

Revealing the extent of the first wave of the COVID-19 pandemic in Kenya based on serological and PCR-test data

John Ojal^{1,2*†}, Samuel P. C. Brand^{1,3,4*†}, Vincent Were⁵, Emelda A Okiro^{6,7}, Ivy K Kombe¹, Caroline Mburu¹, Rabia Aziza^{3,4}, Morris Ogero¹, Ambrose Agweyu¹, George M Warimwe^{1,7}, Sophie Uyoga¹, Ifedayo M O Adetifa^{1,8}, J Anthony G Scott^{1,8}, Edward Otieno¹, Lynette I Ochola-Oyier¹, Charles N Agoti^{1,9}, Kadondi Kasera¹⁰, Patrick Amoth¹⁰, Mercy Mwangangi¹⁰, Rashid Aman¹⁰, Wangari Ng'ang'a¹¹, Benjamin Tsofa¹, Philip Bejon^{1,7}, Edwine Barasa^{5,7}, Matt. J. Keeling^{3#}, and D. James. Nokes^{1,3,4#}

¹Kenya Medical Research Institute (KEMRI) -Wellcome Trust Research Programme (KWTRP), Kilifi, Kenya

²London school of Hygiene and Tropical Medicine (LSHTM), UK

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

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

⁵Health Economics Research Unit, KEMRI-Wellcome Trust Research Programme, Nairobi, Kenya.

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

⁷Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom.

⁸Department of Infectious Diseases Epidemiology, London School of Hygiene and Tropical Medicine, London, United Kingdom.

⁹School of Public Health, Pwani University, Kenya.

¹⁰Ministry of Health, Government of Kenya, Nairobi, Kenya.

¹¹Presidential Policy Strategy Unit, The Presidency, Government of Kenya.

* Correspondence to: John Ojal (JOjal@kemri-wellcome.org) and/or Samuel Brand (S.Brand@warwick.ac.uk).

† John Ojal and Samuel Brand contributed equally.

Matt Keeling and James Nokes contributed equally.

Abstract Policy makers 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, 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

Introduction

The potential risk from SARS-CoV-2 to Africa was identified early in the global pandemic [1]. 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 Africa [2]. The question is whether this is the result of lower risk due to demographic structure (young age [3], either cross-reacting immunity (e.g. pre-existing SARS-CoV-2 cross-reactive T cells [4]) or dampened immunological over-reaction [5], a low reproduction number from rapidly imposed interventions (such as school closures and lockdowns [6]), environmental conditions (e.g. temperature and humidity [7]), 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 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 Mombasa [8]. The major concerns from unmitigated spread were a limited surge capacity of the Kenyan health system [9] 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 hypertension [10]. 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 cause [11]. 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 Kenya. 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 September.

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) [12]. Seroprevalence was higher in Nairobi (7.6%) and Mombasa (8.3%). These levels of seropositivity are comparable to those reported in May in the UK [13], April/May in Spain [14], and March/April in some US cities [15], 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 in first 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) [16]. 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 compartmental mechanistic and data-driven transmission model for Kenya, which integrates three sources of longitudinal data: national time series PCR tests, the Kenyan serological survey and Google mobility behavioural data. The overall modelling approach is similar to Flaxman et al [16]; 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 [16] 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 (Fig. 1 HG). 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). 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 (Fig. 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 (Fig. 2). In Nairobi and Mombasa we predict that reduction in susceptibility of the population (Fig. 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 (Fig. 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 (Fig. 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 (Fig. 1 EF). 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 modelling [17]. 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 PCR [18] (Fig. 1 GH and AB). 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, (Fig. 1 CD). This estimated level of seropositivity is substantially higher than has been estimated in some countries that have been hit hard by the pandemic [13, 15, 14]. However, they are in broad agreement with a study in Niger state, Nigeria, from June 2020 [19], as well as seropositivity rates reported from the hard-hit city of Manaus, Brazil, in May 2020 [20]. 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 (Fig. 1 CD). 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 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 ($IFR_{verity} = 0.26\%$ [21]; 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; Fig. 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 (Fig. 3 BC). 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 2020 [22]. However, SARI admissions were lower in the early phase of the Kenyan epidemic than observed counts from the same months in 2018 and 2019 [22] 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 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. Manaus [20]. 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% at-

tack rate by the end of 2020 [20]. 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 COVID [23].

