Understanding the infections dynamics and parameters that are consistent with the epidemiological data from Yemen

Yuwei Cheng[1], Kevin Van Zandvoort[2], Nada Abdelmagid[2], Ali Ahmed Alwaleedi[3], Abdulkareem Al-Muaalemi[4], Nuha Mahmoud[4], Jeremias Naiene[4], Afaf Naser[4], Samuel Omara[4], Mikiko Senga[4], Natasha Howard[1,2], Hannah Eleanor Clapham[1*]

1: Saw Swee Hock School of Public Health, National University of Singapore, Singapore
2: London School of Hygiene and Tropical Medicine, United Kingdom
3: Ministry of Public Health and Population, Yemen
4: World Health Organization, Yemen Country Office

**Corresponding author*

The first source of epidemiological data is reported cases and deaths over time (Figure 1). This data suggests that the epidemic peak of cases and deaths in Yemen happened between May and July and declined thereafter. We use this information for peak timing, and not the magnitude of cases and deaths for the model, due to a high test positivity rate suggesting under-reporting.

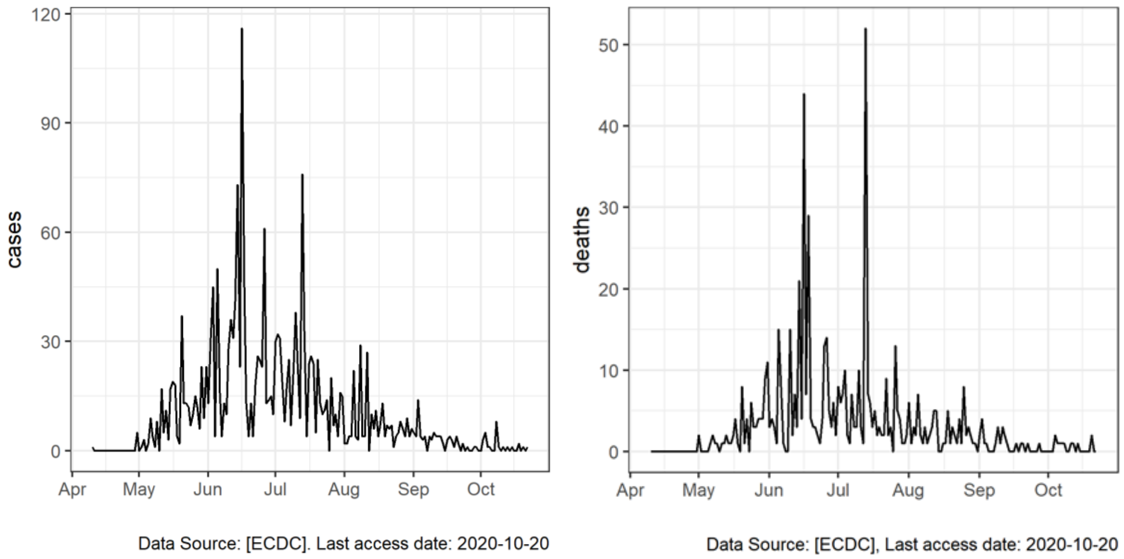
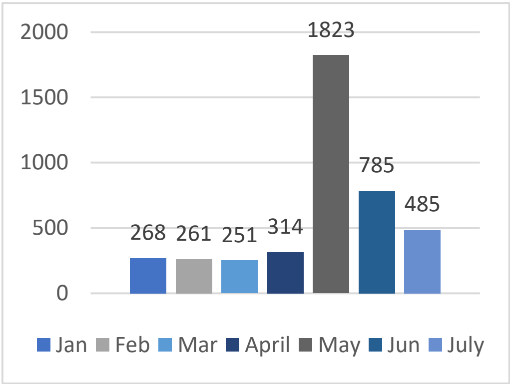


Figure 1: COVID-19 cases and deaths reported in Yemen (European Centre for Disease Prevention and Control (ECDC), 2020). Last date access is Oct 20, 2020. The first COVID-19 case in Yemen was reported on April 10th

A second source of data is number of monthly civil death records from Aden (Figure 2.1) and estimates of the weekly new burials (Figure 2.2) from Besson et al. (2020). In this data, it appears that the peak in deaths in Aden occurred between May and July 2020 in Aden, with an estimate of excess deaths from April 1st to July 6th of 1451 or 1560 deaths depending on the method used. As this data is more complete recording of deaths we are also able to use this data on magnitude of deaths too.



