The relatively young and rural population may limit the spread and severity of COVID-19 in Africa: a modelling study

Binta Zahra Diop 1, Marieme Ngom, Clément Pougué Biyong, John N Pougué Biyong

ABSTRACT

Introduction A novel coronavirus disease 2019 (COVID-19) has spread to all regions of the world. There is great uncertainty regarding how countries’ characteristics will affect the spread of the epidemic; to date, there are few studies that attempt to predict the spread of the epidemic in African countries. In this paper, we investigate the role of demographic patterns, urbanisation and comorbidities on the possible trajectories of COVID-19 in Ghana, Kenya and Senegal.

Methods We use an augmented deterministic Susceptible-Infected-Recovered model to predict the true spread of the disease, under the containment measures taken so far. We disaggregate the infected compartment into asymptomatic, mildly symptomatic and severely symptomatic to match observed clinical development of COVID-19. We also account for age structures, urbanisation and comorbidities (HIV, tuberculosis, anaemia).

Results In our baseline model, we project that the peak of active cases will occur in July, subject to the effectiveness of policy measures. When accounting for the urbanisation, and factoring in comorbidities, the peak may occur between 2 June and 17 June (Ghana), 22 July and 29 August (Kenya) and, finally, 28 May and 15 June (Senegal). Successful containment policies could lead to lower rates of severe infections. While most cases will be mild, we project in the absence of policies further containing the spread, that between 0.78% and 1.03%, 0.61% and 1.22%, and 0.60% and 0.84% of individuals in Ghana, Kenya and Senegal, respectively, may develop severe symptoms at the time of the peak of the epidemic.

Conclusion Compared with Europe, Africa’s younger and rural population may modify the severity of the epidemic. The large youth population may lead to more infections but most of these infections will be asymptomatic or mild, and will probably go undetected. The higher prevalence of underlying conditions must be considered.

INTRODUCTION

Since the first reported severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection in December 2019, the virus has spread to all continents.1 There is still little evidence on the pattern of the spread in Africa. Although the African continent is made up of countries with different infrastructures, health policies and characteristics in the face of this novel coronavirus disease 2019 (COVID-19), some characteristics such as a young population,2 comorbidities (tuberculosis (TB), HIV, anaemia)3,4 and low urbanisation rates transcend these differences and have been seldom considered in the large number of studies published to date. For example, the median age below 20,5 and the low rates of urbanisation, could potentially...
lead to a lower death toll of the epidemic in African countries than elsewhere.

However, having a young population implies that many infected individuals may not display symptoms and will risk infecting more people than would symptomatic individuals. Additionally, the large number of informal settlements could accentuate this phenomenon. It is therefore urgent to develop a framework that could accurately predict the spread of the virus, accounting for the idiosyncrasies of the African context. A country-specific model will provide policymakers with a wide range of prediction scenarios, based on different actions they can take to address the pandemic. With the scarce resources at their disposal, models like these will help target prevention strategies to individuals with comorbidities who might suffer the most from the epidemic. Moreover, with containment policies that can grind economies to a halt, understanding the trade-off between rural and urban spreads could lead to better informed decisions between the short-term impacts of the epidemic, and the long-term looming shortage in the food supply that could stem enforcing strict social distancing measures in rural areas.

This study contributes to the meagre literature on the burden of the virus on African countries; it also adds to the use of differential equation models to predict the spread of epidemics. This paper focuses on three African countries that have received little attention: Ghana, Kenya and Senegal. We chose Kenya to have a comparison point with another in-depth study by Brand et al.; Ghana and Senegal, on the other hand, have had transparent data sharing policies from the start of the epidemic; they made available publicly the number of positive cases, the number tests conducted and a clear outline of the containment measures. Ghana has an extensive testing policy, while Senegal has tested very few individuals comparatively (As of 1 May 2020, Ghana has conducted 3.37 tests per thousand individuals with comorbidities who might suffer the most from the epidemic. Moreover, with containment policies that can grind economies to a halt, understanding the trade-off between rural and urban spreads could lead to better informed decisions between the short-term impacts of the epidemic, and the long-term looming shortage in the food supply that could stem enforcing strict social distancing measures in rural areas.

This study contributes to the meagre literature on the burden of the virus on African countries; it also adds to the use of differential equation models to predict the spread of epidemics. This paper focuses on three African countries that have received little attention: Ghana, Kenya and Senegal. We chose Kenya to have a comparison point with another in-depth study by Brand et al.; Ghana and Senegal, on the other hand, have had transparent data sharing policies from the start of the epidemic; they made available publicly the number of positive cases, the number tests conducted and a clear outline of the containment measures. Ghana has an extensive testing policy, while Senegal has tested very few individuals comparatively (As of 1 May 2020, Ghana has conducted 3.37 tests per thousand individuals while Kenya and Senegal are respectively at 0.4 and 0.7611), we are thus able to see the difference in predictions for two countries that have adopted widely different testing strategies.

