Check for updates

RESEARCH ARTICLE

Using contact data to model the impact of contact tracing and physical distancing to control the SARS-CoV-2 outbreak in Kenya [version 1; peer review: awaiting peer review]

Moritz Wagner [iD] [1,2], Ivy K. Kombe[3], Moses Chapa Kiti[3], Rabia Aziza [iD] [4], Edwine Barasa[3], D. James Nokes [iD] [3,4]

[1]Centre for the Mathematical Modelling of Infectious Diseases, Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, UK
[2]University College London, London, UK
[3]KEMRI-Wellcome Trust Research Programme, Kilifi, Kenya
[4]School of Life Sciences and the Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research (SBIDER), University of Warwick, Coventry, UK

Open Peer Review

Reviewer Status *AWAITING PEER REVIEW*

Any reports and responses or comments on the article can be found at the end of the article.

Abstract

Background: Across the African continent, other than South Africa, COVID-19 cases have remained relatively low. Nevertheless, in Kenya, despite early implementation of containment measures and restrictions, cases have consistently been increasing. Contact tracing forms one of the key strategies in Kenya, but may become infeasible as the caseload grows. Here we explore different contact tracing strategies by distinguishing between household and non-household contacts and how these may be combined with other non-pharmaceutical interventions.

Methods: We extend a previously developed branching process model for contact tracing to include realistic contact data from Kenya. Using the contact data, we generate a synthetic population of individuals and their contacts categorised by age and household membership. We simulate the initial spread of SARS-CoV-2 through this population and look at the effectiveness of a number of non-pharmaceutical interventions with a particular focus on different contact tracing strategies and the potential effort involved in these.

Results: General physical distancing and avoiding large group gatherings combined with contact tracing, where all contacts are isolated immediately, can be effective in slowing down the outbreak, but were, under our base assumptions, not enough to control it without implementing extreme stay at home policies. Under optimistic assumptions with a highly overdispersed R0 and a short delay from symptom onset to isolation, control was possible with less stringent physical distancing and by isolating household contacts only.

Conclusions: Without strong physical distancing measures, controlling the spread of SARS-CoV-2 is difficult. With limited

resources, physical distancing combined with the isolation of households of detected cases can form a moderately effective strategy, and control is possible under optimistic assumptions. More data are needed to understand transmission in Kenya, in particular by studying the settings that lead to larger transmission events, which may allow for more targeted responses, and collection of representative age-related contact data.

Keywords

contact tracing, SARS-CoV-2, mathematical modelling, infectious disease dynamics, public health, Kenya, branching process, social contact data

This article is included in the KEMRI | Wellcome Trust gateway.
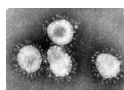
This article is included in the Coronavirus (COVID-19) collection.

Corresponding author: Moritz Wagner (moritz.wagner@lshtm.ac.uk)

Introduction

Contact tracing forms one of the essential public health tools for tackling outbreaks of directly transmitted pathogens. Its effectiveness is governed by the ability to identify infected contacts rapidly, before they can continue infection spread, the feasibility of which becomes progressively more difficult as cases increase. Hence, contact tracing, particularly in resource-scarce settings, is most useful early in the epidemic evolution or when case numbers are low. Across the African continent, the rate of spread of COVID-19 has been relatively slow. In Kenya, following the first reported case on March 12th, cumulative numbers remained below 1000 until May 20th. As of the end of July, total cases have exceeded 20,000 and are doubling roughly every 18 days[1]. The majority of new cases are occurring in Nairobi, overstretching testing capacity. In most other counties, incident cases remain low, but this is likely to be a temporary situation and testing and health services could soon be overstretched and insight is needed on improving strategy efficiency[2].

The effectiveness of contacting close contacts of an infected case depends heavily on the natural history of infection, in particular the proportion of pre-symptomatic and asymptomatic transmission occurring, as has been shown previously[3–8]. For SARS-CoV-2, pre-symptomatic transmission seems to play a significant role in the transmission process[9–11]. In addition, asymptomatic cases[12,13] have been reported, although the overall contribution of these to transmission is still unclear. A further limitation of contact tracing is the amount of resources involved in tracing close contacts. Thus, informed decisions must be made of when and how to best implement or refine contact tracing. Such decisions can be guided by models that simulate transmission across contact networks.

This study extends a previously developed stochastic transmission model[3,4] of contact tracing to include the use of diary-based contact data[12] to form a context-specific picture of the effect of contact tracing. Differences in social structure between developed and developing countries could translate into marked differences in mixing patterns that could have an impact on how much effort needs to go into contact tracing. We utilise data from a diary-based contact survey we undertook within a coastal Kenyan community[14]. Although conducted 10 years ago, such studies are uncommon in developing countries, and we assume that qualitative aspects of mixing patterns have not changed greatly over time. We have information on age-grouped contacts in a single day and the usual frequency of each contact for over 500 study participants in rural and semi-urban settings. We model the effectiveness of non-pharmaceutical interventions (NPI) compared to different strategies for contact tracing, including the isolation of household contacts, under different transmission scenarios. As an example, we choose the setting of the original contact survey, that of Kilifi County, specifically the population of the Kilifi Health and Demographic Surveillance System. This study aims to inform health sector decision-makers (national and county) on how they may be able to effectively implement contact tracing strategies and at which point resources should be focused on other intervention strategies.

