

Regional comparisons of COVID reporting rates, burden, and mortality age-structure using auxiliary data sources

Mollie M. Van Gordon^{a,1}, Lawrence Mwananyanda^{b,c}, Christopher J. Gill^b, and Kevin A. McCarthy^a

^aInstitute for Disease Modeling at the Bill & Melinda Gates Foundation; ^bDepartment of Global Health, Boston University School of Public Health; ^cRight To Care – Zambia

This manuscript was compiled on August 18, 2021

We correct common assumptions about COVID burden and disease characteristics in high-income (HIC) versus low- and middle-income (LMIC) countries by augmenting widely-used surveillance data with auxiliary data sources. We constructed an empirically-based model of serological detection rates to quantify COVID reporting rates in national and sub-national locations. From those reporting rates, we estimated relative COVID burden, finding results that contrast with estimates based on case counts and modeling. To investigate COVID mortality by age in an LMIC context, we utilized a unique morgue study of COVID in Lusaka alongside the population attributable fraction method to account for HIV comorbidity. We calculated the comorbidity-corrected age-adjusted mortality curve in Lusaka and found it significantly skewed toward younger age groups as compared to HICs. This unexpected result recommends against the unexamined use of HIC-derived parameterizations of COVID characteristics in LMIC settings, and challenges the hypothesis of an age-structure protective factor for COVID burden in Africa. Indeed, we found overall COVID burden to be higher in Lusaka than in HICs. Concurrent with high COVID burden, many LMICs have high prevalence of other public health issues such as HIV, which compete for limited health investment resources. Given differences in age-structure, comorbidities, and healthcare delivery costs, we provide a case study comparing the cost efficacy of investment in COVID versus HIV and found that even in a high HIV prevalence setting, investment in COVID remains cost-effective. As a whole, these analyses have broad implications for interpretations of COVID burden, modeling applications, and policy decision-making.

COVID-19 | Surveillance | Reporting rates | Disease burden | Mortality age-structure

Effective counter-measures against the raging COVID-19 pandemic require accurate geographically-specific knowledge, but surveillance and data inequities present major challenges to these efforts. Surveillance systems are generally more robust in high-income countries (HICs), meaning that data from HICs tend to be more readily available and complete than data from low- and middle-income countries (LMICs) (1, 2). For lack of locally-specific data, epidemiological characteristics of COVID are often based on data from HICs, but the transferability to LMIC contexts remains an open question (3). Accounting for differences in surveillance and incorporating auxiliary data sources can help fill these data gaps and inform our understanding of COVID across contexts.

For example, official statistics on regional COVID burden are based on reported case counts (4), despite evidence of substantial case underreporting particularly in LMIC contexts (5, 6). Such differences in reporting rates can significantly alter estimates of relative disease burden across regions (7). COVID reporting rates are difficult to determine, but incorporating

serological data can inform reporting rate estimates. While serology is challenging to work with, it offers some of the best information in data-sparse settings if the limitations of serological data are accounted for (8, 9).

Surveillance and reporting influence our understanding of COVID dynamics in other ways as well. For example, while there is strong evidence that COVID parameters such as infection fatality rate (IFR) vary even within HICs, data challenges in LMICs mean that HIC estimates are often used in LMIC settings (10). This practice has major implications for estimates of COVID burden and risk factors, and subsequently for policies and public health practices targeting COVID. The common understanding of IFRs and the age-structure of COVID mortality has led to hypotheses that Africa's young population distributions have a protective effect against COVID (11), yet questions remain about impacts of comorbidity distributions, differences in disease characteristics across settings, and the role of COVID interventions in the context of other public health concerns.

We address these questions and assumptions through auxiliary data sources. Using serological modeling, we calculate reporting rates for different national and sub-national locations in HICs and LMICs. We then use that data in a reporting rate model to adjust national burden estimates accounting for differences in surveillance. Taking a closer look at a local

Significance Statement

The analyses presented here demonstrate the power of auxiliary COVID data sources to fill information gaps, particularly for LMICs. Our results reveal differences in COVID surveillance and disease dynamics between HICs and LMICs that challenge common perceptions and assumptions about COVID in these respective contexts. We show the divergence of COVID reporting rates between HICs and LMICs and the effects on relative estimated burden. Contradicting common modeling practices, our analysis demonstrates that the age-structure of COVID mortality cannot be accurately generalized from HICs to LMICs. We find higher COVID burden in LMIC contexts than HICs particularly in younger age groups and show that investment in COVID is cost-effective even in light of other public health concerns.

Author contributions are as follows. MVG: conceptualization, literature search, data curation, formal analysis, investigation, methodology, project administration, software, validation, visualization, writing of the original draft, review and editing. LM & CG: writing review and editing. KM: conceptualization, methodology, review and editing.

The authors declare no competing interests.

¹To whom correspondence should be addressed. E-mail: mvangordon@idmod.org

context, we use morgue sampling data to circumvent issues introduced by sampling bias, and statistical approaches to adjust for comorbidity. With this age-disaggregated data we compare mortality age-structure and burden between LMIC and HIC contexts. Cognizant of competing public health issues, particularly in LMIC contexts, we examine relative costs and benefits of investing in COVID vaccine.

