medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070904; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Ongoing outbreak of COVID-19 in Iran: challenges and signs of concern

M. Ghafari[1*], B. Hejazi[2], A. Karshenas[3], S. Dascalu[1,4], L. Ferretti[5], A. Ledda[6], A. Katzourakis[1*]

[1] Department of Zoology, Peter Medawar Building for Pathogen Research, University of Oxford, Oxford, UK. [2] Department of Physics, Wesleyan University, Middletown, Connecticut, USA. [3] Department of Engineering, University of Oxford, Oxford, UK. [4] Avian Influenza Virus, The Pirbright Institute, Woking, UK. [5] Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, UK. [6] Department of Infectious Disease Epidemiology, Imperial College London, UK.

*Corresponding author.
Email: mahan.ghafari@zoo.ox.ac.uk (MG); aris.katzourakis@zoo.ox.ac.uk (AK)

Abstract

Since the first outbreak in China, the Coronavirus Disease 2019 (COVID-19) has rapidly spread around the world. Iran was one of the first countries outside of China to report infections with COVID-19. With nearly 100 exported cases to various other countries, it has since been the epicentre of the outbreak in the Middle east. By examining the age-stratified COVID-19 case fatality rates across the country and 14 university hospitals in Tehran, we find that, in younger age groups, the reported cases on 13/03/2020 only capture less than 10% of symptomatic cases in the population. This indicates significant levels of under-reporting in Iran. Using the 18 full-genome sequences from cases with a travel history or link to Iran, as well as the one full genome sequence obtained from within the country, we estimate the time to the most recent common ancestor of sequences which suggests the likely start of the outbreak on 21/01/2020 (95% HPD: 05/12/2019 - 14/02/2020) with an approximate doubling time of 3.07 (95% HPD: 1.68 - 16.27). Also, based on known exported cases to Oman, Kuwait, Lebanon, and China, we estimate the outbreak size on 25 February and 6 March to be around 13,700 (95% CI: 7,600 - 33,300) and 60,500 (43,200 - 209,200), respectively. Knowing the size of the outbreak at two time points and the typical doubling times associated with the COVID-19 epidemics in countries across Europe and North America, we can independently verify that the likely start of epidemic in Iran is around 15/01/2020 (27/12/2019 - 24/01/2020). Our assessment of the fate of the epidemic based on current levels of non-pharmaceutical interventions implemented by the government suggests upward of 10 million cases (IQR: 6.7×10^6 - 1.8×10^7) and 100,000 ICU beds required (IQR: 7.7×10^4 - 1.4×10^5) during the peak of the epidemic with more than

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070904; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

100,000 cumulative deaths (IQR: 1.8×10^5 - 2.4×10^5). We also predict a peak in demand for ICU beds on 21/04/2020 (IQR: 06/04/2020 - 23/05/2020). The large span of the peak of the ICU demand is a result of two separate peaks, with the first occurring at around 15/4/2020 and the second in approximately a months time. The latter is also expected to last longer and is based on the relatively relaxed social distancing measures in place. The exact magnitude and timing of the peaks strictly depends on levels of interventions and can change significantly upon new information or change of policy. We caution that a lack of, or relaxed, stringent intervention measures, during a period of highly under-reported spread, would likely lead to the healthcare system becoming overwhelmed in the next few months.

Introduction

The current COVID-19 pandemic had a major impact on societies all around the world. While the first outbreak occurred in Wuhan, China, the SARS-CoV-2 virus responsible for the disease spread rapidly to other parts of China and the world. Iran was one of the first countries outside of China to be hit by the COVID-19 pandemic; it has since exported many cases to regional countries in the Middle East, such as Kuwait, Oman, Bahrain, Afghanistan, Iraq, and Pakistan, as well as to more distant countries such as Canada, Australia, Germany, and the UK. The first reported cases in Iran were of two deceased individuals in the city of Qom on 19 February 2020. The reported cases in several provinces around the country started to grow in a matter of a few days following the two cases in Qom (see Fig. 1 and the time lapse video of the COVID-19 outbreak across Iran during the first 32 days of the epidemic in the Supplementary materials). Some of the early reports suggest the first two cases could have been infected by a merchant who had reportedly travelled to China [Robin Wright, 2020]. However, with COVID-19, identifying the so-called 'patient zero' is problematic, due to the high rates of asymptomatic and pauci-symptomatic infections [Li et al., 2020b]. The most likely route of virus spread from China to Iran was via (asymptomatic) air travelers. Infected individuals typically show no symptoms for about 5 days [Li et al., 2020a] or sometimes no symptoms at all [Bai et al., 2020], while silently spreading the virus and therefore hindering control efforts [Ferretti et al., 2020]. Since Qom is the religious capital of the country with millions of pilgrims visiting its shrines every year, transmission chains likely grew rapidly across several parts of the country. In response to the first cases, the government put in place minimal containment strategies which failed to limit the spread of the epidemic, particularly during the early stages, at a time when movement associated with visits to shrines may have contributed to enhancing the spread.