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}) [24]. The SEIR model with effective population size is an extension of the homogeneous SEIR model [25] 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 policies [26]. 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 [27, 28]. 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,

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

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

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

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 [29]). 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 ι_n . For a given set of model parameters ι_n was generated by solving the ODE system (1), giving,

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

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 $Q_{sero}(\tau)$:

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

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

$$\begin{aligned}l((P^+)_1, \dots, (P^+)_T, (S^+)_1, \dots, (S^+)_T, \theta_{OM}) = \\ \sum_{n=1}^T \ln f_{PCR}((ObsP^+)_n | (P^+)_n, \theta_{OM}) \\ + \ln f_{sero}((ObsS^+)_n | (S^+)_n, \theta_{OM}).\end{aligned}\quad (5)$$

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 observation model parameters included bias and over-dispersion parameters for finding PCR positives in the daily testing group compared to an unbiased binomial sample over the population; we used a beta-binomial count data model whenever the number of negative PCR tests for that day was available in the county and a negative binomial count data model whenever it was not. We assumed that the serological test-positives were from an unbiased sample but included over-dispersion in sampling serological positives by using a beta-binomial count data model. *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} .

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_t , we used Hamiltonian Markov-chain Monte Carlo (HMC) [30] 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 trends [31] for the whole period, we sequentially improved our β_t estimate using the expectation-maximisation (EM) algorithm [32]. The E-step corresponding to posterior distribution estimation using HMC, and the M-step corresponding to optimising the daily effective contact rate estimates $(\beta_n)_{n=41, \dots, T}$ using the popular stochastic gradient descent algorithm ADAM [33]. The first 40 days of effective contact rate estimates $(\beta_n)_{n=1, \dots, 40}$ 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} ,

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

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.

Data and materials availability: 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 a highly 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 the supplementary materials. The code base underlying the analysis is accessible at the open github repository <https://github.com/ojal/KenyaSerology>.

Author contributions

DJN, MJK, JO, SPCB, AA, GMW, JAGS, IMO, VW, EB, PB Conceptualised and/or designed the study; JO, SPCB, MJK, IK, RA, CM Formal analysis, methodology, software; SO, LIOO, CNA Investigation; BT, AA, EB, KK, PA, MM, RA, WN Administration; DJN, MJK, EAO, GMW, IMO, KA Supervision; MO, EO, LIOO, SO, CNA Data curation; SPCB, JO, DJN, MJK, PB, EB Writing - Original Draft; All contributors 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 research was funded by: Foreign, Commonwealth and Development Office and Wellcome [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.

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.

References

- [1] Marius Gilbert, Giulia Pullano MSc, Francesco Pinotti PhD, Eugenio Valdano PhD, Chiara Poletto PhD, Prof Pierre-Yves BoÅ lle PhD, Eric D'Ortenzio MD, Prof Yazdan Yazdanpanah PhD, Serge Paul Eholie MD, Mathias Altmann PharmDr, Bernardo Gutierrez MSc, Moritz U G Kraemer DPhil, and Vittoria Colizza PhD. Preparedness and vulnerability of African countries against importations of COVID-19: a modelling study. *The Lancet*, 395(10227):871–877, March 2020.
- [2] Joseph Waogodo Cabore, Humphrey Cyprian Karamagi, Hillary Kipruto, James Avoka Asamani, Benson Droti, Aminata Binetou Wahebine Seydi, Regina Titi-Ofei, Benido Impouma, Michel Yao, Zabulon Yoti, Felicitas Zawaira, Prosper Tumusiime, Ambrose Talisuna, Francis Chisaka Kasolo, and Matshidiso R Moeti. The potential effects of widespread community transmission of SARS-CoV-2 infection in the World Health Organization African Region: a predictive model. *BMJ Global Health*, 5(5):e002647–13, May 2020.
- [3] Binta Zahra Diop, Marieme Ngom, Clémence Pougé Biyong, and John N Pougé Biyong. The relatively young and rural population may limit the spread and severity of COVID-19 in Africa: a modelling study. *BMJ Global Health*, 5(5):e002699–11, May 2020.
- [4] Julian Braun, Lucie Loyal, Marco Frentsch, Daniel Wendisch, Philipp Georg, Florian Kurth, Stefan Hippenstiel, Manuela Dingeldey, Beate Kruse, Florent Fauchere, Emre Baysal, Maïke Mangold, Larissa Henze, Roland Lauster, Marcus A Mall, Kirsten Beyer, Jobst Ramp x000F6 hmel, Sebastian Voigt, Jamp x000FC rgen Schmitz, Stefan Miltenyi, Ilja Demuth, Marcel A Mamp x000FC ller, Andreas Hocke, Martin Witzernath, Norbert Suttrop, Florian Kern, Ulf Reimer, Holger Wenschuh, Christian Drosten, Victor M Corman, Claudia Giesecke-Thiel, Leif Erik Sander, and Andreas Thiel. SARS-CoV-2-reactive T cells in healthy donors and patients with COVID-19. *Nature Publishing Group*, pages 1–23, July 2020.