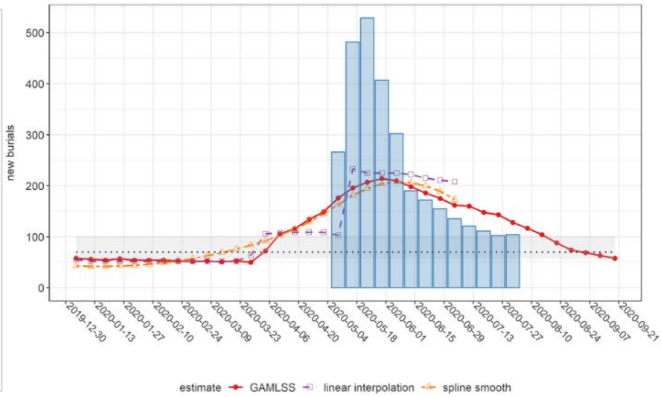


Figure 2.1: Monthly civil death records in Aden (Ghouth, 2020). Figure 2.2: Weekly burials in Aden. Bars represent weekly civil death records and lines indicate estimates of weekly new burials (Besson et al., 2020)

Inputs on movements and contacts and how they change

In order reconstruct the changes in contacts that occurred in the general population over this period due to control measures we use google mobility data from Yemen (Figure 3). There was an initial decrease in movements to all places outside the households, and movements for grocery and pharmacy, transit stations, parks and retail and recreation have been rising since May 25th. We used these to modulate the contacts in the model. We also had information on school closures March ("Tracking Public Health", 2020) to Oct (Qasim, 2020) and therefore modelled no school contacts in this period.

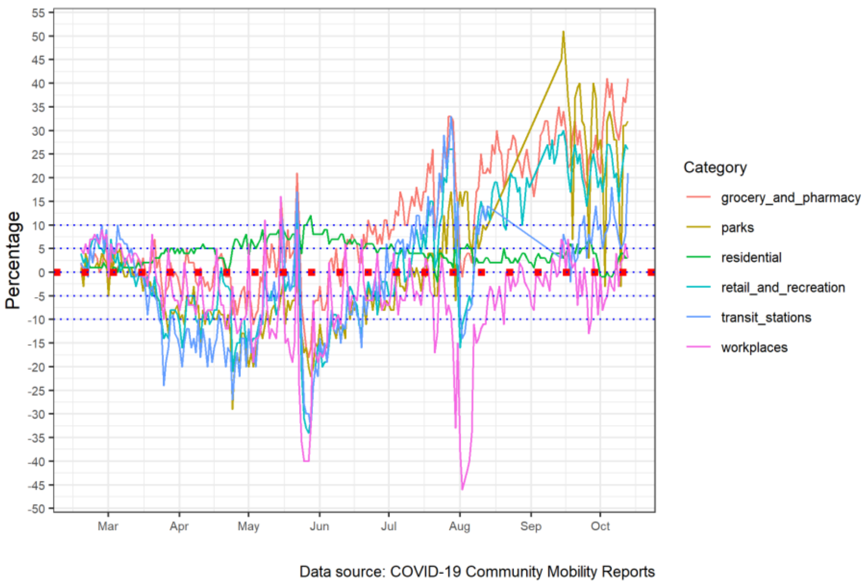


Figure 3: Daily mobility changes in Yemen (Google LLC, 2020). Last date access is Oct-20, 2020. Baseline is the median mobility value of 2020-01-03 to 2020-02-06

Transmission parameters consistent with observed peak timing for Yemen and Aden, and number of deaths in Aden

We ran the transmission model (Davies et al., 2020) with different R_0 and seeding times to find which of these would be consistent with the observed peak timing in the datasets. Timing of peak suggests

R_0 of range 2.8 to 3.5 for Yemen and R_0 of range 3.3 to 3.6 for Aden (depending on the seeding date chosen). There is much uncertainty in the seeding date.

Running the model with the R_0 consistent with the peak timing, without and with the observed interventions (google mobility changing adults contacts, school closures leading to no school contacts) suggests the epidemic has been slightly mitigated, with 13-15% difference in deaths between the two scenarios (Figure 4). For the simulations with these R_0 and seeding dates suggest a high infection rate in the population, of between 60% and 70% of the population infected.

