



RESEARCH ARTICLE

REVISED **Revealing the extent of the first wave of the COVID-19 pandemic in Kenya based on serological and PCR-test data**
[version 2; peer review: 2 approved, 1 approved with reservations]

John Ojal^{1,2}, Samuel P. C. Brand^{id}^{3,4}, Vincent Were⁵, Emelda A. Okiro^{id}⁶, Ivy K. Kombe¹, Caroline Mburu¹, Rabia Aziza^{3,4}, Morris Ogero^{id}⁵, Ambrose Agweyu¹, George M. Warimwe^{id}¹, Sophie Uyoga^{id}¹, Ifedayo M. O. Adetifa^{id}^{1,2}, J. Anthony G. Scott^{1,2}, Edward Otieno^{id}⁵, Lynette I. Ochola-Oyier¹, Charles N. Agoti^{id}^{1,7}, Kadondi Kasera⁸, Patrick Amoth⁸, Mercy Mwangangi⁸, Rashid Aman⁸, Wangari Ng'ang'a⁹, Benjamin Tsofa^{id}¹, Philip Bejon^{id}^{1,10}, Edwine Barasa^{5,10}, Matt J. Keeling^{id}³, D. James Nokes^{id}^{1,3,4}

¹Kenya Medical Research Institute - Wellcome Trust Research programme, Kilifi, Kenya

²Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, UK

³The Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research (SBIDER), University of Warwick, Coventry, UK

⁴School of Life Sciences, University of Warwick, Coventry, UK

⁵Health Economics Research Unit, Kenya Medical Research Institute - Wellcome Trust Research Programme, Nairobi, Kenya

⁶Population Health Unit, Kenya Medical Research Institute - Wellcome Trust Research programme, Nairobi, Kenya

⁷School of Public Health, Pwani University, Kilifi, Kenya

⁸Ministry of Health, Government of Kenya, Nairobi, Kenya

⁹Presidential Policy & Strategy Unit, The Presidency, Government of Kenya, Nairobi, Kenya

¹⁰Nuffield Department of Medicine, University of Oxford, Oxford, UK

v2 **First published:** 25 May 2021, **6**:127
<https://doi.org/10.12688/wellcomeopenres.16748.1>
Second version: 25 Feb 2022, **6**:127
<https://doi.org/10.12688/wellcomeopenres.16748.2>
Latest published: 22 Sep 2022, **6**:127
<https://doi.org/10.12688/wellcomeopenres.16748.3>

Abstract

Policymakers in Africa need robust estimates of the current and future spread of SARS-CoV-2. We used national surveillance PCR test, serological survey and mobility data to develop and fit a county-specific transmission model for Kenya up to the end of September 2020, which encompasses the first wave of SARS-CoV-2 transmission in the country. We estimate that the first wave of the SARS-CoV-2 pandemic peaked before the end of July 2020 in the major urban counties, with 30-50% of residents infected. Our analysis suggests,

Open Peer Review

Approval Status ✓ ? ✓

	1	2	3
version 3			✓
(revision)			view
22 Sep 2022			↑
version 2			✓
(revision)			view
25 Feb 2022			

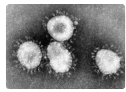
first, that the reported low COVID-19 disease burden in Kenya cannot be explained solely by limited spread of the virus, and second, that a 30-50% attack rate was not sufficient to avoid a further wave of transmission.

Keywords




SARS-CoV-2, Kenya, dynamic model, serology, PCR cases



This article is included in the [KEMRI | Wellcome Trust gateway](#).



This article is included in the [Coronavirus \(COVID-19\)](#) collection.

	1	2	3
version 1	✓	?	
25 May 2021	view	view	
1. Mark Kimathi  , Machakos University, Machakos, Kenya			
2. Amy Wesolowski  , Johns Hopkins Bloomberg School of Public Health, Baltimore, USA			
3. Benjamin J. Cowling  , The University of Hong Kong, Hong Kong, Hong Kong			
Any reports and responses or comments on the article can be found at the end of the article.			

Corresponding authors: John Ojal (John.Ojal@lshtm.ac.uk), Samuel P. C. Brand (S.Brand@warwick.ac.uk)

Author roles: **Ojal J:** Conceptualization, Formal Analysis, Methodology, Software, Writing – Original Draft Preparation, Writing – Review & Editing; **Brand SPC:** Conceptualization, Formal Analysis, Methodology, Software, Writing – Original Draft Preparation, Writing – Review & Editing; **Were V:** Conceptualization, Writing – Review & Editing; **Okiro EA:** Data Curation, Supervision, Writing – Review & Editing; **Kombe IK:** Formal Analysis, Methodology, Software, Writing – Review & Editing; **Mburu C:** Formal Analysis, Methodology, Software, Writing – Review & Editing; **Aziza R:** Formal Analysis, Methodology, Project Administration, Software, Writing – Review & Editing; **Ogero M:** Data Curation, Writing – Review & Editing; **Agweyu A:** Conceptualization, Project Administration, Writing – Review & Editing; **Warimwe GM:** Conceptualization, Supervision, Writing – Review & Editing; **Uyoga S:** Data Curation, Investigation, Writing – Review & Editing; **Adetifa IMO:** Conceptualization, Supervision, Writing – Review & Editing; **Scott JAG:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Otieno E:** Data Curation, Writing – Review & Editing; **Ochola-Oyier LI:** Data Curation, Investigation, Writing – Review & Editing; **Agoti CN:** Data Curation, Investigation, Writing – Review & Editing; **Kasera K:** Project Administration, Writing – Review & Editing; **Amoth P:** Project Administration, Writing – Review & Editing; **Mwangangi M:** Project Administration, Writing – Review & Editing; **Aman R:** Project Administration, Writing – Review & Editing; **Ng'ang'a W:** Project Administration, Writing – Review & Editing; **Tsofa B:** Project Administration, Writing – Review & Editing; **Bejon P:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Barasa E:** Conceptualization, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Keeling MJ:** Conceptualization, Formal Analysis, Funding Acquisition, Methodology, Software, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Nokes DJ:** Conceptualization, Funding Acquisition, Project Administration, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: DJN, VW are members of the National COVID-19 Modelling Technical Committee, for the Ministry of Health, Government of Kenya. KK, PA, MM, RA are from the Ministry of Health, Government of Kenya, and WN from the Presidential Policy & Strategy Unit, The Presidency, Government of Kenya. All other authors declare no competing interests

Grant information: This work was supported by the Foreign, Commonwealth and Development Office and Wellcome Trust [220985/Z/20/Z]; National Institute for Health Research (NIHR) [17/63/82] using UK aid from the UK Government to support global health research; Wellcome Trust Intermediate Fellowship awards [201866, 107568] EAO, LIOO; MRC/DFID African Research Leader Fellowship [MR/S005293/1] IMO, CM; NIHR Global Health Research Unit on Mucosal Pathogens [16/136/46] JO; DFID/MRC/NIHR/Wellcome Trust Joint Global Health Trials Award [MR/R006083/1] AA; Wellcome Trust Senior Research Fellowship [214320] and the NIHR Health Protection Research Unit in Immunisation JAGS. The views expressed in this publication are those of the author(s) and not necessarily those of any of the funding agencies.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2022 Ojal J *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Ojal J, Brand SPC, Were V *et al.* **Revealing the extent of the first wave of the COVID-19 pandemic in Kenya based on serological and PCR-test data [version 2; peer review: 2 approved, 1 approved with reservations]** Wellcome Open Research 2022, **6**:127 <https://doi.org/10.12688/wellcomeopenres.16748.2>

First published: 25 May 2021, **6**:127 <https://doi.org/10.12688/wellcomeopenres.16748.1>

REVISED Amendments from Version 1

Reviewer 1. Typographical errors have been corrected.

Reviewer 2. The manuscript has been updated in the following ways: (i) A brief explanation is provided as to why there are narrow confidence intervals for $R(t)$ early in the pandemic, (ii) Figure 2 is revised to include dates of the lifting of major restrictions, (iii) sections of the Methods previously only in supplementary information have now been added to the Methods main text for easier reading, (iv) diagnostic visualisations have been improved and are available for all Kenya counties <https://github.com/ojal/KenyaSerology>.

Any further responses from the reviewers can be found at the end of the article

Introduction

The potential risk from severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) to Africa was identified early in the global pandemic¹. As the epicenter of transmission moved from East Asia to West Asia and Europe and then to North America, there was speculation as to the likely impact of the pandemic on the African continent with its young populations, high infectious disease burden, undernutrition and fragile health infrastructure. However, as health systems and economies of high-income countries strained, the reported burden of COVID-19 cases and associated deaths in Africa remained low with the exception of South Africa and Northern Africa². The question is whether this is the result of lower risk due to demographic structure (young age³, either cross-reacting immunity (e.g. pre-existing SARS-CoV-2 cross-reactive T cells⁴) or dampened immunological over-reaction⁵, a low reproduction number from rapidly imposed interventions (such as school closures and lockdowns⁶), environmental conditions (e.g. temperature and humidity⁷), or under-reporting. The reason this remains a conundrum is, at least in part, a paucity of good quality data to reveal the probable extent of SARS-CoV-2 spread in African populations.