- [5] Moustapha Mbow, Bertrand Lell, Simon P Jochems, Badara Cisse, Souleymane Mboup, Benjamin G Dewals, Assan Jaye, Alioune Dieye, and Maria Yazdanbakhsh. COVID-19 in Africa: Dampening the storm? *Science*, 369(6504):624–626, August 2020.
- [6] Thomas Hale, Samuel Webster, Anna Petherick, Toby Phillips, and Beatriz Kira. Oxford covid-19 government response tracker. *Blavatnik School of Government*, 25, 2020.
- [7] Yueling Ma, Yadong Zhao, Jiangtao Liu, Xiaotao He, Bo Wang, Shihua Fu, Jun Yan, Jingping Niu, Ji Zhou, and Bin Luo. Effects of temperature variation and humidity on the death of COVID-19 in Wuhan, China. *Science of the Total Environment*, 724(C):138226, July 2020.
- [8] Kenyan Ministry of Health: COVID-19 situation reports.
- [9] Edwine Barasa. Assessing the Hospital Surge Capacity of the Kenyan Health System in the Face of the COVID-19 Pandemic. *medRxiv*, pages 1–24, April 2020.
- [10] Peter M Macharia, Noel K Joseph, and Emelda A Okiro. A vulnerability index for COVID-19: spatial analysis to inform equitable response in Kenya. *medRxiv*, pages 1–26, May 2020.
- [11] Kenyan Ministry of Health. Press statement on the update of the coronavirus in the country and response measure. pages 1–3, March 2020.
- [12] Sophie Uyoga, Ifedayo M O Adetifa, Henry K Karanja, James Nyagwange, James Tuju, Perpetual Wanjiku, Rashid Aman, Mercy Mwangangi, Patrick Amoth, Kadondi Kasera, Wangari Ng'ang'a, Charles Rombo, Christine Yegon, Khamisi Kithi, Elizabeth Odhiambo, Thomas Rotich, Irene Orgut, Sammy Kihara, Mark Otiende, Christian Bottomley, Zonia N Mupe, Eunice W Kagucia, Katherine E Gallagher, Anthony Etyang, Shirine Voller, John N Gitonga, Daisy Mugo, Charles N Agoti, Edward Otieno, Leonard Ndwiga, Teresa Lambe, Daniel Wright, Edwine Barasa, Benjamin Tsoga, Philip Bejon, Lynette I Ochola-Oyier, Ambrose Agweyu, J Anthony G Scott, and George M Warimwe. Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Kenyan blood donors. *Science*, pages eabe1916–11, November 2020.
- [13] H Ward, C Atchison, M Whitaker, K E C Ainslie, J Elliott, Lucy C Okell, R Redd, D Ashby, C A Donnelly, W Barclay, A Darzi, G Cooke, S Riley, and P Elliott. Antibody prevalence for SARS-CoV-2 following the peak of the pandemic in England: REACT2 study in 100,000 adults. *medRxiv*, pages 1–20, August 2020.
- [14] Marina Marina Pollan, Beatriz Perez-Gomez, Roberto Pastor-Barriuso, Jesus Oteo, Miguel A Hernan MD, Mayte Perez-Olmeda PhD, Jose L Sanmartan MEng, Aurora Fernandez-Garca PhD, Israel Cruz PhD, Nerea Fernandez de Larrea MD, Marta Molina MD, Francisco Rodrandez-Cabrera MD, Mariano Martan MEng, Paloma Merino-Amador MD, Jose Leon Paniagua PhD, Juan F Muoz-Montalvo MEng, Faustino Blanco MD, Raquel Yotti MD, Faustino Blanco, Rodrigo Gutierrez Fernandez, Mariano Martan, Saturnino Mezcua Navarro, Marta Molina, Juan F Muoz Montalvo, Matas Salinero Hernandez, Jose L Sanmartan, Manuel Cuenca-Estrella, Raquel Yotti, Jose Leon Paniagua, Nerea Fernandez de Larrea, Pablo Fernandez Navarro, Roberto Pastor-Barriuso, Beatriz Perez-Gomez, Marina Pollan, Ana Avellan, Giovanni Fedele, Aurora Fernandez-Garca, Jesus