291   Supplementary Materials Outline:

292

293

A1 | Reported SARS-CoV-2 case counts, mortality, and testing in sub-Saharan Africa as of June 2020

1.1. 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 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.

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 June 30, 2020 (https://coronavirus.jhu.edu/map.html).

Case fatality ratios (*CFR*s) 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 [38]. As reported confirmed cases are likely to be a significant underestimate of the true number of infections, *CFR*s may be a poor proxy for the infection fatality ratio (*IFR*), defined as the proportion of infections that result in mortality [4].

1.2 Variation in testing and mortality rates

Testing rates among SSA countries varied by multiple orders of magnitude: the number of tests completed per 100,000 population ranged from 6.50 in Tanzania to 13,508.13 in Mauritius (**Figure S1A**). The number of reported infections (i.e., positive tests) was strongly correlated with the number of tests completed (Pearson's correlation coefficient, $r = 0.9667$, $p < 0.001$) (**Figure S1B**). As of June 30, 2020, no deaths due to SARS-CoV-2 were reported to the Africa CDC for five SSA countries (Eritrea, Lesotho, Namibia, Seychelles, Uganda). Among countries with at least one reported death, *CFR* varied from 0.22% in Rwanda to 8.54% in Chad (**Figure S1C**). 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), indicate that the data are insufficient to determine country specific *IFR*s and *IFR* by age profiles. As a result, global *IFR* by age estimates were used for the subsequent analyses in this study.

407 **Table S1**

408 **Sub-Saharan Africa country country codes, case counts, and testing as of June 30, 2020**

Country Name	Country Code	Cases[a]	Deaths[a]	Tests[a]	Population[b]	Cases per 100k[c]	Tests per 100k[c]	Positivity (%)	CFR (%)
Angola	AGO	267	11	22895	32866268	0.81	69.66	1.17	4.12
Benin	BEN	1187	19	20014	12123198	9.79	165.09	5.93	1.60
Botswana	BWA	89	1	36868	2351625	3.78	1567.77	0.24	1.12
Burkina Faso	BFA	959	53	9040	20903278	4.59	43.25	10.61	5.53
Burundi	BDI	170	1	2359	11890781	1.43	19.84	7.21	0.59
Cameroon	CMR	12592	313	80000	26545864	47.43	301.37	15.74	2.49
Cabo Verde	CPV	1165	12	22665	555988	209.54	4076.53	5.14	1.03
Central Africa Republic	CAF	3429	45	23208	4829764	71.00	480.52	14.78	1.31
Chad	TCD	866	74	4633	16425859	5.27	28.21	18.69	8.55
Comoros	COM	293	7	1173	869595	33.69	134.89	24.98	2.39
Côte d'Ivoire	CIV	9101	66	48340	26378275	34.50	183.26	18.83	0.73
Congo (DRC)	COD	6939	167	24657	89561404	7.75	27.53	28.14	2.41
Djibouti	DJI	4656	53	46108	988002	471.25	4666.79	10.10	1.14
Equatorial Guinea	GNQ	2001	32	16000	1402985	142.62	1140.43	12.51	1.60
Eritrea	ERI	191	0	7943	3546427	5.39	223.97	2.40	0.00
Eswatini	SWZ	781	11	11094	1160164	67.32	956.24	7.04	1.41
Ethiopia	ETH	5846	103	250604	114963583	5.09	217.99	2.33	1.76
Gabon	GAB	5209	40	34774	2225728	234.04	1562.37	14.98	0.77
Gambia	GMB	45	2	2947	2416664	1.86	121.94	1.53	4.44
Ghana	GHA	17351	112	294867	31072945	55.84	948.95	5.88	0.65
Guinea	GIN	5291	30	33737	13132792	40.29	256.89	15.68	0.57
Guinea Bissau	GNB	1614	21	8056	1967998	82.01	409.35	20.03	1.30
Kenya	KEN	6190	144	167417	53771300	11.51	311.35	3.70	2.33
Lesotho	LSO	27	0	3000	2142252	1.26	140.04	0.90	0.00
Liberia	LBR	768	34	6125	5057677	15.18	121.10	12.54	4.43

409
410
411

412 (Table S1 continued)

Country Name	Country Code	Cases[a]	Deaths[a]	Tests[a]	Population[b]	Cases per 100k[c]	Tests per 100k[c]	Positivity (%)	CFR (%)
Madagascar	MDG	2138	20	21444	27691019	7.72	77.44	9.97	0.94
Malawi	MWI	1152	13	13369	19129955	6.02	69.89	8.62	1.13
Mali	MLI	2147	113	12869	20250834	10.60	63.55	16.68	5.26
Mauritania	MRT	4149	126	39398	4649660	89.23	847.33	10.53	3.04
Mauritius	MUS	341	10	171792	1271767	26.81	13508.13	0.20	2.93
Mozambique	MOZ	859	5	28586	31255435	2.75	91.46	3.00	0.58
Namibia	NAM	183	0	8706	2540916	7.20	342.63	2.10	0.00
Niger	NER	1074	67	6555	24206636	4.44	27.08	16.38	6.24
Nigeria	NGA	24567	565	130164	206139587	11.92	63.14	18.87	2.30
Congo (ROC)	COG	1245	40	11790	5518092	22.56	213.66	10.56	3.21
Rwanda	RWA	900	2	137751	12952209	6.95	1063.53	0.65	0.22
São Tomé and Príncipe	STP	713	13	17773	219161	325.33	8109.56	4.01	1.82
Senegal	SEN	6698	108	76343	16743930	40.00	455.94	8.77	1.61
Seychelles	SYC	77	0	704	98340	78.30	715.88	10.94	0.00
Sierra Leone	SLE	1427	60	9973	7976985	17.89	125.02	14.31	4.20
Somalia	SOM	2894	90	11807	15893219	18.21	74.29	24.51	3.11
South Africa	ZAF	138134	2456	1567084	59308690	232.91	2642.25	8.81	1.78
South Sudan	SSD	2006	37	10630	11193729	17.92	94.96	18.87	1.84
Tanzania	TZA	509	21	3880	59734213	0.85	6.50	13.12	4.13
Togo	TGO	642	14	30316	8278737	7.75	366.19	2.12	2.18
Uganda	UGA	870	0	186200	45741000	1.90	407.07	0.47	0.00
Zambia	ZMB	1531	21	53370	18383956	8.33	290.31	2.87	1.37
Zimbabwe	ZWE	567	6	66712	14862927	3.81	448.85	0.85	1.06

413

414 [a] Data from Africa CDC as of June 30, 2020 (https://africacdc.org/covid-19/)

415 [b] Data from UN Population Division UNPOP (2019 revision) estimates of population by single calendar year (2020)[38]

416 [c] Rates per 100,000 population

417

418 # Figure S1

419 ## Variation between SSA countries in testing and reporting rates as of June 30, 2020

420 **A**: Reported number of tests completed per country as of June 2020 (source: Africa CDC). **B**: Number of infections (I)

421 per reported number of tests (T); line shows linear regression: $I = 8.454 \times 10^{-2} \times T - 8.137 \times 10^2$ ($R^2 = 0.933$, $p < 0.001$).
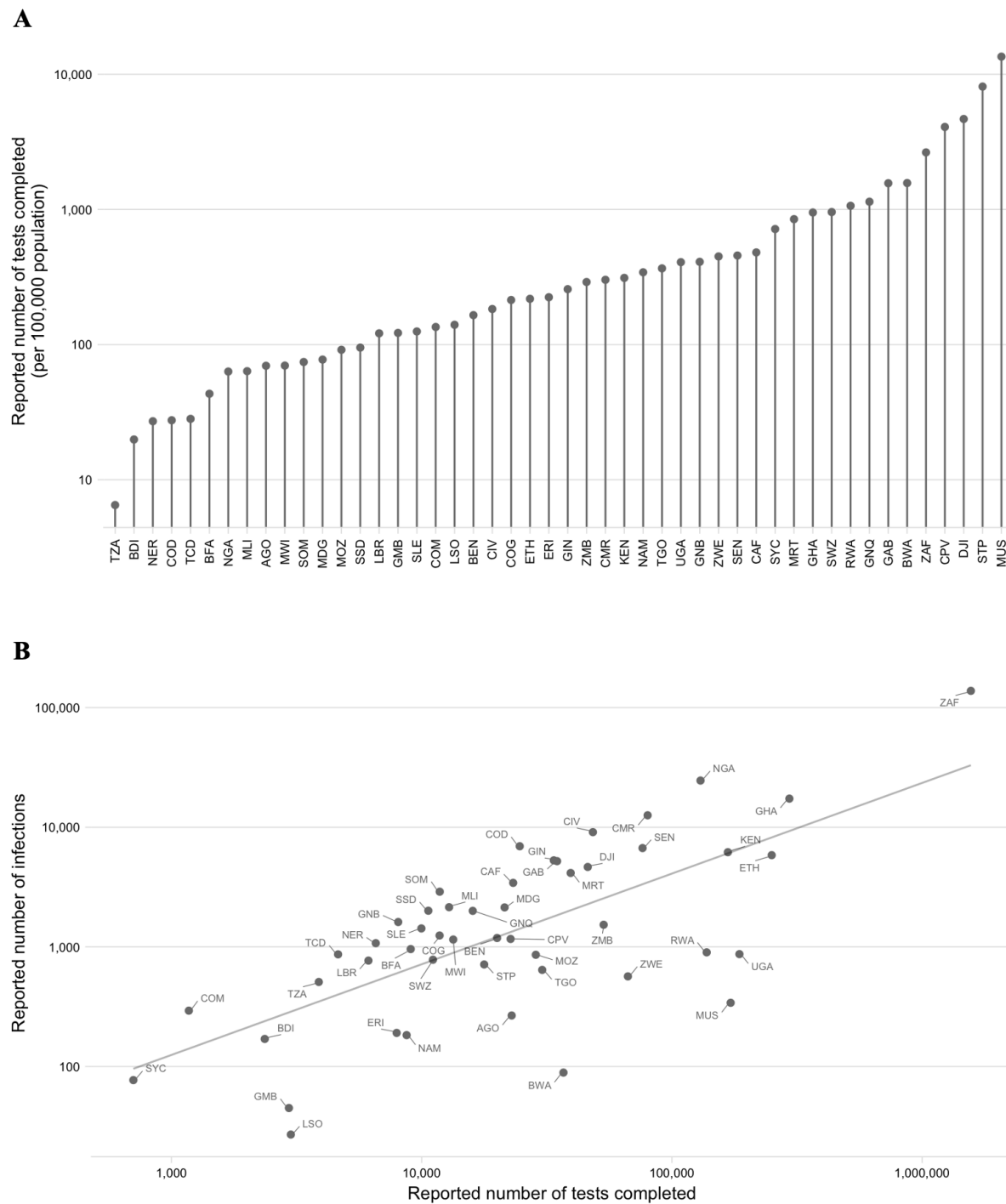
422
423