Following the first confirmed coronavirus disease 2019 (COVID-19) case in Kenya on 13th March 2020, the Kenyan Government moved rapidly, closing international borders, schools, restaurants, bars and nightclubs, banning meetings and social gathering, and imposing a dusk to dawn curfew and movement restrictions in the two major city counties, Nairobi and Mombasa⁸. The major concerns from unmitigated spread were a limited surge capacity of the Kenyan health system⁹ and groups of the Kenyan population identified as potentially highly vulnerable to infection, due to socio-economic factors such as crowded households or lack of access to handwashing, and/or severe disease, due to epidemiological factors such as higher rates of obesity and hypertension¹⁰. Throughout the months of April, May and into June 2020 few people in Kenya were reported SARS-CoV-2 test positive by polymerase chain reaction (PCR), or severely diseased or dying with COVID-19 as the established cause¹¹. There followed a relaxation of some measures in June and July including controlled opening of restaurants and places of worship and the removal of travel restrictions into and out of Mombasa and Nairobi counties. As of 30th September 2020, there were 45,795 laboratory-confirmed

positive swab tests out of over 340,000 tests (about 13.5%), and 749 deaths with a positive test result in Kenya¹¹. This should be compared with the 200–250,000 cases and 30–40,000 deaths attributable to SARS-CoV-2 for similar sized countries in Europe (France, Italy, UK) by the end of September¹².

The reason for this apparently low level of COVID-19 disease in Kenya is unknown; one possible explanation is that SARS-CoV-2 had not widely spread among the Kenyan population by the end of September. However, two pieces of information suggest that SARS-CoV-2 had already spread extensively by the end of September. First, a regionally-stratified seroprevalence study of 3098 Kenyan blood donors sampled between May and June reported a national estimate of 4.3% (adjusted to reflect the population distribution by age, sex and region)¹³. Sero-prevalence was higher in Nairobi (7.6%) and Mombasa (8.3%). These levels of seropositivity are comparable to those reported in May in the United Kingdom (UK)¹⁴, April/May in Spain¹⁵, and March/April in some United States (US) cities¹⁶, where high numbers of PCR-positive cases, hospitalizations and deaths have also been reported, in contrast to Kenya. Second, we noticed that test-positive PCR cases, and daily reported test-positive deaths, were declining first in Mombasa (from early July 2020) and then Nairobi (from early August 2020); respectively Kenya's second and first largest cities. In Europe, declining case and mortality rates have been closely associated with non-pharmaceutical interventions (NPIs)¹⁷. However, in Kenya this went counter to evidence of increased mixing, and hence reproduction potential, arising from Google Mobility data for these cities which showed a steady reversion in mobility towards pre-COVID-19 intervention levels since early April (Fig. S1). These observations, in turn, lead to the conclusion that either a smaller than expected proportion of infected individuals have had severe disease, and/or, that there has been under-reporting of severe disease.

To investigate these findings, we developed a simple SEIR (susceptible-exposed-infectious-recovered) compartmental mechanistic and data-driven transmission model for Kenya, which integrates three sources of longitudinal data: national time series polymerase chain reaction (PCR) tests, the Kenyan serological survey and Google mobility behavioural data. The overall modelling approach is similar to Flaxman *et al.*¹⁷; that is we use time-to-event lag distributions, and the daily incidence time series, and, both models generate the daily incidence time series using a simple deterministic transmission model with the key unknowns being initial numbers of infected individuals and $R(t)$. Where we differ in approach from Flaxman *et al.*¹⁷ is that, instead of using reported test-positive deaths as the most reliable data for inferring underlying transmission patterns, we use a combination of PCR test-positive and serological data. The PCR test-positive data informs the model on the epidemic trajectory but does not account for likely under-detection of cases. This under-detection of cases is inferred from the proportion exposed to SARS-CoV-2 evidenced by the seroprevalence estimates, hence scaling the incidence estimation. Finally, the mobility data, as a proxy for the contact rate, determines the contribution of the intervention (which acts to alter contact patterns) relative to other factors that alter

incidence and the effective reproduction number, the most important of which is the susceptible proportion of the population. Our aim is to derive a coherent picture of the SARS-CoV-2 epidemiology in Kenya in the first wave and reveal the historic and future patterns of spread across the country and by county. Reported deaths are not used as primary data for inference, but rather the trend in changing rates of reported deaths is used as a validation data set for model predictive accuracy (see supporting information for description of model validation). Reported deaths may be subject to substantial under-reporting, and we assume that the bias in under-reporting is consistent over time.

Results

Underlying transmission rates in Mombasa and Nairobi during the first wave

As at 30th September, a substantial proportion of PCR positive tests have been samples from the capital Nairobi (25,182 positive tests), while Kenya's second largest city, Mombasa, has reported the next highest number of PCR positive tests (2,056). We infer that the underlying rate of new infections peaked on May 18th 2020 (CI May 16th - May 21st) in Mombasa and July 9th 2020 (CI July 7th - July 10th) in Nairobi, and subsequently declined from peak transmission (Figure 1 H, G). The model suggests that the PCR test and serology data can be explained by the initial presence of <200 infected individuals in both Mombasa and Nairobi on 21st February, three weeks before the first reported case in Kenya. Thereafter, growth of transmission was rapid in both counties. In early March, the reproductive ratio was estimated to be 1.94 (CI 1.89-1.98) and 2.00 (CI 1.97-2.02) in Mombasa and Nairobi, respectively, with associated doubling-time of 4.84 and 4.59 days, respectively. After March, the transmission curves flattened substantially. This change is consistent with the introduction of containment measures by the Kenyan government, and evidence of substantial reduction in mobility (see Google Mobility data Fig. S1). However, we should note that there was very limited PCR testing available in Kenya before April 2020, and our estimates of $R(t)$ pre-April 2020 rely on the assumption that $R(t)$ dropped by ~45% in late March, in parallel to the drop in mobility data (see *Methods* and *supporting data*).

From late April, through May and June, and into July the evidence suggests movement restrictions became steadily less effective. The waning effectiveness of movement restrictions results in an inferred increase in $R(t)$ across Kenyan counties and an increased rate of epidemic growth (Figure 2). The increasing $R(t)$ estimates are broadly in line with predicted trends from Google mobility data (supporting information), although it should be noted that the $R(t)$ estimates exhibit secondary fluctuations around the increasing mobility trend (Figure 2). In Nairobi and Mombasa we predict that reduction in susceptibility of the population (Figure 1C,D) caused the effective reproductive ratio (R_{eff} ; the mean number of secondary cases accounting for reduced susceptibility) to drop significantly below the basic R value from June onwards (Figure 2). However, other counties, where the epidemic did not establish itself as early as Mombasa and Nairobi, and where a substantial majority of the population are likely to still be susceptible, now have $R(t)$ estimates which we estimate

rebounded to the original levels estimated as occurring before Kenyan public health measures (Figure 2).

By accounting for the delay of an average of 19 days between infection and death (supporting information for details on infection to death distribution) we find the transmission curve, estimated from PCR tests and serology, generates a good prediction of the observed trend in daily deaths in Nairobi and Mombasa (Figure 1 E, F). We did not use mortality data in transmission model inference, therefore the good fit to the observed trend in deaths with a PCR-confirmed test result represents an out-of-sample validation of the modelling¹⁸. Note, it is the distribution of deaths over time, rather than the absolute numbers, that we consider to be a good fit. In accord with observations, we estimate a peak of positive PCR test samples occurred at the end of July or early August in Nairobi and earlier, mid-June, in Mombasa. The lag between transmission peak and positive swab testing peak being explained by both the delay between infection and becoming detectable by PCR, and the period after an infected individual has ceased being actively infectious but remains detectable by PCR¹⁹ (Figure 1 G,H and A,B). As of the end of September 2020 we estimate that about 35.4% (CI 29.0%-40.4%) of the Nairobi population, and 30.3% (CI 23.6-36.7%) of the Mombasa population were serologically positive with SARS-CoV-2, (Figure 1 C,D). This estimated level of seropositivity is substantially higher than has been estimated in some countries that have been hit hard by the pandemic¹⁴⁻¹⁶. However, they are in broad agreement with a study in Niger state, Nigeria, from June 2020²⁰, as well as seropositivity rates reported from the hard-hit city of Manaus, Brazil, in May 2020²¹. Note that these estimates of seropositivity at the end of September assume both that waning seropositivity would not have had a significant effect on serological observations by late September, and furthermore that waning immunity leading to re-infection remained insignificant by late September.