Oteo Iglesias, Mara a Teresa Perez Olmeda, Israel Cruz, Maria Elena Fernandez Martınez, Francisco D Rodrandez Cabrera, Miguel A Hernan, Susana Padrones Fernandez, Jose Manuel Rumbao Aguirre, Jose M Navarro Mara, Begoa Palop Borras, Ana Belan Perez Jimenez, Manuel Rodrandez Iglesias, Ana Mara a Calvo Gascan, Mara a Luz Lou Alcaıne, Ignacio Donate Suarez, Oscar Suarez lvarez, Mercedes Rodrandez Perez, Margarita Cases Sanchas, Carlos Javier Villafanila Gomila, Lluıs Carbo Saladrigas, Adoracian Hurtado Fernandez, Antonio Oliver, Elas Castro Feliciano, Mara a Noema Gonzalez Quintana, Jose Mara a Barrasa Fernandez, Mara a Araceli Hernandez Betancor, Melisa Hernandez Febles, Leopoldo Martan Martan, Luis-Mariano Lopez Lopez, Teresa Ugarte Miota, Inas De Benito Poblacian, Mara a Sagrario Celada Perez, Mara a Natalia Vallas Fernandez, Tomas Matas Enraquez, Miguel Villa Arranz, Marta Domınguez-Gil Gonzalez, Isabel Fernandez Natal, Gregoria Megas Loban, Juan Luis Muoz Bellido, Pilar Ciruela, Ariadna Mas i Casals, Maria Dolada Botas, M Angeles Marcos Maeso, Dania Perez del Campo, Antonio Felix de Castro, Raman Liman Ramerez, Maria Francisca Elas Retamosa, Manuela Rubio Gonzalez, Mara a Sinda Blanco Lobeiras, Alberto Fuentes Losada, Antonio Aguilera, German Bou, Yolanda Caro, Noema Marauri, Luis Miguel Soria Blanco, Isabel del Cura Gonzalez, Montserrat Hernandez Pascual, Roberto Alonso Fernandez, Paloma Merino-Amador, Natalia Cabrera Castro, Aurora Tomas Lizcano, Cristobal Ramerez Almagro, Manuel Segovia Hernandez, Nieves Asuncion Elizaga, Mara a Ederra Sanz, Carmen Ezpeleta Baquedano, Ana Bustinduy Bascaran, Susana Iglesias Tamayo, Luis Elorduy Otazua, Rebeca Benarroch Benarroch, Jesus Lopera Flores, Antonia Vazquez de la Villa, and ENE-COVID Study Group. Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. *The Lancet*, 396(10250):535–544, August 2020.
- [15] Fiona P Havers, Carrie Reed, Travis Lim, Joel M Montgomery, John D Klena, Aron J Hall, Alicia M Fry, Deborah L Cannon, Cheng-Feng Chiang, Aridh Gibbons, Inna Krapinayaya, Maria Morales-Betoulle, Katherine Roguski, Mohammad Ata Ur Rasheed, Brandi Freeman, Sandra Lester, Lisa Mills, Darin S Carroll, S Michele Owen, Jeffrey A Johnson, Vera Semenova, Carina Blackmore, Debra Blog, Shua J Chai, Angela Dunn, Julie Hand, Seema Jain, Scott Lindquist, Ruth Lynfield, Scott Pritchard, Theresa Sokol, Lynn Sosa, George Turabelidze, Sharon M Watkins, John Wiesman, Randall W Williams, Stephanie Yendell, Jarad Schiffer, and Natalie J Thornburg. Seroprevalence of Antibodies to SARS-CoV-2 in 10 Sites in the United States, March 23-May 12, 2020. *JAMA Internal Medicine*, pages 1–11, July 2020.
- [16] Seth Flaxman, Swapnil Mishra, Axel Gandy, H Juliette T Unwin, Thomas A Mellan, Helen Coupland, Charles Whitaker, Harrison Zhu, Tresnia Berah, Jeffrey W Eaton, Melodie Monod, Pablo N Perez-Guzman, Nora Schmit, Lucia Cilloni, Kylie E C Ainslie, Marc Baguelin, Adhiratha Boonyasiri, Olivia Boyd, Lorenzo Cattarino, Laura V Cooper, Zulma Cucunuba, Gina Cuomo-Dannenburg, Amy Dighe, Bimandra Djaafara, Ilaria Dorigatti, Sabine L van