Methods

Contacts

Diary-based contact dataset. Data on daily contact numbers were available from a survey conducted in the northern part of the Kilifi Health and Demographic Surveillance System (KHDSS) on the coast of Kenya[14]. The average household size was 9.2 for rural and 7.0 for urban participants, and ~20% of the residents were aged <5 years. Note that 'household' was defined as members of one or more building units that use a communal cooking facility. A random sample of 568 diaries, stratified by location (rural to semi-urban township), age class (<1, 1–5, 6–15, 16–19, 20–49, and >50 years) and season (rainy and dry), was collected between August 2011 and January 2012. A record was kept over a single day (a randomly selected day of the week) from first waking to going to bed, recording all contacts of a physical nature (including a handshake, hug or kiss). Individuals above 10 years of age recorded their own diary, while the contacts of those aged 10 years and under were recorded by an elder sibling or adult. For each person contacted, a record was made of their age class, the frequency of contacting this individual (mostly daily, once or twice a week, once or twice a month, or less than once a month, and never before), the tally of contacts for that day, and whether a member of the same household. For the 10,042 contacts recorded, the mean contact rate was 17.7 per day, highest in primary school children aged 6-15 years (20.1) and lowest for infants and the elderly (13.9), and higher in the rural compared to the semi-urban setting (18.8 versus 15.6). Raw data and all materials used in this diary study are open access and can be downloaded from the Journal site[14].

Sampling from the contact data. For each outbreak simulation we create a synthetic population of participants and their unique contacts for a given time period by sampling from the original contact data[14]. We use information on the frequency of a contact to determine the probability of contacts being repeated in a given time period and their relative likelihood of becoming infected. As this study included both semi-urban and rural settings, we also allow for sampling with a given urban-rural divide. Thus we can form populations of different sizes and settings, assuming that the contacts are representative of semi-urban/rural contacts across Kenya. The final output is a full contact data set with participant and contact IDs. This defines the susceptible population during the outbreak and for each ID (contact and participant), we keep track of who becomes infected (see S8. of the *Extended data*, Supplementary Appendix[15] for details).

Infection step

Sampling transmission and isolation parameters. The infection and isolation steps are taken directly from a previous stochastic transmission model based on a branching process model, where for each infected individual secondary cases are drawn randomly based on a predefined distribution[3,4]. In the following, we distinguish between infectors: individuals that infect during the infection step, and infecteds: individuals that become infected during the infection step. Parameter distributions are shown in Table 1 and Figure 1.

Table 1. Parameters and distributions used. Base parameters are in bold and parameters used for sensitivity analyses in italics.

Sampled Parameters	Distribution	Median (95% CI) Based on 1000 samples	Source
R0	Distribution: Negative binomial Mean: *2*, **2.5** Overdispersion: 0.58, *0.16*	Base: **2.5 (0–13)*** High dispersion: 2.5 (0–21)* *High R0: 2 (0–10)**	16,17
Incubation period	Distribution: Weibull Shape: **2.32** Scale: **6.49**	**10.54 (6.40-16.45)**	18,19
Generation time	Distribution: Skewed normal Mean: Defined by the sampled incubation period Skew: Defined by proportion of presymptomatic transmission	For incubation period of 5 days: **5.89 (2.57–9.43)**	3,10
Delay from onset to isolation	Distribution: Weibull Medium Shape: **1.651524** Scale: **4.287786** *Short Shape: 1.5 Scale: 1.5*	Medium: **3.43 (0.45-9.54)** *Short: 1.35 (0.12-3.49)*	20
Fixed Parameters	***Value***		
Number of initial cases	**5**		Assumed
Relative susceptibility by age	"<1": **0.49** "1-5": **0.49** "6-14": **0.48** "15-19": **0.47** "20-49": **1.00** ">=50": **1.00**		21
Proportion of infections symptomatic by age	"<1": **0.29** "1-5": **0.29** "6-14": **0.25** "15-19": **0.21** "20-49": **0.33** ">=50": **0.60**		21
Proportion pre-symptomatic transmission	*Low: 0.15* Medium: **0.3** *High: 0.45*		9,18,21
Probability of tracing or quarantining a contact	HH: **0, 1** non-HH: **0, 0.25, 0.5, 0.75, 1**		Assumed