SARS-CoV-2 attack rates in the first wave in Kenyan counties and the estimated crude infection-to-fatality ratio

Accounting for the sensitivity of the serological assay, and the delay between infection and seroconversion, we estimate that the actual exposure of the population to SARS-CoV-2 by September 30th was 43.3% (CI 35.3%-49.5%) in Nairobi and 37.6% (CI 29.2%-45.7%) in Mombasa (Figure 1 C,D). Such levels of population exposure are predicted to be associated with decreased rates of new cases due to reduced numbers of susceptible individuals in these urban populations, although this is also influenced by the estimated reproductive number and effective population size at risk of exposure (P_{eff}). The effective population size accounts for the impact of heterogeneity in the susceptibility, transmissibility and social interactivity in the population (supporting information for more details on effective population size in transmission modelling); for Nairobi it was inferred as 81.8% of actual population size (CI 66.7%-93.2%), for Mombasa 71.9% (CI 56.3%-86.5%). The effective population size estimates rest upon inferred variation in risk across the population. There remains a possibility of future increase in transmission if population mobility continues to rise, if population mixing patterns alter leading to changed

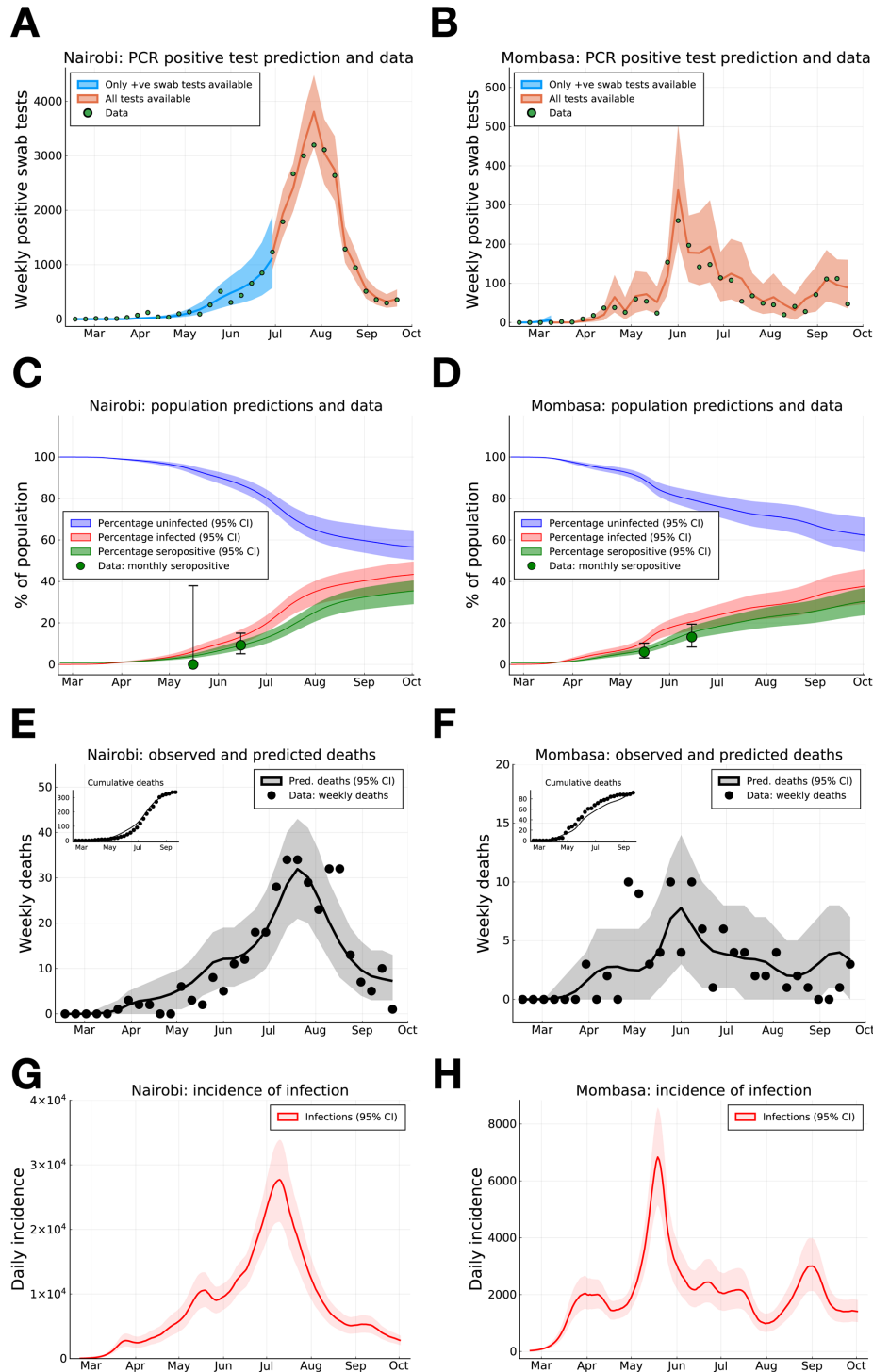


Figure 1. SARS-CoV-2 PCR positive swab tests, seroprevalence and deaths in Nairobi and Mombasa, Kenya, with model forecasting. (A) and (B) Weekly reported positive PCR positive swab tests (green dots) for Nairobi (A) and Mombasa (B), model prediction of mean weekly detection during both sampling periods when negative PCR test data was unavailable (blue curve), and available (orange curve). (C) and (D) Monthly seropositivity of Kenya National Blood Transfusion Service (KNBTS) blood donors in Nairobi (C) and Mombasa (D) (green dots), model predictions for population percentage of seropositivity (green curve), exposure to SARS-CoV-2 (red curve), and uninfected (blue curve). (E) and (F) Daily deaths with a positive SARS-CoV-2 test in Nairobi (E) and Mombasa (F) by date of death (black dots), and model prediction for daily deaths (black curve). Inset plots in (E) and (F) indicate cumulative reported deaths and model prediction. (G) and (H) Model estimates for rate of new infections per day in Nairobi (G) and Mombasa (H). Background shading indicates 95% central credible intervals. Dates for all graphs mark the 1st of each month.

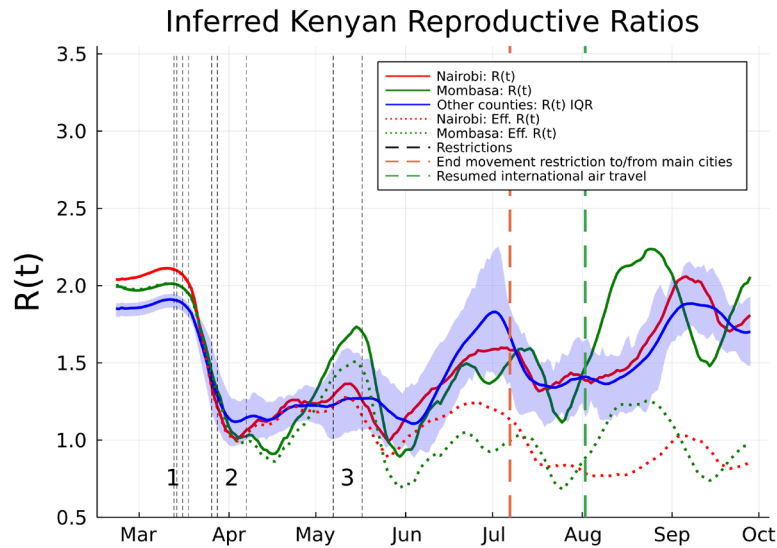


Figure 2. Estimated basic and effective reproductive numbers in Kenya since Feb 21st 2020. The posterior mean reproductive number for Nairobi (red curves), Mombasa (green curves), and the inter-quartile range (IQR) over mean reproductive number estimates for all other Kenyan counties (blue curve and shading). Shown are both the basic reproductive numbers (expected secondary infections in a susceptible population adjusted for mobility changes since the epidemic start; solid curves), and effective reproductive numbers (expected secondary infections accounting for depletion of susceptible prevalence in the population; dotted curves). The effective reproductive number varied significantly from county to county and is not shown except for Mombasa and Nairobi. Restrictions aimed at reducing mobility in risky transmission settings (black dotted lines) are labelled in groups. The chronologically ordered restrictions in each group are: 1) First PCR-confirmed case in Kenya, suspension of all public gatherings, closure of all schools and universities, and retroactive quarantine measures for recent returnees from foreign travel, 2) suspension of all inbound flights for foreign nationals, imposition of a national curfew, and regional lockdowns of Kilifi, Kwale, Mombasa and Nairobi counties, and 3) additional no-movement restriction of worst affected areas within Mombasa and Nairobi, and, closure of the border with Somalia and Tanzania. There were two relaxation of measures in this time frame: the end of no-movement restriction to Mombasa and Nairobi, and, the resumption of international air travel.

risk heterogeneity or if immunity is short lived, leading to a rebound in reported cases. One or more of these factors could lead either to lengthening the tail after the first peak in cases/deaths, or even to a secondary increase in cases and/or deaths.