To project the trends of the epidemic, we augment the canonical Susceptible-Infected-Recovered (SIR) model with an exposed compartment that comprises infected individuals who do not yet have symptoms and who are not infectious. Danon et al. also use a SEIR model but split the infected compartment into two subcompartments: mild symptoms and symptomatic. Finally, Arenas et al. study a model composed of susceptible, exposed, asymptomatic infections, infected, hospitalised to intensive care unit (ICU), dead and recovered compartments; however, they assume that all asymptomatic infectious individuals cannot recover before they ever develop symptoms.

There is early evidence that a large number of individuals infected with COVID-19 will recover without ever developing symptoms and that asymptomatic individuals are contagious to varying degrees. Based on these findings, our model assumes that individuals are contagious from the moment they get infected. We define an SIR model with vital dynamics (see online supplementary appendix figure 1). The known natural progression of the disease is (1) asymptomatic, (2) mild symptomatic, (3) moderate symptomatic, (4) severe, and (5) critical. There are benefits in understanding the heterogeneity among infected individuals, namely those carriers without symptoms (asymptomatic), carriers with symptoms (mild and moderately symptomatic) and severe cases who might seek medical attention (severe and critical). We therefore propose to divide the infected compartment into three subcompartments: asymptomatic infectious, mildly (and moderately) symptomatic infectious, and severely (and critically) infected requiring medical attention.

We introduce some notations. S is the share of susceptible, that is, individuals who are exposed to the virus but not immune. Ls, Lm, and L are respectively the shares of asymptomatic, mildly symptomatic and severely symptomatic individuals. R is the share of immune individuals. D is the share of deceased individuals (due to COVID-19 and other non-related causes). All numbers are expressed in terms of percentages of the total population. We note I = Is + Im + I, the total share of infected and N = S + I + R the share of individuals alive. We suppose that borders are closed. Moreover, all compartments experience natural vital dynamics via the birth rate $\mu_{birth}$ and the death rate $\mu_{death}$ from causes unrelated to the virus (eg, long-term diseases, accidents). Daily epidemic transmission is described by equations (1)–(5):

$$\frac{dS}{dt} = -\beta SI + \mu_{birth} (S + R) - \mu_{death} S$$

$$\frac{dI}{dt} = \beta SI - (\beta + \delta) I$$

$$\frac{dR}{dt} = \delta I$$

$$\frac{dD}{dt} = \delta I$$

where $\beta SI$ is the contact rate between asymptomatic infected individuals and susceptible ones. Because asymptomatic individuals are not aware of their infection, their rate of contact with susceptible individuals is the same as the rate of contact within the group of susceptible individuals. This contact rate will vary with containment measures that are enforced within each country.

$\beta_{as}$ is the contact rate between mildly symptomatic individuals and susceptible ones. It is assumed to be lower than $\beta_s$ because symptomatic individuals tend to self-isolate, either because they are bedridden due to their

**METHODS**

**Compartmental epidemiological model**

Several models have been used to predict the spread of the virus. Read et al. use a standard Susceptible-Exposed-Infected-Recovered (SEIR) model with an exposed compartment

**BMJ Global Health**

symptoms or simply because they want to limit contacts with susceptible individuals.

\( \beta \) is the contact rate between severely symptomatic and susceptible individuals. Individuals who experience severe symptoms may seek medical care and get admitted as inpatients at a hospital. They might not get hospital care for various reasons (e.g., health facilities are overwhelmed). This rate accounts for contacts between hospitalised patients and healthcare workers, but can also be interpreted as contacts between severely symptomatic individuals and any caregiver (at home for instance, if the health services are overloaded). It also accounts for the contacts between hospitalised severely symptomatic individuals and susceptible individuals outside of their caretakers. It remains unclear how contacts other than healthcare workers affect the value of \( \beta \).

\[
\frac{dI_{ss}}{dt} = \beta SI - \frac{\tau_{ms}}{T_{ms}} I_{ss} - \frac{(1-\tau_{ms})}{T_{ms}} I_{ss} - \mu_{death} I_{ss} \tag{2}
\]

where \( \tau_{ms} \) is the probability of recovery without ever developing symptoms, \( T_{ms} \) is the recovery time of an asymptomatic individual and \( T_{ms} \) is the incubation period during which an individual is infected and infectious, but does not have symptoms. We define the asymptomatic effective reproduction number as the average number of secondary cases per asymptomatic case at time \( t \).