The genomic data collected from cases with a history of travel to Iran during the last week of February and early March suggest the emergence of a clade of SARS-CoV-2 [Eden et al., 2020, Rambaut et al., 2020]. These sequences show signs of extreme genetic

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070904; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

similarity possibly suggesting the outbreak may have started from a single transmission event as suggested by some early analysis from NextStrain [Bell et al., 2020, Hodcroft et al., 2020]. Although this analysis cannot rule out the possibility of separate introductions from unsampled sources in the population. We note that, as of 14/04/2020, there has been only one full-genome sequence from inside Iran. There are 20 sequences collected from individuals that flew out of Iran and were tested and had their viral genomes determined in other countries, which can serve to determine the parameters of the epidemic in Iran. Two of these sequences are epidemiologically linked and cannot be used for these purposes. Genomic data originating from Iranian travellers should contain the genetic signals that characterised the local epidemic growth of the country they travelled from. Early analyses of genomic data from China and across the world have revealed aspects of the timing and transmission of the virus [Lu et al., Volz et al., 2020].

In this study we investigate several aspects of the COVID-19 outbreak in Iran. We compare the reported number of cases and deaths in Iran to other countries that have a comparable outbreak size. The comparison of data shows that the numbers from Iran do not follow the expected global pattern for this disease which we argue is caused by the limited testing facilities available in the country. Studying the age-specific case fatality ratio in Iran provides further evidence that the outbreak in Iran is likely much larger than what the official numbers suggest. We infer the start date of the epidemic in Iran as well as the rate of exponential growth from genetic data largely from travellers from Iran to other countries. We also predict an approximate date for when the outbreak started in the country based on the analysis of travel data on the number of cases that have been exported to other countries from Iran. We suggest that the outbreak had started about a month before the first deaths related to COVID-19 were reported. Other studies have shown that sustained transmission may exist up to a month prior to the first reported death [Lourenço et al., 2020]. Furthermore, we examine the burden that this epidemic has on the healthcare system under different intervention scenarios and how different measures can help limit the surge in demand for ICU beds and curtail the size of the outbreak. In the appendix we briefly examine the possibility of data manipulation using Benford's law and find no evidence in support of such a claim. We also give a detailed explanation of our analysis of estimating the outbreak size using air travel data. We have also created a timeline of all major events related to the COVID-19 outbreak in Iran in the Supplementary materials.

Reports from the COVID-19 outbreak in Iran

Six days after the first two deaths were announced by Iran's Ministry of Health, various sources from hospitals around the country told BBC Persian [2020e] that there have been at least 210 deaths related to COVID-19 with most cases coming from Tehran while official reports only showed 34 deaths at the time. On the same day, Radio Farda [2020b] reported that the World Health Organization Director-General Tedros Adhanom Ghebreyesus said

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070904; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

that 97 cases in 11 countries originated in Iran while the total number of cases in the country at that time were 281. Similar reports from Iran International [2020] on 2 March state there were a total of 428 deaths in the country while official numbers reported by the government were 66. Iran International further reported on the same day that according to the president of the University of Health Sciences of Golestan province, their medical team was of the opinion that 594 patients that were hospitalised in the province were infected with COVID-19. He added that between 18 to 29 February they had sent 135 samples of COVID-19 tests from patients in Golestan to Tehran for testing and only 26 results were reported back with 12 confirmed cases, of which 5 had died and the other 7 had recovered and been sent home. He stated that according to the ministries protocols, not all patients with severe breathing difficulties were going to receive COVID-19 testing. These reports and various similar stories from other provinces highlight that insufficient testing equipment is one of the major challenges the country is facing since the start of the outbreak which is likely to have led to significant underestimation of active cases, thereby delaying appropriate public health interventions. These issues would have been exacerbated due to limited governmental support for employers and businesses affected by the outbreak. Indeed. fears over economic uncertainty in the country have prompted many citizens not to follow recommendations from the government which might have impacted them financially such as working from home or temporarily closing their businesses [Basravi, 2020, Reality Check team, 2020, Wintour, 2020]. As a consequence, many activities followed as usual, and, at best, the impact of early public health measures in Iran only had a marginal effect.