The inferred IFR_{crude} values for both Nairobi ($IFR_{crude} = 0.019\%$ (CI 0.014%-0.024%)) and Mombasa ($IFR_{crude} = 0.022\%$ (CI 0.016%-0.027%)) are substantially lower than the age-adjusted IFR expected for Kenya under full ascertainment from the age-specific IFR estimated given by Verity *et al.* ($IFR_{verity} = 0.26\%$ ²²; and supporting information). This is a crude observational value for the infection to fatality ratio, since we do not currently have an estimate of the reporting bias of deaths of individuals infected with SARS-CoV-2. Therefore, our estimate of IFR_{crude} potentially reflects lower detection in Kenya compared to China, as well as any lower mortality risk due to fewer comorbidities.

We extended our model-based inference to each of the 47 counties in Kenya (see dataset S1 for parameter estimates, peak time estimates and IFR_{crude} estimates for each county). We find that, in addition to the two main Kenyan city counties, more than 25–30% of the population in each of the semi-urban counties neighbouring Nairobi (Kiambu, Kajiado, and Machakos) had been infected. However, the infection rate is predicted to be either lower than 25% and/or subject to high uncertainty in

other counties (with high uncertainty defined as a prediction standard error of > 10% of county population size; Figure 3).

Due to the lag between infection and the observability of the infected person (whether by swab PCR test, serology test, or death), we estimate that both daily PCR positive test detections and daily observed deaths attributed to COVID-19 across the two main cities, and semi-urban counties neighbouring Nairobi had a peak in early August 2020 (Figure 3 B,C). Hospitalisation rates are not available for all Kenyan hospitals. However, sentinel clinical surveillance of severe acute respiratory infection (SARI), with or without a PCR test for SARS-CoV-2, at 14 county hospitals suggests an increasing rate of adult admissions in June and July 2020²³. However, SARI admissions were lower in the early phase of the Kenyan epidemic than observed counts from the same months in 2018 and 2019²³ and the apparent rise in SARI admissions could represent a reversion towards pre-COVID numbers; this observation underlines the difficulties in using hospital data to understand the penetration of SARS-CoV-2 in Kenya.

Conclusions and discussion

Our modelling analysis provides a coherent account of the SARS-CoV-2 pandemic in Kenya up to end September 2020. Limitations include lacking information on the PCR testing denominators for the full time frame, the limited serological

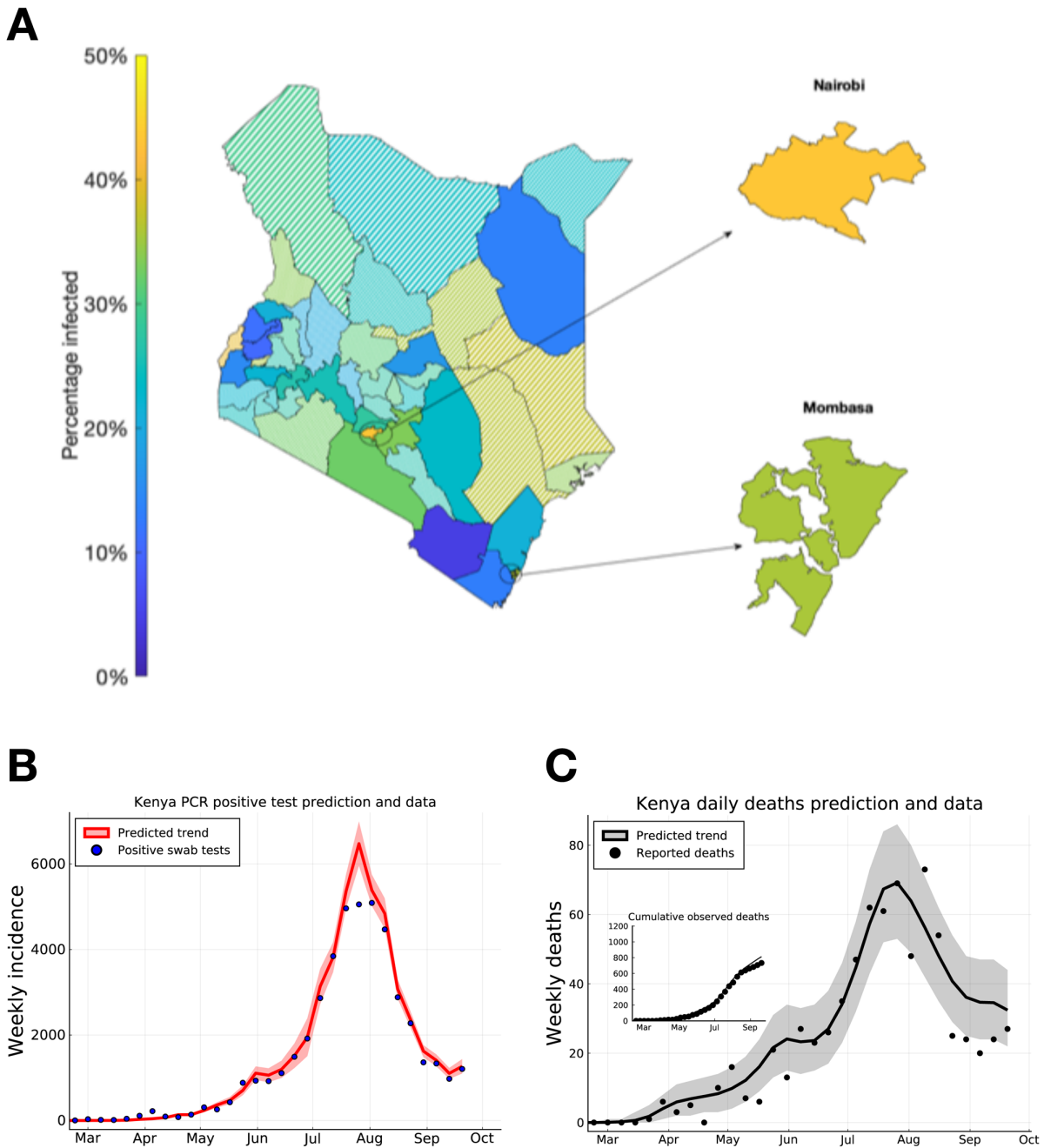


Figure 3. Predicting peak timing of transmission rate by Kenyan county, and forecasting of Kenya-wide PCR positive swab tests and reported deaths. (A) Posterior mean estimates for the attack rate (% of population) in each county. Solid shaded counties have a posterior standard deviation in their attack rate estimate of less than 10%, candy-stripe shaded counties have greater uncertainty associated with their attack rate estimate. (B) Kenya total positive swab tests collected by day of sample (blue dots) with model prediction of daily positive swab test trend (red curve). (C) Kenya total reported deaths with a positive swab test (black dots), with model prediction of reported death rates (black curve). Inset plot indicates cumulative reported deaths with model prediction of cumulative deaths. Dates on (B) and (C) mark 1st of the month.

survey and that we have applied a simple dynamic model. In mitigation similar results were obtained when excluding all negative tests, and the dynamic model is transparently a fit to the data where the availability of the latter is a key strength of our study.

Our analysis suggests that 30–50% of the urban population were already exposed by the end of September, and that the first wave of the Kenya epidemic peaked in the urban and semi-urban counties during a period of relatively little restrictions or physical distancing. This infers a burden of infection in Nairobi and Mombasa similar to extremely hard-hit cities in South America at the same time, e.g. Manaus²¹. However, in Manaus there was a substantial first epidemic wave through April–June 2020, with about 50% of the population infected, followed by a long epidemic tail leading to an estimated total 75% attack rate by the end of 2020²¹. In Kenya, the second wave came shortly after the first (October to December 2020), which suggests that the causes of the Kenyan second wave are complex.

Whilst the full picture of the epidemiology in Kenya will not be established until cause-specific mortality data become available (e.g. from resumption of Demographic Surveillance System and verbal autopsy activities), our model, fitted to three sources of nationwide longitudinal data, suggests that the number of symptomatic COVID-19 cases reported and the mortality attributed to the SARS-CoV-2 epidemic are substantially lower in Kenya than in Europe and the USA at a similar stage of the epidemic. This would remain the case even if reported deaths accounted for just 1/10th of the true value. However, there is insufficient data for speculating on the degree of under-reporting and previous estimates of 1 in 4 deaths occurring in hospital may not be generalizable to the hospital access during the COVID-19 pandemic²⁴.

Late 2020 saw the spread of COVID-19 to more rural areas of Kenya, with less infrastructure and access to public health facilities and a second wave of SARS-CoV-2. This second wave needs to be dissected and understood. Policy makers need to balance the direct and indirect health and socio-economic consequences of any control measures; a balance that becomes more precise as we develop a better understanding of SARS-CoV-2 dynamics in Kenya.

