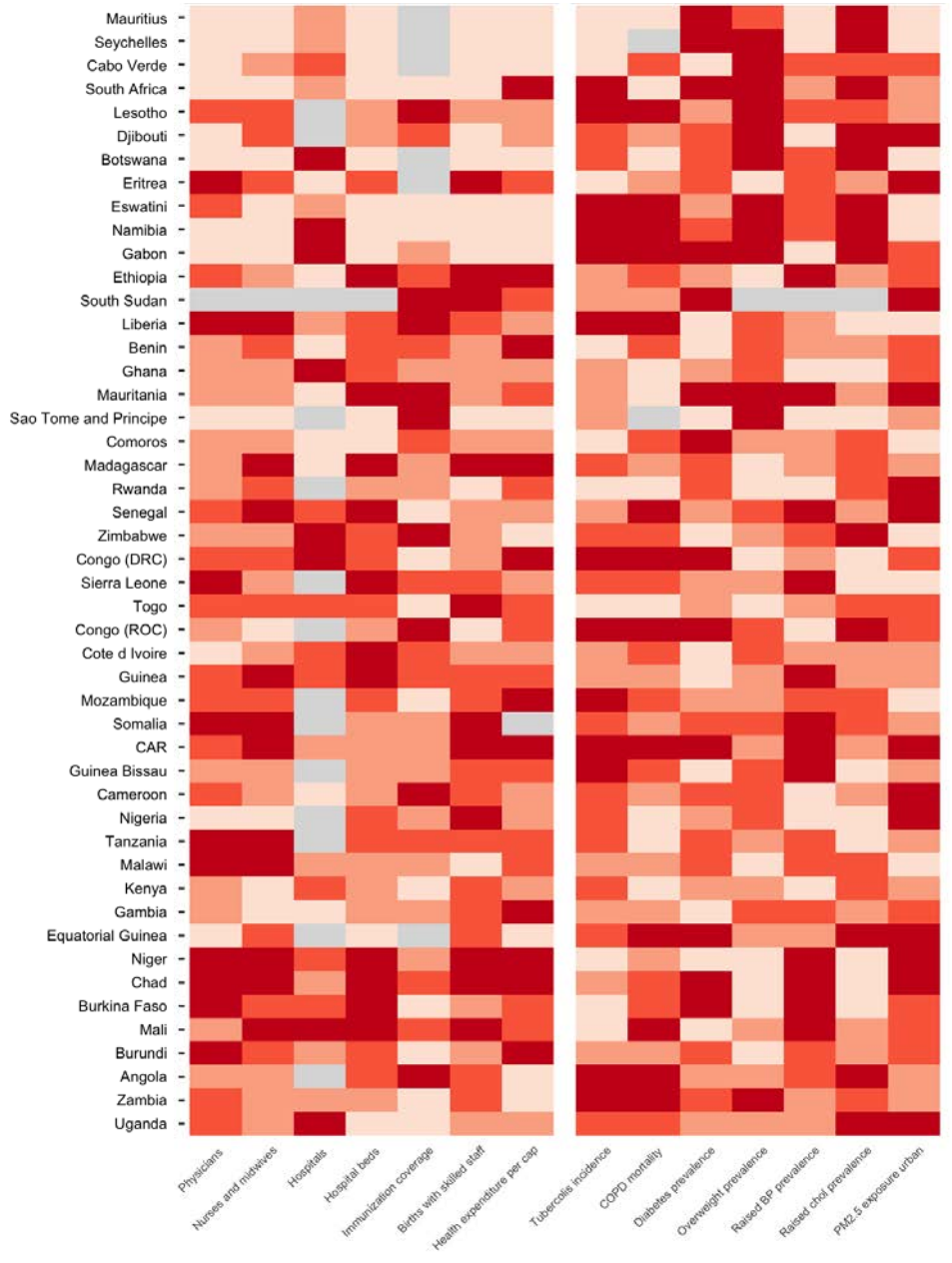


C | Range in mortality under simulated *IFR* scenarios



D | Indicators of access to care at national level

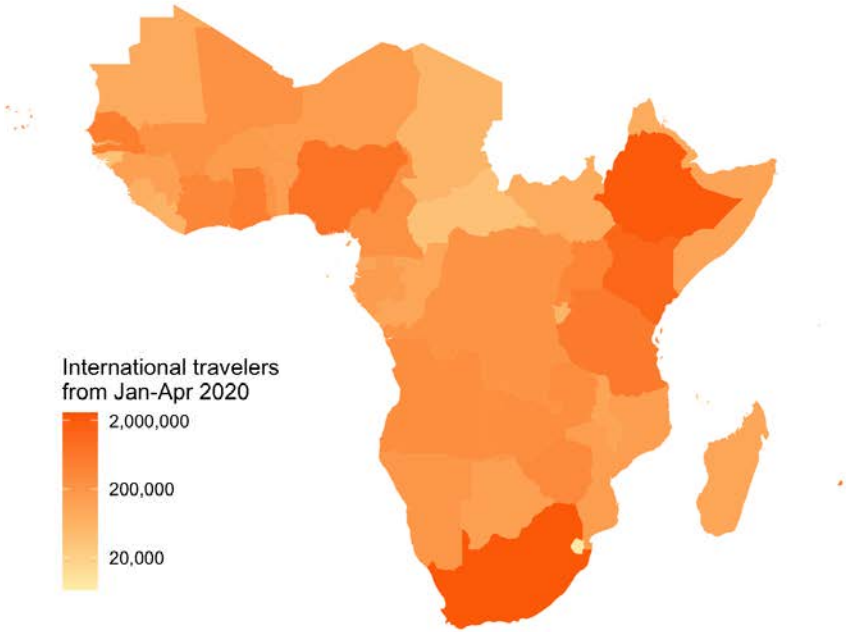
E | Indicators of comorbidity burden at national level