Cloud of uncertainty over reported numbers from Iran

One of the main characteristics of countries with major COVID-19 outbreaks is the exponential increase of number of cases over time. By comparing the reported cases and deaths in several countries with sizable outbreaks, we can see that Iran was also initially on an exponential trajectory (see Fig. 2a and Fig. 2b). After reaching around 7,000 reported cases by mid-March, we see a significant deviation from the exponential trajectory which may partly be associated with school and university closures announced by the government on 5 March. Despite the fact that first cases in several European countries were reported in mid to late February and that similar or stricter intervention policies were implemented around the same time, it is surprising to see the reported numbers from Iran are much lower than those countries. In particular, the constant accumulation of new deaths from 21/03/2020 to 10/04/2020 in Iran suggests that post-mortem testing is likely capped by the limited test kits available in the country. Figure 2c compares daily Case Fatality Ratio (CFR) between several countries (daily CFR = daily reported deaths / daily reported cases). We note that the sharp decline of daily CFR from late February to early March in Iran is due to having 100% CFR on 19 February which dropped as new tests were taken and the number of cases started to grow. We also see a second peak in CFR on 23 March which, again, could be due to testing capacity not catching up with the rate at which the epidemic is

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070904; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

growing in the country. Another possible explanation for the second peak is that changes of population behaviour (e.g. due to control or awareness of increasing deaths) can lead to a reduction in transmission, which, naturally, leads to a drop in reported infections while incidence in deaths follows this reduction with some delay in time.

Results

Age-stratified analysis of fatality rates in Iran: evidence of significant under-reporting

We examine the only nationwide data available for age-specific naïve Case Fatality Ratio (nCFR) obtained from the Iranian Health Ministry on 13/03/2020 [The Islamic Republic News Agency(IRNA), 2020] and compare it to the reports on 04/04/2020 from 14 hospitals associated with Tehran University of Medical Sciences [Tehran university school of medicine COVID-19 response team, 2020] and the Chinese Center for Disease Control reported on 11/02/2020 [The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team, 2020]. According to the Ministry of Health, there were a total of 11,364 reported cases and 514 total deaths across the country on 13/03/2020. Tehran University of Medical Sciences hospitals reported 8,840 cases with 654 deaths and the study by CDC China reported 44,672 confirmed cases and 1,023 deaths in the Mainland. We see a noticeable difference between nCFR in Iran and China across all age groups (see Fig. 3a). In particular, Iran's nCFR for < 50 years-old is the same as those reported from hospitalised patients in Tehran and more than 10 times higher than China which suggests that Iran is only reporting severe cases of the infection. There also seems to be an underestimation of deaths in the > 80 age-group in Iran compared to those in China. We also note that nCFR will likely be an underestimate of the true CFR in Iran because the results from all cases are not known yet and the epidemic is still ongoing. To adjust for such biases in CFR, e.g. using a CFR correction method developed by Russell et al. [2020], we need to have temporal data, but we were not able to obtain more than one data point as of 10/04/2020.

A recent analysis by researchers at Imperial College London [Triggle, 2020] showed that the age-specific risk of dying from COVID-19 greatly overlaps with 'normal' annual death rate in males and females. This suggests those dying from COVID-19 could have died within that same year and that SARS-CoV-2 has accelerated this. However, we see a completely different picture in Iran. Figure 3b shows a significant increase in death rate due to COVID-19 compared to the annual rates in almost all age categories which is also likely due to under-reporting of cases in Iran.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070904; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

When did the epidemic start and how fast did it grow in Iran?