Methods

Transmission model definition

The dynamics of transmission in each Kenyan county were assumed to follow a SEIR transmission model with an effective population size parameter (P_{eff})²⁵. The SEIR model with effective population size is an extension of the homogeneous SEIR model²⁶ with the additional flexibility that $P_{eff}N$ out of a total population size N in each county is at risk of contracting SARS-CoV-2. $P_{eff} = 1$ recovers the homogeneous SEIR model, whereas, $P_{eff} < 1$, recovers the effect of underlying heterogeneity in the transmission potential and risk in the population of the county on the aggregate dynamics of epidemic. This aspect of heterogeneous models of transmission has been widely investigated, for example, in the context of comparing vaccination

coverage thresholds for elimination between uniform and targeted vaccination policies²⁷. In the context of the SARS-CoV-2 pandemic modelling literature, the possible role of population heterogeneity in decoupling estimates of R_0 from predictions of the “herd-immunity” threshold and final attack rate has again been identified^{28,29}. In this study, rather than make strong assumptions about the mechanism of population heterogeneity, e.g. differential susceptibility, differential rates of social mobility etc., we have taken a phenomenological approach; the effect of heterogeneity in the population was encoded in the effective population parameter P_{eff} and this parameter was inferred jointly with R_0 . Our *a priori* belief was that the most probable value was $P_{eff} = 1$. We assumed that P_{eff} was constant over the period of inference.

The model dynamics for each Kenyan county were represented as a system of ordinary differential equations,

$$\begin{aligned}\dot{S}(t) &= -\gamma R_t \frac{S(t)I(t)}{P_{eff}N}, \\ \dot{E}(t) &= \gamma R_t \frac{S(t)I(t)}{P_{eff}N} - \sigma E, \\ \dot{I}(t) &= \sigma E(t) - \gamma I(t), \\ \dot{R}(t) &= \gamma I(t), \\ \dot{C}(t) &= \gamma R_t \frac{S(t)I(t)}{P_{eff}N}.\end{aligned}\tag{1}$$

With initial conditions (time 0 is the calendar date 21st Feb 2020 and all rates are per day),

$$\begin{aligned}S(0) &= P_{eff}N - E_0 - I_0, \\ E(0) &= E_0, I(0) = I_0, R(0) = 0, C(0) = 0.\end{aligned}\tag{2}$$

Where the dynamic variables $S(t)$, $E(t)$, $I(t)$, $R(t)$ were the numbers of susceptibles-at-risk, exposed (but not yet infectious), infectious, and, recovered individuals in the county. The full number of susceptibles in the county at any time was $(1 - P_{eff})N + S(t)$. $C(t)$ was the cumulative numbers of infected individuals in the county.

The incubation-to-infectious rate was $\sigma = 1/3.1$ per day, and the recovery rate was $\gamma = 1/2.4$ per day, implying a mean generation time of 5.5 days (see *Supporting information* for a comparison to the generation distribution inferred by Ferretti *et al.*³⁰). The instantaneous reproductive ratio $R_t = R_0\beta_t$ decomposed into a basic reproductive ratio R_0 and an effective contact rate β_t , where $\beta_t = 1$ represents a pre-pandemic baseline contact rate in the population.

Transmission model inference

We used a mixed Bayesian and maximum a-posteriori (MAP) approach to parameter inference for each of the 47 Kenyan counties, based on daily observations of positive and negative PCR and serology tests in each county. The likelihood of individuals being detectable on any given day was based on whether they had been infected before that day, and, the number of days since their infection. The number of new infections

on each day n , was denoted I_n . For a given set of model parameters I_n was generated by solving the ODE system (1), giving,

$$I_n = C(n+1) - C(n), \quad (3)$$

for each day n . Given the daily numbers of new infections, the number of people in the county on each day n who are detectable by PCR testing, denoted $(P^+)_n$, and serological testing, $(S^+)_n$, were given by convolving the new infection time series with the probability of (respectively) being detectable by a PCR or serological test τ days after infection, $Q_{PCR}(\tau)$ and $Q_{sero}(\tau)$:

$$\begin{aligned} (P^+)_n &= [I * Q_{PCR}](n), \\ (S^+)_n &= [I * Q_{sero}](n). \end{aligned} \quad (4)$$

The log-likelihood function for each county has the form,

Where, $\ln f_{PCR}((ObsP^+)_n | (P^+)_n, \theta_{OM})$, and, $\ln f_{sero}((ObsS^+)_n | (S^+)_n, \theta_{OM})$, were, respectively, the log-probability of observing $(ObsP^+)_n$ PCR test-positives and $(ObsS^+)_n$ serological test positives on days $n = 1, \dots, T$ given the model prediction of numbers of PCR and serological detectable people in the population, and the observation model parameters θ_{OM} . Day $n = 1$ corresponded to the calendar date 21st February 2020, and, day $n = T = 223$ corresponded to 30th September 2020.

The underlying transmission prediction depended only on parameters relevant to infection (e.g. basic pre-measures reproductive ratio etc), however, the statistical modelling of the observation of evidence of these infections varied by type of test and availability of negative PCR test data. Together these form a likelihood function, which integrates the different data sources, since they are all, ultimately, generated by the same underlying infection process. The three statistical models of observation data were:

- **Serological tests:** On each day that serological samples were collected, the log-probability of the observed number of positive tests ($\ln f_{sero}((ObsS^+)_n | (S^+)_n, \theta_{OM})$) was assumed to be that of a Beta-Binomial distribution with unbiased sampling of the underlying proportion of serologically detectable people in the county ($(S^+)_n/N$). The extra dispersion compared to a Binomial sample being due to uncertainty in the underlying sensitivity of the serological assay (see *supporting information in supporting data*).
- **PCR swab positive tests when no data on negative PCR tests was available:** Negative PCR swab tests were not available in every county on every day of simulation. When negative swab tests were not available we assumed that the log-probability of the daily observed PCR test positives was from a Negative-binomial distribution:

$$\begin{aligned} \mu_n &= p_{test} TR(n) (P^+)_n \\ (ObsP^+)_n &\sim NegBin(\hat{\mu} = \mu_n, \hat{\alpha} = \alpha). \end{aligned} \quad (5)$$

Where the mean number of daily observed test positives, conditional on the model prediction of PCR-detectable people in the population, is based on sampling a fraction $p_{test} TR(n)$. p_{test} was an observation parameter that was jointly

inferred during inference, and $TR(n)$ was a normalized testing rate based on nationally reported data (see *supporting information in supporting data*). α was a clustering factor for negative-binomial sampling, jointly inferred with other model parameters.

- **PCR swab positive tests when data on negative PCR tests was available:** When both positive and negative PCR test data was available, we assumed that the fraction of positive samples reflected a biased observation of the underlying true fraction of PCR-detectable individuals in the population, e.g. being infected with SARS-CoV-2 could be expected to influence the odds of someone seeking a PCR test. We assumed that the daily detection of PCR test positives could be modelled as samples from a Beta-Binomial distribution with two parameters to infer: 1) The bias of a PCR-detectable individual being PCR tested compared to a PCR-undetectable individual (χ), and, 2) the effective sample size parameter (M_{PCR}). $M_{PCR} \rightarrow \infty$ recovered a Binomial distribution for the number of positive PCR tests were observed among the tests conducted that day, $M_{PCR} < \infty$ allowed the model to infer much greater variance in daily proportion of test positives than would be expected from a Binomial distribution. On days where negative swab tests were available, we connect the observable status of epidemic to the data thus,

$$p_n = \frac{\chi (P^+)_n}{(\chi - 1)(P^+)_n + N} \quad (6)$$

$$(ObsP^+)_n \sim BetaBin(\hat{N}_s = N_{PCR,n}, \hat{p} = p_n, \hat{M} = M_{PCR}).$$

Where $N_{PCR,n}$ is the total number of PCR swab samples collected on day n and p_n is the proportion of tests performed returning positive expected by the model, accounting for bias in the sampling regime. The bias parameter $\chi = 1$ recovers an unbiased sample of PCR positives from the underlying population.

Supporting information gives further details on the data sources and the log-likelihood calculation including a full description of all observation model parameters and the functional forms and underlying evidence for Q_{PCR} and Q_{sero} . The data sources used were: The Kenya Ministry of Health National linelist, the Kenya Medical Research Institute Wellcome Trust Research Programme (KEMRI-WTRP) serological surveillance programme and Google mobility data³¹. The full Kenyan SARS-CoV-2 line list contains sensitive personal information that could potentially allow the identification of individual cases. The analysis performed in this study only required an aggregated dataset derived from the Kenyan linelist. Other data used in this paper was openly available. All data is available in the main text or as underlying data³².