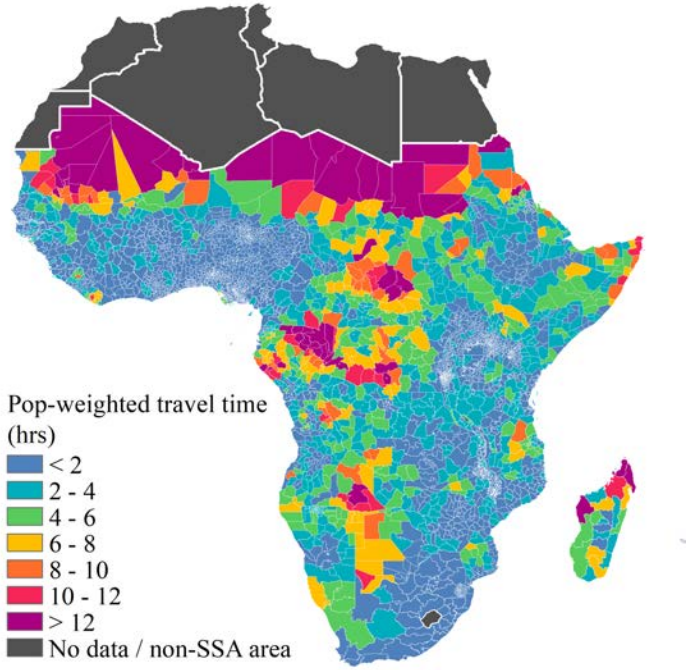
National level risk (quartiles):

NA

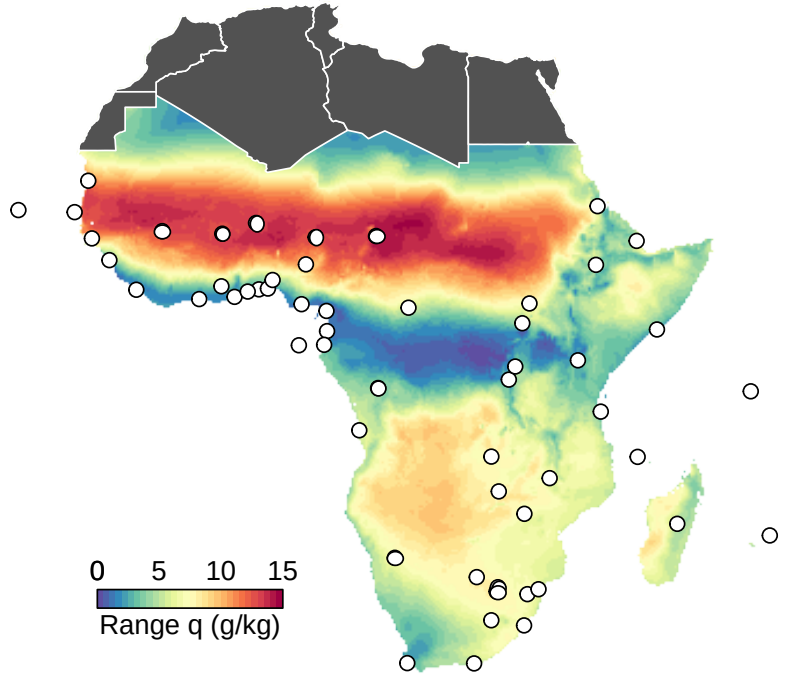
A | International travel by country



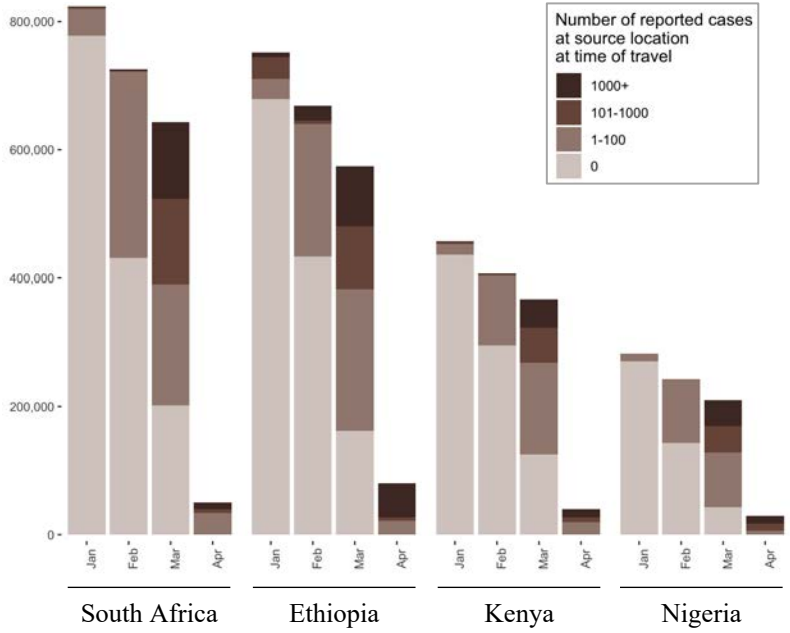
C | Connectivity (pop-weighted mean travel time to nearest city)