In this article, we first present our results on reporting rates across locations and relative regional COVID burden informed by serology. Next, we use auxiliary data from Zambia to examine mortality age-structure in an LMIC-specific location, and conduct an impact comparison of COVID intervention with other public health interventions. We follow with a discussion describing broader implications of our study results for programmatic and modeling applications. Finally, we describe in detail our data and methods, including serological and reporting rate modeling, mortality burden quantification, statistical correction of COVID mortality for comorbidity factors, and the cost modeling used to compare public health investments.

Results

Serology across World Health Organization (WHO) regions.

Reporting rates in the European region and the region of the Americas (EURO and AMRO) are higher than elsewhere. Matching seroprevalence estimates with reported cases in a particular location allows calculation and comparison of reporting rates (Figure 1). We take the cumulative reported cases for a location, up until the time of the serostudy, and adjust the case rate by a model of seroconversion and reversion. This case rate is lower than the raw reported cumulative case rate, as serosurveillance does not detect all infections. Reporting rates follow the dashed gray lines on the plot. In general, countries in the AMRO and EURO regions have the highest reporting rates, shown on a log scale. Note that seroprevalence cannot be directly compared across locations because of the differences in the dates of the serostudies.

Because serostudy data from the African region (AFRO) are sparse, we include sub-national prevalence estimates, and match those with case data from the particular sub-national locations only. Kenya and Zambia have data for multiple sub-national locations, and in both countries reporting rates vary by more than a factor of ten. While there is a clear correlation between seroprevalence and reported cases, there are notable exceptions that appear to align with national wealth. For example, Niger State and Cape Town both have similar seroprevalence rates, but differ a thousand-fold in reported cases. South Africa is a country with considerable resources, expertise, and infrastructure for disease surveillance, and there is reason to expect that capacity to test and report COVID cases is likely better in Cape Town than across Niger State. A similar argument can be hypothesized for other low income countries on this matrix.

Ranked estimated infections show lowest burden in EURO. To identify the location-specific effects of reporting rates on reported case burden, we estimated infection burden across locations on the same date. Serostudies provide estimates of reporting rates for a specific window of time, but we don't assume reporting rates are constant in time. To address this challenge, we used our point estimates of reporting rates to build an model for

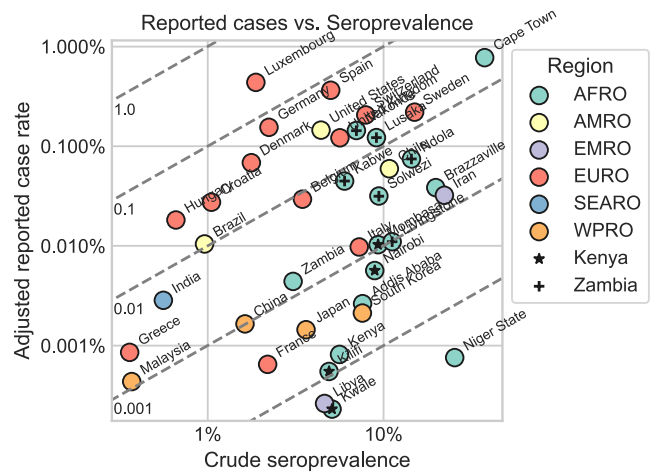


Fig. 1. Adjusted reported cases vs. crude seroprevalence by location. Dashed gray lines indicate reporting rates, labeled on the inside of the y-axis. WHO region is shown in color; sub-national locations in Kenya and Zambia are indicated with symbols.

the relationship between reporting rate and testing rate, for which we have continuous time series data. We then used the testing rate time series to model dynamic reporting rates and unify the dates of estimated infections.

Figure 2 shows a ranking of COVID burden across locations on November 12th, 2020, the most recent date with continuous available testing data for all locations. The top plot shows the ranking based on reported cases; the bottom plot shows the ranking of estimated infections calculated using reporting rates. In the reported case burden ranking, the EURO region is high relative to other regions; however in the estimated infections ranking, EURO countries have low burden relative to other regions.

We note that increased estimated cases relative to reported cases occurs mainly, but not exclusively, in lower resource country settings. As with the analysis presented in Figure 1, these data support the concern that underreporting, perhaps as a consequence of resource limitations, may lead to a significant under-counting of cases in certain parts of the world.

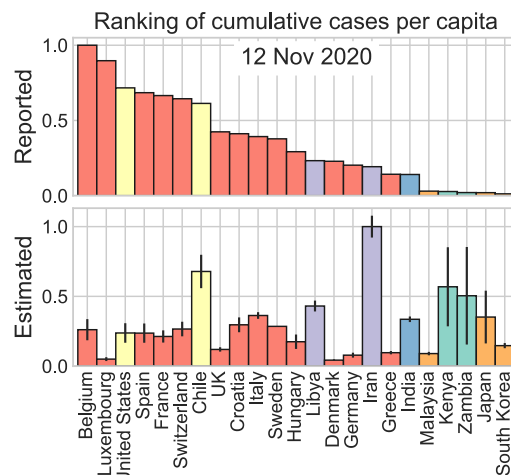


Fig. 2. Ranking of cumulative COVID reported cases (top) and estimated infections (bottom) as of November 12th, 2020.