We assumed that β_t was piece-wise constant on days, and, therefore, could be reconstructed from daily effective contact rates $(\beta_n)_{n=1, \dots, T}$. For any fixed estimate of the effective contact rate β , we used Hamiltonian Markov-chain Monte Carlo (HMC)³³ to estimate the posterior distribution for the transmission model parameters; that is the initial condition values (E_0, I_0) and

fixed parameters (P_{eff} , R_0) jointly with the observation model parameters θ_{OM} . Prior distributions for parameters were chosen for groups of counties (e.g. largely rural counties had different priors to major urban conurbations like Nairobi and Mombasa; see *supporting information* for further details). Starting from an initial estimate that β_t followed daily Google mobility trends³¹ for the whole period, we sequentially improved our β_t estimate using the expectation-maximisation (EM) algorithm³⁴. The E-step corresponding to posterior distribution estimation using HMC, and the M-step corresponding to optimising the daily effective contact rate estimates (β_n) _{$n=41, \dots, T$} using the popular stochastic gradient descent algorithm ADAM³⁵. The first 40 days of effective contact rate estimates (β_n) _{$n=1, \dots, 40$} were assumed to be fixed to their Google estimate; this improved identifiability jointly with R_0 and captured the observed sharp drop in mobility in response to Kenyan public health measures following the first identified case on 13th March 2020. See *supporting information* for further details on the use of Google mobility data and the EM algorithm method used in this study.

After inference of transmission parameters, the model implied a prediction of the expected number of daily deaths due to COVID, $\mathbb{E}(X^+)_n$, based on an overall population infection-to-fatality ratio (IFR), and, the delay distribution between infection and death, p_{ID} ,

$$\mathbb{E}(X^+)_n = IFR[t * p_{ID}](n). \quad (6)$$

In this study, we assume that the IFR is constant for each county over the period of inference, which allows us to construct a Bayesian estimator of the crude IFR, IFR_{crude} , by fitting to the observed daily numbers of test-positive deaths, $(ObsX^+)_n$ (see *supporting information* for details and background data informing p_{ID}). Because the observed test-positive deaths were not used in inferring model parameters, we treat the log-predictive density of deaths from the model as an out-of-sample validation metric for the model. However, we emphasise that the out-of-sample comparison is to the **trend** of daily deaths, because this is invariant to the IFR_{crude} estimator, which is itself sensitive to under-reporting of COVID deaths. *Supporting information* gives full details on the Bayesian model validation used in this study.

This study was approved by the Kenya Medical Research Institute Scientific and Ethics Review Unit (KEMRI-SERU) with approval numbers KEMRI/SERU/CGMR-C/203/4085 and KEMRI/SERU/CGMR-C/203/3426 for the modelling and serosurvey studies respectively.

Data availability

Underlying data

Zenodo: Revealing the extent of the first wave of the COVID-19 pandemic in Kenya based on serological and PCR-test data. <https://doi.org/10.5281/zenodo.4705244>³² This project contains the following underlying data:

- Data S4 (The number of positive, and negative where available, PCR-confirmed swab tests for each county by date of sample collection (21st Feb to 30th September)).
- Data S5. (The number of positive and negative sero-logical results for each county by date of sample collection (21st Feb to 6th August)). This is from the **Kenyan Ministry of Health National linelist**.
- Data S6. (The number of deaths with a PCR-confirmed swab test for each county by recorded date of death (21st Feb to 30th September)).
- Data S7. (Summary data of Kenyan epidemic, including reported total number of test performed in Kenya.
- supp material.docx (A more detailed description of the data)

Software availability

The analysis code was written in Julia language version 1.4.

- The code base underlying the analysis is accessible at the open github repository <https://github.com/ojal/KenyaSerology>. For the analysis presented here we developed a module in the Julia programming language called *KenyaSerology*. Tutorial notebooks for running *KenyaSerology* and analysing the underlying data are available in the repository.
- Archived source code at time of publication: <https://doi.org/10.5281/zenodo.4705244>³²
- License: [GNU General Public License v3.0](https://www.gnu.org/licenses/gpl-3.0.html)

Acknowledgements

We thank all members of Kenya's county rapid response teams (who collected swab samples), and testing centres (who conducted the laboratory PCR assays) and of Kenya National Blood Transfusion Service Centres. This study is published with the Permission of the Director of the Kenya Medical Research Institute.

An earlier version of this article can be found on medRxiv (doi:<https://doi.org/10.1101/2020.09.02.20186817>)

References

1. Gilbert M, Pullano G, Pinotti F, *et al.*: **Preparedness and vulnerability of African countries against importations of COVID-19: a modelling study.** *Lancet.* 2020; **395**(10227): 871–877. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
2. Cabore JW, Karamagi HC, Kipruto H, *et al.*: **The potential effects of widespread community transmission of SARS-CoV-2 infection in the World Health Organization African Region: a predictive model.** *BMJ Glob Health.* 2020; **5**(5): e002647. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
3. Diop BZ, Ngom M, Biyong CP, *et al.*: **The relatively young and rural**

- population may limit the spread and severity of COVID-19 in Africa: a modelling study. *BMJ Glob Health*. 2020; 5(5): e002699.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
4. Braun J, Loyal L, Frentsch M, *et al.*: **SARS-CoV-2-reactive T cells in healthy donors and patients with COVID-19.** *Nature*. 2020; 587(7833): 270–274.
[PubMed Abstract](#) | [Publisher Full Text](#)
 5. Mbow M, Lell B, Jochems SP, *et al.*: **COVID-19 in Africa: Dampening the storm?** *Science*. 2020; 369(6504): 624–626.
[PubMed Abstract](#) | [Publisher Full Text](#)
 6. Hale T, Webster S, Petherick A, *et al.*: **Oxford covid-19 government response tracker.** Blavatnik School of Government. 2020; 25.
[Reference Source](#)
 7. Ma Y, Zhao Y, Liu J, *et al.*: **Effects of temperature variation and humidity on the death of COVID-19 in Wuhan, China.** *Sci Total Environ*. 2020; 724: 138226.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 8. Kenyan Ministry of Health: **COVID-19 situation reports.** 2021.
[Reference Source](#)
 9. Barasa E, Ouma P, Okiro E: **Assessing the Hospital Surge Capacity of the Kenyan Health System in the Face of the COVID-19 Pandemic.** *medRxiv*. 2020; 1–24.
[Publisher Full Text](#)
 10. Macharia PM, Joseph NK, Okiro EA: **A vulnerability index for COVID-19: spatial analysis to inform equitable response in Kenya.** *medRxiv*. 2020; 1–26.
[Publisher Full Text](#)
 11. Kenyan Ministry of Health: **Press statement on the update of the coronavirus in the country and response measure.** 2020; 1–3.
[Reference Source](#)
 12. Roser M, Ritchie H, Ortiz-Ospina E, *et al.*: **Coronavirus pandemic (COVID-19).** Our Worldin Data. 2020.
[Reference Source](#)
 13. Uyoga S, Adetifa IMO, Karanja HK, *et al.*: **Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Kenyan blood donors.** *Science*. 2021; 371(6524): 79–82.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 14. Ward H, Atchison C, Whitaker M, *et al.*: **Antibody prevalence for SARS-CoV-2 following the peak of the pandemic in England: REACT2 study in 100,000 adults.** *medRxiv*. 2020; 1–20.
[Publisher Full Text](#)
 15. Pollán MM, Pérez-Gómez B, Pastor-Barriuso R, *et al.*: **Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study.** *Lancet*. 2020; 396(10250): 535–544.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 16. Havers FP, Reed C, Lim T, *et al.*: **Seroprevalence of Antibodies to SARS-CoV-2 in 10 Sites in the United States, March 23-May 12, 2020.** *JAMA Intern Med*. 2020; 1–11.
[PubMed Abstract](#) | [Publisher Full Text](#)
 17. Flaxman S, Mishra S, Gandy A, *et al.*: **Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe.** *Nature*. 2020; 584(7820): 257–261.
[PubMed Abstract](#) | [Publisher Full Text](#)
 18. Gelman A, Hwang J, Vehtari A: **Understanding predictive information criteria for Bayesian models.** *Stat Comput*. 2014; 24(6): 997–1016.
[Publisher Full Text](#)
 19. Zhou F, Yu T, Du R, *et al.*: **Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study.** *Lancet*. 2020; 395(10229): 1054–1062.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 20. Majiya H, Aliyu-Paiko M, Balogu VT, *et al.*: **Seroprevalence of COVID-19 in Niger State.** *medRxiv*. 2020; 1–24.
[Reference Source](#)
 21. Buss LF, Prete CA Jr, Abraham CMM, *et al.*: **Three-quarters attack rate of SARS-CoV-2 in the Brazilian Amazon during a largely unmitigated epidemic.** *Science*. 2021; 371(6526): 288–292.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 22. Verity R, Okell LC, Dorigatti I, *et al.*: **Estimates of the severity of coronavirus disease 2019: a model-based analysis.** *Lancet Infect Dis*. 2020; 20(6): 669–677.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 23. KEMRI-Wellcome Trust Research Programme: **Status of the COVID-19 Pandemic in Kenya: Evidence from serological and clinical surveillance, and predictive modelling.** Technical report, 2020.
[Reference Source](#)
 24. Ong'ayo G, Ooko M, Wang'onduru R, *et al.*: **Effect of strikes by health workers on mortality between 2010 and 2016 in Kilifi, Kenya: a population-based cohort analysis.** *Lancet Glob Health*. 2019; 7(7): e961–e967.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 25. Li M, Dushoff J, Bolker BM: **Fitting mechanistic epidemic models to data: A comparison of simple Markov chain Monte Carlo approaches.** *Stat Methods Med Res*. 2018; 27(7): 1956–1967.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 26. Keeling MJ, Rohani P: **Modeling Infectious Diseases in Humans and Animals.** Princeton University Press, 2008.
[Publisher Full Text](#)
 27. Anderson RM, May RM, Anderson B: **Infectious Diseases of Humans: Dynamics and Control.** Oxford University Press, Oxford, 1991.
 28. Aguas R, King JG, Gonçalves G, *et al.*: **Herd immunity thresholds for SARS-CoV-2 estimated from unfolding epidemics.** *medRxiv*. 2020; 1–42.
[Publisher Full Text](#)
 29. Tkachenko AV, Maslov S, Elbanna A, *et al.*: **Persistent heterogeneity not short-term overdispersion determines herd immunity to COVID-19.** *medRxiv*. 2020; 1–10.
[Reference Source](#)
 30. Ferretti L, Wymant C, Kendall M, *et al.*: **Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing.** *Science*. 2020; 368(6491): eabb6936.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 31. Google LLC: **Google COVID-19 Community Mobility Reports.**
[Reference Source](#)
 32. Brand S, Ojal: **ojal/KenyaSerology: First release (Version v1.0.0).** *Zenodo*. 2021.
<http://www.doi.org/10.5281/zenodo.4705244>
 33. Betancourt M: **A Conceptual Introduction to Hamiltonian Monte Carlo.** arXiv.org. 2017.
[Reference Source](#)
 34. Dempster AP, Laird NM, Rubin DB: **Maximum likelihood from incomplete data via the EM algorithm.** *J R Stat Soc Series B*. 1977; 39(1): 1–38.
[Reference Source](#)
 35. Kingma DP, Ba J: **Adam: A method for stochastic optimization.** CoRR abs/1412.6980. 2014.
[Reference Source](#)