424 (Figure S1 continued)
425 **C**: Reported infections and deaths for sub-Saharan African countries with case fatality ratios (*CFR*s) shown as
426 diagonal lines. **D**: Date of first detection per number of reported tests
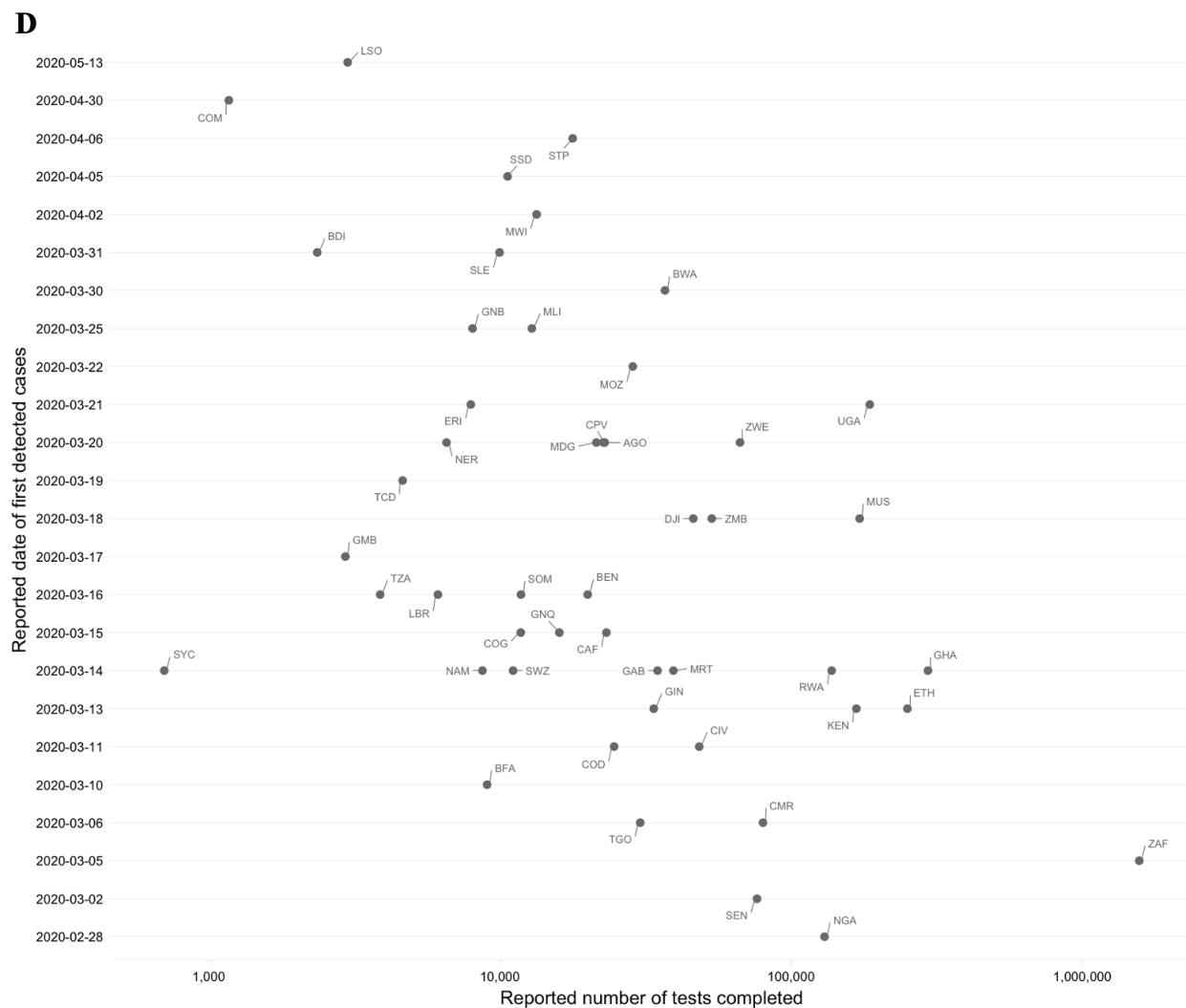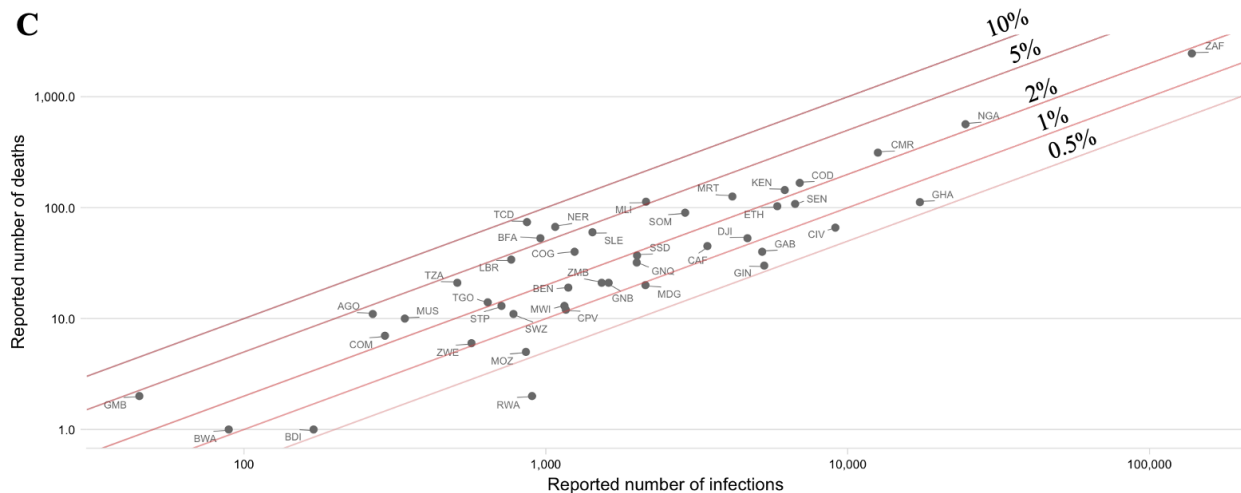
427

428 **A2 | Methods: Synthesizing factors hypothesized to increase or**
429 **decrease SARS-CoV-2 epidemic risk in SSA**
430
431 *2.1. Variable selection and data sources for variables hypothesized to associate with an*
432 *increased probability of severe clinical outcomes for an infection*
433
434 To characterize epidemic risk, defined as potential SARS-CoV-2 related morbidity and mortality,
435 we first synthesized factors hypothesized to influence risk in SSA settings (**Table S2**). Early
436 during the pandemic, evidence suggested that age was an important risk factor associated with
437 morbidity and mortality associated with SARS-CoV-2 infection [39], a pattern subsequently
438 confirmed across settings [2,9,40]. Associations between SARS-CoV-2 mortality and comorbidities
439 including hypertension, diabetes, and cardiovascular disease emerged early [39]; and have been
440 observed across settings, with further growing evidence for associations with obesity [9,41], severe
441 asthma [9], and respiratory effects of pollution [42].
442
443 Many possible sources of bias complicate interpretation of these associations [43], and while they
444 provide a useful baseline, inference is also likely to change as the pandemic advances. To
445 reflect this, our analysis combines a number of high level variables likely to broadly encompass
446 these putative risk factors (e.g., non-communicable disease (NCD) related mortality and health
447 life expectancy) with more specific measures encompassed in evidence to date (e.g.,
448 prevalence of diabetes, obesity, and respiratory illness such as Chronic Obstructive Pulmonary
449 Disease (COPD)). We also include measures relating to infectious diseases, undernourishment,
450 and anemia given their interaction and effects in determining health status in these settings [44].
451
452 Data on the identified indicators were sourced in May 2020 from the World Health Organization
453 (WHO) Global Health Observatory (GHO) database (https://www.who.int/data/gho), World Bank
454 (https://data.worldbank.org/), and other sources detailed in **Table S3**. National level
455 demographic data (population size and age structure) was sourced from United Nations World
456 Population Prospects (UNPOP) [38] and data on subnational variation in demography was
457 sourced from WorldPop [25]. Household size data was defined by the mean number of individuals
458 in a household with at least one person aged > 50 years, taken from the most recently available
459 demographic health survey (DHS) data [45]. All country level data for all indicators can be found
460 online at the **SSA-SARS-CoV-2-tool** (https://labmetcalf.shinyapps.io/covid19-burden-africa/).
461
462 Comparisons of national level estimates sourced from WHO and other sources are affected by
463 variation within countries and variation in the uncertainty around estimates from different
464 geographical areas. To assess potential differences in data quality between geographic areas
465 we compared the year of most recent data for variables (**Figure S2**). The mean (range varied
466 from 2014.624 to 2014.928 by region) and median year (2016 for all regions) of the most recent
467 data varied little between regions. To account for uncertainty associated in the estimates
468 available for a single variable, we also include multiple variables per category (e.g.,
469 demographic and socio-economic factors, comorbidities, access to care) to avoid reliance on a
470 single metric. This allows exploring variation between countries across a broad suite of
471 variables likely to be indicative of the different dimensions of risk.

472

473 Although including multiple variables that are likely to be correlated (see PCA methods below
474 for further discussion) would bias inference of cumulative risk in a statistical framework, we do
475 not attempt to quantitatively combine risk across variables for a country, nor project risk based
476 on the variables included here. Rather, we characterize the magnitude of variation among
477 countries for these variables (see **Figure 2** in the main text for a subset of the variables; **Figure**
478 **3B** for bivariate risk maps following [46]) and then explore the range of outcomes that would be
479 expected under scenarios where *IFR* increases with age at different rates (see **Figure 3** in the
480 main text).