123 COVID mortality in Lusaka.

124 **COVID mortality burden is substantial among younger age groups in**
125 **Lusaka and higher overall as compared to HICs.** Based on a morgue
126 study in Lusaka, Figure 3, top left, shows the percentage of
127 all-cause deaths that tested positive for COVID postmortem,
128 adjusted to exclude HIV-attributable comorbid deaths. The
129 frequency of COVID deaths by age-bin hovers between 9% and
130 19% for all age groups up to age 70. COVID death frequencies
131 in the 70s and 80+ age bins are 24% and 34%, respectively.
132 Using age structures of population and life expectancy, COVID
133 mortality frequencies can be translated into years of life lost
134 (YLL) by age group, which we present for Lusaka alongside
135 the United States for a comparison of LMIC and HIC contexts
136 (Figure 3, top right). While as expected, YLL per capita due
137 to COVID is highest in older age groups in the US, the pattern
138 in Lusaka is drastically different: young age groups, instead
139 of older age groups, suffer the highest COVID burden. In
140 addition, total years of life lost per capita across all age-bins
141 is higher overall in Lusaka than in the United States.

142 Plotted in Figure 3, bottom right, are age-adjusted COVID
143 mortality rates in Lusaka vs. HICs (log scale). In all age-
144 groups, age-adjusted COVID mortality rates are higher in
145 Lusaka than in HICs. Fitting an exponential model to each
146 age-adjusted mortality curve provides an estimate of the rate
147 of increase of mortality with age. The slope of this model
148 is defined as the growth constant of the curve, plotted for
149 multiple locations in Figure 3, bottom left. The significantly
150 lower growth rate of mortality with age in Lusaka indicates
151 age-adjusted mortality is skewed toward higher mortality in
152 younger age groups as compared to HICs. While Lusaka is not
153 necessarily representative of all LMICs, these results highlight
154 the pitfall of generalizing COVID disease parameters from
155 HIC to LMIC contexts.

156 Costing mortality reduction.

157 **Investment in COVID mortality reduction should be prioritized even**
158 **in the context of other public health challenges.** Building on the
159 age-specific mortality dataset from Lusaka, we used basic cost
160 modeling to demonstrate the utility of investment in COVID
161 mortality reduction. As HIV is the number one cause of death
162 in Zambia, we examine HIV investment alongside COVID
163 (12). We include for comparison the United States, a low HIV-
164 burdened HIC setting. Based on the age structures of population
165 as well as COVID and HIV mortality, we calculate total YLL
166 per capita for different levels of combined mortality reduction
167 of COVID and HIV across age groups. The heat maps in
168 Figure 4 show these results for Zambia (left) and the United
169 States (right). Model input data are detailed in the table in
170 Figure 4. Costing model results are shown as white isotropic
171 lines on the heat maps, indicating cost per percent mortality
172 reduction for COVID and HIV combined.

173 In both Zambia and the United States, the cost of HIV
174 mortality reduction outstrips that of COVID mortality reduc-
175 tion. As a result, the gradients across the cost model isotropic
176 lines are dominated by HIV mortality reduction in both plots.
177 The underlying structure of YLL per capita, however, differs
178 between the two countries. In the United States, the mag-
179 nitude of COVID-attributed YLL per capita is twice that of
180 HIV-attributed YLL per capita, as annotated by the white
181 circles on the right plot. As a result, the gradient direction

of YLL per capita in the United States is dominated by the
COVID mortality reduction axis. In Zambia, HIV as the
number one cause of death might be expected to dominate
the gradient of YLL per capita, but COVID-attributed YLL
per capita is nearly 70% that of HIV. This skews the gradi-
ent direction of total combined YLL per capita to a roughly
diagonal orientation, from lower left to upper right.

The overlay of the cost/percent mortality reduction
isotropic lines and the YLL per capita heat map indicate the
cost efficacy of investment in HIV vs. investment in COVID.
In both the United States and Zambia on a constant per capita
budget, maximizing investment in COVID mortality reduction
relative to HIV investment minimizes YLL per capita, the
desired outcome. Note that there are many components to
decision-making about public health investment, and we do not
claim that our model results establish exact cost for mortality
reduction or that COVID investment is definitively indicated.
Rather, this analysis serves to unseat any a priori assumptions
that COVID burden is insignificant in settings with other
high-burden public health issues. Rather, COVID investment
should be considered as a possible avenue for cost-effective
reduction in total disease burden.

Discussion

Using surveillance data auxiliary to reported COVID cases
and deaths, we demonstrate that common assumptions about
regional COVID burden must be reconsidered. We calculate
reporting rates across locations and show the impact of regional
differences on perceived COVID burden. Further, contrary
to impressions derived from case counts, we establish higher
burden of COVID in the African context as compared to
HICs, particularly in younger age groups. This challenges
predominant assumptions about the age structure of mortality
rates and the protective effects of younger populations in
Africa (3, 8, 13).