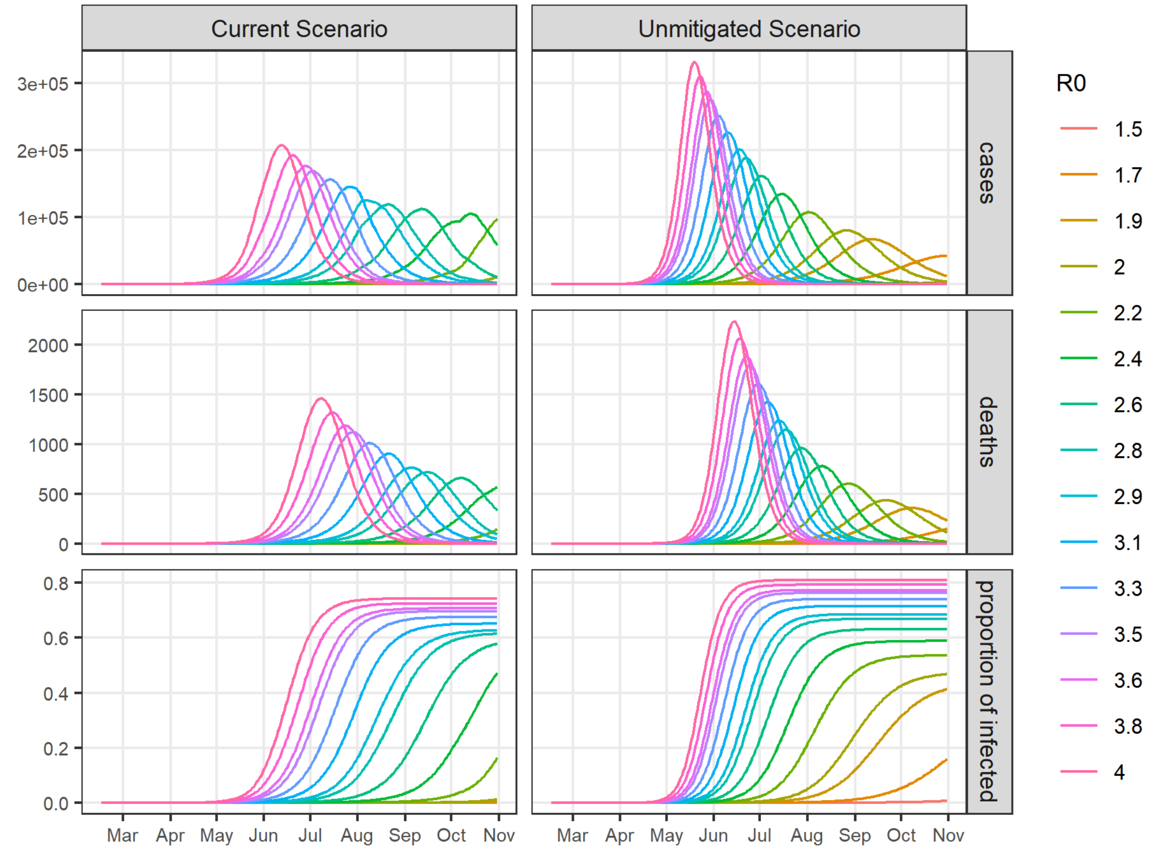


Figure 4: Simulated COVID-19 cases, deaths, and proportion of population immunity by different scenarios and R_0 values in Yemen when seeding date is 2020-02-16

For the Aden model we simulate the deaths under the mitigated scenario, and the magnitude of excess deaths between April and July is also consistent R_0 of (3.3, 3.6), again suggesting 60%-70% of the population have now been infected.

R0	2020-02-01	2020-02-16	2020-03-01
2.6	328	88	20
2.8	641	212	45
2.9	812	311	68
3.1	1120	585	144
3.3	1361	917	281
3.5	1547	1238	494
3.6	1627	1380	627
3.8	1769	1622	928
4	1898	1814	1241

Table 1: Model output estimated simulated COVID-19 deaths in Aden from April 1st to July 6th

There are several limitations to this analysis, we cannot include impact of other control measures and the assumption about the impact of school closures is subject to uncertainties in the role children play in transmitting infection. The excess deaths in this period may not only be due to COVID-19 but may be due to other knock-on impacts of COVID, so other scenarios with lower excess deaths are still possible, and the result is sensitive to our assumptions about age-stratified infection fatality rate (IFR) for which there is limited data in Yemen.

Age-stratified reported deaths for Yemen

There is some data on age-distribution of deaths reported to WHO Yemen (Table 2). This can be of some use for looking at parameter ranges for Yemen infection rates and IFR by age. The age group of 30-44 had the highest reported number of deaths, followed by the age group of 60+ and 45-59. The current output of the model is not consistent with the observed distribution. There are three possible explanations for this, the model needs additional differences in (i) risk of getting infected by age (simulated by changing the contact patterns for different group), (ii) risk of death by age and (iii) seeking hospital care by age.