481

482 *2.2. Variable selection and data sources for variables hypothesized to modulate the rate of viral*
483 *spread*

484

485 In addition to characterizing variation among factors likely to modulate burden, we also
486 synthesize data sources relevant to the rate of viral spread, or pace, for the SARS-CoV-2
487 pandemic in SSA. Factors hypothesized to modulate viral transmission and geographic spread
488 include climatic factors (e.g., specific humidity), access to prevention measures (e.g.,
489 handwashing), and human mobility (e.g., international and domestic travel). **Table S2** outlines
490 the dimensions of risk selected and references the previous studies relevant to the selection of
491 these factors.

492

493 Climate data was sourced from the global, gridded ERA5 dataset [47] where model data is
494 combined with global observation data (see **Section A7** for details).

495

496 International flight data was obtained from a custom report from OAG Aviation Worldwide (UK)
497 and included the departure location, airport of arrival, date of travel, and number of passenger
498 seats for flights arriving to 113 international airports in SSA (see **Section A5**).

499

500 As an estimate of connectivity within subregions of countries, the population weighted mean
501 travel time to the nearest city with a population greater than 50,000 was determined; details are
502 provided in **Section A6**. To obtain a set of measures that broadly represent connectivity within
503 different countries in the region, friction surfaces from ref[24] were used to obtain estimates of the
504 connectivity between different administrative level 2 units within each country. Details of this,
505 alongside the metapopulation model framework used to simulate viral spread with variation in
506 connectivity are in **Section A6**.

507

508 **Figure 2** in the main text shows variation among SSA countries for four of the variables; **Figure**
509 **S3** shows variation for all variables. **Figure 3** in the main text shows variation for a subset of the
510 comorbidity and access to care indicators as a heatmap; **Figure S4** shows variation for all the
511 variables (both also available online at the **SSA-SARS-CoV-2-tool**
512 (https://labmetcalf.shinyapps.io/covid19-burden-africa/)).

Table S2

Hypothesized dimensions of risk and expected direction of effect on SARS-CoV-2 transmission or burden in sub-Saharan Africa (SSA) relative to higher latitude countries

Dimension of risk	Factors hypothesized to **decrease** transmission or burden in sub-Saharan Africa relative to other geographic areas	Factors hypothesized to **increase** transmission or burden in sub-Saharan Africa relative to other geographic areas
(A) Demographic and socio-economic characteristics in SSA	Younger populations, and thus a smaller proportion of individuals in the older age groups that experience the highest mortality [2–4]	A larger proportion of urban populations living in dense settings, which may result in higher transmission [48]; higher contact with older individuals as a result of multi-generation households [14]
(B) Comorbidities in SSA	Lower rates of some comorbidities that have been associated with risk of worse outcomes, e.g., obesity [9,41]	Higher rates of NCDs such as hypertension or COPD[39], which are associated with worse outcomes; and a potential role for as yet undescribed interactions e.g., with anemia, or high prevalence infectious diseases
(C) Climate in SSA	Warmer, wetter climates on average driving reduced transmission [1,49]	
(D) Capacity to deploy prevention measures in SSA	Experience with previous outbreak response which may yield more rapid and nimble approaches to reducing transmission [21,22]	Lower access to handwashing [50,51] and other prevention options such as self-isolation [52], increasing transmission
		Subregions of countries with reduced governance infrastructure [53]
(E) Access to healthcare in SSA		Larger variation in access to and coverage of health systems [54] including fewer medical staff and facilities such as hospital beds [14] increasing burden
		Increased vulnerability to disruption of routine health services (e.g.,[31])
		Limited testing capacity [55] reducing the capacity to identify and interrupt chains of transmission
(F) Human mobility and travel in SSA	Fewer viral importations due to reduced frequency of international travel [29,30]	
	Decreased rate of internal spread due to less connectivity within countries [56]	

Table S3

Variables and data sources for indicators of SARS-CoV-2 epidemic risk in sub-Saharan Africa

ID	Variable	Source	Hypothesized association(s) with SARS-CoV-2 outcomes
(A) Demographic and socio-economic characteristics			
A1	Human population size; Proportion of population over age 50 (%) from the UN Population Division UNPOP (2019 revision) estimates of population by single calendar year (2020), age, and country	UN[38]	Morbidity and mortality observed to increase with age (e.g, [2–4])
A2	Subnational spatial variation in the distribution of the human population and age structure	WorldPop	
A3	Household size: Mean household size for households with an individual over age 50	DHS	Proxy for social contact rate for the elderly population at higher risk for SARS-CoV-2 morbidity and mortality [14]
A4	Proportion of households with an individual over age 50	DHS	
A5	Health life expectancy (HALE) at age 60 (years)	WHO	Proxy for baseline health status of elderly population
A6	Proportion of population below the poverty line (%)	World Bank	More severe clinical outcomes associated with poverty; A proxy for access to advanced care [57,58]
A7	Proportion of the urban population living in crowded, low quality housing (defined as households lacking one or more of the following conditions: access to improved water, access to improved sanitation, sufficient living area, and durability of housing) (%)	World Bank	Indicator of capacity for prevention (e.g., through handwashing); Transmission observed to increase with crowding [48]
A8	Gross domestic product (GDP) per capita	World Bank	Used in PCA analysis (see below) as an indicator of socio-economic status at the national level
A9	GINI index, a measure of inequality in the distribution of income	World Bank	
(B) Comorbidities: General and nutrition related non-communicable diseases (NCDs)			
B1	NCDs overall mortality per 100 000 popn, age-standardized	WHO	Indicator of NCD burden in population; Comorbidities increase probability of severe clinical outcomes
B2	Cardiovascular disease related mortality per age group (annual deaths attributable per 100,000 population)	GDB 2017[59]	
B3	Diabetes prevalence among ages 20-79 (%)	World Bank	Increases probability of severe clinical outcomes
B4	Diabetes related mortality per age group (annual deaths attributable per 100,000 population)	GDB 2017[59]	

(Table S3 continued)

ID	Variable	Source	Hypothesized association(s) with SARS-CoV-2 outcomes
(B) Comorbidities: General and nutrition related non-communicable diseases (NCDs)			
B5	Raised glucose prevalence, age-standardized (%)	WHO	Indicator of metabolic disease risk; Metabolic disease increases probability of severe clinical outcomes
B6	Raised blood pressure prevalence, age-standardized (%)	WHO	
B7	Raised cholesterol prevalence, age-standardized (%)	WHO	
B8	Overweight prevalence among adults, age-standardized (%)	WHO	
B9	Anemia prevalence among non-pregnant women (%)	WHO	Indicator of poor nutritional status; Poor nutritional status may increase probability of severe clinical outcomes
B10	Undernourishment prevalence (%)	WHO	
(B) Comorbidities: NCDs related to respiratory system and pollution			
B11	Annual mean PM2.5 exposure in urban areas (ug/m3)	WHO	Exposure to air pollution increases mortality [42]
B12	Lung, tracheal, and esophageal cancer mortality per 100 000 popn, age-standardized	WHO	Indicator of prevalence and management of chronic disease and inflammation affecting the respiratory tract
B13	Chronic respiratory diseases (excluding asthma) related mortality per age group (annual deaths attributable per 100,000 population)	GDB 2017[59]	
B14	COPD mortality per 100 000 popn, age-standardized	WHO	
(B) Comorbidities: Infectious diseases			
B15	Respiratory infections mortality per 100 000 popn, age-standardized	WHO	Indicator of prevalence and management of infectious disease affecting the respiratory tract
B16	TB incidence per 100 000 popn	World Bank	Indicator of susceptibility to respiratory infections and immune suppression
B17	HIV prevalence among ages 15-49 (%)	World Bank	Indicator of immunosuppressed population
(C) Climate			
C1	Seasonal change in specific humidity (in selected urban centers)	ERA5[47]	Transmission rate of coronaviruses may decline with humidity
(D) Capacity to deploy prevention measures			
D1	Proportion of urban popn with basic handwashing facilities with water and soap at home (%)	WHO	Handwashing observed to reduce infection rates for respiratory pathogens
D2	Proportion of the population with access to a handwashing station with soap and water in 2019	Ref[60]	

(Table S3 continued)

| ID	Variable	Source	Hypothesized association(s) with SARS-CoV-2 outcomes
(D) Capacity to deploy prevention measures			
D3	Proportion of 1 year olds receiving full immunization coverage (%)	WHO	Proxy for coverage of routine health services
D4	Reported number of completed tests reported for SARS-CoV-2 infection as of June 30, 2020	Africa CDC	Indicator of surveillance capacity
(E) Access to healthcare in SSA			
E1	Proportion of children with pneumonia symptoms taken to a health facility (%)	WHO	Proxy for access to medical care and care seeking
E2	Subnational spatial variation in the probability of seeking treatment for fever at public facilities	Ref[61]	
E3	Proportion of births attended by skilled staff (%)	World Bank	
E4	Nurses and midwives per 100 000 popn	World Bank	Indicators of treatment capacity
E5	Physicians per 100 000 popn	World Bank	
E6	Hospitals per 100 000 popn	World Bank	
E7	Hospital beds 100 000 popn	World Bank	
E8	Health expenditure per capita in (USD)	WHO	Proxy for health system resources; A significant predictor of intensive care unit (ICU) capacity [62]
E9	Proportion of health expenditures that are out-of-pocket (%)	WHO	
(F) Human mobility and travel: International			
F1	Estimated number of international passengers arriving at SSA airports from January-April 2020	OAG	Indicator of the timing and number of introductions of SARS-CoV-2
F2	Estimated number of international passengers arriving at SSA airports from January-April 2020 by SARS-CoV-2 status at departure location	OAG	

(Table S3 continued)

ID	Variable	Source	Hypothesized association(s) with SARS-CoV-2 outcomes
(F) Human mobility and travel: Domestic			
F3	National population-weighted mean travel time to the nearest city (national mean of indicator F4)	Ref[63]	Indicator of connectivity within countries; A proxy for the rate of human mobility
F4	Population-weighted mean travel time to the nearest city (population > 50,000) for administrative level 2 units	Ref[63]	
F5	Relative costs of travel between centroids of administrative level 2 derived from friction surfaces obtained by integrating data on travel infrastructure (Open Street Map, land cover types, etc).	Ref[24]	

533   **Figure S2**

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

535   Dotted vertical line shows regional median; solid vertical line shows regional mean. Note that most data comes from

536   2015-2019 (median = 2016, mean = 2014.62-2014.93).