*These represent the mean and 95% CI values

For each infector, we sample an incubation period and a delay to isolation. These determine, when the infector shows symptoms and, following this, becomes isolated. Isolation in this case does not represent isolation based on contact tracing, but rather isolation based on the individual self-isolating or seeking health care and thus being isolated. In addition, we sample the basic reproduction number *R0* to determine how many new infecteds the infector produces. Based on the infector's incubation period, we sample a generation time for each infected, which we define as the time from the infector's exposure to the infected's
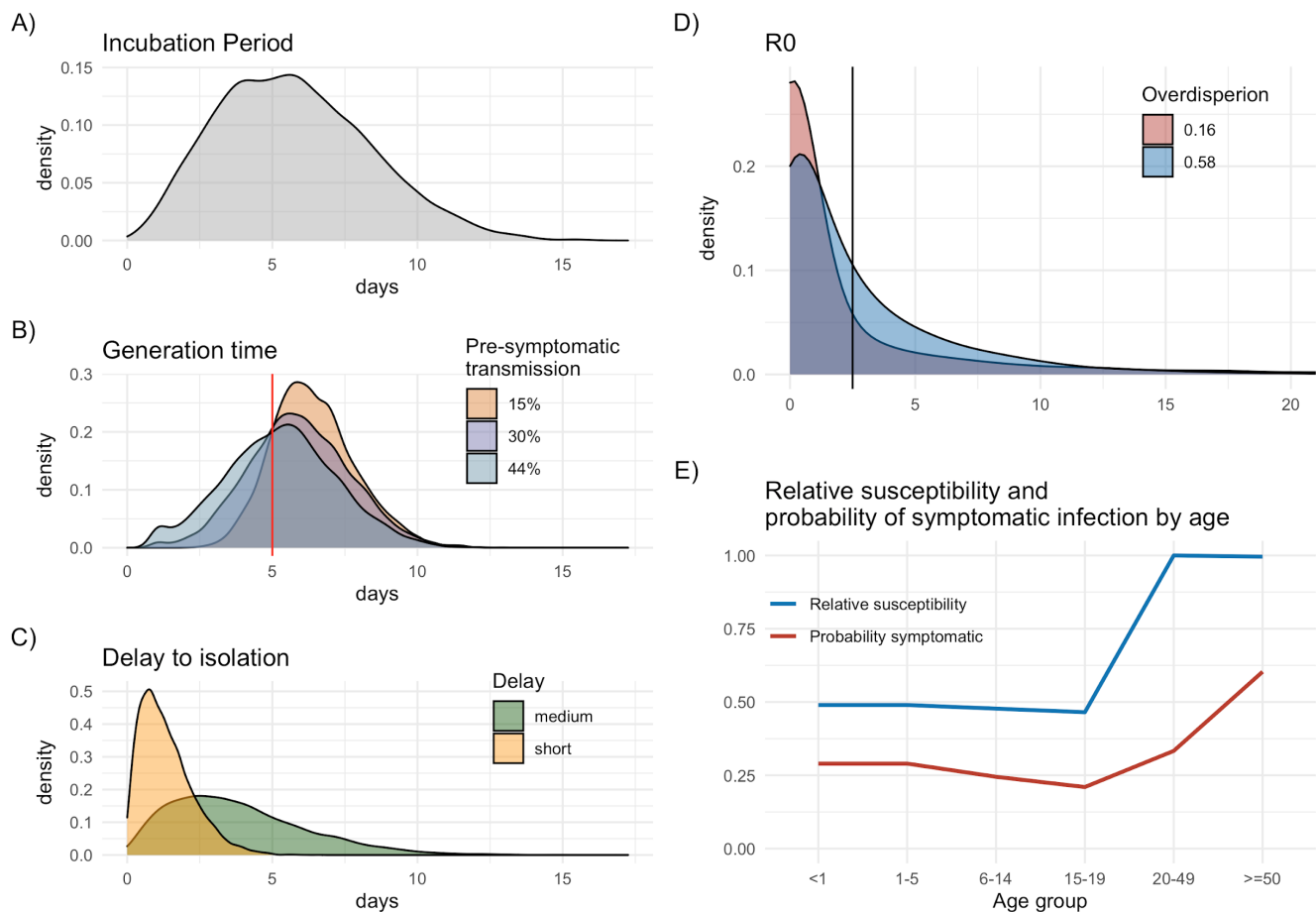
Figure 1. Parameters used. (**A**) Distribution of incubation periods. (**B**) Distributions of generation times based on sampling an incubation period of 5 days (red line) for different assumptions of pre-symptomatic transmission. (**C**) Distributions of the delay to isolation following symptom onset. (**D**) Distributions for the basic reproduction number for different dispersion parameters. (**E**) Relative susceptibility and the probability of developing symptoms that warrant isolation by age taken from Davies *et al.*[21].

exposure. Here a predefined proportion of generation times are smaller than the incubation period to model pre-symptomatic transmission events[3–8]. If a generation time is chosen such that infection occurs after the infector's isolation time, the infection does not occur. In addition, we allow for a proportion of cases, which we define as asymptomatic and for which isolation never occurs. These are based on a previous modelling study and can represent paucisymptomatic cases that are too mild to warrant self- or health care based-isolation, or completely asymptomatic cases[21]. For the intervention scenarios, each infected case also has a certain probability of being missed by tracing, in which case they will continue transmitting until they are isolated with a delay following symptom onset (or not isolated if asymptomatic).

Thus, for each infection step we determine, which individuals become infected and/or isolated and/or traced. Once a contact becomes isolated, their *R0* reduces to 0 and they can no longer infect other individuals. Different scenarios are illustrated in detail in Figure S8 in the Supplementary Appendix of the LSHTM study[3].

Extending the model to include realistic contact data. We extend the LSHTM model to include realistic contact data stratified by age and household. Given an infector with a number of potential infecteds (based on a sample from *R0*), infections are matched to the infector's contacts. Contacts are infected based on their frequency of contact as well as their relative susceptibility given by age using estimates from a previous modelling study[17]. Given the use of a highly overdispersed *R0*, there will be random draws of *R0* that are larger than the total contacts available to an infector. In this case we assume the occurrence of a super-spreading event (SSE), where the available contacts are not captured by the contact data, e.g. larger group gatherings (LGG). We extend the number of contacts available to match R0 by sampling further contacts from participants of the same age. In addition to assigning infecteds based on the infector's contacts, we keep track of the number of HH and non-HH contacts each traced/quarantined individual has. This allows us to count the number of contacts that need to be traced. If an infected individual was infected through the HH, we set their HH contacts to be traced to zero to avoid double counting, as these will have been traced already through the infector. We

ignore double counting of non-HH contacts, which assumes that non-HH contacts are not shared amongst individuals (see S9. of the *Extended data*, Supplementary Appendix[15] for more details).

Interventions

We consider five different intervention types. *Isolation:* Symptomatic individuals are isolated or isolate themselves after a delay following onset of symptoms and are no longer able to infect their contacts. Asymptomatic individuals are never isolated; *Ban LGGs:* SSEs are not allowed to occur; *Physical distancing:* Non-HH contacts are limited to an absolute maximum number of unique contacts within a given time period; *School closures:* Physical distancing applied to children only; *Contact tracing:* Contact tracing is implemented in two forms: a) *Tracing:* Symptomatic individuals are isolated or isolate themselves after a delay following onset of symptoms and are no longer able to infect their contacts. Once isolated a set proportion of their contacts are traced. Any successfully traced contacts that are infected become isolated immediately when they develop symptoms. Asymptomatic contacts are missed. b) *Quarantine:* Symptomatic individuals are isolated or isolate themselves after a delay following onset of symptoms and are no longer able to infect their contacts. Once isolated a set proportion of their contacts are traced. All traced contacts that are infected become quarantined immediately when the infector becomes isolated regardless of symptom status. Note that for simplicity in the model we do not quarantine contacts that are traced, but did not get infected.