\[
\frac{dI_{ms}}{dt} = \left( \frac{1-\tau_{ms}}{T_{ms}} \right) I_{ms} - d_{ms} \frac{T_{ms}}{T_{ms} + \tau_{ms}} I_{ms} - \left( 1 - d_{ms} \right) \left( \frac{1-\tau_{ms}}{T_{ms} + \tau_{ms}} \right) I_{ms} - \mu_{death} I_{ms} \tag{3}
\]

where \( d_{ms} \) is the probability of dying from a fast deterioration, \( T_{d} \) is the time elapsed between the appearance of first symptoms and the death of the individual, \( \tau_{ms} \) is the probability to recover from mild symptoms, \( T_{ms} \) is the recovery time associated with \( \tau_{ms} \) and \( T_{ms} \) is the time for severe symptoms to develop. We deviate for the recovery rates of the mildly symptomatic compartment \( \tau_{ms} \) by taking the weighted average of age-grouped fatality rates of COVID-19 found in Hubei, Hong Kong and Macau:\n
\[
\tau_{ms} = \sum_{ag} w_{ag} \left( 1 - f_{ag} \right)
\]

where the sum is over the age groups \( ag \in \{0, 9\}, \{10, 19\}, \ldots, \{70, 79\}, 80+ \}, \( w_{ag} \) is the share of the population in age group \( ag \) and \( f_{ag} \) is the fatality rate found in earlier studies for the population in age group \( ag \).

\[
\frac{dI_{ms}}{dt} = \frac{(1-\tau_{ms})(1-d_{ms})}{T_{ms}} I_{ms} - d_{ms} \frac{T_{ms}}{T_{ms} + \tau_{ms}} I_{ms} - (1-d_{ms}) \left( \frac{1-\tau_{ms}}{T_{ms} + \tau_{ms}} \right) I_{ms} - \mu_{death} I_{ms} \tag{4}
\]

where \( d_{ms} \) is the probability of dying after progressively developing severe symptoms that require hospitalisation, \( T_{ms} \) is the recovery time of the severely symptomatic. \( aT_{I} \) is the time to death from the start of severe symptoms, for individuals who pass away from severe, progressively developing symptoms. Intuitively, if most severe cases are hospitalised, \( a \) should be higher than 1 as health professionals will slow down the evolution of the disease.

\[
\frac{dR}{dt} = \frac{\tau_{ms}}{T_{ms}} I_{ms} + (1-d_{ms}) \frac{\tau_{ms}}{T_{ms}} I_{ss} + (1-\tau_{ms}) \frac{T_{ms}}{T_{ms} + \tau_{ms}} I_{ss} - \mu_{death} R \tag{5}
\]

In our simulations, we include fatalities, but we do not include the outcomes in the results. We make this choice because of the high uncertainty around the capacity of the healthcare systems of each individual country to absorb the increased demand from patients who are severely ill. For instance, with the same predictions, a country that has a stock of ventilators of 1000 will likely have less fatalities than a country with no ventilators. Because we do not have data on healthcare capacities, we chose not to present these results. Our predicted number of fatalities is however subtracted from the number of susceptibles.

Patient and public involvement
Our study does not involve the participation of patients or any members of the public. All data used for the purpose of this study are aggregated and publicly available.

RESULTS
Baseline simulations
We use publicly available data from the European Centre for Disease Prevention and Control, and from daily press releases made by the Senegalese Ministry of Health and Social Protection.21,22 We also do checks using the Ghana Health Service and the Kenya Ministry of Health websites.

Whenever possible, we use values of parameters drawn from the literature to fit the model (see table 1).

Although there are reports that as many as 80% of active cases are asymptomatic,23 these reports are based on cases that are still active and include presymptomatic individuals. We thus use 40% as the share of individuals infected with COVID-19 who recover without ever developing symptoms.16–19

Ghana, Kenya and Senegal have extensive communication strategies including in local languages to ensure that communities are able to detect the symptoms of COVID-19 such as a cough and fever and would report any person with those symptoms. We therefore use the share of individuals who do not have a cough as a proxy for the rate of symptomatic infected individuals who can leave their home without being reported. Wang et al24 find that 59% of individuals who test positive for COVID-19 have a cough which implies \( \beta_{as} = 0.41 \beta_{ss} \). We set \( \beta_{as} = \beta_{ss} \) because we presume that individuals would be most at risk of infecting other susceptible individuals during their transport to the hospital and so this ratio that a patient who is severely ill would infect as many individuals over the course of their being in the severely symptomatic subcompartment as an asymptomatic but infectious patient would in 1 day.