Combining seroprevalence data with seroconversion and
reversion modeling, we calculate reporting rates across WHO
regions and present estimated infections across locations. Our
analysis shows high reporting rates in EURO and AMRO
relative to other regions, a data-based result consistent with
anecdotal understanding (1, 7). Heterogeneity of reporting
rates at the sub-national level adds to the common under-
standing of geographic heterogeneity of COVID prevalence
(14, 15). By identifying a relationship between reporting rate
and testing rate, we unify the date of COVID prevalence es-
timates and rank countries according to relative estimated
prevalence. The relatively low burden in EURO countries
contradicts geographic burden distributions based on reported
cases as well as modeling estimates (13, 16).

Randomly sampled morgue-based COVID testing data pro-
vides the opportunity to evaluate mortality dynamics in an
African context, without the challenges associated with report-
ing systems. The age structure of mortality in the African
setting is significantly different from HIC settings, with unex-
pectedly higher burden in younger age groups. In addition,
overall mortality burden in the African setting outstrips that
in HICs. This poses serious risks for LMIC countries where
age distributions are skewed younger, directly contradicting
the age-protection hypothesis. We use the mortality data
alongside a simple cost model to show that even in a context
with substantial other public health concerns such as HIV,

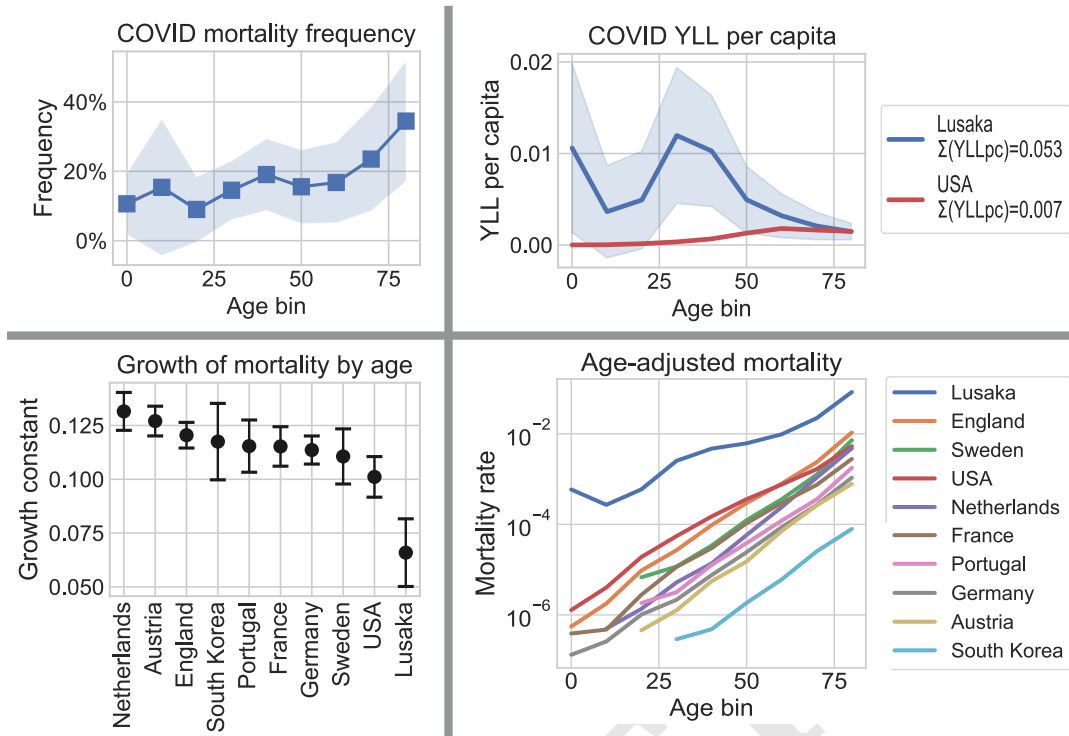


Fig. 3. COVID mortality for Lusaka and HICs. Upper left: Frequency of COVID-positive deaths among all-cause deaths by age bin (Lusaka). Upper right: Years of life lost per capita by age bin for Lusaka and the United States. Lower right: Age-adjusted mortality for Lusaka and HICs by age bin. Lower left: Exponential rate of increase of age-adjusted mortality with age across locations. All mortality data is from early August, 2020. COVID mortality in Lusaka is adjusted to exclude HIV-attributable comorbid deaths. Shading and error bars represent 95% confidence intervals.