537
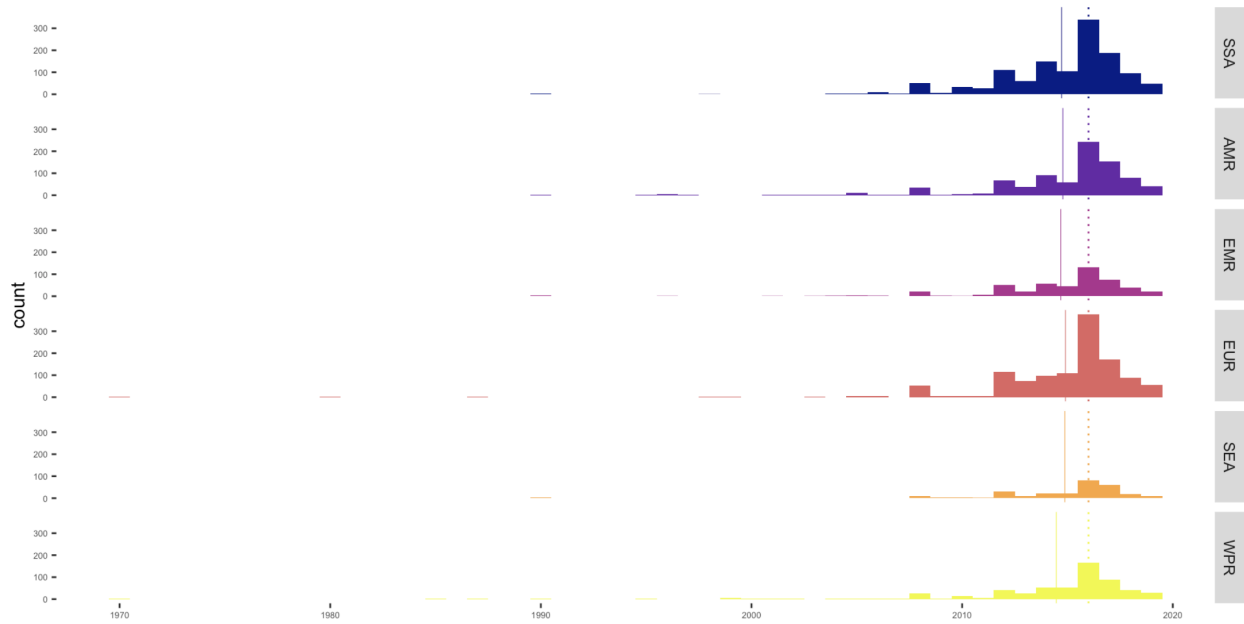
538
539
540
541

542 **Figure S3**

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

545 A subset of variables is shown in Figure 2A-D in the main text, the remaining variables are

546 shown in supplementary file "Figure S3 compiled.pdf" and available online: **SSA-SARS-CoV-2-**

547 **tool** (https://labmetcalf.shinyapps.io/covid19-burden-africa/)

Figure S4

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

A subset of variables is shown in Figure 3D-E in the main text, the remaining variables are shown and available 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 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)
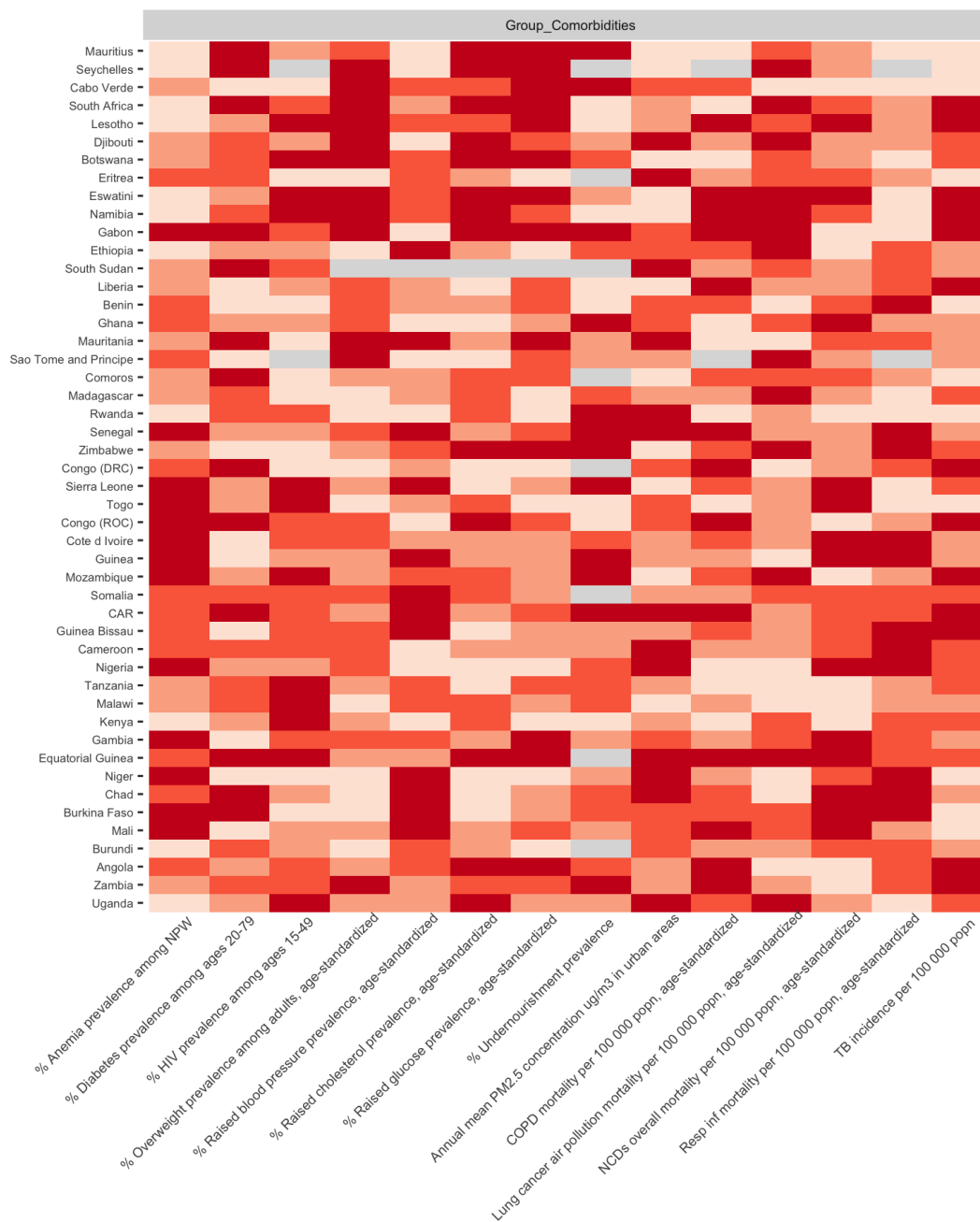
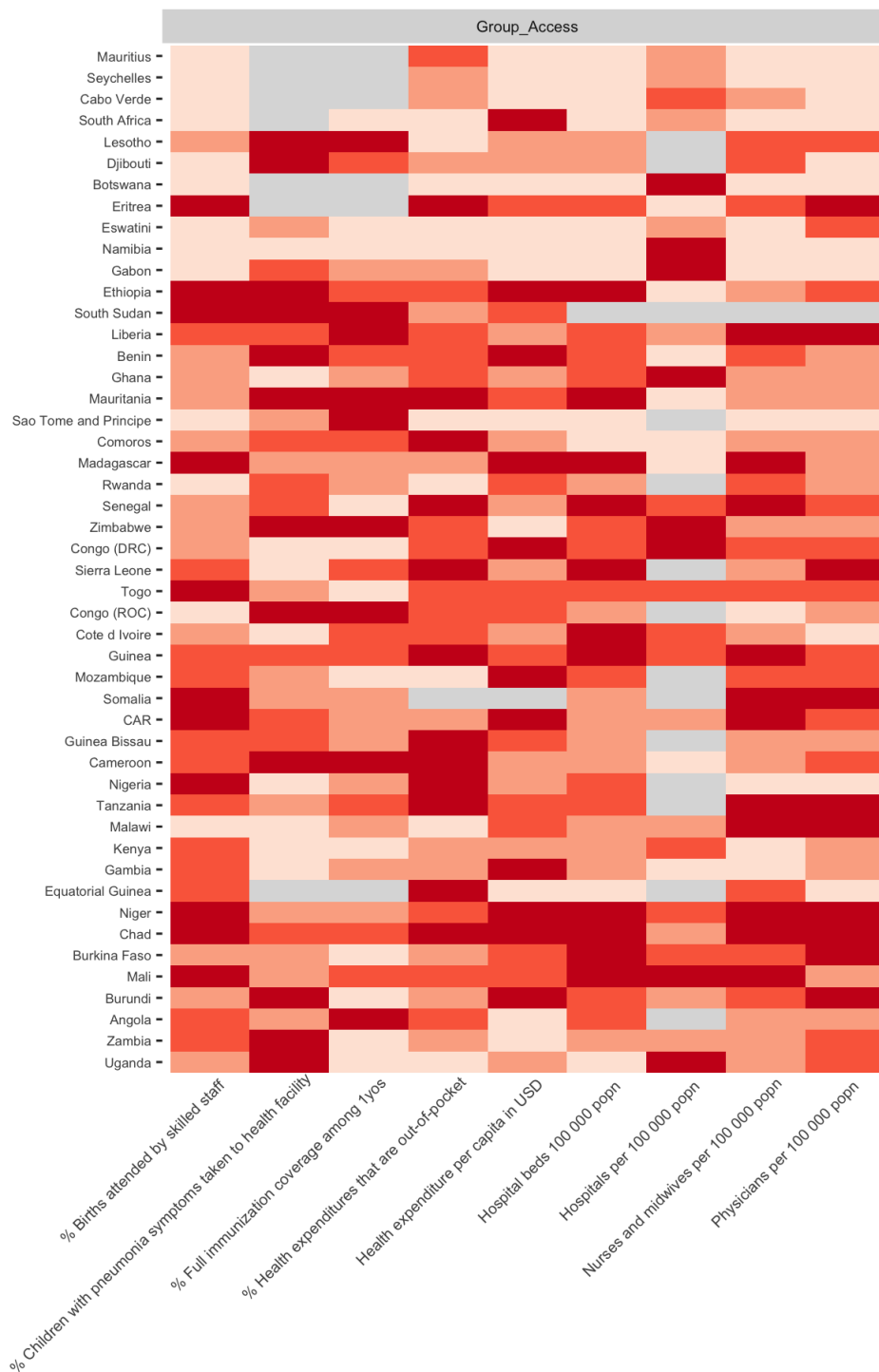559   **B**: Select national level indicators; estimates of reduced access to care (e.g., fewer hospitals) shown with darker red
560   for higher risk quartiles Countries missing data for an indicator (NA) are shown in gray. For comparison between
561   countries, estimates are age-standardized where applicable (see **Table S3** for details)

562

A3 | Principal component analysis (PCA) of variables considered

3.1 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.