Using these interventions, we explore six different transmission scenarios detailed in Table 2. with more detailed descriptions in Table 3 of the *Extended data*, Supplementary Appendix[15].

The two layers of tracing and quarantine effectively reduce the delay from exposure to isolation. Tracing reduces the delay from onset to isolation, while quarantine can reduce some of the pre-symptomatic transmission events, as individuals may

be isolated before symptom onset, as well as asymptomatic transmission. For the tracing and quarantine scenarios, we distinguish between HH and non-HH contacts and set the probability of tracing a HH contact to one while we alter the probability of tracing a non-HH contact.

Results

We sample 200,000 participants from the contact data and after accumulating contacts over a 7 day period, this results in a contact dataset of approximately 300,000 individuals. This constitutes our susceptible population, and approximates to the current size of the Kilifi HDSS. For each intervention scenario we run 100 simulations of an outbreak seeded with five initial infectors for 8 weeks.

We compared a number of different base intervention strategies. Figure 2 shows the results for the weekly number of cumulative cases (Figure 2A), the weekly number of HH and non-HH contacts that are isolated through tracing/quarantine (Figure 2B, C), the effective reproduction number (Figure 2D), and the proportion of outbreaks that go extinct or have less than 1000 cumulative cases in the 8 weeks (Figure 2E, F) for each intervention scenario. The boxplots represent the median, and the range in which 50% and 90% of simulations lie.

The "No interventions" scenario results in a median of 6271 (281-25883) cumulative cases in 8 weeks. Allowing for the isolation of symptomatic cases, and tracing of all HH and 25% non-HH contacts, provide only limited reductions in overall case numbers with 4740 (174-18,150), and 4596 (167-15,153) cumulative cases in 8 weeks, respectively. School closures, physical distancing measures, and quarantining HH members and 25% non-HH members of any traced case have the potential to significantly reduce the number of infections, with 1594 (0-7413), 1551 (58-5518), and 732 (0-5932) cumulative cases in 8 weeks, respectively. Tracing HH and 25% of non-HH contacts may involve having to trace and isolate 4151 (155-23,685)

Table 2. Intervention scenarios modelled. More detailed descriptions in the Extended data, Supplementary Appendix[15].

Scenario	Isolation	LGGs allowed?	Physical distancing	Contact tracing
No intervention	No	Yes	No	No
Isolation of all symptomatic cases	Yes	No	No	No
School closures	Yes	No	HH contacts only for children	No
Physical distancing	Yes	No	5 unique non-HH contacts per week across all ages*	No
Tracing	Yes	No	No	Tracing HH + 25% non-HH
Quarantine	Yes	No	No	Quarantine HH + 25% non-HH

* The figure of 5 unique non-HH contacts per week is an assumption.
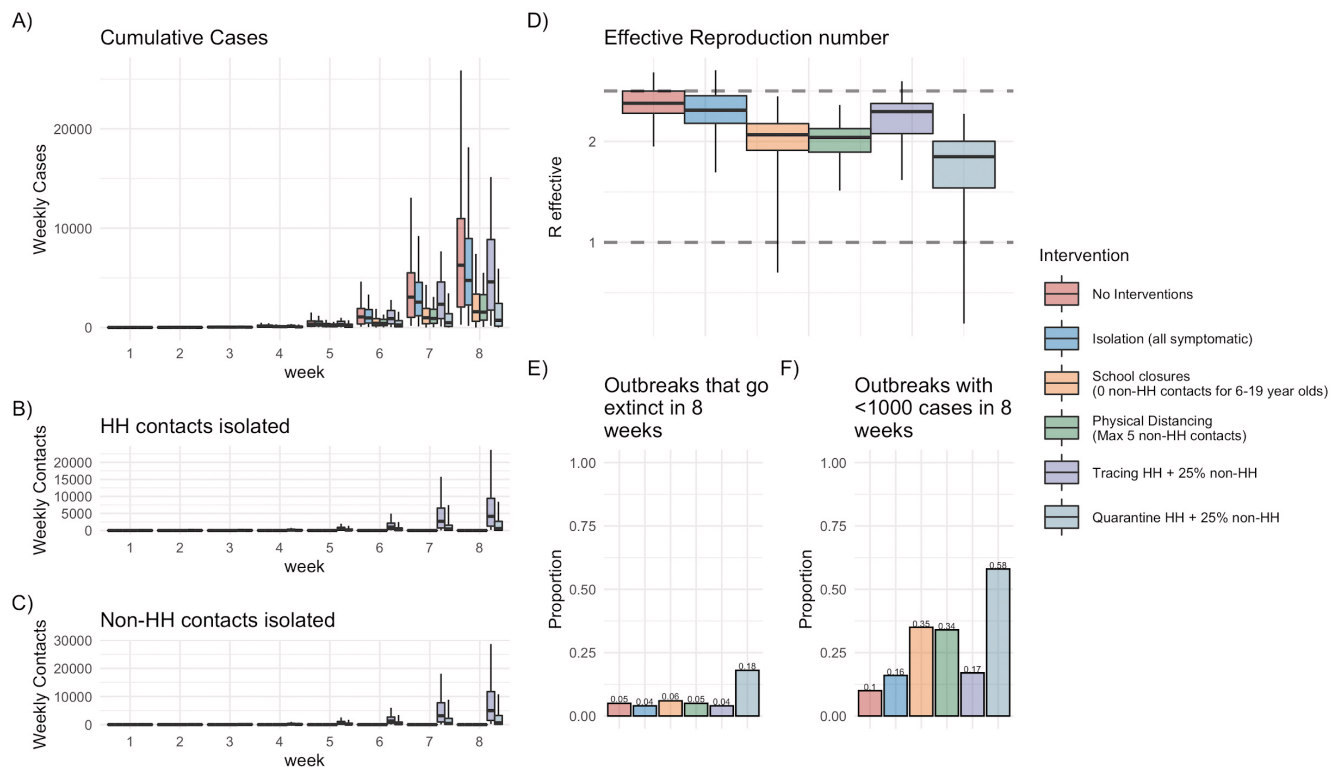
Figure 2. Base scenarios. (**A**) Boxplots of cumulative cases for each week. (**B**) HH contacts that need to be isolated in a given week based on tracing or quarantine. (**C**) Non-HH contacts that need to be isolated in a given week based on tracing or quarantine. (**D**) Boxplots of effective reproduction number estimates for each scenario. (**E**) Proportion of outbreaks that die out for each scenario. (**F**) Proportion of outbreaks that remain below 1000 cumulative cases for each scenario.