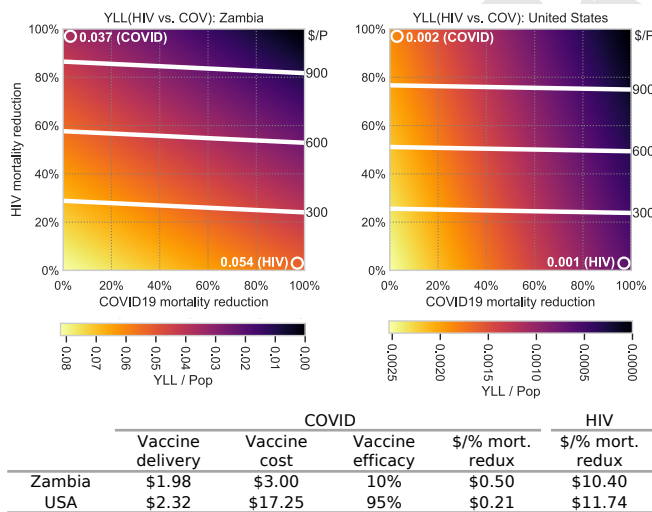


Fig. 4. Cost models of reduction of COVID and HIV mortality for Zambia (left) and the United States (right). Axes indicate model inputs of percent mortality reduction for HIV and COVID. Heat map illustrates remaining YLL/total population for HIV and COVID combined. Cost per capita for combined HIV and COVID mortality reduction are shown in white isotropic lines, labeled on the right y-axis of each plot. White circles on the plot are annotated with the total YLL per capita for one disease only, as indicated. Table shows model inputs for COVID and HIV cost per percent mortality reduction, sources cited in the Materials and Methods section.

COVID may be a cost-effective investment for disease burden reduction.

Our study is subject to a number of limitations, particularly as we grapple with understanding COVID dynamics in LMIC contexts. LMICs are subject to limited data availability and substantial uncertainty, which we address in part by making use of sub-national data sources including serology and COVID testing from morgue sampling. Challenges when working with serology include inconsistencies in testing protocols and sampling frameworks alongside the impacts of seroconversion and reversion on results. To address these hurdles, we focus on serostudies that do not target particular populations, and adjust estimates for seroconversion and reversion.

Necessitated by data and uncertainty limitations, some of the models we present rely on broad approximations. Modeling reporting rate as a function of testing rate, for example, is an approximation made to include countries where more detailed auxiliary data are not available. We do not attempt to estimate the magnitude of COVID burden in different locations, only their relative ranking. Finally, cost modeling is presented as a ballpark framework to evaluate COVID in the context of other public health concerns, rather than a comprehensive costing model. We use HIV as an example to compare with COVID, recognizing that there are other sources of burden and other approaches to public health investment than single disease-focused strategies. We do not attempt to model the complexities of mortality reduction dynamics, rather seeking to demonstrate that COVID should be considered as part of a public health investment portfolio.

The substantial differences in reporting rates across lo-

272 cations and the subsequent influence on perceived burden
273 highlight the pitfalls of making programmatic decisions based
274 on reported cases alone. The differences in mortality and bur-
275 den between Africa and HIC contexts calls for caution when
276 translating epidemiological age structure assumptions from
277 HIC to LMIC contexts, and rethinking assumptions about pro-
278 tective factors for perceived lower burden in African contexts.
279 Cost-effectiveness of COVID mortality reduction recommends
280 investment even in contexts with other entrenched public
281 health concerns and limited vaccine efficacy.

282 Materials and Methods

283 **Data.** Reporting rate analysis relies on serostudies and reported
284 cases. In order to reduce bias in the serology data, only studies
285 that do not target a particular population (e.g. health care
286 workers) were included. For WHO regions other than AFRO,
287 nationally pooled seroprevalence estimates were used (2, 17).
288 Because serostudy data in AFRO are sparse, both national
289 estimates and sub-national estimates were included (14, 18–
290 22). National case rates were obtained from Our World in Data
291 (16). Sub-national case rates were calculated from sub-national
292 reported case counts and population data (14, 15, 19, 23–31).

293 Because vital reporting systems in Africa can be limited
294 (32), we used a morgue sampling study for COVID and HIV
295 mortality rates in Lusaka instead of reported numbers (6), all-
296 cause mortality rate for Lusaka (32), and HIV prevalence data
297 for Zambia (33). Zambia HIV prevalence data was extrapo-
298 lated to older age bins corresponding to other data sources via
299 a Gaussian model. For HICs, where vital reporting systems
300 are strong, reported age-binned mortality rates for COVID
301 and HIV were used (34, 35). Years of life lost and age-adjusted
302 mortality require life expectancy and population by age, which
303 were pulled from the UN World Population Prospects and the
304 2018 Zambia Demographic and Health Survey (36, 37). Input
305 data for cost models included vaccine delivery, cost and efficacy
306 for COVID, and an established costing model for HIV (38–42).
307 Cost data from the literature were adjusted for inflation to
308 2021 dollars.

309 Methods.

310 **Seroconversion and reversion.** Seropositivity provides an indica-
311 tor of the number of past infections, which can be compared
312 with cumulative reported cases to estimate reporting rate.
313 The probability of testing seropositive, however, is a function
314 of the elapsed time from infection because of seroconversion
315 and reversion dynamics. In other words, a serostudy does not
316 detect all past infections. For a fair comparison of detected
317 seropositivity with reported cases, we adjusted reported cases
318 down to the number of those cases that could be detected by
319 serosampling based on a model of seroconversion and reversion.
320 We tested two models for calculating this case adjustment,
321 each with different data requirements.