3.2 Principal Component Analysis

The PCA was conducted on each of the three subsets described above, using the scikitlearn library [64]. 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 [65]. A link to the code for this analysis is available online at the **SSA-SARS-CoV-2-tool** (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).

607   *3.3 PCA Results*

608

609   The first two principal components from the analysis of 29 variables explain 32.6%, and 13.1%
610   the total variance, respectively, in the dataset. Countries with higher numbers of completed
611   SARS-CoV-2 tests reported tended to associate with an increase in principal component 1
612   (Pearson correlation coefficient, r = 0.67, p = 1.1e-7, **Figure S5A**). Similarly, high GDP
613   countries seem to associate with an increase in principal component 1 (Pearson correlation
614   coefficient, r = 0.80, p = 6.02e-12), **Figure S5B**). In contrast, countries with greater wealth
615   inequality (as measured by the GINI index) are associated with a decrease in principal
616   component 2 (Pearson correlation coefficient, r = -0.42, p = .0042, **Figure S5C**). Despite these
617   correlations, a relatively low percentage of variance is explained by each principal component:
618   for the 29 variables, 13 of the 29 principal components are required to explain 90% of the
619   variance (**Figure S5D**). When only the access to care subset of variables is considered, the first
620   two principal components explain 50.7% and 19.1% of the variance, respectively, and five of
621   eight principal components are required to explain 90% of the variance. When only the
622   comorbidities subset is considered, the first two principal components explain 27.9% and 17.8%
623   of the variance, respectively, and nine of 14 principal components are required to explain 90%
624   of the variance (**Figure S4D**).

625

626   *3.4 PCA Discussion*

627

628   These data suggest that inter-country variation in this dataset is not easily explained by a small
629   number of variables. Moreover, though correlations exist between principal components and
630   high-level explanatory variables (testing capacity, wealth), their magnitude is modest. These
631   results highlight that dimensionality reduction is unlikely to be an effective analysis strategy for
632   the variables considered in this study. Despite this overall finding, the PCA on the access to
633   care subset of variables highlights that the variance in these variables is more easily explained
634   by a small number of principal components, and hence may be more amenable to
635   dimensionality reduction. This finding is unsurprising as, for example, the number of hospital
636   beds per 100,000 population is likely to be directly related to the number of hospitals per
637   100,000 population (indeed r = 0.60, p = 5.7e-6 for SSA). In contrast, for comorbidities, the
638   relationship between different variables is less clear. Given the low percentages of variation
639   captured by each principal component, and the high variability between different types of
640   variables, these results motivate a holistic approach to using these data for assessing relative
641   SARS-CoV-2 risk across SSA.

Figure S5

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)

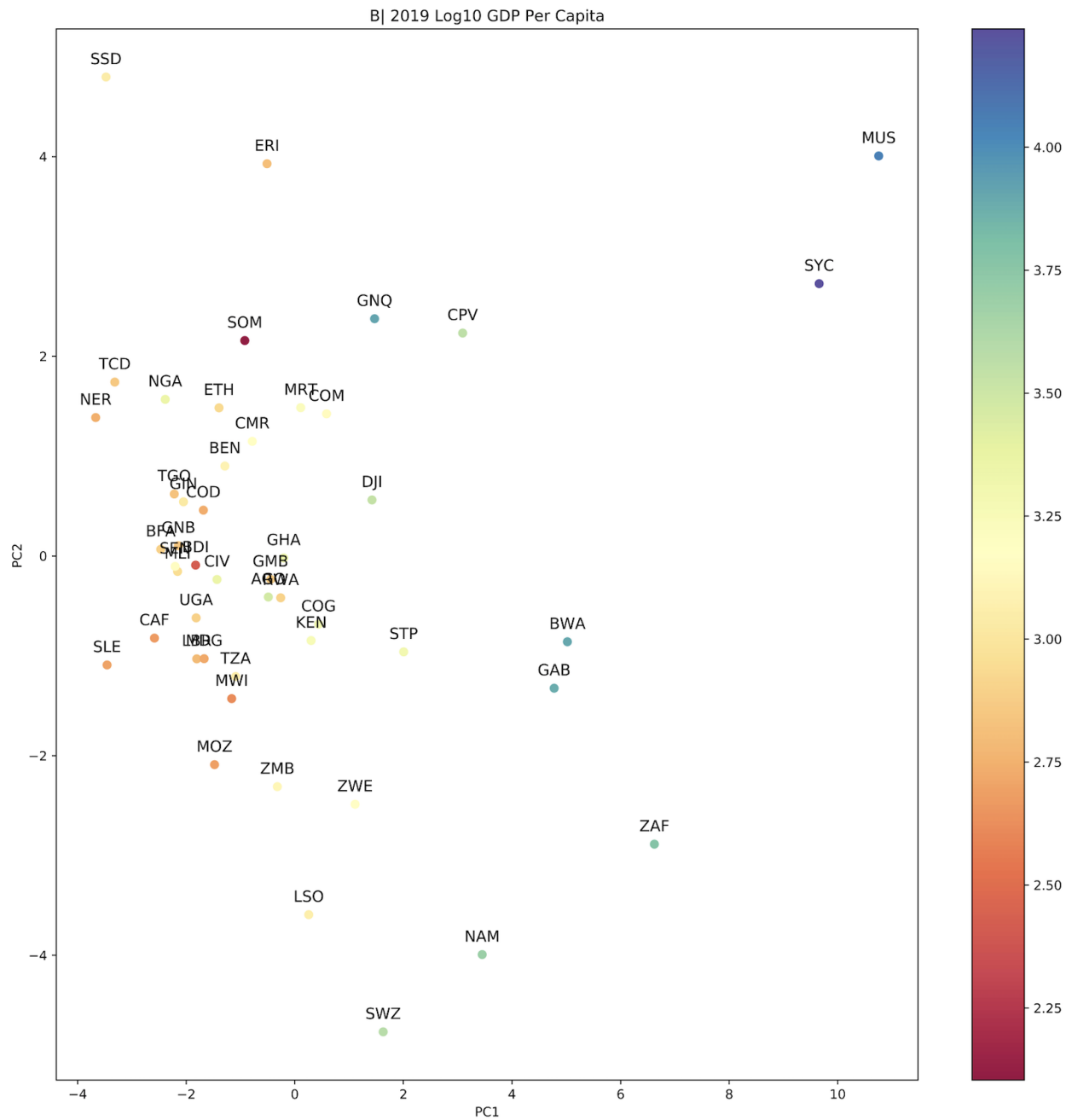A| Log10 of COVID-19 Tests per 100k Completed as of June 30
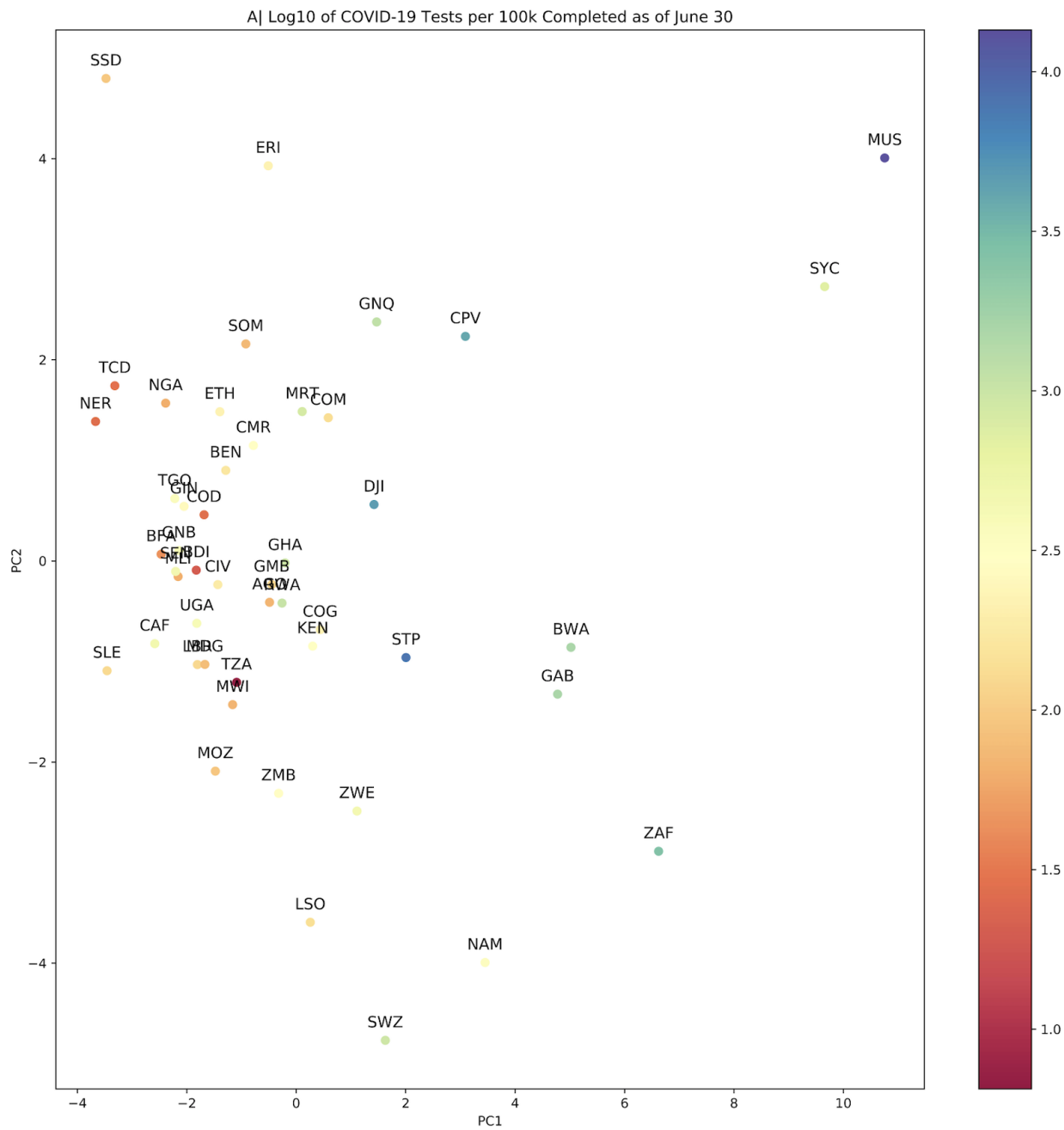
648   (Figure S5 continued)