- Elsland, Richard G FitzJohn, Katy A M Gaythorpe, Lily Geidelberg, Nicholas C Grassly, William D Green, Timothy Hallett, Arran Hamlet, Wes Hinsley, Ben Jeffrey, Edward Knock, Daniel J Laydon, Gemma Nedjati-Gilani, Pierre Nouvellet, Kris V Parag, Igor Siveroni, Hayley A Thompson, Robert Verity, Erik Volz, Caroline E Walters, Haowei Wang, Yuanrong Wang, Oliver J Watson, Peter Winskill, Xiaoyue Xi, Patrick G T Walker, Azra C Ghani, Christl A Donnelly, Steven Riley, Michaela A C Vollmer, Neil M Ferguson, Lucy C Okell, and Samir Bhatt. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature Publishing Group*, 584(7820):1–15, August 2020.
- [17] Andrew Gelman, Jessica Hwang, and Aki Vehtari. Understanding predictive information criteria for Bayesian models. *Stat Comput*, 24(6):997–1016, August 2013.
- [18] Fei Zhou, Ting Yu, Ronghui Du, Guohui Fan, Ying Liu MD, Zhibo Liu MD, Jie Xiang MS, Yeming Wang MD, Bin Song MS, Xiaoying Gu PhD, Lulu Guan MD, Yuan Wei MS, Hui Li MD, Xudong Wu MS, Jiuyang Xu MD, Shengjin Tu MD, Yi Zhang MD, Hua Chen, and Bin Cao. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet*, 395(10229):1054–1062, March 2020.
- [19] Hussaini Majiya, Mohammed Aliyu-Paiko, Vincent Tochukwu Balogu, Dickson Achimugu Musa, Ibrahim Maikudi Salihu, Abdullahi Abubakar Kawu, Ishaq Yakubu Bashir, Aishat Rabiuni Sani, John Baba, Amina Tako Muhammad, Fatima Ladidi Jibril, Ezekiel Bala, Nuhu George Obaje, Yahaya Badeggi Aliyu, Ramatu Gogo Muhammad, Hadiza Mohammed, Usman Naji Gimba, Abduljaleel Uthman, Hadiza Muhammad Liman, Sule Alfa Alhaji, Joseph Kolo James, Muhammad Muhammad Makusidi, Mohammed Danasabe Isah, Ibrahim Abdullahi, Umar Ndagi, Bala Waziri, Chindo Ibrahim Bisallah, Naomi John Dadi-Mamud, Kolo Ibrahim, and Abu Kasim Adamu. Seroprevalence of COVID-19 in Niger State. *medRxiv*, pages 1–24, August 2020.
- [20] Lewis F Buss, L F Buss, Carlos A Prete, C A Prete, CMM Abraham, Claudia M M Abraham, A Mendrone, Alfredo Mendrone, M C Fitzpatrick, Tassila Salomon, Cesar de Almeida-Neto, Rafael F O França, Maria C Belotti, Maria P S S Carvalho, Allyson G Costa, Myuki A E Crispim, Suzete C Ferreira, Nelson A Fraiji, Susie Gurzenda, Charles Whittaker, Leonardo T Kamaura, Pedro L Takecian, Pedro da Silva Peixoto, Marcio K Oikawa, Anna S Nishiya, Vanderson Rocha, Nanci A Salles, Andreza Aruska de Souza Santos, Martirene A da Silva, Brian Custer, Kris V Parag, Manoel Barral-Netto, Moritz U G Kraemer, Rafael H M Pereira, Oliver G Pybus, Michael P Busch, Márcia C Castro, Christopher Dye, Vitor H Nascimento, Nuno R Faria, and Ester C Sabino. Three-quarters attack rate of SARS-CoV-2 in the Brazilian Amazon during a largely unmitigated epidemic. *science.sciencemag.org*, 371(6526):288–292, January 2021.