We chose South Korea as a benchmark to validate \( \beta_{ms} \), \( \beta_{as} \) and other parameters of our model because it is cited as an example for its extensive testing, tracking and tracing of infections. We calibrate \( K \) by allowing it to change at each new containment measure taken by South Korean authorities until the number of identified cases reaches...
Table 1  Parameters of the model

<table>
<thead>
<tr>
<th>First case</th>
<th>Ghana</th>
<th>Kenya</th>
<th>Senegal</th>
<th>South Korea</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_{mth}$</td>
<td>9.01E-05</td>
<td>9.71E-05</td>
<td>1.00E-04</td>
<td>1.60E-05</td>
<td>CountryMeters</td>
</tr>
<tr>
<td>$\mu_{eth}$</td>
<td>9.01E-05</td>
<td>9.71E-05</td>
<td>1.00E-04</td>
<td>1.60E-05</td>
<td>CountryMeters</td>
</tr>
<tr>
<td>$T_{as}$</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>Nishiura et al</td>
</tr>
<tr>
<td>$T_{rec,as}$</td>
<td>14 days</td>
<td>14 days</td>
<td>14 days</td>
<td>14 days</td>
<td>Hubei: McIntosh</td>
</tr>
<tr>
<td>$T_{rec,ms}$</td>
<td>5 days</td>
<td>5 days</td>
<td>5 days</td>
<td>5 days</td>
<td>McIntosh</td>
</tr>
<tr>
<td>$T_{rec,mss}$</td>
<td>14 days</td>
<td>14 days</td>
<td>14 days</td>
<td>14 days</td>
<td>McIntosh</td>
</tr>
<tr>
<td>$T_{rec,ss}$</td>
<td>16 days</td>
<td>16 days</td>
<td>16 days</td>
<td>16 days</td>
<td>McIntosh</td>
</tr>
<tr>
<td>$T_{rec,ms}$</td>
<td>5 days</td>
<td>5 days</td>
<td>5 days</td>
<td>5 days</td>
<td>McIntosh</td>
</tr>
<tr>
<td>$T_{rec,mss}$</td>
<td>18 days</td>
<td>18 days</td>
<td>18 days</td>
<td>18 days</td>
<td>McIntosh</td>
</tr>
<tr>
<td>$d_{ms}$</td>
<td>2%</td>
<td>2%</td>
<td>2%</td>
<td>0.03%</td>
<td>CountryMeters</td>
</tr>
<tr>
<td>$d_{ss}$</td>
<td>98.34%</td>
<td>98.53%</td>
<td>98.62%</td>
<td>94.33%</td>
<td>Verity et al</td>
</tr>
<tr>
<td>$d_{as}$</td>
<td>98.34%</td>
<td>98.53%</td>
<td>98.62%</td>
<td>94.33%</td>
<td>Verity et al</td>
</tr>
<tr>
<td>$\beta_{ms}$</td>
<td>0.41</td>
<td>0.41</td>
<td>0.41</td>
<td>0.41</td>
<td>Author's estimate using Wang et al</td>
</tr>
<tr>
<td>$\beta_{as}$</td>
<td>0.41$T_{rec}$</td>
<td>0.41$T_{rec}$</td>
<td>0.41$T_{rec}$</td>
<td>0.41$T_{rec}$</td>
<td>Author's estimate using Wang et al</td>
</tr>
<tr>
<td>$\beta_{ms}$</td>
<td>0.41$T_{rec}$</td>
<td>0.41$T_{rec}$</td>
<td>0.41$T_{rec}$</td>
<td>0.41$T_{rec}$</td>
<td>Author's estimate using Wang et al</td>
</tr>
<tr>
<td>$\beta_{as}$</td>
<td>0.41$T_{rec}$</td>
<td>0.41$T_{rec}$</td>
<td>0.41$T_{rec}$</td>
<td>0.41$T_{rec}$</td>
<td>Author's estimate using Wang et al</td>
</tr>
</tbody>
</table>

$R_t$ in parenthesis are $R_t$ values for optimistic and pessimistic scenarios, respectively.

We solve an optimisation problem constrained by equations (1)–(5) using the MATLAB optimisation codes of D’Errico. The values of $R_t$ obtained and other parameters are summarised in table 1. The number of infections computed with our model accurately approximates the positive cases in South Korea (see figure 1).

For the fatality rates $d_{ms}$ and $d_{ss}$ in Ghana, Kenya and Senegal, we split the reported regional fatality rate (As of 25 April 2020, the Western African fatality rate is 2.49% while the Eastern African fatality rate is 2.25%), we take the average of both these numbers. Note that we use the regional figures because the number of cases at the national level is still low in the three countries. Western Africa has 8034 cases and Eastern Africa has 3319 cases as of 27 April 2020 (Africa CDC). Although these numbers are much lower than the Africa-wide (32 182 cases) and the worldwide ones, we find them more appropriate as they are more faithful to the standards of living and
the age pyramid of the countries we study. Particularly, Algeria (12.57% fatality rate) and Egypt (6.99%), among others, raise the Africa-wide fatality rate to 4.44% but are structurally different from Ghana, Kenya and Senegal. That being said, we acknowledge that our choice relies on the testing capacity in both Western and Eastern African regions and might underestimate the true fatality rate as a consequence) 2.37% between $d_m=0.03\%$ and $d_m=1.04\%$ as most deaths occurred for the severely symptomatic. (Note that there are two ways to compute the fatality rate, either (1) as the ratio deaths/total cases, or (2) deaths/closed cases. While the former is likely to be an underestimate because lots of open cases can still end up in death, the latter is an overestimate because it is likely that deaths are closed quicker than recoveries. As critical cases are more likely to be detected than mild infections, it is also likely that the number of true cases is underestimated by official numbers but the number of COVID-19-related deaths is relatively well captured by official reports. Therefore, the ratio (1) is likely to be a better estimate of the true fatality rate than (2), and we use this definition of fatality rate.)