649   **Figure S5**

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

651   **B:** Principal Component 1 and 2, countries colored by Log10 scaled GDP per capita

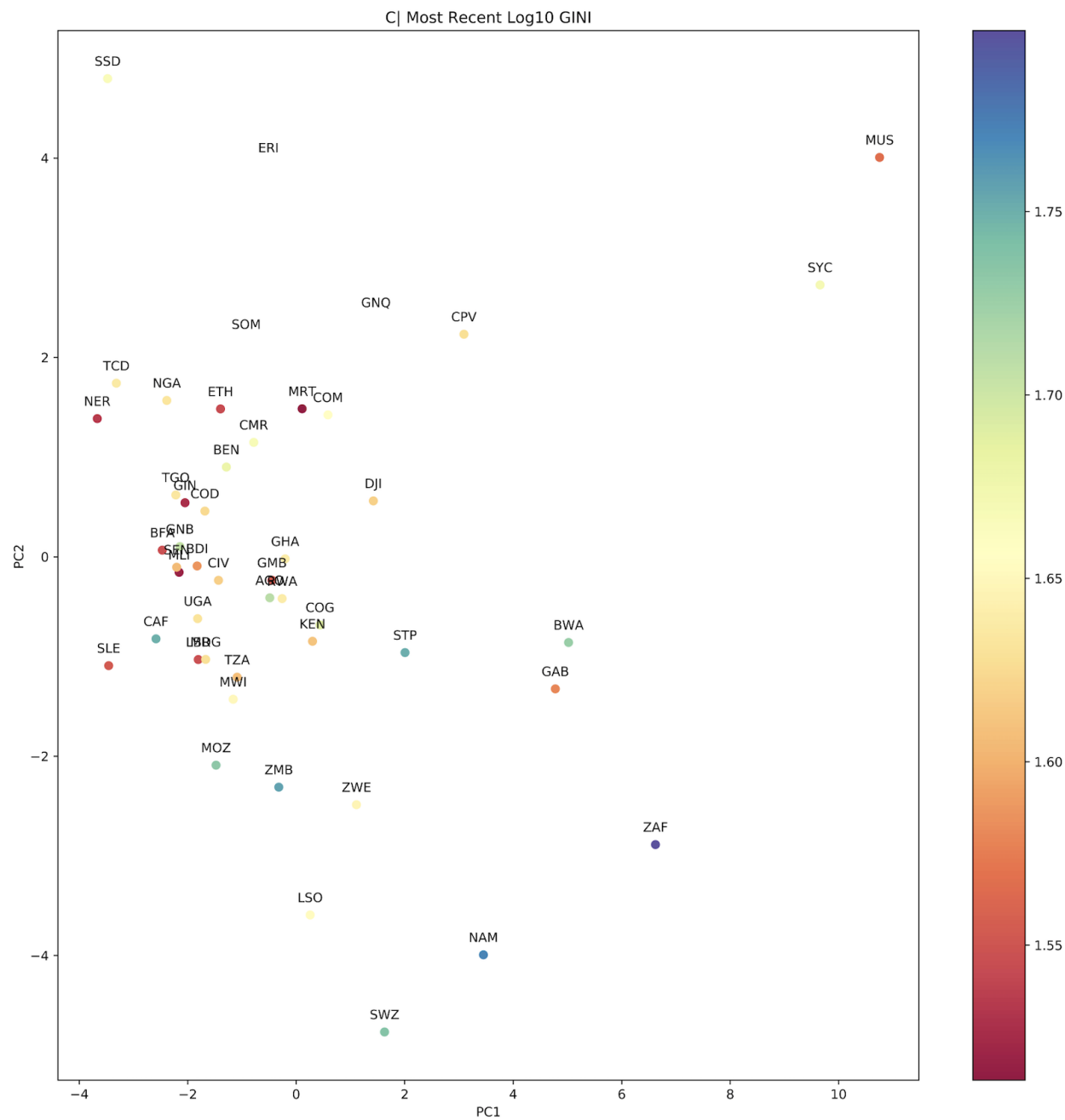B| 2019 Log10 GDP Per Capita

652

653

654

655   (Figure S5 continued)

656   **Figure S5**

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

658   **C**: Principal Component 1 and 2, countries colored by the GINI index (a measure of wealth disparity)

C| Most Recent Log10 GINI

659

660

661

662 (Figure S5 continued)

663 # Figure S5

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

665 **D**: Scree plot showing the cumulative proportion of variance explained by principal component for analysis done
666 using all variables (blue, 29 variables), comorbidity indicators (green, 14 variables, Section B in **Table S3**)), and
667 access to care indicators (orange, 8 variables, Section E in **Table S3**)

668

669
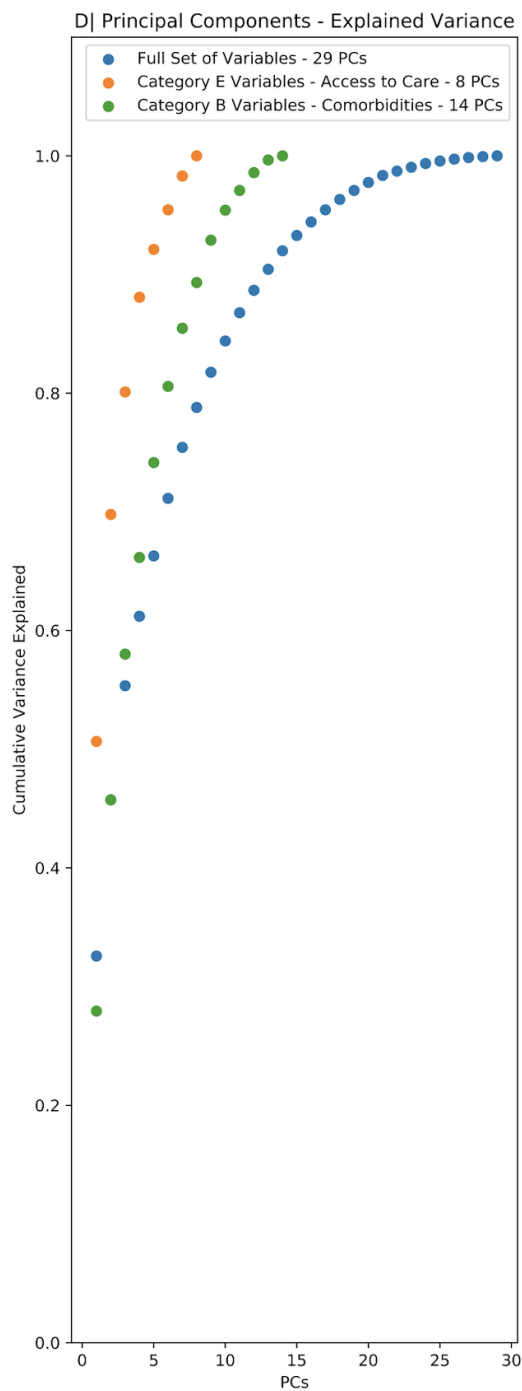
A4 | Evaluating the burden emerging from the severity of infection outcome

4.1 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 [66]. 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**).

Table S4: Sources of age-stratified IFR estimates

Study	Population	Methods
Salje et al. 2020 [2]	Deaths and hospitalizations due to COVID-19 in French public and private hospitals across the country between 13 March - 11 May	Combined data from France with data from Diamond Princess Cruise ship to estimate age-stratified IFR, case severity, and hospitalization probabilities accounting for asymptomatic cases and underascertainment.
Verity et al. 2020 [4]	Deaths due to COVID-19 in Hubei province, China	Combined data from Hubei with data from PCR testing of repatriated citizens under quarantine to estimate age-stratified IFR accounting for asymptomatic cases and underascertainment.
Rinaldi et al. 2020 [3]	Deaths due to COVID-19 reported in Lombardia, Italy	Analyzed deaths in the Lombardia region, one of the hardest hit regions in Italy, and used seroprevalence surveys of the region to estimate that 30% of the population was infected to estimate age-stratified IFR.

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 [67] in R version 4.0.2 [68]. 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 *IFRs* 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.

697   *4.2 Data sourcing: Comorbidities over age from IHME*

698

699   Applying these *IFR* estimates to the demographic structure of SSA countries provides a
700   baseline expectation for mortality, but depends on the assumption that mortality patterns in sub-
701   Saharan Africa will be similar to those from where the *IFR* estimates were sourced (France,
702   China, and Italy). Comorbidities have been shown to be an important determinant of the severity
703   of infection outcomes (i.e., *IFR*); to assess the relative risk of comorbidities across age in SSA,
704   estimates of comorbidity severity by age (in terms of annual deaths attributable) were obtained
705   from the Institute for Health Metrics and Evaluation (IHME) Global Burden of Disease (GBD)
706   study in 2017 [69]. Data were accessed through the GBD results tool for cardiovascular disease,
707   chronic respiratory disease (not including asthma), and diabetes, reflecting three categories of
708   comorbidity with demonstrated associations with risk (**Table S2**). We make the assumption that
709   higher mortality rates due to these NCDs, especially among younger age groups, is indicative of
710   increased severity and lesser access to sufficient care for these diseases - suggesting an
711   elevated risk for their interaction with SARS-CoV-2 as comorbidities. While there are significant
712   uncertainties in these data, they provide the best estimates of age specific risks and have been
713   used previously to estimate populations at risk[18].