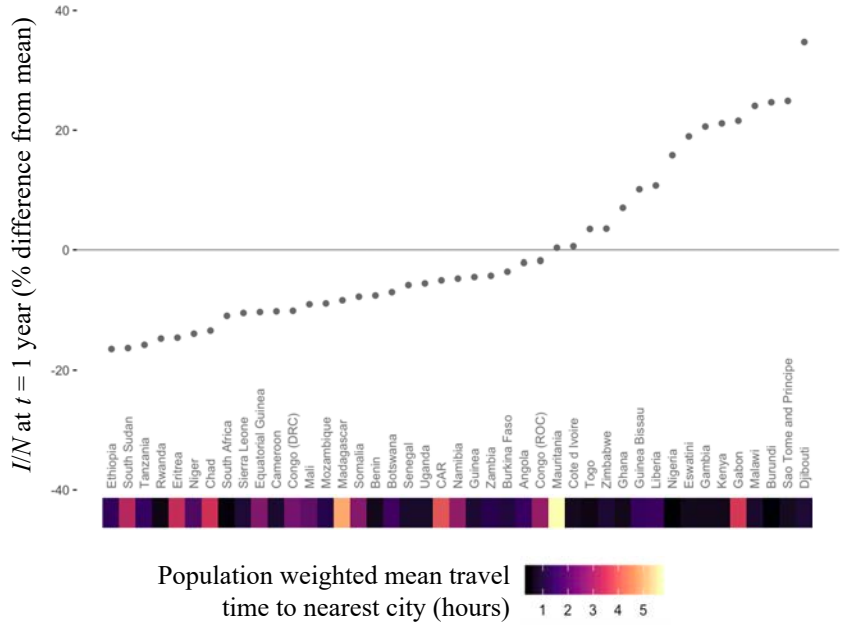
E | Seasonal variation in humidity



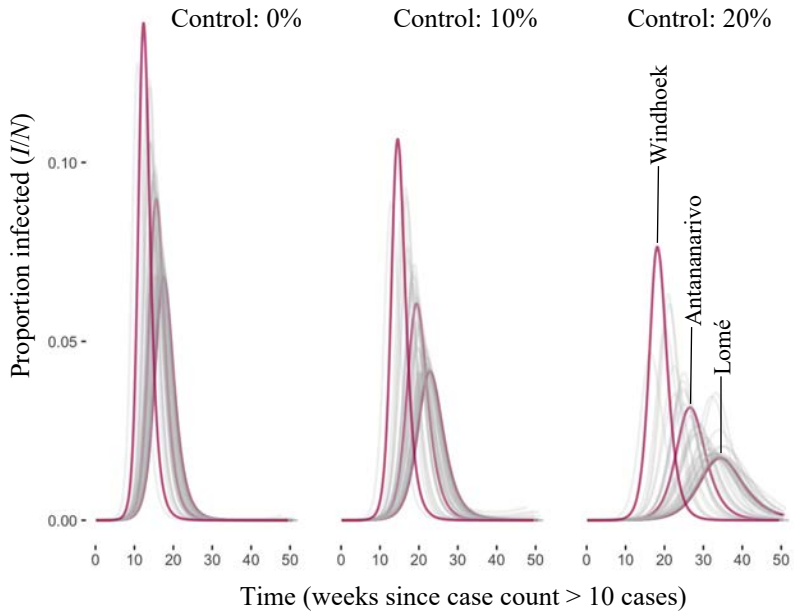
B | International travelers in 2020 by departure location



D | Connectivity vs proportion infected at 1 year



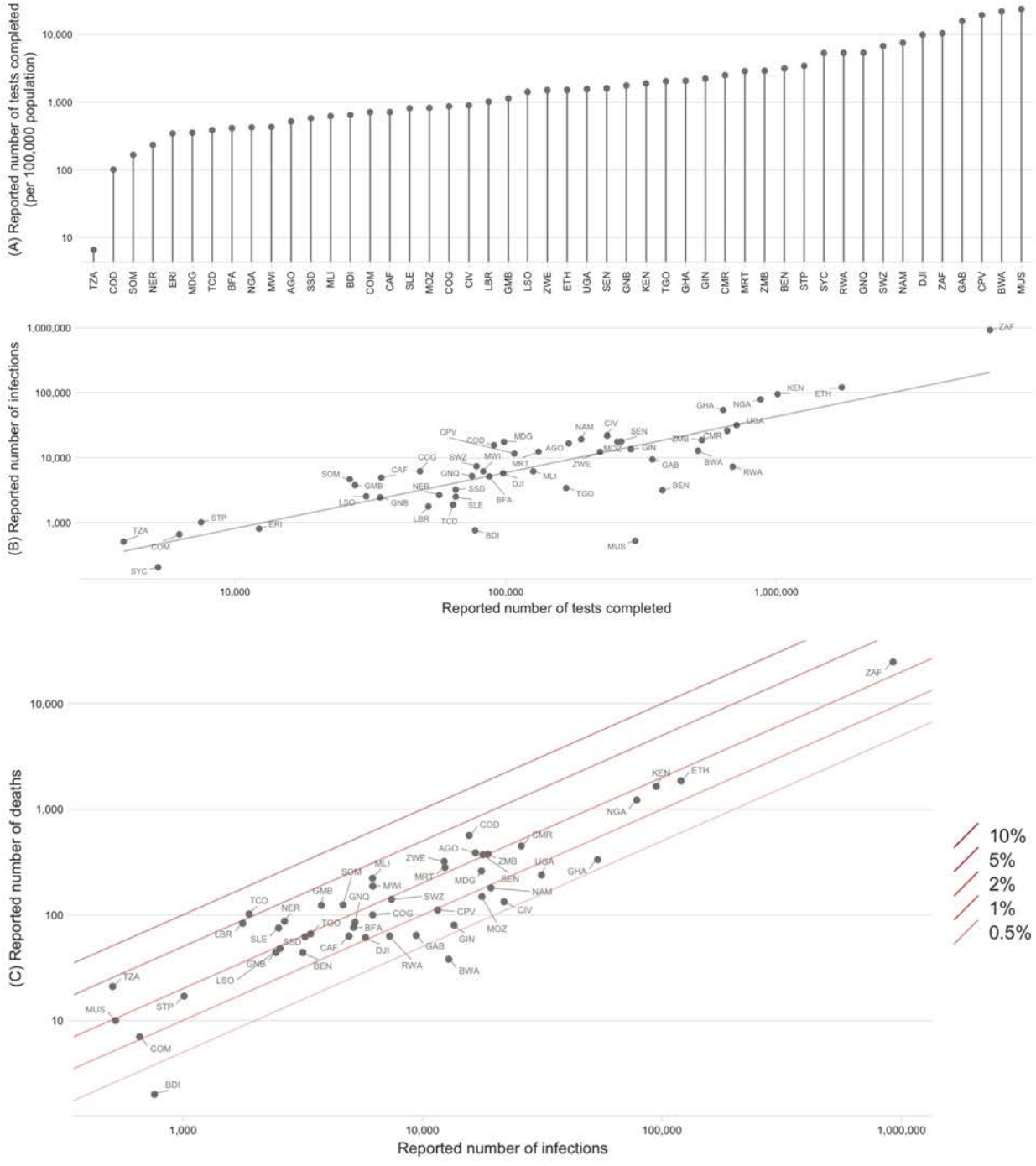
F | Infection time series assuming model with climate forcing