We report predictions for a year (see figure 3). Under the assumptions of the baseline model and their limitations, we predict that the peak of the epidemic will occur in July for all three countries as detailed in table 2. For Ghana, Kenya and Senegal, respectively, this peak should lead to approximately 11.1, 18.9 and 5.8 million active infections (including asymptomatic and symptomatic cases) at the peak of the epidemic, with 308 000, 465 000 and 138 000 individuals severely ill needing medical attention (see figure 3).

These long-term scenarios should be interpreted with great caution as they do not consider future policies or actions that could drastically reduce the contact rates and subsequently, flatten the curve further. (For instance, wearing a mask in public space is now mandatory in Kenya (since 15 April), Senegal (since 20 April) and Ghana (since 25 April). Additionally, the government of Ghana lifted its partial lockdown on 20 April.)

### Testing the sensitivity of the simulations to $R_t$

We perform a sensitivity analysis for $R_t$ on our baseline model. We perturb it 100 times by $\varepsilon$ drawn uniformly in $[-5\%R_t, +5\%R_t]$. The number of infections at the peak fluctuates between 27% and 40% of the total population in the active infection compartment for Ghana, between 28% and 40% for Kenya and between 25% and 37% for...
Figure 3  Projection of active infections.

Senegal (see online supplementary appendix figure 2). In countries that enforced strict social distancing measures, predictions were significantly updated down—from about 2.2 million deaths on 16 March12 to about 60 000 on 30 March.27 A similar update can be expected from the outputs of our model as authorities take effective measures to reduce $R_t$ and/or people in these countries gradually adopt behaviour that would minimise contacts.

**Population density and rate of reproduction**

As the population density increases, the rate of transmission of infectious diseases increases.28 With respectively 43.3%, 72.2% and 50.6% as a share of their population living in rural areas, Ghana, Kenya and Senegal have sparsely populated areas outside of their main metropolitan areas, compared with countries like South Korea (18.5% of rural population).

There is little information on the relative rate of transmission of COVID-19 between rural and urban areas but we draw on other diseases, for which there are available data. During the 2014 Ebola outbreak in Sierra Leone,29 we find that the basic reproduction number in Kambia (the least densely populated district of Sierra Leone) is 0.56 times the one in the Western Area Urban district (the most densely populated district that comprises the capital.
Freetown). (We chose the most and the least populated districts because all these districts include urban areas.) We take that to mean that the $R_t$ in rural areas was 0.56 times that in urban areas. Other mostly rural districts had a higher $R_t$. To mirror the range of ratios of reproduction rates observed in mostly rural and mostly urban districts for the Ebola epidemic in Sierra Leone, we run two simulations, one in which the rural to urban ratio is 0.50, and one where it is 0.75. These two ratios bound the difference between mostly rural districts and mostly urban areas for Ebola in Sierra Leone.29 Using these ratios, we calibrate the rural and urban reproduction rates so their population weighted average is equal to the national $R_t$ which we keep constant across our baseline scenario and this scenario:

$$
\begin{align*}
R_t^{rur} &= \gamma R_t^{urb} \\
R_t &= R_t^{urb} + (1 - \gamma) R_t^{rur}
\end{align*}
$$

where $R_t^{rur}$ and $R_t^{urb}$ are the rural and urban reproduction rates, respectively; $\gamma = 0.50$, or $\gamma = 0.75$; $rurb$ is the national urbanisation rate; $R_t$ is the national reproduction rate listed in table 1. We use the first day of the first community transmission as the day of the first case in the rural area. Results are compiled in table 3. Effectively, we see in figure 4 that when accounting for rural areas, we observe two peaks. The first peak is driven by the spread in urban areas while the second peak, delayed in time, is driven by the spread in rural areas. Kenya, with a rural share of the population of over 70%, has the most noticeable split across its rural and urban areas.