- [21] Robert Verity, Lucy C Okell, Iliaria Dorigatti, Peter Winskill, Charles Whittaker, Natsuko Imai, Gina Cuomo-Dannenburg, Hayley Thompson, Patrick G T Walker, Han Fu PhD, Amy Dighe MRes, Jamie T Griffin PhD, Marc Baguelin PhD, Sangeeta Bhatia PhD, Adhiratha Boonyasiri MD, Anne Cori PhD, Zulma Cucunubá PhD, Rich FitzJohn PhD, Katy Gaythorpe PhD, Will Green MSc, Arran Hamlet MSc, Wes Hinsley PhD, Daniel Laydon PhD, Gemma Nedjati-Gilani PhD, Prof Steven Riley DPhil, Sabine van Elsland PhD, Erik Volz PhD, Haowei Wang MSc, Yuanrong Wang, Xiaoyue Xi MSc, Christl A Donnelly, Azra C Ghani, and Neil M Ferguson. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet. Infectious diseases*, 20(6):1–9, March 2020.
- [22] KEMRI-Wellcome Trust Research Programme. Status of the COVID-19 Pandemic in Kenya: Evidence from serological and clinical surveillance, and predictive modelling. Technical report, August 2020.
- [23] Gerald Ong’ayo, Michael Ooko, Ruth Wang’ondu, Christian Bottomley, Amek Nyaguara, Benjamin K Tsofa, Thomas N Williams, Philip Bejon, Anthony Scott, and Anthony O Etyang. Articles Effect of strikes by health workers on mortality between 2010 and 2016 in Kilifi, Kenya: a population-based cohort analysis. *The Lancet Global Health*, 7(7):e961–e967, May 2019.
- [24] Michael Li, Jonathan Dushoff, and Benjamin M Bolker. Fitting mechanistic epidemic models to data: A comparison of simple Markov chain Monte Carlo approaches. *Statistical Methods in Medical Research*, 27(7):1956–1967, November 2016.
- [25] Matt J Keeling and Pejman Rohani. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, 2008.
- [26] Roy M Anderson, Robert M May, and B Anderson. *Infectious Diseases of Humans: Dynamics and Control*. dynamics and control. Oxford University Press, Oxford, March 1991.
- [27] Ricardo Aguas, Rodrigo M Corder, Jessica G King, Guilherme Goncalves, Marcelo U Ferreira, and M Gabriela M Gomes. Herd immunity thresholds for SARS-CoV-2 estimated from unfolding epidemics. *medRxiv*, pages 1–42, July 2020.
- [28] Alexei V Tkachenko, Sergei Maslov, Ahmed Elbanna, George Wong, Zachary Weiner, and Nigel Goldenfeld. Persistent heterogeneity not short-term overdispersion determines herd immunity to COVID-19. *medRxiv*, pages 1–10, July 2020.
- [29] Luca Ferretti, Chris Wymant, Michelle Kendall, Lele Zhao, Anel Nurtay, Lucie Abeler-Dörner, Michael Parker, David Bonsall, and Christophe Fraser. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science*, 368(6491):eabb6936–9, May 2020.
- [30] Michael Betancourt. A Conceptual Introduction to Hamiltonian Monte Carlo. *arXiv.org*, January 2017.
- [31] Google LLC. *Google COVID-19 Community Mobility Reports*.
- [32] A P Dempster, N M Laird, and Rubin D B. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B.*, pages 1–38, January 1977.
- [33] D P Kingma, J Ba URL <http://arxiv.org/abs/1412.6980>, and 2014. Adam: A method for stochastic optimization. CoRR [abs/1412.6980](https://arxiv.org/abs/1412.6980).