Extended Data Figure 1

Variation between SSA countries in testing and reporting rates

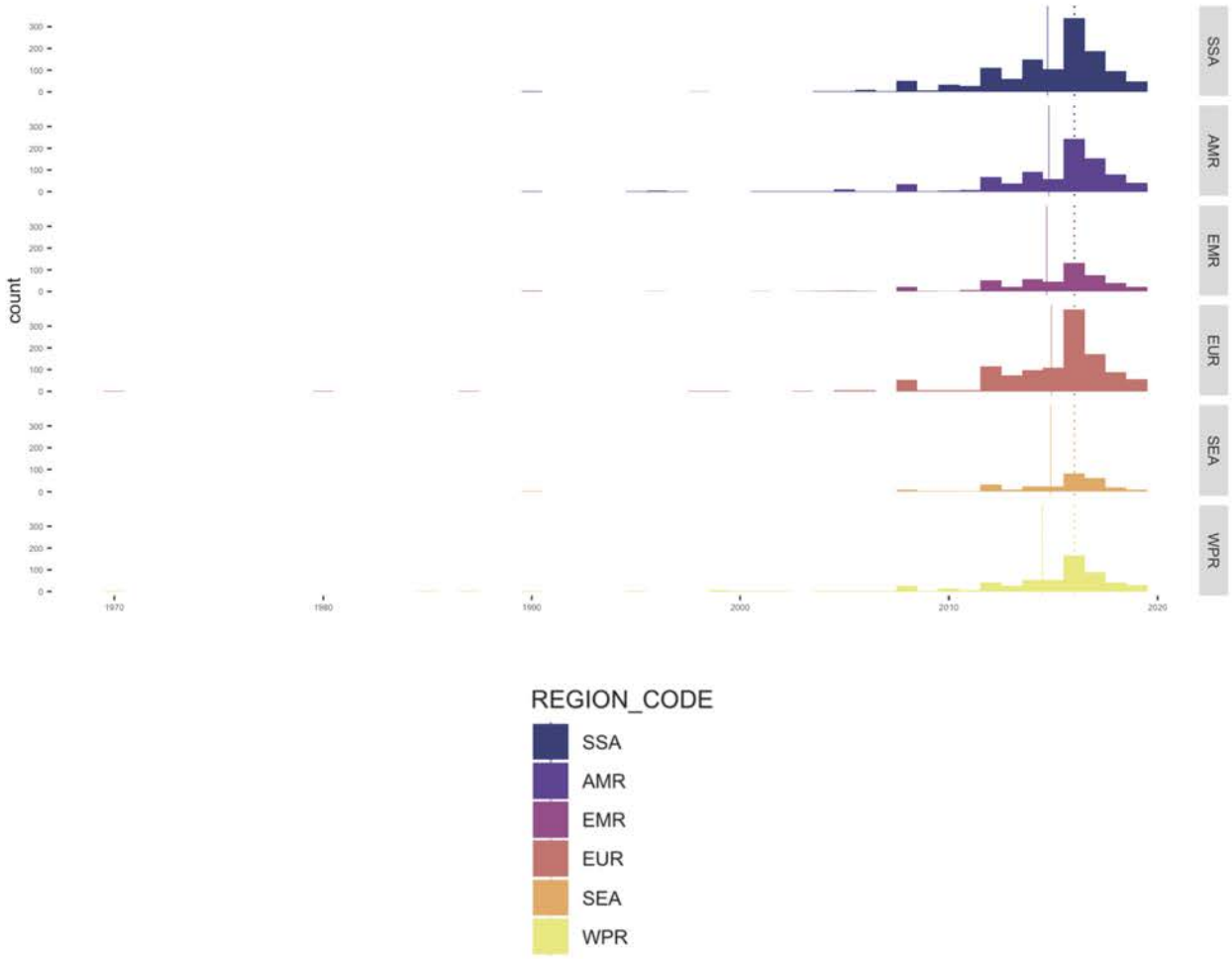
A: Reported number of tests completed per country as of December 20, 2020. **B**: Number of infections (I) per reported number of tests (T); line shows linear least squares regression: $I = 1.422\times10^{-1}\times T - 1.912\times10^{4}$ ($df = 46$, adjusted $R^2 = 0.9496$, Pearson's correlation coefficient, $r = 0.9750$, $p < 0.001$). **C**: Reported infections and deaths for sub-Saharan African countries with case fatality ratios (CFRs) shown as diagonal lines.



Extended Data Figure 2

Year of most recent data available for variables compared between global regions

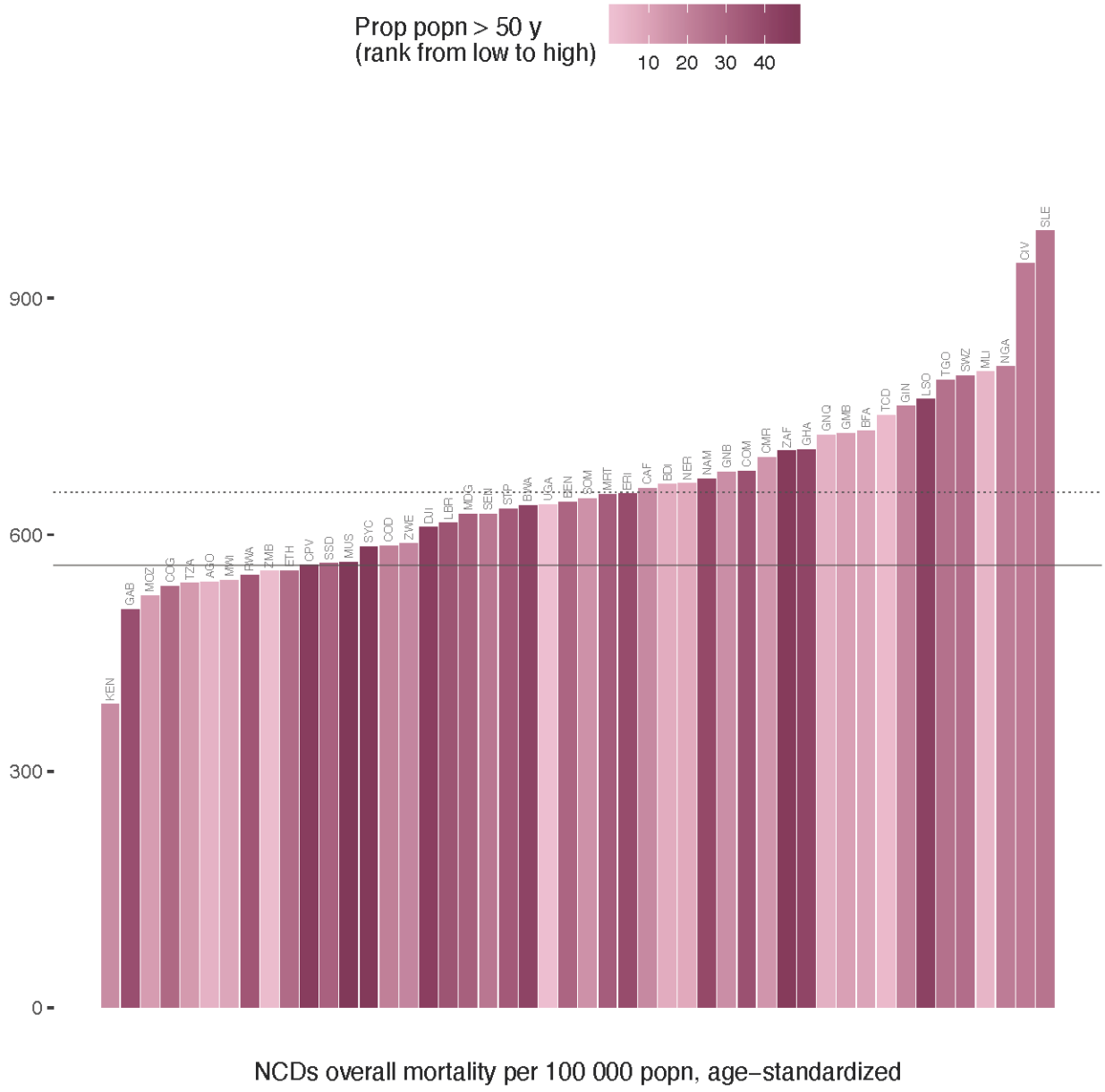
Dotted vertical line shows regional median; solid vertical line shows regional mean. Note that most data comes from 2015-2019 (median = 2016, mean = 2014.62-2014.93).