### Table 2 Projections of active cases at the peak of the epidemic for each infected compartment

<table>
<thead>
<tr>
<th>Country</th>
<th>Days since first case</th>
<th>Active cases at peak</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Severe symptoms</td>
</tr>
<tr>
<td>Ghana</td>
<td>Low policy effectiveness</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>Baseline</td>
<td>114</td>
</tr>
<tr>
<td></td>
<td>High policy effectiveness</td>
<td>250</td>
</tr>
<tr>
<td>Kenya</td>
<td>Low policy effectiveness</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>Baseline</td>
<td>116</td>
</tr>
<tr>
<td></td>
<td>High policy effectiveness</td>
<td>252</td>
</tr>
<tr>
<td>Senegal</td>
<td>Low policy effectiveness</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>Baseline</td>
<td>128</td>
</tr>
<tr>
<td></td>
<td>High policy effectiveness</td>
<td>254</td>
</tr>
</tbody>
</table>

The peaks of each infection subcompartment might not align.

### Comorbidity and rise in the occurrence of severe symptoms

Comorbidity can impact the share of mild cases that develop severe symptoms.30 In Asia and Europe, hypertension, obesity, diabetes and coronary heart diseases have been drivers of adverse health outcomes.30 31 Because the combined prevalence of diabetes, hypertension and obesity is not higher in Ghana, Kenya and Senegal than they are in regions we use to derive the recovery rates, the baseline simulations already account for them. However, Ghana, Kenya and Senegal have persistent and high rates of anaemia and TB.3 To our knowledge, there is no study on the magnitude of the impacts of anaemia, TB or HIV on the recovery of patients who have contracted the virus. We simulate two scenarios, with 25% and 75% of the recovery rate of otherwise healthy individuals for

### Table 3 Projection of infections accounting for Africa-specific factors

<table>
<thead>
<tr>
<th>Comorbidity</th>
<th>% of the survival rate of healthy patients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>75%</td>
</tr>
<tr>
<td></td>
<td>Severe</td>
</tr>
<tr>
<td>Ghana</td>
<td>50%</td>
</tr>
<tr>
<td></td>
<td>75%</td>
</tr>
<tr>
<td>Kenya</td>
<td>50%</td>
</tr>
<tr>
<td></td>
<td>75%</td>
</tr>
<tr>
<td>Senegal</td>
<td>50%</td>
</tr>
<tr>
<td></td>
<td>75%</td>
</tr>
</tbody>
</table>

The peaks of each infection subcompartment might not align.

*Days of total infection peak, since the first case tested positive. March 2 for Senegal, March 13 for Ghana and March 14 for Kenya.
individuals with one of these underlying conditions\textsuperscript{1}. In comparison, Zhou \textit{et al}\textsuperscript{30} find that in Wuhan, China, patients with comorbidities (hypertension, diabetes, coronary heart disease, chronic obstructive lung disease, carcinoma, chronic kidney diseases and others) have a recovery rate equal to 73.2\% of their otherwise healthy counterparts. Though uncertain for HIV, anaemia and TB, the impact of these underlying conditions on the recovery of individuals will likely lie between these two bounds. This translates into adjusting $r_{\text{ms}}$ for individuals with TB, HIV and anaemia. Comorbidities have age-specific incidence rates; anaemia affects women of child-bearing age primarily, while HIV affects young adults at higher rates. The prevalence of HIV, TB and anaemia is extracted from the open database Global Burden Disease.\textsuperscript{32} We account for these age-based differences to compute the recovery rate of the population accounting for comorbidities:

$$r_{\text{ms,morb}} = \sum_{\text{ag}} w_{\text{ag}} r_{\text{ag}} (1 - i_{\text{morb}} x_{\text{ag,morb}})$$

where $r_{\text{ms,morb}}$ is the rate of recovery for infected individuals who develop mild symptoms and have one of the three comorbidities, $r_{\text{ag}}$ is the recovery rate of the otherwise healthy individuals in the age group, $i_{\text{morb}}$ is either 1–0.25 or 1–0.75 depending on the scenario and $x_{\text{ag,morb}}$ is the share of individuals in each age group with the comorbidity.

We report the results in figure 4 and table 3. As expected, the predictions are higher in the case where we assume that individuals with comorbidities have a rate of recovery that is 25\% that of otherwise healthy individuals. In the scenario with $R_{t,\text{rur}}=0.75R_{t,\text{urb}}$, the number of active severe cases at the peak is 0.242 million with a 75\% recovery scenario for Ghana (0.308 million for the 25\% scenario), 0.313 million for Kenya (0.631 million) and 0.104 million (0.145 million) for Senegal. Kenya's large impact is driven by its larger HIV-positive population.