HH and 4969 (205-11734) non-HH contacts in week 8 alone, while quarantining HH and 25% of non-HH contacts is estimated to result in 576 (0-8459) HH and 660 (0-10799) non-HH contacts having to be traced and isolated in week 8.

The effective reproduction number boxplots in Figure 2D represent the median, 50%, and 90% prediction intervals of the mean effective reproduction numbers for each simulation. For all scenarios, the majority of simulations yield effective reproduction numbers above 1 (Figure 2D, E). While school closures and physical distancing may reduce the effective reproduction number and thus the size of the initial outbreak, only quarantine increases the likelihood of extinction from approximately 5% to 18% and results in over half of the outbreaks remaining below 1000 cumulative cases in 8 weeks (Figure 2D, E).

Figure 3 illustrates scenarios where levels of quarantine are combined with differing levels of physical distancing showing the cumulative number of cases in week 8, the number of HH and non-HH contacts to be traced in week 8, the effective reproduction number, and the proportion of outbreaks that go extinct or remain below 1000 cumulative cases in 8 weeks. Note that all scenarios include quarantine of HH members. In addition, we do not allow for SSEs to occur for any scenario.

Allowing either for no or a maximum number of one non-HH contacts per week significantly reduces the number of cases regardless of how many non-HH contacts are traced. Cumulative cases in week 8 range from 21 (8–56) cases with maximum physical distancing and tracing of all contacts to 136 (11-438) cases when allowing for one weekly non-HH contact and tracing of all HH contacts only (Figure 3A). Additionally, these require little to no effort in terms of tracing such that for the majority of scenarios 0 HH and non-HH contacts need to be isolated in week 8 with a maximum of 1 (0-404) non-HH contacts having to be traced and isolated if one non-HH contact is allowed and 50% of non-HH contacts are traced (Figure 3B, C). Only extreme physical distancing results in the median of effective reproductions number below 1 and at least 80% of outbreaks going extinct (Figure 3D, E).

Less stringent physical distancing measures range from 88 (7-1455) cumulative cases in 8 weeks with physical distancing of a maximum of 5 non-HH contacts and tracing of all non-HH contacts; to 1732 (9-10510) cumulative cases in 8 weeks with no physical distancing and tracing of only HH contacts. In general, the more aggressive tracing of non-HH contacts is, the less HH contacts need to be traced and isolated as there are less cases overall (Figure 3B). For non-HH contacts, a more complex
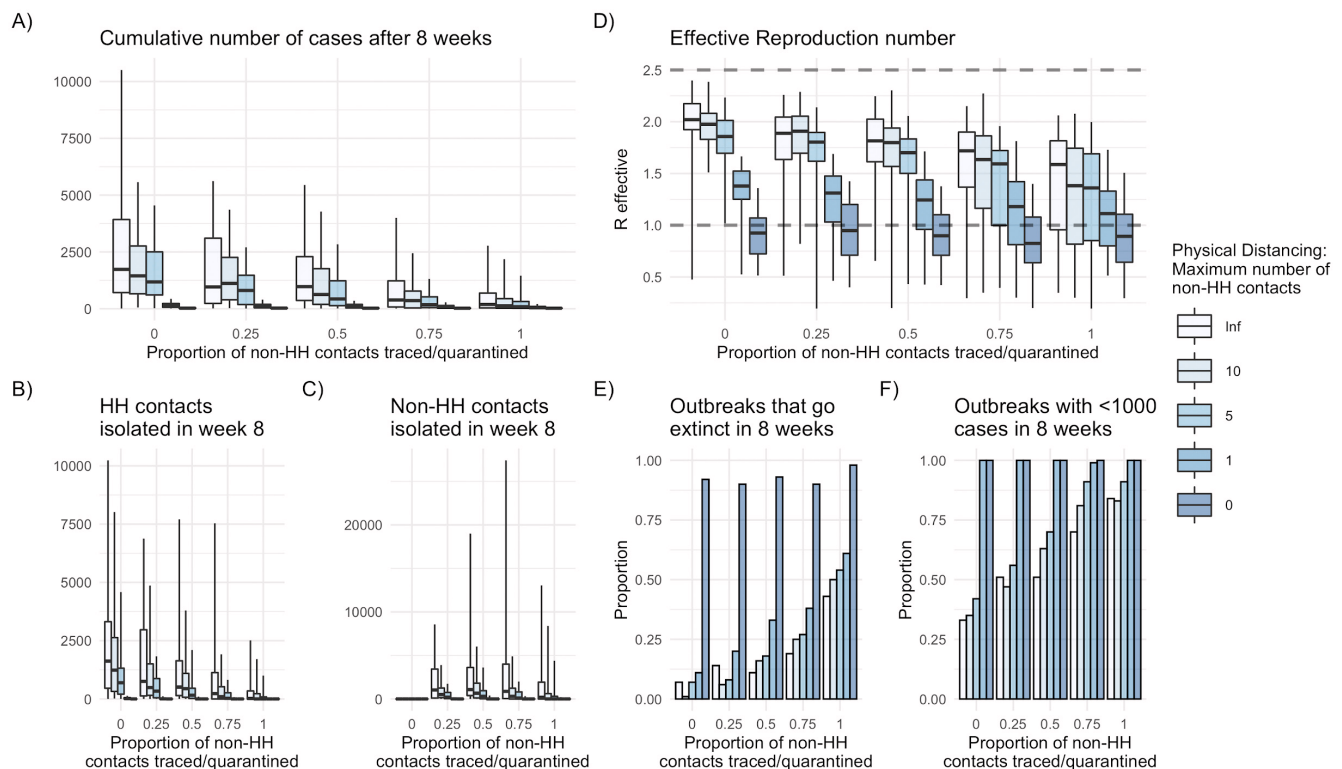
Figure 3. Combining SD and quarantine for base scenario. (**A**) Boxplots of cumulative cases after 8 weeks. (**B**) HH contacts that need to be isolated in week 8 based on tracing or quarantine. (**C**) Non-HH contacts that need to be isolated in week 8 based on tracing or quarantine. (**D**) Boxplots of effective reproduction number estimates for each scenario. (**E**) Proportion of outbreaks that die out for each scenario. (**F**) Proportion of outbreaks that remain below 1000 cumulative cases for each scenario.