Extended Data Figure 3

Variation among sub-Saharan African countries in determinants of SARS-CoV-2 risk by variable

A subset of variables is shown in Figure 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/).

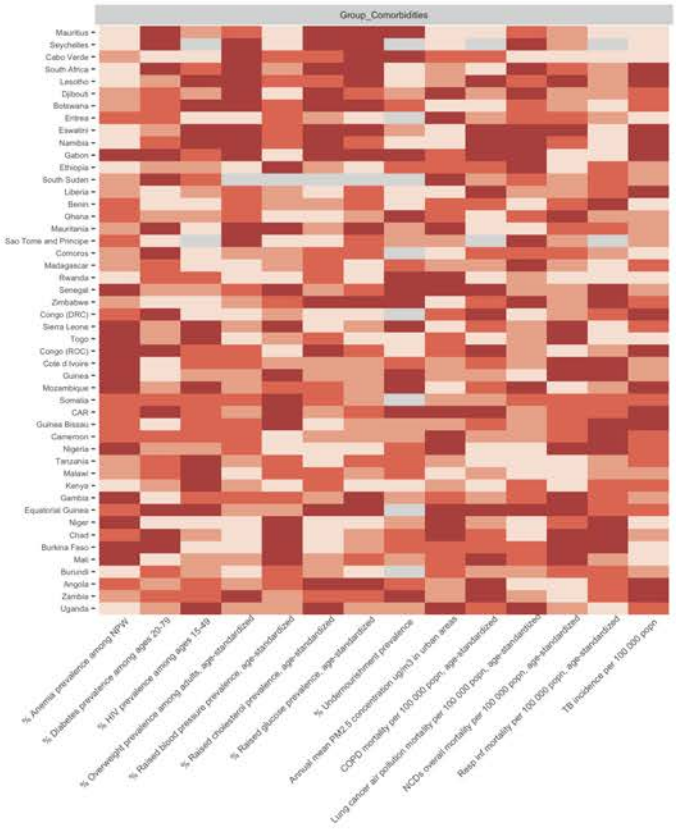


Extended Data Figure 4

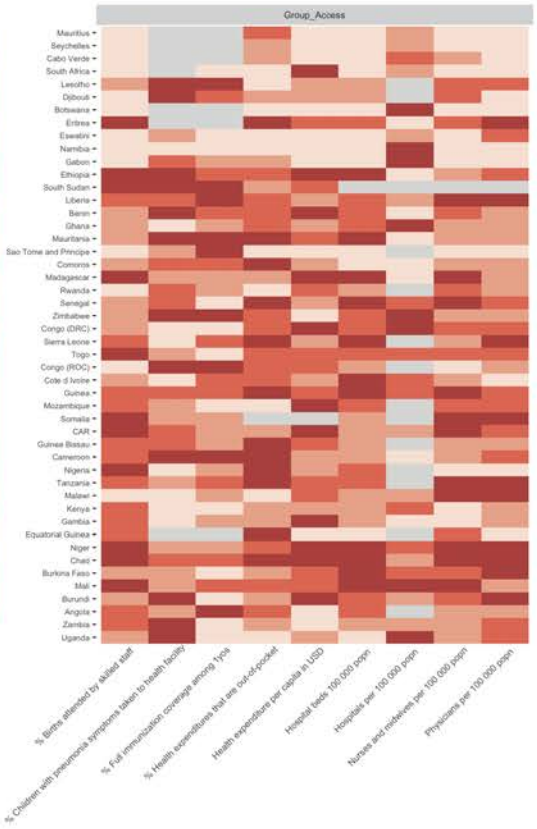
Variation among sub-Saharan African countries in determinants of SARS-CoV-2 mortality risk by category

A subset of variables is shown in Figure 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 (e.g., 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 (e.g., 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 **Table S3** for details).