Phylogenetic analysis

We used 20 whole genome sequences obtained from patients that had travelled out of Iran during the current epidemic (and their infections were reported outside the country), combined with the first whole genome sequence obtained from Iran, in order to estimate parameters of the epidemic using population genetic models implemented in BEAST [Suchard et al., 2018]. We reconstructed a maximum liklelihood phylogenetic tree of all sequences that were linked to Iran using PhyML [Guindon et al., 2010] (see Fig. 4). Two of the sequences were from epidemiologically linked patients, and were therefore removed from the dataset as they would bias estimators based on the coalescent, resulting in 19 sequences overall including the genome from Iran. We have highlighted those in the metadata in the supplementary files. We used the known sampling times, with a CTMC rate reference prior on the rate, in order to estimate the evolutionary rate from this particular sample from serial samples. We implemented the exponential growth model, with a lognormal prior on population size (mean=1, stedev=2), and the growth rate implementation with a laplace prior (scale=100), and an HKY+G model of substitution. Our inferred substitution rate is consistent with other studies [Boni et al., 2020, Rambaut, 2020] at 1.66×10^{-3} (HPD: 2.18×10^{-4} - $3.31 \times 10^{-}3$), and the exponential growth rate was 82 in units of years (HPD: 15.55-150.34), corresponding to a doubling time of around 3 days (95% HPD: 1.68-16.27). The age of the root is placed on 21/01/2020 (95% HPD: 05/12/2019-14/02/2020).

Epidemiological analysis

Two recent studies by Tuite et al. [2020] and Zhuang et al. [2020] estimated the size of the outbreak in Iran on 25 February to be approximately 17,000 (4,800 - 44,500). The study by Tuite et al. [2020] used the method developed by Fraser et al. [2009] to estimate the outbreak size based on 3 exported cases to the UAE, Lebanon, and Canada. Their estimates relies on information about the average length of stay of visitors (taken from Encyclopedia of Nations [2003]), proportion of international travels who are residents of Iran [World Tourism Organisation, 2020], and the 'infectious disease vulnerability index', a proxy for measuring the country's ability to detect cases. Apart from Imam Khomeini airport (the one with the highest number of international passengers), they only included Rasht and Arak airports with a total of less than 10,000 international passenger per year and exclude from their analysis major airports in Mashhad, Shiraz, and Isfahan with more than 10 million international travellers. The other study by Zhuang et al. [2020] used the Binomial sampling method [Imai et al., 2020] to estimate the outbreak size based on 5 exported cases to Oman, UAE, and Lebanon. They used the VariFlight [2020] platform to estimate the weekly number of passengers to Iran. Although we can not independently verify those numbers, our estimates are fairly similar (see Table 3). The method used by Zhuang et al. [2020] demands fewer parameters and produces similar results as Tuite et

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070904; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

al. [2020], but it still does not account for the variation in catchment population size of international airports in Iran and the expected exposure time of those exported passengers.

Here, we estimate the outbreak size in late February and early March using a detailed list of flight information from all airlines and try to accurately estimate the flux of passengers to/from four international airports in Iran with the highest number of weekly international flights to countries reporting exported cases from Iran (see Table 3, and supplementary materials) before the Gulf Cooperation Council countries suspended their flights to Iran. To capture the variation that exposure time, t, and catchment population size, M, create on the probability of having an exported case on a flight, $p = \frac{tD}{M}$, where D is the daily passenger flux, we use a Beta-binomial distribution where the compound Beta function is defined on the random variable $\zeta = \frac{t}{M}$ and a support of $(\frac{t_{min}}{M_{max}}, \frac{t_{max}}{M_{min}})$ where $t_{min} = 20$, $t_{max} = 50$, $M_{min} = 4 \times 10^7$, and $M_{max} = 5.56 \times 10^7$. Then, for each country, i, in Table 3 with n_i cases and corresponding success probability p_i of finding a case, we can calculate the expected outbreak size $\hat{\lambda}$ using the likelihood function given by

$$\hat{\lambda}_i = \arg\max_{\lambda_i} \binom{\lambda_i}{n_i} \int_{\zeta_{min}}^{\zeta_{max}} p_i^{n_i} (1 - p_i)^{\lambda_i - n_i} g(p_i) dp_i,$$