Age Group	Number of Deaths	Percentage
0-4	3	1.7%
15-14	4	2.2%
15-29	34	19.5%
30-44	56	31.3%
45-59	38	21.2%
60+	44	24.6%
Total	179	100%

Table 2: Age-stratified reported deaths (WHO Yemen data personal communication, last case date in file is July 25th)

We find that to produce the proportion of deaths consistent with the observed data, contacts in age group 40+ needs to be reduced by at least 75%. Figure 5 plots the epidemic curve under the scenario with age-stratified reported deaths consistent with the data. Under these parameters, we would estimate that (23,376 - 25,155) deaths occurred between Feb 16th and Oct 31st and 66% of the population is infected, compared to mitigated scenario previously (51,313 deaths, and 70% of population infected).

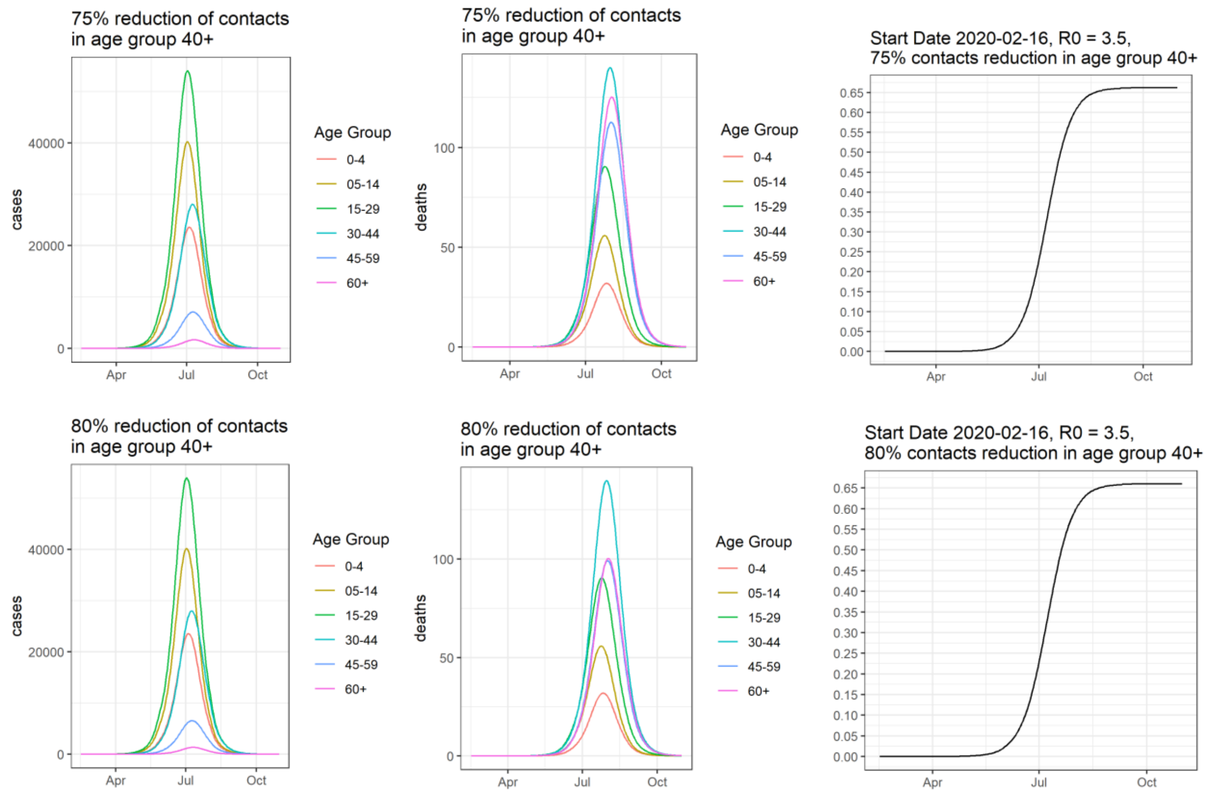


Figure 5: Simulated epidemic curve whose output of age-stratified deaths is consistent with the observed data. Seeding date is 2020-02-16, R_0 is 3.5, and contacts for age group 40+ are reduced by 75% or 80%.

For disease severity, we find that to produce the proportion of deaths consistent with the observed data on the proportion of deaths in each age group, disease severity in age group 20-40 needs to be increased at least by 4.6 times, and age group 40-60 by 2.4 times compared to the value used in previous modelling work (estimated from the outbreak setting in Asia and adjusted for LMIC settings) (see Table 3 for both values). Figure 6 plots the epidemic curve under this scenario and we estimate 82,262 deaths occurred between Feb 16th to Oct 31st under this scenario compared to 51, 313 deaths in the mitigated scenario previously. Under this scenario the same proportion of the population is infected.