A. Comorbidities



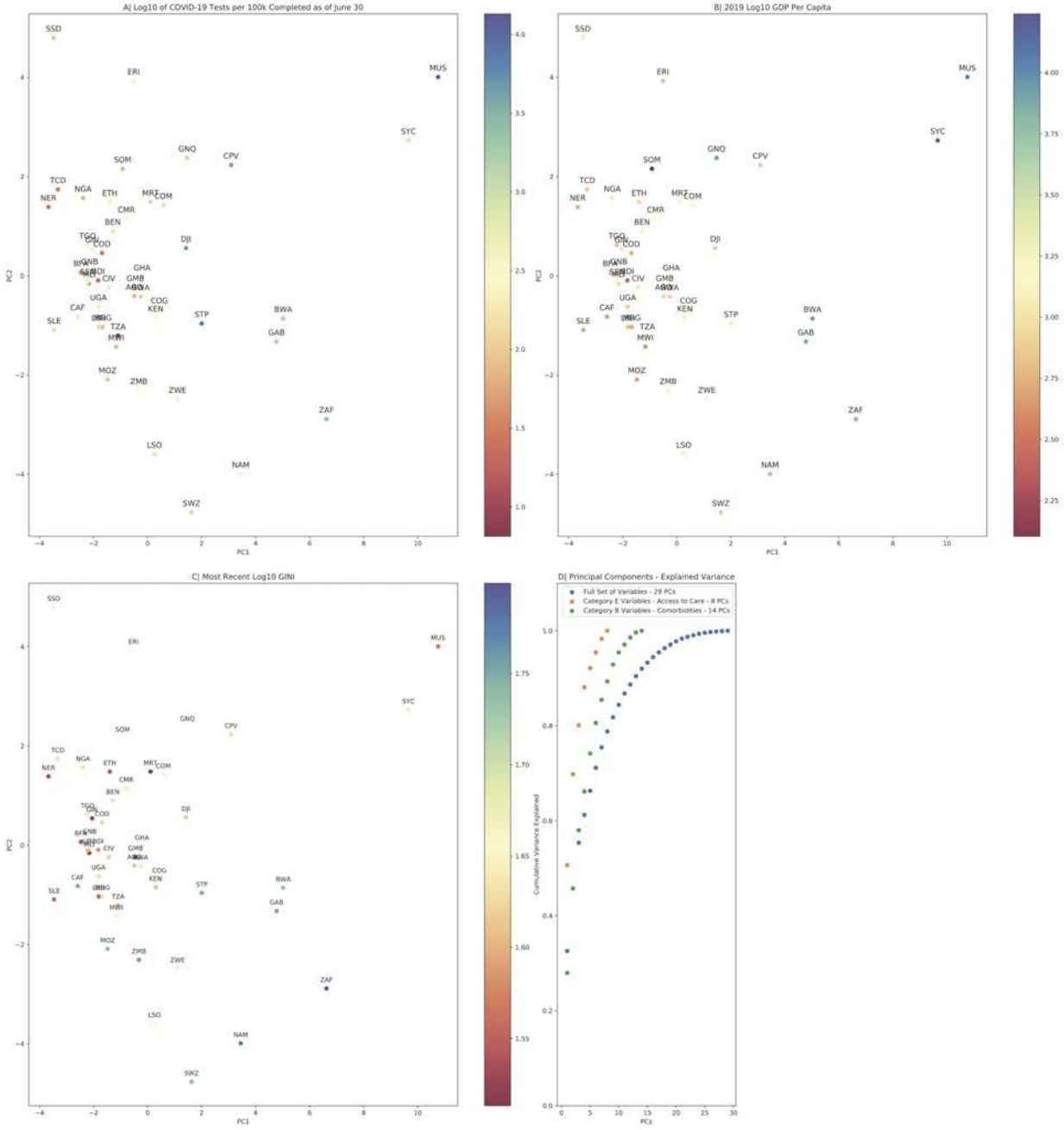
B. Access to care



Extended Data Figure 5

Principal Component Analysis of all variables and category specific subsets of variables

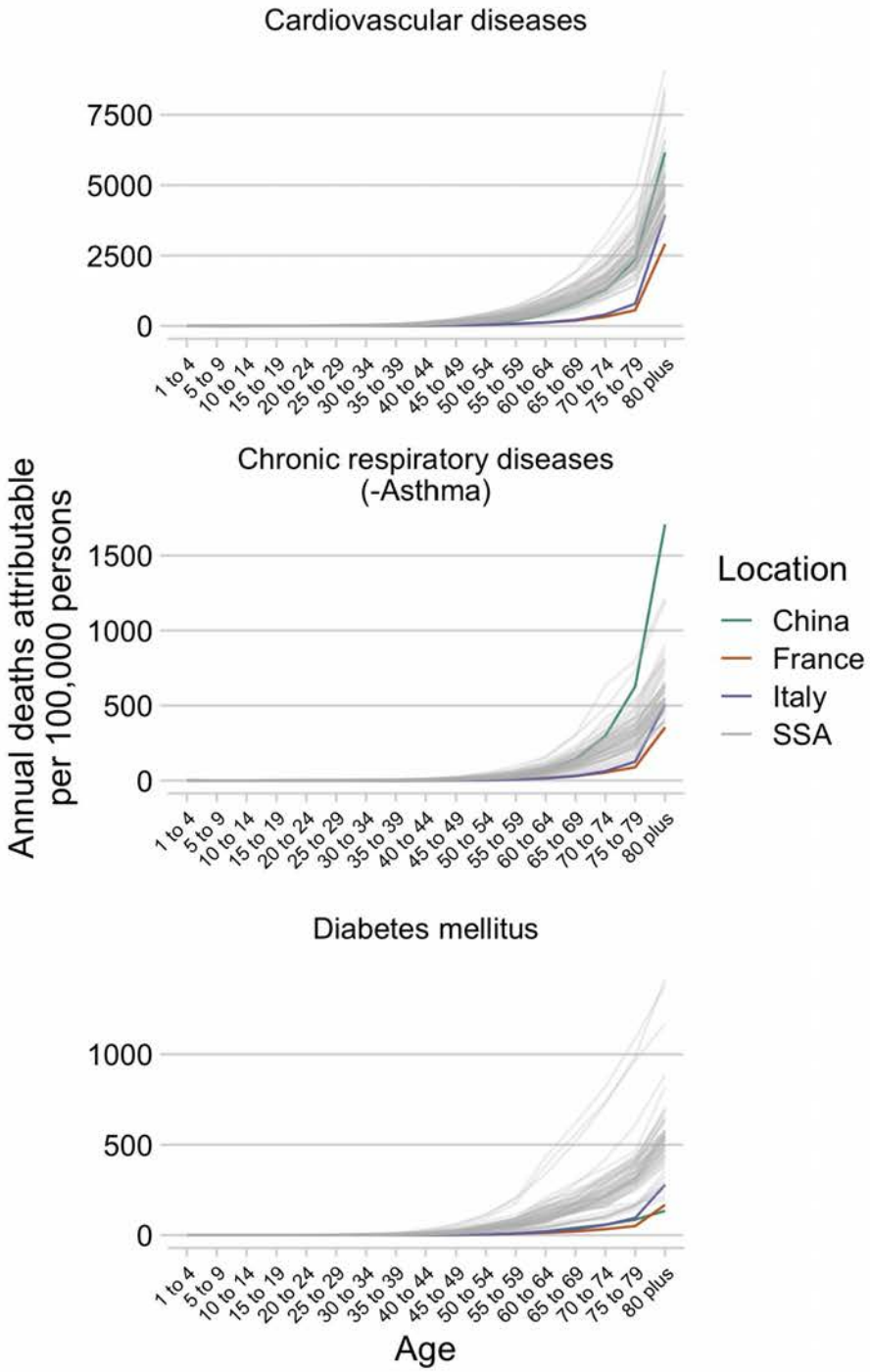
A: Principal Component 1 and 2, countries colored by Log10 scaled tests per 100,000 population (as of June 30, 2020). **B:** Principal Component 1 and 2, countries colored by Log10 scaled GDP per capita. **C**: Principal Component 1 and 2, countries colored by the GINI index (a measure of wealth disparity). **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 **Table S3**)), and access to care indicators (orange, 8 variables, Section E in **Table S3**)