interplay emerges with more aggressive tracing means having to trace and isolate more non-HH contacts but only up to a certain threshold until the number of contacts to trace is offset by the impact of more aggressive tracing efforts and thus lower case numbers. For our base case this threshold seems to be quite high, requiring tracing of at least 75% or more non-HH contacts (Figure 3C). None of the less stringent physical distancing measures are able to reduce our estimates of the median effective reproduction below 1 (Figure 3D). Keeping at least 50% of outbreak simulations below 1000 cumulative cases in 8 weeks requires either physical distancing of a maximum of one non-HH contact or tracing of 50% of non-HH contacts (Figure 3F).

Figure 4 represents an optimistic scenario for the combined effect of quarantine and physical distancing measures assuming a short delay to isolation of 1.35 (0.12-3.49) days and a highly dispersed R0 (SARS-like). Quarantining HH members of an isolated case with no physical distancing (apart from preventing SSEs) alone can reduce cumulative cases in 8 weeks to 208 (5-1720) (Figure 4A). Tracing efforts remain relatively low with a maximum of 0 (0-1175) non-HH contacts having to be traced and isolated in week 8 if no physical distancing is in place and 50% non-HH contacts are traced (Figure 4C).

For the majority of scenarios, median estimates of the effective reproduction number are below 1, although the high dispersion results in a large amount of uncertainty dominated by stochasticity (Figure 4D). For the majority of scenarios at least 50% of outbreaks become extinct and over 90% remain below 1000 cumulative cases in 8 weeks (Figure 4E, F).

Discussion
Under the present assumptions, it becomes clear that controlling the spread of SARS-CoV-2 can be difficult even with extreme physical distancing measures in place. Nevertheless, some of these measures can successfully reduce the caseload and thus the burden on health systems at any single point in time. We estimate that school closures and general physical distancing are able to reduce the cumulative number of cases within 8 weeks by over 50% compared to strategies involving isolation of symptomatic cases. The impact of school closures is likely an overestimate as this assumes that children limit their contacts to HH contacts only, whilst Google mobility data suggests that residential mobility has increased as a result of school closures[22]. Even fairly rapid isolation of symptomatic cases does not prevent enough onwards transmission, as a large proportion of cases are either asymptomatic or transmit prior to symptom onset (S1).
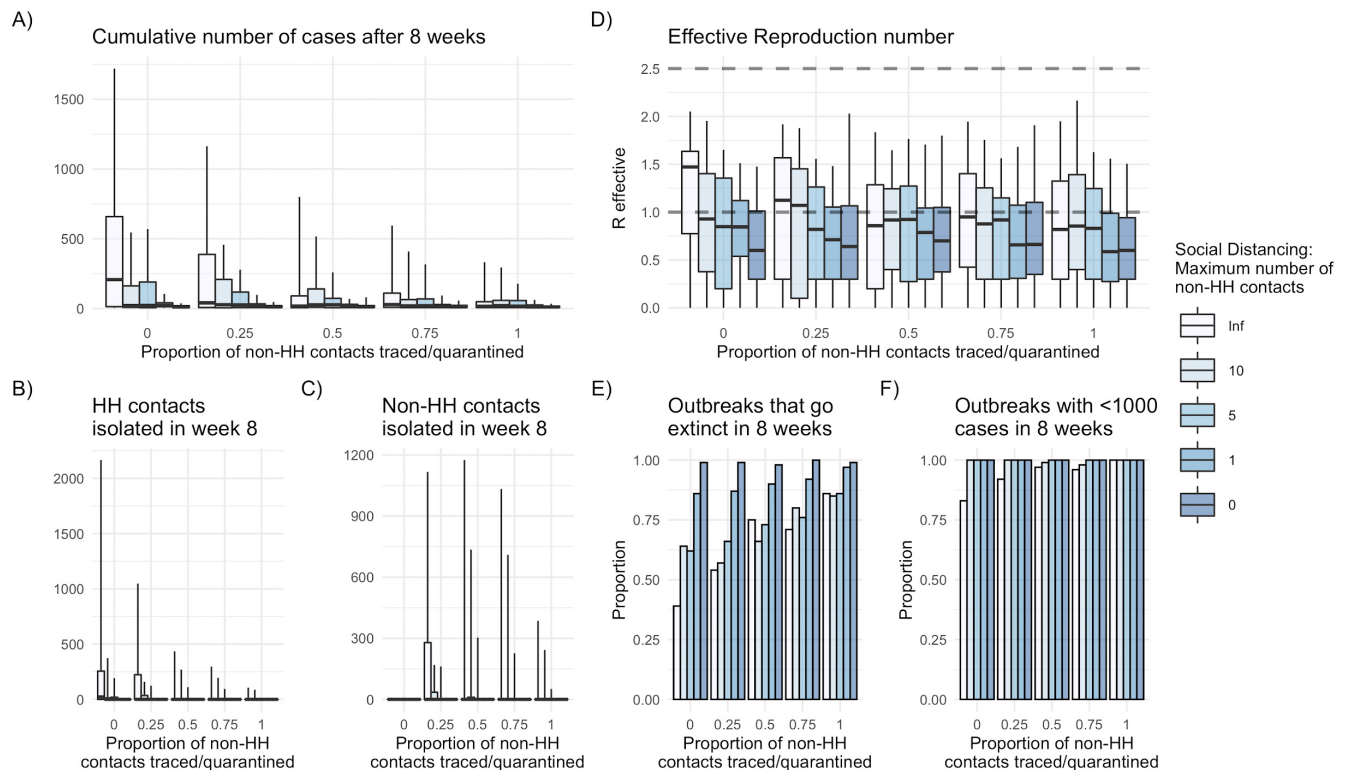
Figure 4. Combining SD and quarantine for optimistic scenario with high dispersion and short delay to isolation following symptom onset. (**A**) Boxplots of cumulative cases after 8 weeks. (**B**) HH contacts that need to be isolated in week 8 based on tracing or quarantine. (**C**) Non-HH contacts that need to be isolated in week 8 based on tracing or quarantine. (**D**) Boxplots of effective reproduction number estimates for each scenario. (**E**) Proportion of outbreaks that die out for each scenario. (**F**) Proportion of outbreaks that remain below 1000 cumulative cases for each scenario.