322 The more data-intensive model requires continuous time
323 series of daily reported case data. The model provides an
324 estimate of the probability for an infection to appear seroposi-
325 tive at time t , where t is the time between infection and the
326 serostudy. To construct this model, we combined empirical
327 models for the probabilities of seroconversion (43) and rever-
328 sion (44), Figure S1. Where τ is the time between infection and
329 seroconversion, $conv$ indicates seroconversion, rev indicates

330 seroreversion, $P(t)$ indicates the probability of seroconversion
331 or reversion at time t , and $p(t)$ indicates the probability of
332 seroconversion or reversion by time t , the probability of an
333 infection being seropositive at time t is as follows:

$$334 P_{pos}(t) = p_{conv}(t) - \int_0^t P_{conv}(\tau)p_{rev}(t - \tau)d\tau \quad 334$$

335 Because the empirical models from the literature include
336 only about four months of data, we used the last month of the
337 combined empirical model to extrapolate out to a year using
338 a log-linear regression. The full P_{pos} model is shown in Figure
339 S2.

340 The calculation for a point estimate of reporting rate using
341 this model of seropositivity is then as follows, where T_k is the
342 date the serostudy k was conducted, R is reporting rate, c is
343 daily reported cases, and S is seropositivity rate:

$$344 R_{T_k} = \frac{\int_0^{T_k} c(t) * P_{pos}(t)dt}{S_{T_k}} \quad 344$$

345 While continuous daily case reporting data largely exists
346 for HICs, LMICs do not necessarily report data at a daily
347 frequency, particularly for sub-national locations. In order
348 to be able to include more LMIC serology studies, we also
349 developed a less data-intensive model for the case count adjust-
350 ment to approximate seroconversion and reversion dynamics.
351 In this model, only two data points for cumulative cases are
352 required: cumulative cases at 21 days and 60 days before the
353 serostudy. These time delays were selected heuristically to
354 account for seroconversion and reversion, respectively. Only
355 cases reported within these time bounds are then used for the
356 reporting estimation. Where C is cumulative cases:

$$357 R_{T_k} = \frac{C_{(T_k - 21days)} - C_{(T_k - 60days)}}{S_{T_k}} \quad 357$$

358 We compared results from the two different models for
359 adjusting cases and estimating reporting rate. We found
360 the less data-intensive model to be a good approximation of
361 the more complete probability-based model, see Figure S3.
362 We used the second, less data-intensive model for Figure 1
363 to include sub-national locations, and the probability-based
364 model for Figure 2.

365 **Dynamic reporting rate modeling.** With reporting rate estimates
366 based on serostudies each conducted at a different time T_k , it
367 remains to unify dates of estimated infection rates for com-
368 parison across locations. To allow reporting rates to vary over
369 time, we constructed a hybrid reporting rate model based on
370 the reporting rate at the time of a serostudy and the log-log
371 relationship between testing rate and reporting rate in the
372 serostudy locations.

373 As the serostudies we used are relatively early in the pan-
374 demic, we approximated reporting rate up until the time of
375 a serostudy as the reporting rate at the time of the seros-
376 tudy. This offsets under-estimation of reporting rates early in
377 the pandemic when testing policies were largely symptomatic
378 and testing rates were low. For dates after the serostudy,
379 we allowed the reporting rate to vary with testing rate. The
380 parameters of the relationship between these two variables
381 were established by a log-log regression, illustrated in Figure

S4. Where α and β are parameters of the regression fit and E is testing rate:

$$\log(R_{T_k}) = \alpha + \beta(\log(E_{T_k}))$$

Figure S5 illustrates output from the hybrid reporting rate model for a few example countries. The modeled reporting rate over time was then used to calculate cumulative estimated infections per capita at time T :

$$I(T) = \int_0^{T_k} \frac{c(t)}{R_{T_k}} dt + \int_{T_k}^T \frac{c(t)}{e^\alpha E(t)^\beta} dt$$

Mortality and years of life lost. Formulations for mortality and years of life lost in Lusaka were determined by available data. Despite the challenges that the number of deaths in Lusaka at large is difficult to determine and the catchment of the Lusaka morgue study is unknown, the randomly sampled morgue data made it possible to create formulations for age-adjusted mortality and YLL that use only available information.