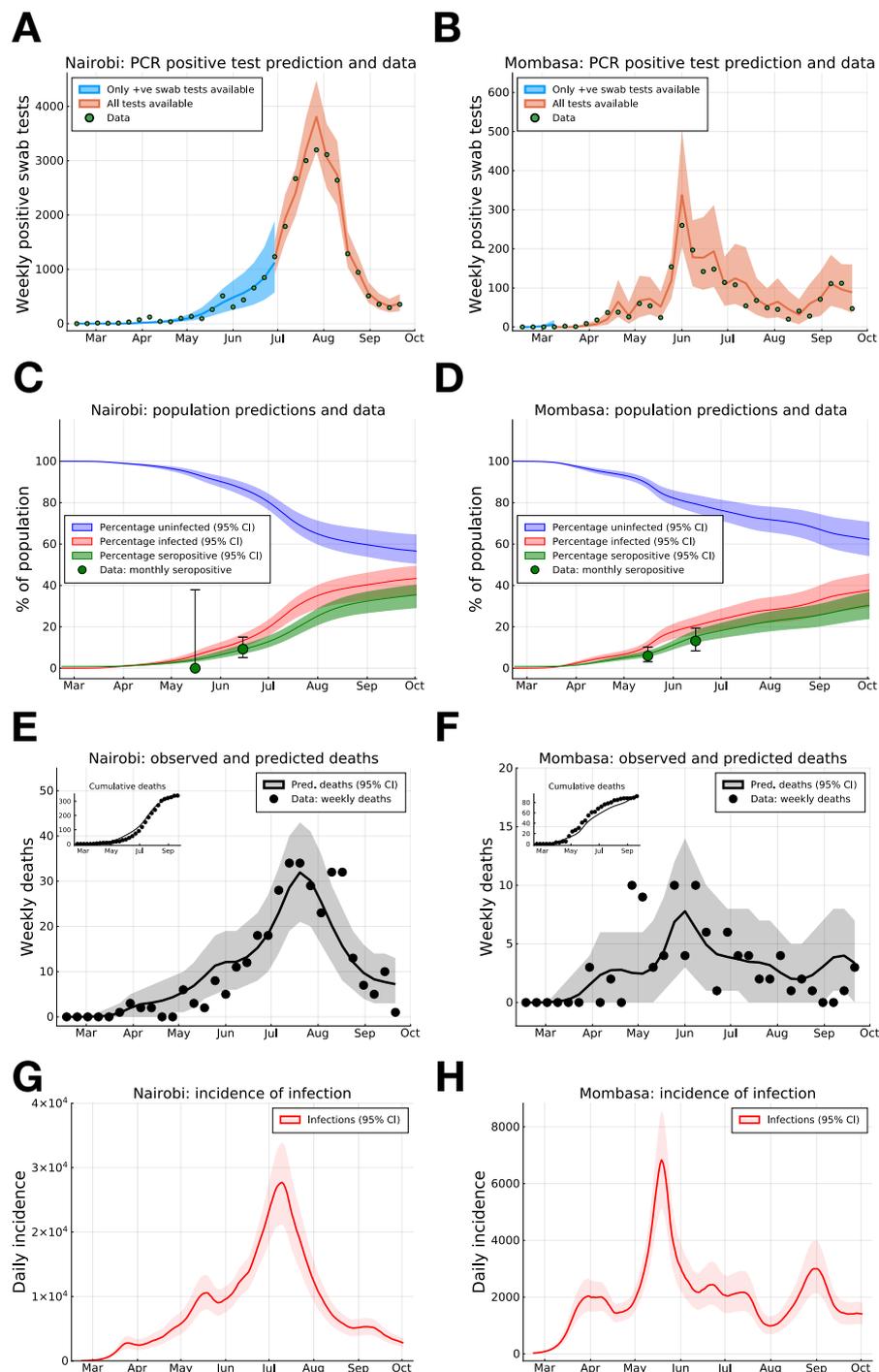


Figure 1. SARS-CoV-2 PCR positive swab tests, seroprevalence and deaths in Nairobi and Mombasa, Kenya, with model forecasting. (A) (B) Weekly reported positive PCR positive swab tests (green dots) for Nairobi (A) and Mombasa (B), model prediction of mean weekly detection during both sampling periods when negative PCR test data was unavailable (blue curve), and available (orange curve). (C) (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) (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) (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.