Our estimate of school closures is very optimistic, as we assume that children reduce their contacts to the HH only without any non-HH contacts. This reduces the caseload, but also shifts the relative burden towards older age groups, where severity is significantly higher (S7)[23]. General physical distancing where non-HH contacts are reduced to five unique contacts per week has a similar effect on overall cases without the shift in age-specific burden (S7). Quarantine strategies that involve isolating all HH members and any successfully traced non-HH contacts immediately may still be effective even given the large proportion of cases that are missed due to being asymptomatic. This assumes no delay in tracing these contacts, which may apply to HH members, but is unlikely for non-HH contacts. Nevertheless, the main benefit does seem to arise through isolation of HH members whilst any additional benefit of tracing non-HH contacts is offset by the large number of contacts that have to be traced and the effort that may be involved in this. It is worth noting that a HH-only quarantine would not necessarily require tracing, if the population is told to isolate HHs as soon as an individual shows symptoms and provided that adherence is high. For model simplicity contacts that never get infected are not quarantined. In reality, infection status would not be known immediately and there may be a delay following testing until non-infected contacts are removed from quarantine. These few additional susceptibles in the model are unlikely to play a major

role in the overall transmission dynamics. We also explored shielding of elderly in our model (see *Extended data*, Supplementary Appendix[15]), which was not particularly effective as the majority of contacts are made by younger age groups so transmission continues. In addition, the burden to older age groups is not significantly reduced as they can still become infected through the HH by mixing with other age groups (S7). This assumed the elderly shield by remaining at home as opposed to physical isolation through rehousing as has been suggested and modelled elsewhere[24].

The most effective strategies that we explored involve a combination of general physical distancing and quarantining of household members and as many non-household contacts as possible of any symptomatic case that is reported. As to be expected, the higher the proportion of non-HH contacts that are traced and isolated and the lower the maximum number of non-HH contacts allowed, the lower the caseload. In general, implementing more extreme physical distancing measures seems to be more effective at reducing case numbers than putting efforts into quarantining a higher proportion of non-HH contacts, although this does not consider the potential societal costs involved in physical distancing. With our base parameter assumptions, however, control is only possible under extreme physical distancing measures, i.e. stay-at-home policies.

Under optimistic parameter assumptions which involve a highly dispersed R0 (SARS-like) and a short delay from symptom onset to isolation of 1.35 (0.12-3.49) days, even moderate physical distancing and quarantine of HH and some non-HH contacts could keep the effective reproduction number below 1, although this also comes with a significant amount of uncertainty.

Limitations

Whilst using realistic contact data is a major strength of this approach, it is also limited by the data available. The contact data is relatively old and whilst it might still be representative of the population in rural and semi-urban settings across Kenya, it cannot with any real confidence be used to represent a current urban area like Nairobi.

Simulations are seeded with an initial number of five imported cases, after which we assume that the population remains closed. In reality, further importations throughout the outbreak are likely. Including these would affect the final caseload and reduce the proportion of extinctions, especially if there are a number of importations early on in the outbreak. Thus, exact estimates should be interpreted with caution.

The age-specific proportion of symptomatic infection estimates used are based on a previous model that was fit to data from multiple countries. These figures represent the clinical fraction of cases by age, i.e. cases that warrant symptoms such that the infected person may seek clinical care. As the model was fit to data from high income countries only, these estimates may not translate directly to the health seeking behaviour in a Kenyan setting, where public health resources are limited. If these are indeed overestimates, any symptom-based interventions such as isolation, tracing or quarantine become more difficult, as more cases are missed. The impact of any broader behavioural changes around health-seeking behaviour due to the socioeconomic pressures of the outbreak are also not taken into account, which are difficult to predict and quantify.

More rigorous contact tracing strategies have not been considered such as tracing of secondary contacts or backward tracing. These may be more effective to avoid further pre-symptomatic transmission events, but are also likely to quickly become unfeasible with growing numbers of cases. With high overdispersion of *R0*, however, a combination of backward and forward tracing may be particularly effective as any given infection is much more likely to have come from a SSE, whilst forward tracing alone will find many contacts that never get infected.