Extended Data Figure 6

Comorbidity burden by age in sub-Saharan Africa

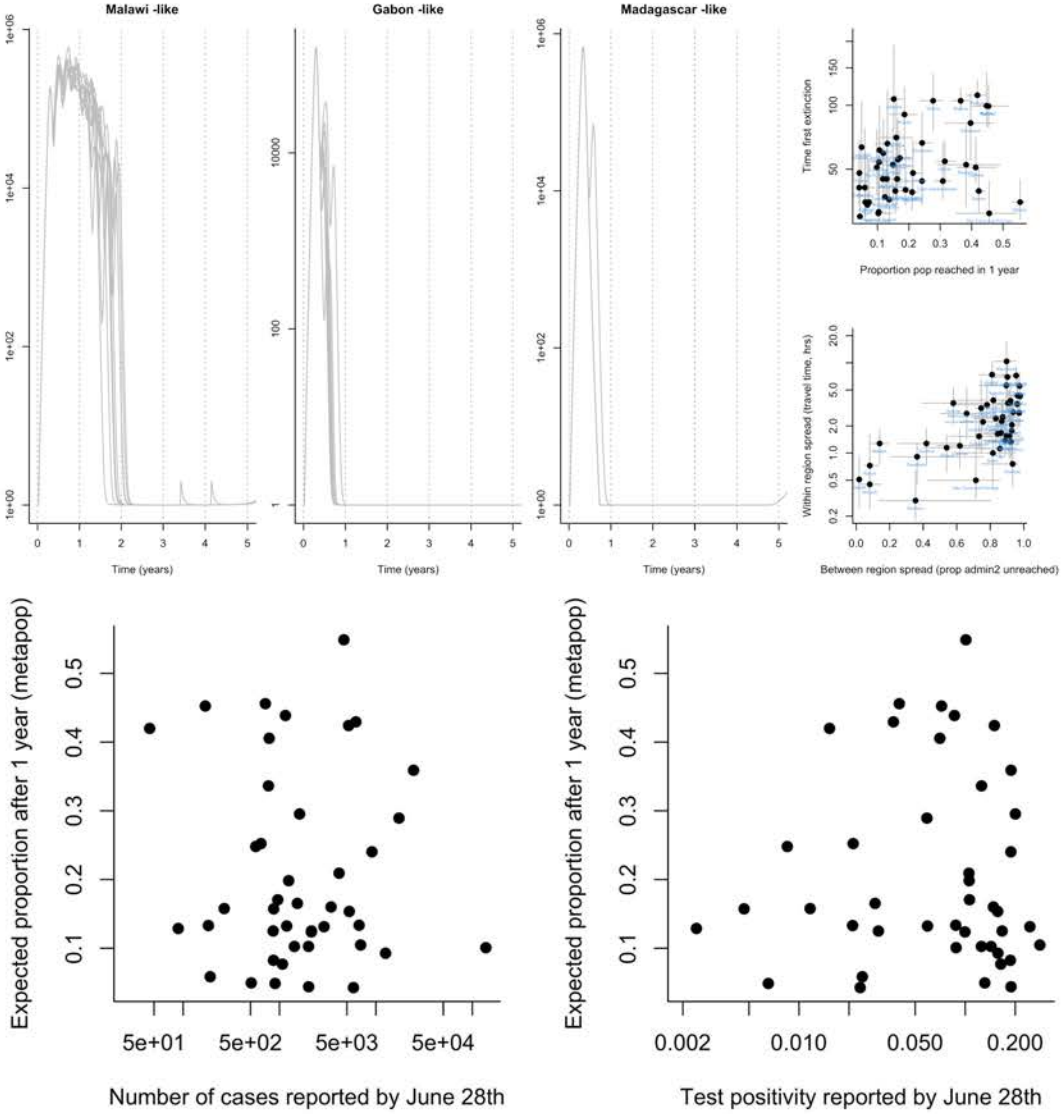
Estimated mortality per age group for sub-Saharan African countries (gray lines) compared to China, France, and Italy (the countries from which estimates of SARS-CoV-2 infection fatality ratios (IFRs) by age are available) for three NCD categories (cardiovascular diseases, chronic respiratory diseases excluding asthma, and diabetes).