Age group	**Risk of death given symptomatic infection in original model***	**Adjusted risk of death given symptomatic infection in updated results**
0-10	0.142%	0.142%
10-20	0.152%	0.152%
20-30	0.185%	0.851%
30-40	0.346%	1.592%
40-50	1.010%	2.424%
50-60	2.810%	2.810%
60-70	6.201%	6.201%
70+	11.5%	11.5%

Table 3: Risk of death given infection *(Zandvoort et al., 2020)

Heterogeneity of healthcare seeking behaviours in age groups is the final possible explanation we explore. We find that to generate consistent death results, at least 58% of deaths in the age group of 45-59 and 72% of deaths in the age group of 60+ would not be captured by the surveillance system.

Under this scenario, we estimate of the 51, 313 deaths that occurred, only 26, 912 deaths would be captured by the surveillance system. This difference will not alter the proportion of the population that is infected.

In reality the age-distribution of deaths could be caused by a combination of these parameters, and this will be considered in future work, though ultimately different data sources will also be needed. Seroprevalence data and any information on contact patterns by age will be informative about whether the risk of being infected is different by age. It would also be informative to understand the age-distribution of co-morbidities known to be associated with severe disease and death for COVID-19, to assess whether the patterns estimated here are consistent with the parameters. Finally, data on age-stratified cases in hospitals would be very useful to assess healthcare seeking behaviour.

Overall conclusions

Piecing together information on cases and deaths from Yemen from a number of sources we have attempted to reconstruct the dynamics. We estimate a high level of transmission, but with the epidemic having been slightly mitigated by the control measures that were put in place. There are number of key uncertainties that will impact the future transmission scenarios from now on in Yemen, the key one being the number of individuals who have been infected, and how long any immunity lasts. Age-stratified seroprevalence will be highly informative about the number of individuals who have been infected, the age-stratified death rate, and therefore possible future scenarios and optimal future interventions.

References:

Besson, E.K., Norris, A., Ghouth, B., Freemantle, T., AlHaffar, M., Checchi, F. (2020). Estimates of COVID-19 attributable mortality in Aden governorate, Yemen: a geospatial and statistical analysis.

"COVID-19 data". (2020). *European Centre for Disease Prevention and Control*. Retrieved from: https://www.ecdc.europa.eu/en/covid-19/data

Davies, N. G., Kucharski, A. J., Eggo, R. M., Gimma, A., Edmunds, W. J., Jombart, T., O'Reilly, K., Endo, A., Hellewell, J., Nightingale, E. S., Quilty, B. J., Jarvis, C. I., Russell, T. W., Klepac, P., Bosse, N. I., Funk, S., Abbott, S., Medley, G. F., Gibbs, H., … Liu, Y. (2020). Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. *The Lancet Public Health, 5(7), e375–e385*. Doi: https://doi.org/10.1016/s2468-2667(20)30133-x

Ghouth, A.S. (2020). Excess deaths within the context of COVID-19 pandemics in Aden in May 2020.
Google LLC "Google COVID-19 Community Mobility Reports". Retrieved from: https://www.google.com/covid19/mobility/ Accessed: <Oct-20, 2020>.

Qasim, A. (2020, Oct 7th). SCHOOLS REOPEN IN YEMEN AFTER MONTHS OF CLOSURE DUE TO COVID-19. *Unicef*. Retrieved from: https://www.unicef.org/yemen/stories/schools-reopen-yemen-after-months-closure-due-covid-19

"Tracking Public Health and Social Measures". (2020). *World Health Organization*. Retrieved from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/phsm

Zandvoort, K., Jarvis, C., Pearson, C., Davies, N., Group, C., Russell, T., . . . Checchi, F. (2020, January 01). Response strategies for COVID-19 epidemics in African settings: A mathematical modelling study. Retrieved from: https://www.medrxiv.org/content/10.1101/2020.04.27.20081711v1

Funding statement:

This material has been funded by UK aid from the UK government; however the views expressed do not necessarily reflect the UK government's official policies. YC, KvZ, NA, NH and HEC are funded by UK aid from the UK government, under the Support for COVID-19 epidemic control in the Middle East and North Africa project, number: 40123957. UK aid did not have any role in the design of the study and collection, analysis, and interpretation of data and in writing the report.