Concluding remarks

A combination of strategies that involve banning of large gatherings, general physical distancing, and quarantining of HHs and as many non-HH contacts of an infected person, may be effective in preventing or at least significantly slowing down the spread of SARS-CoV-2 in Kenya. The number of non-HH contacts to be traced can quickly become infeasible as the caseload grows with limited added benefit. Thus, focusing efforts on more targeted testing and reducing the delay until infected individuals are isolated, whilst isolating their HH, may be more effective than attempting to trace more non-HH contacts. Isolation of infected individuals, school closures, or shielding of elderly alone are not feasible in containing spread, although combinations of these may be effective.

There is the need for a better understanding of the main drivers of SARS-CoV-2 (e.g. what role do school children play) and the role that large gatherings (i.e. with potential for superspreading) may play in overall transmission. In particular, estimates of reproduction numbers and their dispersion in African settings would be useful. Furthermore, investigations into setting specific transmission events would help in order to establish interventions that are more targeted, less resource intensive, and carry less of a socioeconomic burden. Finally, data on age-related contact patterns in different settings and under-different interventions should be viewed as a high priority for collection.

Data availability

Underlying data

All data underlying the results are available as part of the article and no additional source data are required.

Extended data

The supplementary appendix, relevant code, and data are available at: https://github.com/moritz-wagner/COVID19-contact-tracing/.

Archived version at time of publication: http://doi.org/10.5281/zenodo.4004732[15].

License: MIT License.

References

1. **Kenya**. Worldometer. 2020.
   Reference Source

2. Barasa E, Ouma PO, Okiro EA, *et al.*: **Assessing the hospital surge capacity of the Kenyan health system in the face of the COVID-19 pandemic.** *PLoS One.* 2020; **15**(7): e0236308.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

3. Hellewell J, Abbott S, Gimma A: **Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts.** *Lancet Glob Health.* 2020; **8**(4):e488–e496.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

4. Epiforecasts: **Epiforecasts/Ringbp**. GitHub. 2020.
   Reference Source

5. Fraser C, Riley S, Anderson RM , *et al.*: **Factors that make an infectious**

disease outbreak controllable. *Proc Natl Acad Sci U S A.* 2004; **101**(16): 6146–6151.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

6. Klinkenberg D, Fraser C, Heesterbeek H: **The effectiveness of contact tracing in emerging epidemics.** *PLoS One.* 2006; **1**(1): e12.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

7. Kretzschmar ME, Rozhnova G, Michiel van Boven: **Effectiveness of Isolation and Contact Tracing for Containment and Slowing down a COVID-19 Epidemic: a Modelling Study**. 2020.
Publisher Full Text

8. Peak CM, Kahn R, Grad Y, *et al.*: **Modeling the Comparative Impact of Individual Quarantine vs. Active Monitoring of Contacts for the Mitigation of COVID-19**. 2020.
Publisher Full Text

9. Zou L, Ruan F, Huang M, *et al.*: **SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients.** *N Engl J Med.* 2020; **382**(12): 1177–1179.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

10. Ganyani T, Kremer C, Chen D, *et al.*: **Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020.** *Euro Surveill.* 2020; **25**(17): 2000257.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

11. He X, Lau EHY, Wu P, *et al.*: **Temporal dynamics in viral shedding and transmissibility of COVID-19.** *Euro Surveill.* 2020; **26**(5): 672–675.
PubMed Abstract | **Publisher Full Text**

12. Tabata S, Imai K, Kawano S, *et al.*: **Non-Severe vs Severe Symptomatic COVID-19: 104 Cases from the Outbreak on the Cruise Ship 'Diamond Princess' in Japan**. 2020.
Publisher Full Text

13. Li R, Pei S, Chen B, *et al.*: **Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2).** *Science.* 2020; **368**(6490): 489–493.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

14. Kiti MC, Kinyanjui TM, Koech DC, *et al.*: **Quantifying age-related rates of social contact using diaries in a rural coastal population of Kenya.** *PLoS One.* 2014; **9**(8): e104786.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

15. Wagner MI: **moritz-wagner/COVID19-contacttracing (Version v1.0).** *Zenodo.* 2020.
http://www.doi.org/10.5281/zenodo.4004732

16. Imai N, Dorigatti I, Cori A, *et al.*: **Estimating the potential total number of novel Coronavirus cases in Wuhan City, China.** Imperial College London (22-01-2020).
Publisher Full Text

17. Lloyd-Smith JO, Schreiber SJ, Kopp PE, *et al.*: **Superspreading and the effect of individual variation on disease emergence.** *Nature.* 2005; **438**(7066): 355–359.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

18. Backer JA, Klinkenberg D, Wallinga J, *et al.*: **Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20-28 January 2020.** *Euro Surveill.* 2020; **25**(5): 2000062.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

19. Li Q, Guan X, Wu P, *et al.*: **Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia.** *N Engl J Med.* 2020; **382**(13): 1199–1207.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

20. Donnelly CA, Ghani AC, Leung GM, *et al.*: **Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong.** *Lancet.* 2003; **361**(9371): 1761–1766.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

21. Davies NG, Klepac P, Liu Y, *et al.*: **Age-dependent effects in the transmission and control of COVID-19 epidemics.** *Nat Med.* 2020; **26**: 1205–1211.
Publisher Full Text

22. **COVID-19 Community Mobility Report**. Google.
Reference Source

23. Wu Z, Mcgoogan JM: **Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention.** *JAMA.* 2020.
PubMed Abstract | **Publisher Full Text**

24. Zandvoort KV, Jarvis C, Pearson C, *et al.*: **Response Strategies for COVID-19 Epidemics in African Settings: a Mathematical Modelling Study**. 2020.
Publisher Full Text