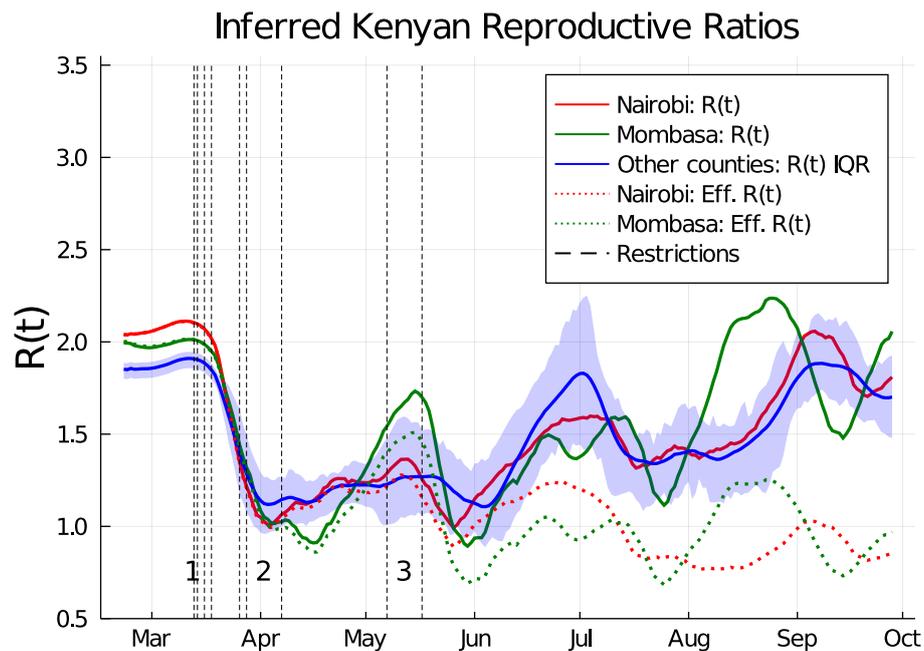


Figure 2. Estimated basic and effective reproductive numbers in Kenya since Feb 21st 2020. The posterior mean reproductive number for Nairobi (red curves), Mombasa (green curves), and the inter-quartile range (IQR) over mean reproductive number estimates for all other Kenyan counties (blue curve and 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.

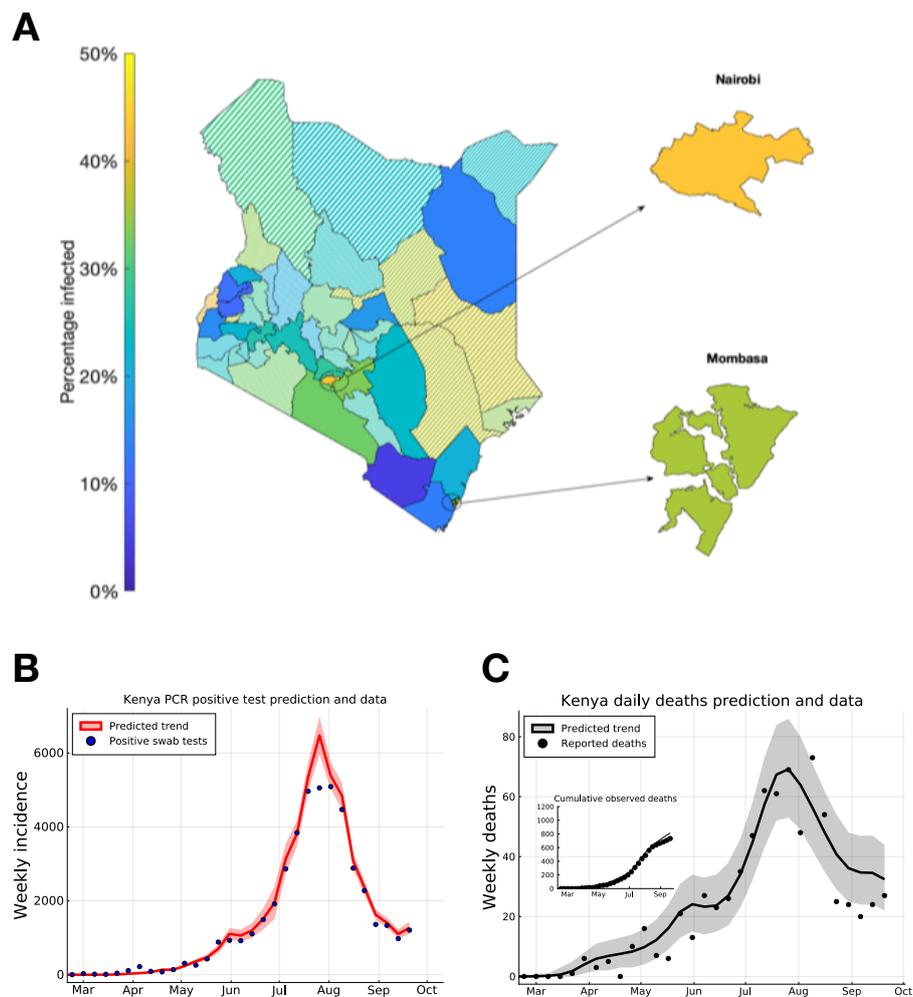


Figure 3. Predicting peak timing of transmission rate by Kenyan county, and forecasting of Kenya-wide PCR positive swab tests and reported deaths. (A) Posterior mean estimates for the attack rate (% of population) in each county. Solid shaded counties have a posterior standard deviation in their attack 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) (C) mark 1st of the month.