By definition, where $pop(age)$ is population by age bin; Pop is total population; $deaths(age)$ is deaths by age bin; $Deaths$ is total deaths; and $e(age)$ is life expectancy by age:

$$\text{Age-adjusted mortality} = \frac{deaths(age)}{pop(age)}$$

$$YLL(age) = deaths(age) * e(age)$$

Where $morgue(age)$ indicates the number of deaths by age bin from the morgue sample, $Morgue_{AC}$ indicates the total all-cause deaths in the morgue sample, and $Deaths_{AC}$ indicates the total all-cause deaths in Lusaka, we approximated $\frac{deaths(age)}{Deaths_{AC}}$ as $\frac{morgue(age)}{Morgue_{AC}}$. We used the crude death rate for Lusaka from (32): $\frac{Deaths_{AC}}{Pop} = \frac{12.2}{1000}$. For $\frac{pop(age)}{Pop}$, we used the urban population distribution from (37). Age-adjusted mortality then becomes:

$$\frac{deaths(age)}{pop(age)} = \frac{deaths(age)}{Deaths_{AC}} * \frac{Deaths_{AC}}{Pop} * \frac{Pop}{pop(age)}$$

YLL is calculated per capita:

$$\frac{YLL}{Pop} = \sum_{age} \frac{deaths(age)}{Deaths_{AC}} * \frac{Deaths_{AC}}{Pop} * e(age)$$

See Figure S6 for a breakdown of the frequencies of COVID and HIV within age-binned deaths. The methodology for the calculations of the increase of mortality rates with age are based on Goldstein & Lee (45).

Population attributable fraction. For deaths that are both COVID+ and HIV+, we attributed some to COVID and some to HIV based on the population attributable fraction method (46). Where $deaths_{(COV+)}$ denotes all COVID+ deaths, $deaths_{(COV+|HIV+)}$ denotes COVID+ deaths that were also HIV+, $deaths_{(COV+|HIV-)}$ denotes COVID+ deaths with a negative or undetermined HIV status, and HIV indicates HIV prevalence, the population attributable fraction, calculated by age bin, is formulated as follows:

$$PAF = \frac{deaths_{(COV+|HIV+)}}{deaths_{(COV+)}} - \left(\frac{deaths_{(COV+|HIV-)}}{deaths_{(COV+)}} * \frac{HIV}{1 - HIV} \right)$$

And deaths attributed to COVID by age bin are as follows:

$$deaths_{COV_{attr}} = deaths_{(COV+)}(1 - PAF)$$

Cost modeling. Cost per percent mortality reduction for COVID was formulated as follows:

$$\text{cost}_{(COV)} = (\text{delivery}_{vax} + \text{cost}_{vax}) * \frac{0.01}{\text{efficacy}_{vax}}$$

Existing literature informed vaccine delivery cost estimates (38). Vaccine cost and efficacy was calculated for the United States based on mRNA vaccines against wild type COVID (39, 40). For Zambia, vaccine cost and efficacy was calculated based on AstraZeneca against variant B.1.351 (39, 41). Cost per percent mortality reduction for HIV came directly from the literature (42).

ACKNOWLEDGMENTS. We gratefully acknowledge the team responsible for the collection and processing of the morgue data from Zambia, namely, William MacLeod, Geoffrey Kwenda, Rachel Pieciak, Zachariah Mupila, Rotem Lapidot, Francis Mupeta, Leah Forman, Luunga Ziko, Lauren Etter, and Donald Thea (6). Thanks also to Mandy Izzo for her editorial support.