Open Peer Review

Current Peer Review Status:   

Version 2

Reviewer Report 27 June 2022

<https://doi.org/10.21956/wellcomeopenres.19588.r51227>

© 2022 Cowling B. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

 **Benjamin J. Cowling** 

WHO Collaborating Centre for Infectious Disease Epidemiology and Control, School of Public Health, The University of Hong Kong, Hong Kong, Hong Kong

This is a nice study. The authors used national surveillance PCR test data, serological data and mobility data to develop and fit a county-specific transmission model for Kenya up to the end of September 2020, which encompasses the first wave of SARS-CoV-2 transmission in the country. Authors estimated that the first wave of the SARS-CoV-2 pandemic peaked before the end of July 2020 in the major urban counties, with 30-50% of residents infected.

This is an important study and likely has implications for other neighbouring countries in Africa as well. Data on COVID-19 from the African continent are very limited. I encourage indexing of this revised submission and I just had two minor comments:

- Abstract final phrase - "further wave of transmission" do you mean specifically a further wave of transmission with an antigenically-different strain such as a new variant which can escape the population immunity that has built up? Or you mean that the first wave was controlled before population immunity reached a herd immunity threshold? Even so, if the same strain circulates again, one wouldn't expect a large wave because of the existing immunity from the first wave. Waning immunity in medium-term could also play a role in allowing subsequent epidemics.
- Conclusions - I don't find the Manaus estimate of 75% particularly compelling due to methodological issues in that study. There should be other locations with less extreme first-year serological data?

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Infectious disease epidemiology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 05 Sep 2022

James Nokes, KEMRI-Wellcome Trust Research Programme, Kenya

We thank the reviewer for useful review and comments.

Point 1 "Abstract final phrase - "further wave of transmission" do you mean specifically a further wave of transmission with an antigenically-different strain such as a new variant which can escape the population immunity that has built up? Or you mean that the first wave was controlled before population immunity reached a herd immunity threshold? Even so, if the same strain circulates again, one wouldn't expect a large wave because of the existing immunity from the first wave. Waning immunity in medium-term could also play a role in allowing subsequent epidemics."

We were not, at this early stage in the pandemic, suggesting a further wave from a new variant. Instead, we inferred some heterogeneity in population susceptibility, transmissibility or social interactivity, encapsulated by the phenomenological term, the effective population size at risk of exposure (Peff). This population heterogeneity put a break on virus spread in the first wave but made possible a second wave that moved into less infected sections of the population. This was unmeasured and not well understood at the time. However, in our subsequent paper (DOI: [10.1126/science.abk0414](https://doi.org/10.1126/science.abk0414)) we were able to explicitly account for this heterogeneity as differences in mobility of lower (high transmission in wave 1) and higher (low transmission in wave 1) socio-economic classes, particularly in the urban setting.

No change has been made to the manuscript.

Point 2. "Conclusions - I don't find the Manaus estimate of 75% particularly compelling due to methodological issues in that study. There should be other locations with less extreme first-year

serological data?"

We agree with the reviewer and revise the section of text referring to the estimate of 75% (see appended). We do already make other comparisons from serosurveys from that period from other settings including Spain, England, United States and Niger.

The revised text is 'Our analysis suggests that 30–50% of the urban population were already exposed by the end of September, and that the first wave of the Kenya epidemic peaked in the urban and semi-urban counties during a period of relatively little restrictions or physical distancing. This level of exposure however was not sufficient to prevent a second wave which came shortly after the first (October to December 2020), which we assume to have resulted from heterogeneous spread of the virus, perhaps due to variation in population susceptibility, transmissibility or social interactivity.'

Competing Interests: none to disclose

Version 1

Reviewer Report 01 November 2021

<https://doi.org/10.21956/wellcomeopenres.18470.r46411>

© 2021 Wesolowski A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Amy Wesolowski 

Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

In 'Revealing the extent of the first wave of the COVID-19 pandemic in Kenya based on serological and PCR-test data' the authors use a range of statistical and mechanistic approaches to investigate the first wave of the pandemic in Kenya using both serological and PCR-test data. Overall, the authors have done an excellent job at integrating appropriate methods with available data sets to provide a useful investigation into transmission patterns in Kenya. Given the dearth of analyses from Sub-Saharan Africa countries, this manuscript is a welcome addition. Overall, I have few comments, the majority of which are minor and outlined below.

The authors have analyzed transmission patterns across all counties in Kenya, yet only show those for the two most populous (Nairobi and Mombasa) and a country-level result. While this is reasonable for the main text, I believe that adding in results by county to the supplementary information would help the interpretation of the country-level results. Particularly given likely differences in testing and reporting by county, a better understanding of the overall estimates of model predictions and R_t estimates by county would be incredibly valuable. For example, the

authors provide the percentage infected by county in Figure 3, however these values are different to interpret in context, particularly without seeing the data and estimates by county.

In Figure 1 (C,D) the model prediction CIs are incredibly narrow, which seems surprising. Does this occur across the country? Or is this mainly due to the higher quality data in both Nairobi and Mombasa?

In Figure 2, it would be helpful (perhaps in the supplement or in this main figure) to provide context on when restrictions were lifted in addition to when they were put in place. Further, it is interesting that the IQR for the R_t estimates early in the pandemic seem exceptionally narrow. Additional elaborations on these points (is it likely due to overfitting? Some factors associated with the model fitting? Etc.) should be included. In addition, the R_t estimates for Mombasa in August/September seem substantially different than the rest of the country. Can the authors provide additional context? And do they see similar patterns across the coastal counties during this time?

The authors do an excellent job appropriately combining different data sets, which is well explained in the supplementary information. Some of these details would be incredibly helpful to move to the main text, in particular additional detail on how the authors treat the serological versus PCR data (and when there were both negative and positive PCR results) and how these two pieces of evidence are integrated.

Finally, in the supplement the authors use the phrase 'Chinese epidemic', but a small point that it may be more appropriate to say 'epidemic in China'.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: I am an infectious disease epidemiology focused on transmission modeling.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 03 Feb 2022

James Nokes, KEMRI-Wellcome Trust Research Programme, Kenya

In 'Revealing the extent of the first wave of the COVID-19 pandemic in Kenya based on serological and PCR-test data' the authors use a range of statistical and mechanistic approaches to investigate the first wave of the pandemic in Kenya using both serological and PCR-test data. Overall, the authors have done an excellent job at integrating appropriate methods with available data sets to provide a useful investigation into transmission patterns in Kenya. Given the dearth of analyses from Sub-Saharan Africa countries, this manuscript is a welcome addition. Overall, I have few comments, the majority of which are minor and outlined below.

Thank you for your kind comments on this piece of research.

The authors have analyzed transmission patterns across all counties in Kenya, yet only show those for the two most populous (Nairobi and Mombasa) and a country-level result. While this is reasonable for the main text, I believe that adding in results by county to the supplementary information would help the interpretation of the country-level results. Particularly given likely differences in testing and reporting by county, a better understanding of the overall estimates of model predictions and R_t estimates by county would be incredibly valuable. For example, the authors provide the percentage infected by county in Figure 3, however these values are difficult to interpret in context, particularly without seeing the data and estimates by county.