**Mirroring South Korea's effectiveness**

Unlike countries in Europe, Ghana, Kenya and Senegal have taken containment measures very early in the
progression of the disease. The policies could have had impacts similar to the ones in South Korea. We present results of simulations mirroring the $R_t$ for South Korea. Specifically, we decrease $R_t$ for each country 3 weeks after the last recorded policy to 0.88, and then again at 6 weeks to 0.3. We find that the peak is much lower, with a number of active severe infections at the peak between 166 and 214 individuals for Ghana, 208 and 286 individuals for Kenya and 140 and 189 individuals for Senegal; with the two bounds being for recovery rates of respectively 75% and 25% of that of the otherwise healthy patients. These peaks will occur 2–3 months after the first case (see figure 5 and table 4). This scenario is attainable only if these countries are able to maintain effective policies for an extended period.

**DISCUSSION**

In this study, we account for the age structure of the population in each country, the burden of potential comorbidities, and the differential spreads of the virus in rural and urban areas. We find that the relatively young population may limit the severity of the epidemic by lowering the number of infections that lead to severe symptoms. We also find that sparsely populated areas may limit the spread of the epidemic. Rural areas effectively may lead to staggered peaks; this has important implications for policymakers who may be faced with two waves, and so may need to adapt their responses to adaptively deploy personnel on their territory as these peaks occur.

High rates of comorbidities however may lead to more individuals developing severe symptoms relative to a scenario with no comorbidities. We find that at the peak, Ghana, Kenya and Senegal are predicted to have respectively between 0.78% and 1.03%, 0.89% and 1.22% and, finally, 0.60% and 0.84% active clinical severe cases of COVID-19 with a peak of total infections predicted to occur between 2 June and 17 June (Ghana), 22 July and 29 August (Kenya) and 28 May and 15 June (Senegal), respectively, against a July timeline for our baseline specification. Successful containment policies could lead to even lower rates of severe infections.

Though recent models look at a few countries in Africa,33 or at the continent as a whole,34 35 there are little to no studies predicting the spread of COVID-19 in Ghana and Senegal while incorporating specificities of these two countries. In Kenya, however, Brand et al10 account for age-based population mixing and assume that asymptomatic individuals are as infectious as symptomatic individuals to predict that by the end of the year, 46.1 million (ie, 89% of the public) of infections will have occurred. This prediction is comparable with the baseline results of our study in the absence of further containment policies. In that scenario, we find that about 47.7 million (93% of the public) individuals may be infected cumulatively.

### Table 4 Projection of active cases at peak accounting for comorbidities, with South Korea’s $R_t$

<table>
<thead>
<tr>
<th>Comorbidity</th>
<th>$R_t$(rural)/$R_t$(urban)</th>
<th>% of the survival rate of healthy patients</th>
<th>Days*</th>
<th>75%</th>
<th>Severe</th>
<th>Mild</th>
<th>Asymptomatic</th>
<th>25%</th>
<th>Severe</th>
<th>Mild</th>
<th>Asymptomatic</th>
<th>Days*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ghana</td>
<td>100%</td>
<td>166</td>
<td>2428</td>
<td>57</td>
<td>214</td>
<td>3110</td>
<td>3110</td>
<td>57</td>
<td>189</td>
<td>3183</td>
<td>3183</td>
<td>69</td>
</tr>
<tr>
<td>Kenya</td>
<td>100%</td>
<td>208</td>
<td>3206</td>
<td>60</td>
<td>286</td>
<td>4134</td>
<td>4134</td>
<td>60</td>
<td>189</td>
<td>3183</td>
<td>3183</td>
<td>69</td>
</tr>
<tr>
<td>Senegal</td>
<td>100%</td>
<td>140</td>
<td>2661</td>
<td>69</td>
<td>189</td>
<td>3183</td>
<td>3183</td>
<td>69</td>
<td>189</td>
<td>3183</td>
<td>3183</td>
<td>69</td>
</tr>
</tbody>
</table>

The peak of each active case for each subcompartment might not align.

* Days of total infection peak, since the first case tested positive. March 2 for Senegal, March 13 for Ghana and March 14 for Kenya.
Containment measures will be successful only if the public comply; however, measuring compliance is complex and has not been rigorously studied in the context of COVID-19 in these countries. In Ghana, Kenya and Senegal poverty is the main challenge to compliance, with official unemployment rates reaching 68.7%, 51.3% and 64.6%, respectively.\(^{36}\) As a response, authorities have implemented emergency transfer programmes in cash and in-kind to the most vulnerable households partly to address compliance but also to avoid a humanitarian crisis (Senegal, Ghana). In urban areas, officials have required buses and taxis to reduce their number of passengers (Kenya, Senegal) and have mandated the use of masks (Ghana, Senegal).

Looking at how spike in cases was met by various healthcare systems in Europe and Asia, it is likely that most asymptomatic and mild cases may remain undetected.