1. YA Adebisi, et al., SARS-CoV-2 diagnostic testing in Africa: Needs and challenges. *Pan Afr Med J* **35** (2020).
2. A Rostami, et al., SARS-CoV-2 seroprevalence worldwide: A systematic review and meta-analysis. *Clin. Microbiol. Infect.* **0** (2020).
3. BL Rice, et al., Variation in SARS-CoV-2 outbreaks across sub-Saharan Africa. *Nat Med* **27** (2021).
4. WHO, COVID-19 Weekly Epidemiological Update, 15 Dec 2020, (World Health Organization), Technical report (2020).
5. NK Ibrahim, Epidemiologic surveillance for controlling Covid-19 pandemic: Types, challenges and implications. *J. Infect. Public Heal.* **13** (2020).
6. L Mwananyanda, et al., Covid-19 deaths in Africa: Prospective systematic postmortem surveillance study. *BMJ* **372** (2021).
7. F Kobia, J Gitaka, COVID-19: Are Africa's diagnostic challenges blunting response effectiveness? *AAS Open Res* **3** (2020).
8. JM Maeda, JN Nkengasong, The puzzle of the COVID-19 pandemic in Africa. *Science* **371** (2021).
9. AN Ndaye, et al., Challenges in interpreting SARS-CoV-2 serological results in African countries. *The Lancet Glob. Heal.* **9** (2021).
10. AT Levin, et al., Assessing the age specificity of infection fatality rates for COVID-19: Systematic review, meta-analysis, and public policy implications. *Eur J Epidemiol* **35** (2020).
11. D Ghosh, JA Bernstein, TB Mersha, COVID-19 pandemic: The African paradox. *J Glob Heal.* **10** (2020).
12. WHO - Global Health Observatory, Global health estimates: Leading causes of death (2021).
13. MRC Centre for Global Infectious Disease Analysis, Imperial College London, Imperial College COVID-19 LMIC Reports, Technical report (2020).
14. LB Mulenga, et al., Prevalence of SARS-CoV-2 in six districts in Zambia in July, 2020: A cross-sectional cluster sample survey. *The Lancet Glob. Heal.* **0** (2021).
15. WC Government, Western Cape Covid-19 Dashboard | Covid-19 Response (2021).
16. M Roser, H Ritchie, E Ortiz-Ospina, J Hasell, Coronavirus Pandemic (COVID-19), (Our World in Data), Technical report (2020).
17. MV Murhekar, et al., Prevalence of SARS-CoV-2 infection in India: Findings from the national serosurvey, May-June 2020. *Indian J. Med. Res.* **152** (2020).
18. BN Alemu, et al., Sero-prevalence of anti-SARS-CoV-2 Antibodies in Addis Ababa, Ethiopia. *bioRxiv* **0** (2020).
19. AL Batchi-Bouyou, et al., High SARS-CoV-2 IgG/IGM seroprevalence in asymptomatic Congolese in Brazzaville, the Republic of Congo. *Int. J. Infect. Dis.* **106** (2021).
20. M Hsiao, et al., SARS-COV-2 SEROPREVALENCE IN THE CAPE TOWN METROPOLITAN SUB-DISTRICTS AFTER THE PEAK OF INFECTIONS, (National Institute for Communicable Diseases), Supplemental Issue 5 Volume 18 (2020).
21. H Majiya, et al., Seroprevalence of COVID-19 in Niger State. *medRxiv* **0** (2020).
22. S Uyoga, et al., Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Kenyan blood donors. *Science* **371** (2021).
23. M Kagwe, COVID-19 Pandemic Situation Update as at 1/05/2020 - Kenya, (Kenya Ministry of Health), Technical report (2020).
24. Kenya Ministry of Health, COVID-19 Outbreak in Kenya, Daily Situation Report - 132, Technical report (2020).
25. Ministre de la Sante - Congo, SITREP COVID-19 N°111 du 16 Novembre 2020, Technical report (2020).
26. Nigeria Centre for Disease Control, Confirmed COVID19 Cases - Nigeria 2021 (2021).
27. Nigeria National Bureau of Statistics, Demographic Statistics Bulletin 2017, Technical report (2018).
28. SS Africa, Provinces at a Glance: Community Survey 2016, Technical report (2016).
29. ZNPH Institute, Zambia Situation Report No. 59, Technical report (2020).
30. ZNPH Institute, Zambia Situation Report No. 69, Technical report (2020).
31. ZNPH Institute, Zambia Situation Report No. 75, Technical report (2020).
32. SS Mudenda, et al., Feasibility of using a World Health Organization-standard methodology for Sample Vital Registration with Verbal Autopsy (SAVVY) to report leading causes of death in Zambia: Results of a pilot in four provinces, 2010. *Popul Heal. Metr* **9** (2011).
33. Ministry of Health, Zambia, Zambia Population-based HIV Impact Assessment (ZAMPHIA) 2016: Final Report, (Lusaka, Ministry of Health), Technical report (2019).

- 506 34. Institut national d'études démographiques, Data & metadata (2021).
- 507 35. Centers for Disease Control and Prevention, HIV Surveillance Report 2018 (updated), Technical report (2020).
- 508
- 509 36. United Nations Department of Economic and Social Affairs, World Population Prospects - Population Division - United Nations (2019).
- 510
- 511 37. ZSA Zsa, MoH Moh, UTHVL Uth-VI, ICF, Zambia Demographic and Health Survey 2018, Technical report (2020).
- 512
- 513 38. A Portnoy, et al., Costs of vaccine programs across 94 low- and middle-income countries. *Vaccine* **33** (2015).
- 514
- 515 39. UNICEF, COVID-19 Vaccine Market Dashboard (2020).
- 516 40. SA Madhi, et al., Efficacy of the ChAdOx1 nCoV-19 Covid-19 Vaccine against the B.1.351 Variant. *N. Engl. J. Med.* **0** (2021).
- 517
- 518 41. FP Polack, et al., Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *N. Engl. J. Med.* **383** (2020).
- 519
- 520 42. E Lamontagne, M Over, J Stover, The economic returns of ending the AIDS epidemic as a public health threat. *Heal. Policy* **123** (2019).
- 521
- 522 43. AS Iyer, et al., Persistence and decay of human antibody responses to the receptor binding domain of SARS-CoV-2 spike protein in COVID-19 patients. *Sci. Immunol.* **5** (2020).
- 523
- 524 44. D Herrington, TCCRPS Group, Duration of SARS-CoV-2 Sero-Positivity in a Large Longitudinal Sero-Surveillance Cohort: The COVID-19 Community Research Partnership. *medRxiv* **0** (2021).
- 525
- 526
- 527 45. JR Goldstein, RD Lee, Demographic perspectives on the mortality of COVID-19 and other epidemics. *PNAS* **117** (2020).
- 528
- 529 46. SA Western Cape Department of Health, National Institute for Communicable Diseases, Risk Factors for Coronavirus Disease 2019 (COVID-19) Death in a Population Cohort Study from the Western Cape Province, South Africa. *Clin. Infect. Dis.* **0** (2020).
- 530
- 531

DRAFT