where $g(p)$ is the Beta distribution for the country i with shape parameters $\alpha = \beta = 2$ (empirically fitted). To account for asymptomatic cases going undetected at the airports, we further assume that the 'true' number of exported cases, accounting for asymptotic or mildly symptomatic cases, is twice those reported – we note that this is a conservative approximation in that the true percentage of symptomatic cases is likely higher. We carried out the likelihood test for each country separately (see Appendix B) and, based on our analysis, we conclude that due to large passenger flow from Iran to Iraq and the UAE, numerous cases may have gone undetected. Therefore, we only include exported cases to Oman, Lebanon, and Kuwait to calculate the outbreak size on 25 February. We take a similar approach to estimate the outbreak size on 6 March based on 28 exported cases to China (see Fig. 5). We estimate the outbreak size on 25 February to be 13,700 (95% CI: 7,600 - 33,300) and on 6 March 60,500 (95% CI: 43,200 - 209,200), using a range of typical doubling times 2.5 - 4 days we can extrapolate the number of cases backwards to find the start of the epidemic in Iran. Assuming an 18 days delay from the onset of illness to death [Linton et al., 2020, Verity et al., 2020, Wu et al., 2020], we can also estimate the number of deaths in the country and compare it to reports from Iran International [2020], Radio Farda [2020a], and official numbers on those dates (see Fig. 6b).

Burden of the epidemic on the healthcare system

To study different scenarios for how the epidemic in Iran may unfold in the coming weeks and months and its burden on the healthcare system, we use the SEIR model developed by Neher et al. [2020]. We summarise all the model parameters and their corresponding values in Table 1 and Table 2. We assume that a few (1-2) infected individuals sparked

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070904; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

the COVID-19 outbreak in early January and the rest of the population consists of susceptible individuals who become exposed to SARS-CoV-2 after an effective contact with an infectious person. After an incubation period of 4 days, exposed individuals of age-group a will develop either a clinical infection with probability $1 - m_a$ and become hospitalised, or experience a subclinical infection with probability m_a where they recover or become severely ill after 5 days. Those hospitalised may recover or be moved to the ICU stage after 8 days where they may either die or stabilise after 10 days. We assume that removed individuals are immune to reinfection over the period we simulated the epidemic (1 year) – we still do not know how long-lasting immunity will be to COVID-19, but we used similar assumptions to other studies. Deaths are assumed to occur only among severe cases (in the ICU). The fractions of hospitalisation, ICU use, and death are estimated using data from China, and any differences in Iran's population could affect our estimates of health care demand. Our assessment of the expected number of imported cases from China in December and January from three airports in Beijing, Shanghai, Guangzhou, to Tehran suggests that the contribution from importation is likely going to be very small.

In our modelling analysis, we considered a range of scenarios of the epidemic from moderate to high levels of intervention efficacy according to set dates in policy announcements by the government (Table 4). Our finding suggests that the current policy measures are likely not going to be enough to push the effective reproduction number below one (Fig. 7a) and the outbreak is going to continue across the country in the coming months (Figures 7c, 7d, and 7e). It also suggests a very large surge in patients requiring intensive care from early April to late August (Fig. 7b) which is going to overwhelm the healthcare system given the 3,790 ICU beds currently available in the country [Phua et al., 2020]. Unless extreme intervention measures are in place to bring the epidemic under control, we project 221,000 cumulative deaths (IQR: 1.8×10^5 - 2.4×10^5) (see Fig. 7b). We also find there is likely going to be a second peak 40-80 days following the first peak in early April (Fig. 7e). However, the likely occurrence of the second peak is strictly dependent on the current and future intervention policies and can, as a result, vary considerably in magnitude and duration.

Discussion

In this study we provide a timeline of what has likely occurred since the start of the COVID-19 epidemic in Iran. Using the available genetic data, as on 15/04/2020, from infected cases with a travel link to Iran and the one full-genome sequence sample from inside the country, we estimated the doubling time associated with early stages of the COVID-19 outbreak in Iran to be around 3 days (95% HPD: 1.68-16.27) which is close to the estimates from other countries in Europe and Northern America. Our phylogenetics and epidemiological analysis places the likely start of transmissions back to 21/01/2020 (95% HPD: 05/12/2019-14/02/2020). This suggests the virus had been circulating for,