714

715   The comorbidity by age curves for SSA countries were compared to those for the three
716   countries from which SARS-CoV-2 *IFR* by age estimates were sourced. Attributable mortality
717   due to all three NCD categories is higher at age 50 in all 48 SSA countries when compared to
718   estimates from France and Italy and for 42 of 48 SSA countries when compared to China
719   (**Figure S5**).

720

721   Given the potential for populations in SSA to experience a differing burden of SARS-CoV-2 due
722   to their increased severity of comorbidities in younger age groups, we explore the effects of
723   shifting *IFRs* estimated by the GAM of *IFR* estimates from France, Italy, and China younger by
724   2, 5, and 10 years (**Figure 3** in main text).

725 **Figure S6**

726 **Comorbidity burden by age in sub-Saharan Africa**

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

730
731
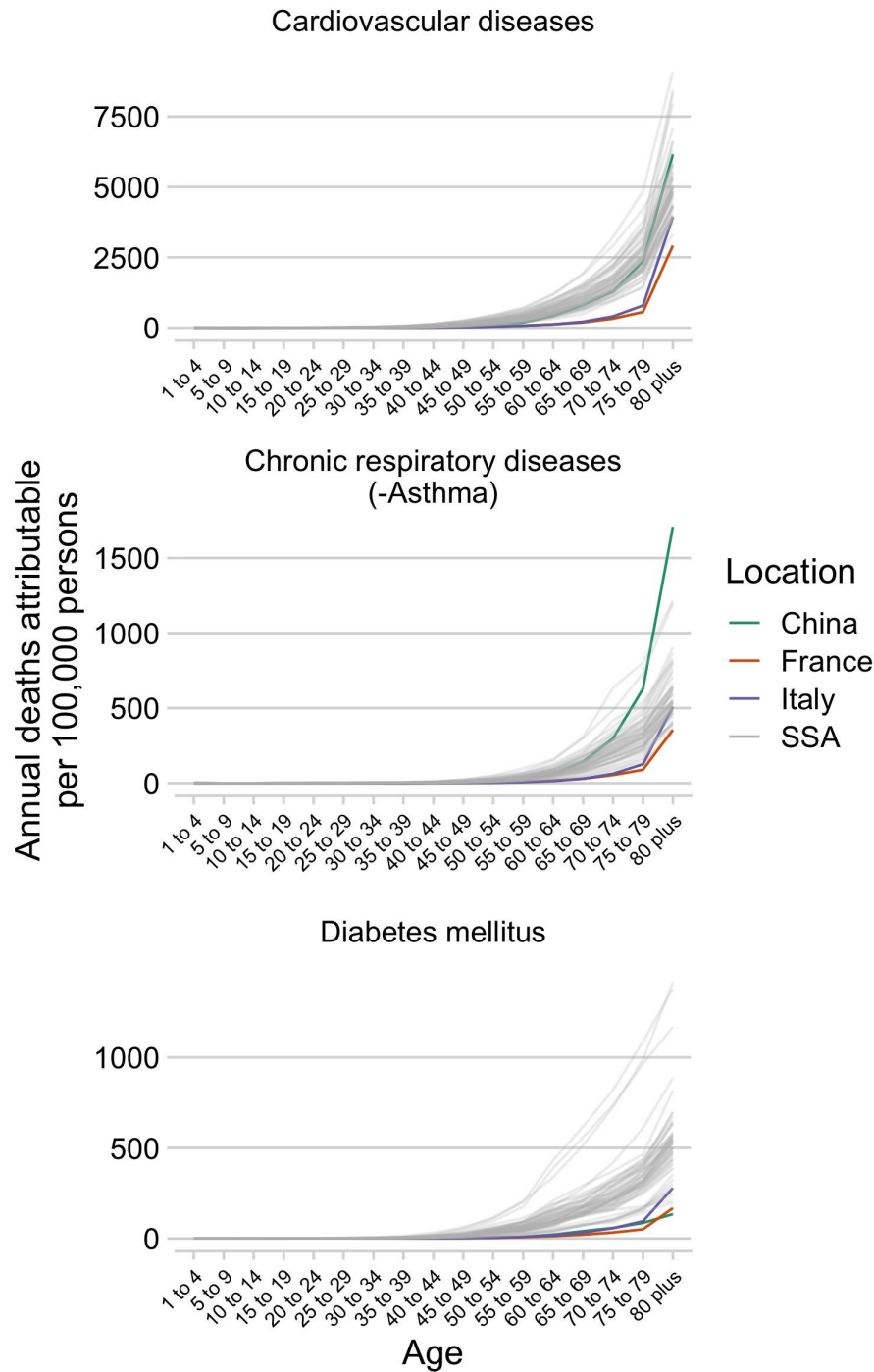
A5 | International Air Travel to SSA

The number of passenger seats on flights arriving to international airports were grouped by
country and month for January 2020 to April 2020 (**Table S5**) - the months when the
introduction of SARS-CoV-2 to SSA countries was likely to have first occurred. The first
confirmed case reported from a SSA country, per the Johns Hopkins Coronavirus Research
Center was in Nigeria on February 28, 2020. By March 31, 2020, 43 of 48 SSA countries had
reported SARS-CoV-2 infections and international travel was largely restricted by April. Lesotho
was the last SSA country to report a confirmed SARS-CoV-2 infection (on May 13, 2020);
however, given difficulties in surveillance, the first reported detections were likely delayed
relative to the first importations of the virus.

The probability of importation of the virus is defined by the number of travelers from each source
location each date and the probability that a traveler from that source location on that date was
infectious. Due to limitations in surveillance, especially early in the SARS-CoV-2 pandemic,
empirical data on infection rates among travelers is largely lacking. To account for differences in
the status of the SARS-CoV-2 pandemic across source locations, and thus differences in the
importation risk for travelers from those locations, we coarsely stratified travelers arriving each
day into four categories based on the status of their source countries:

   i.   Travelers from countries with zero reported cases (i.e., although undetected
         transmission was possibly occurring, SARS-CoV 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 [70] 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

776   source location). Consequently, the risk for viral importation is likely systematically
777   underestimated. However, as the relative risk for viral importation will still scale with the number
778   of travelers, comparisons among SSA countries can be informative (e.g., SSA countries with
779   more travelers from countries with confirmed SARS-CoV-2 transmission are at higher risk for
780   viral importation).
781
782

783   **Table S5**
784   **Arrivals to SSA airports by the number of passenger seats and status of the SARS-CoV-2**
785   **pandemic at the origin at the time of travel**
786
787   (see csv file: "Table S5 International Airtravel to SSA.csv")
788
789   Data Fields:
790
791   1.   country: Name of country
792   2.   n_airports: Number of airports with flight data
793   3.   month: January, February, March, April 2020; or total for all 4 months
794   4.   total_n_seats: Total number of passenger seats on arriving aircraft
795   5.   From source with cases > 0: Number of passengers arriving from source locations with 1 or more reported
796   SARS-CoV-2 infection by the date of travel
797   6.   From source with cases > 100: Number of passengers arriving from source locations more than 100
798   reported SARS-CoV-2 infection by the date of travel
799   7.   From source with cases > 1000: Number of passengers arriving from source locations with more than 1000
800   reported SARS-CoV-2 infection by the date of travel
801
802

803 ## A6 | Subnational connectivity among countries in sub-Saharan Africa

804

805 *6.1. Indicators of subnational connectivity*

806

807 To allow comparison of the relative connectivity across countries, we use the friction surface
808 estimates provided by Weiss et al.[24] as a relative measure of the rate of human movement
809 between subregions of a country. For connectivity within subregions of a country (e.g., transport
810 from a city to the rural periphery), we use as an indicator the population weighted mean travel
811 time to the nearest urban center (i.e., population density > 1,500 per square kilometer or a
812 density of built-up areas > 50% coincident with population > 50,000) within administrative-2
813 units [63]. For some countries, estimates at administrative-2 units were unavailable (Comoros,
814 Cape Verde, Lesotho, Mauritius, Mayotte, and Seychelles); estimates at the administrative-1
815 unit level were used for these cases (these were all island nations, with the exception of
816 Lesotho).

817

818 *6.2. Metapopulation model methods*

819

820 Once SARS-CoV-2 has been introduced into a country, the degree of spread of the infection
821 within the country will be governed by subnational mobility: the pathogen is more likely to be
822 introduced into a location where individuals arrive more frequently than one where incoming
823 travellers are less frequent. Large-scale consistent measures of mobility remain rare. However,
824 recently, estimates of accessibility have been produced at a global scale [24]. Although this is
825 unlikely to perfectly reflect mobility within countries, especially as interventions and travel
826 restrictions are put in place, it provides a starting point for evaluating the role of human mobility
827 in shaping the outbreak pace across SSA. We use the inverse of a measure of the cost of travel
828 between the centroids of administrative level 2 spatial units to describe mobility between
829 locations (estimated by applying the costDistance function in the *gdistance* package in R to the
830 friction surfaces supplied in ref[24]). With this, we develop a metapopulation model for each
831 country to develop an overview of the possible range of trajectories of unchecked spread of
832 SARS-CoV-2.

833

834 We assume that the pathogen first arrives into each country in the administrative 2 level unit
835 with the largest population (e.g., the largest city) and the population in each administrative 2
836 level (of size N_j) is entirely susceptible at the time of arrival. We then track spread within and
837 between each of the administrative 2 level units of each country. Within each administrative 2
838 level unit, dynamics are governed by a discrete time Susceptible (S), Infected (I) and Recovered
839 (R) model with a time-step of ~ 1 week, which is broadly consistent with the serial interval of
840 SARS-CoV-2. Within the spatial unit indexed *j*, with total size N_j, dynamics follow:

841

842 $$I_{j,t+1} = \beta I^{\alpha}_{j,t} S_{j,t} / N_j + \iota_{j,t}$$

843 $$S_{j,t+1} = S_{j,t} - I_{j,t+1} + b$$

844

845 where β captures the magnitude of transmission over the course of one serial interval (and is set
846 to 2.5 to approximately represent the R_0 of SARS-CoV-2); the exponent $\alpha = 0.97$ is used to
847 capture the effects of discretization[71], $\iota_{j,t}$ captures the introduction of new infections into site j at
848 time t, and b reflects the introduction of new susceptible individuals resulting from the birth rate,
849 set to reflect the most recent estimates for that country from the World Bank Data Bank
850 (https://data.worldbank.org/indicator/SP.DYN.CBRT.IN).