**Limitations**

Our model does not incorporate changes in the survival rate of the virus due to weather or humidity, and in that regard, our simulations are a worst-case scenario.\(^{37} 38\) Additionally, the model assumes homogeneous mixing of individuals within rural and urban areas which is an unlikely assumption. In a future iteration of our model, we plan to use a spatially structured model in order to relax the homogeneous mixing assumption by leveraging phone data.\(^{20} 39 40\)

The model also excludes international population flows. All countries in our sample have closed their international borders—airports and roads—before or a few days after their first confirmed imported case (see figure 2). However, it is possible that COVID-19 was spreading undetected for days in the respective countries. If that is true, the peak of active cases might be delayed in comparison to the true peak. Furthermore, the spread of this disease is highly dependent on the reproduction number \(R_t\). Since this number is contingent on many factors (policies, individuals’ behaviour, and so on), its value in the long run is subject to large uncertainties. The projected number of infections in the medium to long term could thus be considerable overestimates (or underestimates) of the true number of infections (depending on the scenarios).

These predictions aggregate infections in rural and urban areas, however, in practice, the peaks in urban areas, due to higher reproduction rates, will occur earlier. In rural areas, however, the peaks will be delayed due to their lower \(R_t\). This distinction is important for policymakers who can target their resources accordingly.

The use of the data also comes with limitations such as the inaccuracy of the data collection. For example, one person was tested positive for COVID-19 on 4 March but entered Senegal on 24 February. We expect that all the countries in our sample are dealing with similar delays, however, we did not find a consistent way to address this issue. Additionally, given the low number of tests performed to detect the virus, we cannot ex ante measure the accuracy of our model in Ghana, Kenya and Senegal.

Because outcomes of individuals with critical needs are highly dependent on the capacity of healthcare systems, having data on healthcare capacity is important in predicting the number of fatalities. In our simulations, information such as the number of ICU beds would inform the fatality rate of individuals with severe symptoms (\(d_m\)). Unfortunately, we do not have access to such data and we thus choose not to show the results for fatality rates in these three countries. Finally, we use an SIR, which assumes perpetual immunity—however, there are still uncertainties regarding the possibility of reinfection.\(^{41}\)

**CONCLUSION**

In conclusion, containment measures, age structures, low urbanisation and comorbidity may lead Ghana, Kenya and Senegal to having different trajectories from the USA, and from Asian and European countries. This study is a first attempt at accounting for rural densities and comorbidity, and it suggests that rural areas will slow down the spread of the epidemic, and that relatively young population will keep the number of severe cases low compared with the nearly 3.5% hospitalisation rate in Europe and Central Asia.\(^{42}\) Our findings also show how sensitive these results are to different assumptions on the effectiveness of policies, assumptions on comorbidities and differential effective rates of reproduction in rural and urban areas.

**Acknowledgements**  We thank Martin J. Williams, Kevin Marsh, Mike English, Renaud Lambiotte, Philip Bejon, Douglas Gollin and Isabel Ruiz for reviewing the manuscript.

**Contributors**  The authors contributed equally to this study.

**Funding**  Binta Zahra Diop’s PhD is funded by the Center for the Study of African Economies and the Global Challenges Research Fund. Marienne Ngom’s work was supported by the U.S. Department of Energy, Office of Science, under Contract DE-AC02-06CH11357. Clémence Pougué Biyong’s PhD is funded by the French Ministry of Higher Education and Ecole Normale Superieure de Rennes. John N. Pougué Biyong’s PhD is funded by the UK Engineering and Physical Sciences Research Council (EPSRC).

**Competing interests**  None declared.

**Patient consent for publication**  Not required.

**Provenance and peer review**  Commissioned; externally peer reviewed.

**Data availability statement**  Data are available in a public, open access repository. We are using publicly available data exclusively. The sources of these data are the following: (1) the European Centre for Disease Prevention and Control: https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide; (2) the Senegalese Ministry of Health: http://www.sante.gov.sn/Pf%3A%5B71; (3) the Ghana Health Service: https://ghanawhathelthservice.org/covid19/; (4) the Kenyan Ministry of Health: https://www.health.go.ke/; (5) Our World in Data: https://ourworldindata.org/coronavirus; (6) Worldometers: https://www.worldometers.info/

**Open access**  This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.
REFERENCES


Correction: The relatively young and rural population may limit the spread and severity of COVID-19 in Africa: a modelling study


This article has been corrected since it published online to include the below funding information:

Binta Zahra Diop’s PhD is funded by the Center for the Study of African Economies and the Global Challenges Research Fund. Marieme Ngom’s work was supported by the U.S. Department of Energy, Office of Science, under Contract DE-AC02-06CH11357. Clémence Pougué Biyong’s PhD is funded by the French Ministry of Higher Education and Ecole Normale Supérieure of Rennes. John N. Pougué Biyong’s PhD is funded by the UK Engineering and Physical Sciences Research Council (EPSRC).

Open access. This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.