This is a very good point. In our original analysis we plotted model prediction intervals against actual data by county as part of model diagnostic. We have now improved our diagnostic visualisations to match the format of the main manuscript plots for Nairobi and Mombasa, and generated county-specific plots for model-based prediction/credible intervals for (i) PCR swab test positives, (ii) population exposure, (iii) deaths, (iv) $R(t)$ against data (where available). All 188 plots (4 x 47 counties) are available in the data and code repository associated with this paper <https://github.com/ojal/KenyaSerology> .

In Figure 1 (C,D) the model prediction CIs are incredibly narrow, which seems surprising. Does this occur across the country? Or is this mainly due to the higher quality data in both Nairobi and Mombasa?

This is correct, the reasonably high model certainty about seroprevalence in Nairobi and Mombasa was because of higher data quality in the main cities in Kenya. Other counties had much wider CIs for model predicted seroprevalence, in Figure 3 we attempted to visualise this by candy-striping the county shading for counties with a posterior standard deviation in model prediction of population exposure of greater

than 10%; that is the counties where a >10% deviation from the posterior mean estimate of population exposure would not be highly unexpected. We have now added population exposure plots for every county, including credible intervals for the population seropositivity.

In Figure 2, it would be helpful (perhaps in the supplement or in this main figure) to provide context on when restrictions were lifted in addition to when they were put in place.

Over the time scale this paper is concerned with (February – October 2020) there were only two significant relaxations, however, this included lifting the movement restrictions on travel out of Nairobi and Mombasa (6th July 2020), and, therefore, was an oversight to not include in Figure 2. We have now revised Figure 2 to include the timing of relaxation of targeted movement restrictions and the reopening of international flight into Kenya (1st August 2020). These have been added to a revised Figure 2 and mentioned in the caption of figure 2.

Further, it is interesting that the IQR for the R_t estimates early in the pandemic seem exceptionally narrow. Additional elaborations on these points (is it likely due to overfitting? Some factors associated with the model fitting? Etc.) should be included.

The early tight estimate for R_t reflected (i) fairly tight estimates that $R_t \sim 1.1-1.2$ in most counties in late April 2020, and (ii) our assumption that for the first 40 days of the simulation (20th Feb 2020 – 31st March 2020) R_t was proportional to Google data derived estimates of mixing in indoor settings outside the home. In the early stages of the epidemic in Kenya there was very limited testing capacity for detection of SARS-CoV-2 transmission in the community (first positive test result was on 12th March 2020). This means that we were forced to make an assumption about the effective relative contact rates, because we could not infer them from epidemiological data. The Google data suggested a fairly uniform 40-45% decrease in mixing in inside settings (e.g., the workplace, etc) outside the home by mid-April 2020, by which time $R_t \sim 1.1 - 1.2$ in most counties. Because of our modelling assumption, this confidence in R_t in April-May 2020 was translated into confidence in $R_t \sim 1.8 - 2.0$ during an unobserved epidemic in February 2020.

We have added this sentence to the opening paragraph of the *Underlying transmission rates in Mombasa and Nairobi during the first wave* section:

“However, we should note that there was very limited PCR testing available in Kenya before April 2020, and our estimates of $R(t)$ pre-April 2020 rely on the assumption that $R(t)$ dropped by ~45% in late March, in parallel to the drop in mobility data (see Methods and supporting data).”

In addition, the R_t estimates for Mombasa in August/September seem substantially different than the rest of the country. Can the authors provide additional context? And do they see similar patterns across the coastal counties during this time?

This is an interesting observation. Upon inspection, other counties in coastal province (Kilifi, Lamu, Tana River and Taita Taveta, but not the Tanzanian border county Kwale)

also have a spike in Rt estimates in late August/early September 2020 (see county specific plots). A possibility is that this is connected to the relaxation of movement into Mombasa, however, there is a delay of greater than 4 weeks between that relaxation and the Rt increase.

The authors do an excellent job appropriately combining different data sets, which is well explained in the supplementary information. Some of these details would be incredibly helpful to move to the main text, in particular additional detail on how the authors treat the serological versus PCR data (and when there were both negative and positive PCR results) and how these two pieces of evidence are integrated.

Thank you very much, the aim was not to overwhelm a non-specialist audience, whilst providing full details within the supporting information. We have now added a further paragraph in the Methods section which we hope makes our methodology clearer to the reader.

“The underlying transmission prediction depended only on parameters relevant to infection (e.g. basic pre-measures reproductive ratio etc), however, the statistical modelling of the observation of evidence of these infections varied by type of test and availability of negative PCR test data. Together these form a likelihood function, which integrates the different data sources, since they are all, ultimately, generated by the same underlying infection process. The three statistical models of observation data were:

- **Serological tests:** On each day that serological samples were collected, the log-probability of the observed number of positive tests ($\ln f_{sero}((ObsS^+)_n | (S^+)_n, \theta_{OM})$) was assumed to be that of a Beta-Binomial distribution with unbiased sampling of the underlying proportion of serologically detectable people in the county ($(S^+)_n/N$). The extra dispersion compared to a Binomial sample being due to uncertainty in the underlying sensitivity of the serological assay (see *supporting information in supporting data*).
- **PCR swab positive tests when no data on negative PCR tests was available:** Negative PCR swab tests were not available in every county on every day of simulation. When negative swab tests were not available we assumed that the log-probability of the daily observed PCR test positives was from a Negative-binomial distribution:

$$\mu_n = p_{test} TR(n) (P^+)_n$$

$$(ObsP^+)_n \square NegBin(\mu^{\wedge} = \mu_n, \alpha^{\wedge} = a) \quad (5)$$

Where the mean number of daily observed test positives, conditional on the model prediction of PCR-detectable people in the population, is based on sampling a fraction p_{test} $TR(n)$, p_{test} was an observation parameter that was jointly inferred during inference, and $TR(n)$ was a normalized testing rate based on nationally reported data (see *supporting information in supporting data*). a was a clustering factor for negative-binomial sampling, jointly inferred with other model parameters.

- **PCR swab positive tests when data on negative PCR tests was available:** When both positive *and* negative PCR test data was available, we assumed that the fraction of positive samples reflected a biased observation of the underlying true fraction of PCR-detectable individuals in the population, e.g. being infected with SARS-CoV-2

could be expected to influence the odds of someone seeking a PCR test. We assumed that the daily detection of PCR test positives could be modelled as samples from a Beta-Binomial distribution with two parameters to infer: 1) The bias of a PCR-detectable individual being PCR tested compared to a PCR-undetectable individual (χ), and, 2) the effective sample size parameter (M_{PCR}). $M_{PCR} \rightarrow \infty$ recovered a Binomial distribution for the number of positive PCR tests were observed among the tests conducted that day, $M_{PCR} < \infty$ allowed the model to infer much greater variance in daily proportion of test positives than would be expected from a Binomial distribution. On days where negative swab tests were available, we connect the observable status of epidemic to the data thus,

$$p_n = \chi(P^+)_n / (\chi - 1)(P^+)_n + N$$

$$(ObsP^+)_n \sim \text{BetaBin}((N_s)^{\wedge} = N_{PCR,n}, p^{\wedge} = p_n, M^{\wedge} = M_{PCR}). \quad (6)$$

Where $N_{PCR,n}$ is the total number of PCR swab samples collected on day n and p_n is the proportion of tests performed returning positive expected by the model, accounting for bias in the sampling regime. The bias parameter $\chi=1$ recovers an unbiased sample of PCR positives from the underlying population."

Finally, in the supplement the authors use the phrase 'Chinese epidemic', but a small point that it may be more appropriate to say 'epidemic in China'.

Noted, and we have changed our language in the supporting information.

Competing Interests: No competing interests were disclosed.

Reviewer Report 14 July 2021

<https://doi.org/10.21956/wellcomeopenres.18470.r44753>

© 2021 Kimathi M. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Mark Kimathi 

Department of Mathematics, Machakos University, Machakos, Kenya

The ideas presented in the paper are plausible, resulting to a model, although a simple SEIR model, that is robust and adequately reproduces the observed trajectory of infections.

It was captivating to see the authors express the force of infection in terms of the case data time series (which is as a result of "successful" contact between susceptible and infected individuals). This is arguably the only way a compartmental model can reliably reproduce the waves of infections observed in data. Also the notion of, only a proportion of the population is at risk, is a quite realistic.

There are a few typos, as indicated below:

- On page 3, column 2, in line 10 write as...declining, first in...
- On page 3, column 1, in line 20 write as ...SARS-CoV-2.
- On page 3, column 1, in third paragraph at line 5 write as ...already...

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Mathematical modeling of infectious disease dynamics, fluid flows and other natural phenomena using differential equations.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 03 Feb 2022

James Nokes, KEMRI-Wellcome Trust Reseach Programme, Kenya

We thank the reviewer for these comments. The typographical errors noted have been corrected in the revised manuscript.

Competing Interests: No competing interests were disclosed.