851

852 We make the simplifying assumption that mobility linking locations i and j, denoted $c_{i,j}$, scales
853 with the inverse of the cost of travel between sites i and j evaluated according to the friction
854 surface provided in [24]. The introduction of an infected individual into location j is then defined by
855 a draw from a Bernouilli distribution following:

856
$$\iota_{j,t} \sim Bern(1 - exp(-\sum_{1}^{L} c_{i,j} I_{i,t}/N_i))$$

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

860

861 Some countries show rapid spread between administrative units within the country (e.g., a
862 country with parameters that broadly reflect those available for Malawi, **Figure S7**), while in
863 others (e.g., reflecting Madagascar), connectivity may be so low that the outbreak may be over
864 in the administrative unit of the largest size (where it was introduced) before introductions
865 successfully reach other poorly connected administrative units. The result is a hump shaped
866 relationship between the fraction of the population that is infected after 5 years and the time to
867 the first local extinction of the pathogen (**Figure S7**, right top). In countries with lower
868 connectivity (e.g., that might resemble Madagascar), local outbreaks can go extinct rapidly
869 before travelling very far; in other countries (e.g., that might resemble Gabon), the pathogen
870 goes extinct rapidly because it travels rapidly and rapidly depletes susceptible individuals
871 everywhere.

872

873 The impact of the pattern of travel between centroids is echoed by the pattern of travel within
874 administrative districts: countries where the pathogen does not reach a large fraction of the
875 administrative 2 units within the country in 5 years are also those where within administrative
876 unit travel is low (**Figure S7**, right bottom).

877

878 These simulations provide a window onto qualitative patterns expected for subnational spread
879 of the pandemic virus, but there is no clear way of calibrating the absolute rate of travel between
880 regions of relevance for SARS-CoV-2. Thus, the time-scales of these simulations should be
881 considered in relative, rather than absolute terms. Variation in lockdown effectiveness, or other
882 changes in mobility for a given country may also compromise relative comparisons. Variability in
883 case reporting complicates clarifying this (**Figure S8**).

884 **Figure S7**

885 **Pace of the outbreak**

886 Each grey line on the left hand panels indicates the total infected across all administrative units in a metapopulation
887 simulation with parameters reflecting the country indicated by the plot title, assuming interventions are constant.
888 Increases after the first peak indicate the pathogen reaching a new administrative 2 unit. In Malawi-like settings
889 (higher connectivity), more administrative units are reached rapidly, whereas in Madagascar-like settings (lower
890 connectivity), a lower proportion of the administrative units are reached by a given time, as fewer introductions occur
891 before the outbreak has burned out in the administrative 2 unit with the largest population. More generally, rapid
892 disappearance of the outbreak (top right hand plot, y axis shows time to extinction) could either indicate rapid spread
893 with a high proportion of the countries' population reached (top right hand plot, x axis) or slow spread, with many
894 administrative units unreached, and therefore remaining susceptible. The pattern of between-administrative unit travel
895 also echoes travel time within administrative units (lower panel, right hand side, x axis indicates fraction of
896 administrative units unreached, and upper panel indicates travel time in hours to the nearest city of 50,000 or more
897 people).
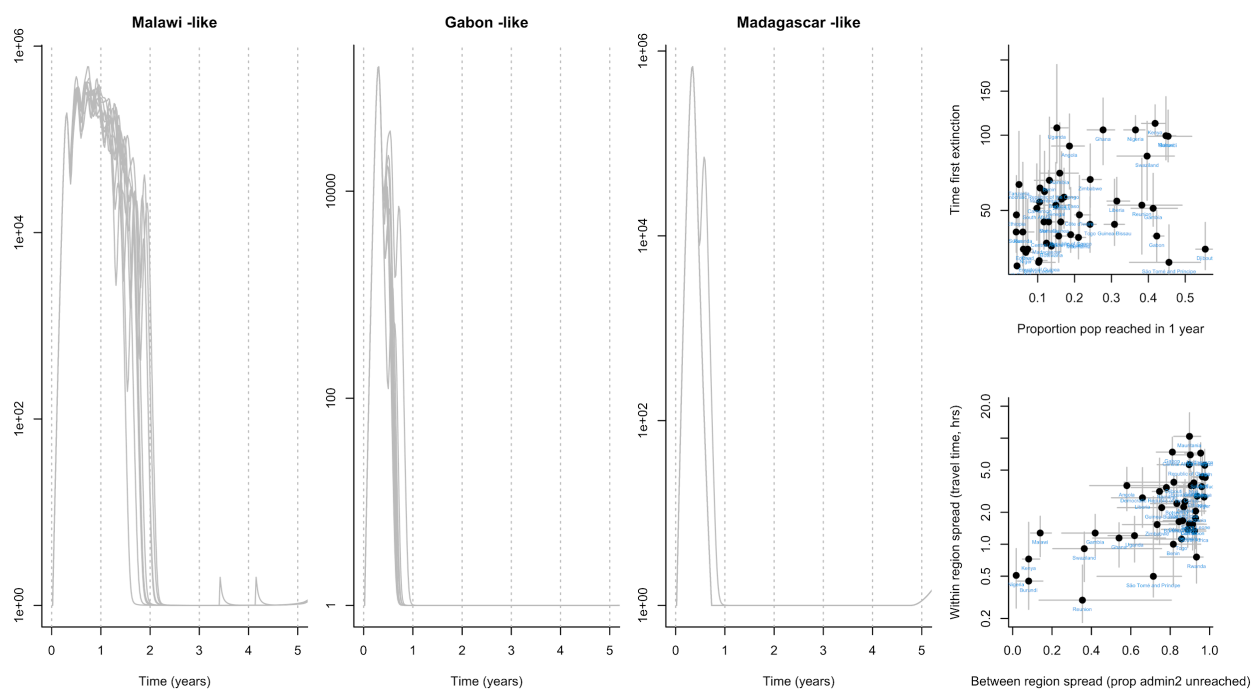
898
899

900 ## Figure S8

901 ### Cases and testing vs. the pace of the outbreak
902 The total number of confirmed cases reported by country (x axis, left, as reported for June 28th by Africa CDC) and
903 the test positivity (x axis, right, defined as the total number of confirmed cases divided by the number of tests run, as
904 reported by Africa CDC, likewise) show no significant relationship with the proportion of the population estimated to
905 be infected after one year using the metapopulation simulation described in A6 (respectively, $\rho = -0.04, p > 0.5, df = 41$
906 41 and $\rho = 0.02, p > 0.5, df = 41$). All else equal, a positive relationship is expected; however, both uncertainty in
907 case numbers, and uncertainty associated with the simulation might both drive the absence of a signal.
908

909
910
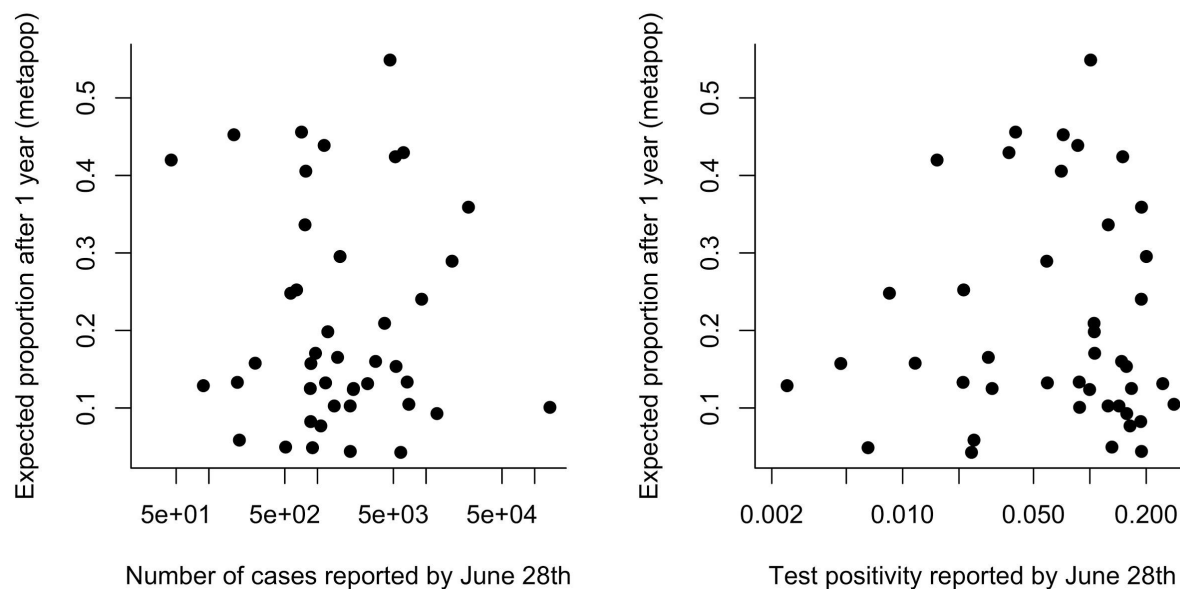
A7 | Modeling epidemic trajectories in scenarios where transmission rate depends on climate

7.1 Climate data sourcing: Variation in humidity in SSA

Specific humidity data for selected urban centers comes from ERA5 using an average climatology (1981-2017)[47]; we do not consider year-to-year climate variations. Selected cities (*n* = 58) were chosen to represent the major urban areas in SSA. The largest city in each SSA country was included as well as any additional cities that were among the 25 largest cities or busiest airports in SSA.

7.2 Methods for climate driven modelling of SARS-CoV-2

We use a climate-driven SIRS (Susceptible-Infected-Recovered-Susceptible) model to estimate epidemic trajectories (i.e., the time of peak incidence) in different cities in 2020, assuming no control measures are in place [23,72]. The model is given by:

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

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

where S is the susceptible population, I is the infected population and N is the total population. D is the mean infectious period, set at 5 days following ref[23,49]. To investigate the maximum possible climate effect, we use parameters from the most climate-dependent scenario in ref[23], based on betacoronavirus HKU1. 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).

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:

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

where $q(t)$ is specific humidity [47] and a is set at -227.5 based on estimated HKU1 parameters [23]. R_{0max} and R_{0min} are 2.5 and 1.5 respectively. We assume the same time of introduction for all cities, set at March 1st, 2020 (consistent with the first reported cases in SSA, **Figure S1D**)