Extended Data Figure 7

Pace of the outbreak and cases and testing vs. the pace of the outbreak

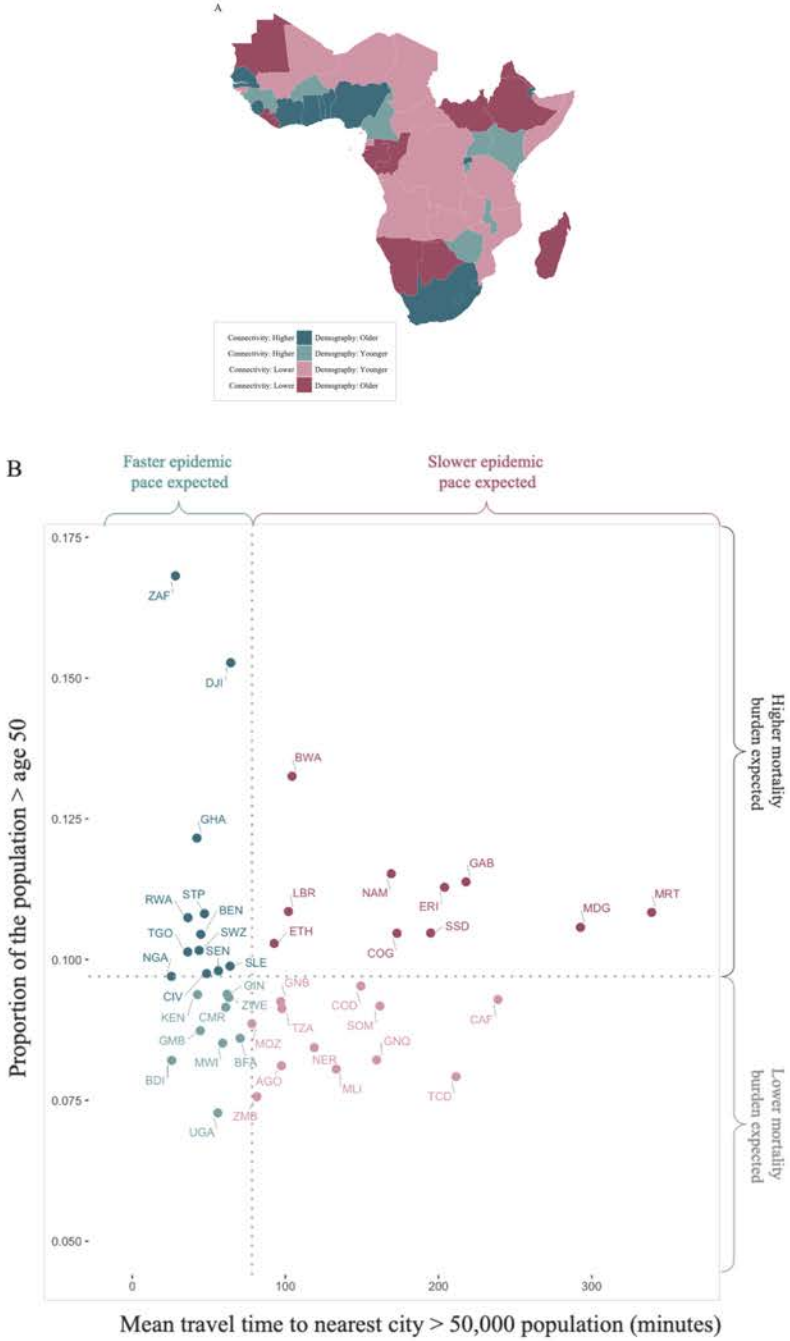
Top: Each grey line on the left-hand panels indicates the total infected across all administrative units in a metapopulation simulation with parameters reflecting the country indicated by the plot title, assuming interventions are constant, and that immunity does not wane. Simulations with parameters 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 Figure 8

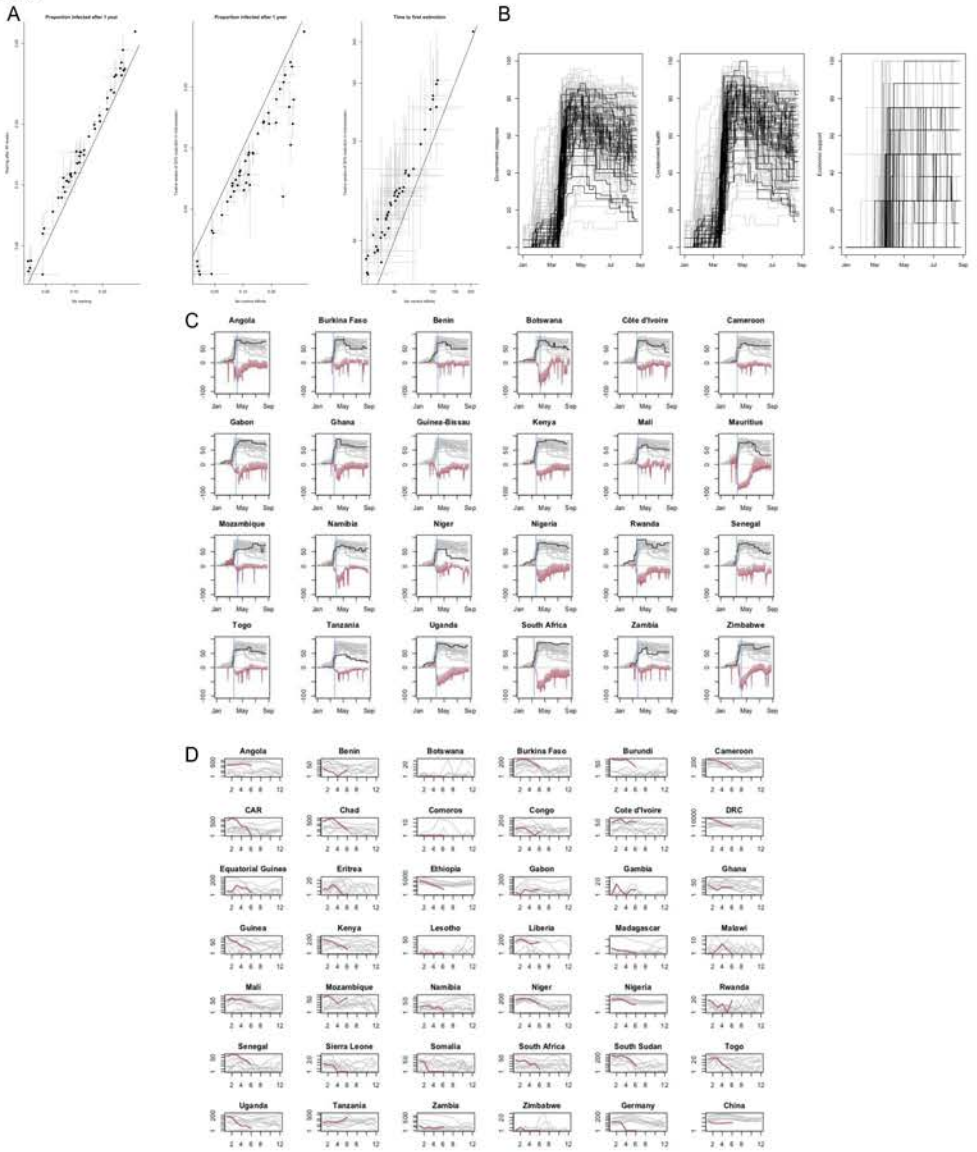
Bivariate example of expected pace versus expected burden at the national level in SARS-CoV-2 outbreaks in sub-Saharan Africa

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 (i.e., less synchronous outbreaks) relative to the median among SSA countries; in blue, countries with more connectivity; darker colors show countries with older populations (i.e., 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 Figure 9

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, i.e., 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 [78] 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 [79] 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 (e.g., dynamics in Madagascar are largely dominated by a honeymoon outbreak that occurred in 2018-2019 [80]) for countries that slowed the SARS-CoV-2 pandemic, signatures of reduction in measles can be identified (e.g., Germany and China; similar patterns are seen in Viet Nam).



Extended Data Figure 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 $1.0 < R_{0min} < 1.5$.