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070904; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

at least, a month prior to first official reports in Iran. We further estimate that, based on travel data, by 25 February there were approximately 13,700 (95% CI: 7,600 - 33,300) and by 6 March 60,500 (43,200 - 209,200) cases in Iran, which is about a hundred times higher than the official numbers at that time. The large discrepancy is partly due to late detection as well as limitations in testing capacity particularly during the early stages of the epidemic. Many of the symptomatic cases have not had access to testing and most tests were done only on clinical patients in hospitals and intensive care units and are processed in a few centralised labs. Therefore, potentially a large number of mild and subclinical cases in Iran are missed. By comparing the nationwide age-stratified nCFR with those from hospitals in Tehran, we see that those two are almost identical for <50 years-old individuals indicating significant levels of under-reporting of cases in those age-groups. We note that the nCFR has an internal bias when calculated during the exponential growth of the epidemic [Baud et al., Spychalski et al., Verity et al., 2020] as the number of deaths lag behind confirmed cases by 1-2 weeks. However, correcting for this bias likely increases the CFR estimates in Iran suggesting that the under-reporting is even higher when compared to CFR estimates from China.

One of the first major interventions implemented by the government was the announcement of indefinite school and university closures on 5 March which was followed by a 14-day period of national New Year's holiday where most businesses were closed. However, several reports indicate that nearly 8.5 million citizens travelled during this period [BBC Persian, 2020c] which suggests that lack of enforced travel restrictions and social distancing measures allowed the transmission to go almost unchecked in various parts of the country. In addition, on 5 April, the government announced that workers may return to work and there will be no lockdown measures in the country. The quick return to normal life while the epidemic seems to be slowing down may result in a loss of control of the epidemic spread and likely result in SARS-CoV-2 becoming endemic in Iran and potentially creating subsequent waves of the epidemic in the country in autumn or winter which can also become a significant burden on the healthcare system of the country. This can potentially be a major issue for many countries particularly if they are attempting to control the outbreak and have to prevent the resurgence of new cases in their countries [Kraemer et al., 2020]. Despite the fact that several European airlines have cancelled their flights to Iran indefinitely, passengers can still travel in and out of Iran through intermediary countries like Qatar and UAE.

So far, despite the local spread from Iran to other countries in the region such as Pakistan, Afghanistan, Oman, Kuwait, Lebanon, and Bahrain there has been no strong indication of a major outbreak in those countries. We cannot rule out under-reporting in those countries, but it has also been suggested that temperature and humidity could play a role in sustaining an epidemic in a country, although the effect of these is debated and may be small [Ficetola and Rubolini, 2020, Notari, 2020, Sajadi et al., 2020] Any effects of climate are likely to be most relevant to particular outbreaks in specific locations and are unlikely to be relevant to the overall pandemic outbreak given a susceptible population

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070904; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

[Baker et al., 2020]. These nearby countries may have benefited in the early stages of the pandemic due to their climate, whereas Iran has more ideal conditions for COVID-19 spread. Nevertheless, significant cases are expected in the coming months in warmer countries and cases originating in Iran could further exacerbate the spread of the pandemic within the broader region. Thus, Iran could become a regional reservoir and contribute to a second wave of epidemics in the Middle East.

Supporting materials

We enclosed a supplementary file to this paper.

Acknowledgement

The authors would like to thank Faramarz Faridi and Amirali Abbaspourmani for providing us with valuable information from various news organisations regarding current developments and policy decisions relating to the COVID-19 outbreak in Iran. We also thank José Lourenço for helpful comments on the manuscript. MG and SD are funded by the Biotechnology and Biological Science Research Council (BBSRC), grant number BB/M011224/1.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070904; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Table 1: Model parameters

Parameter	Description	Expected value	Reference
t_l	Latency from infection to infectiousness	4 days	[Bi et al., 2020],[Nishiura et al., 2020]
t_i	Time to recover/falling severely ill	5 days	[Davies et al., 2020]
t_h	Time to recover/move to a critical state for a severe patient	8 days	[NHS, 2018-19]
t_c	Time to stabilise/die for a critical patient	10 days	[Cao et al., 2020]
m_a	Fraction of sub-clinical cases*	Table 2	[Imperial College COVID-19 Response Team, 2020b]
c_a	Fraction of severe cases that turn critical*	Table 2	[Imperial College COVID-19 Response Team, 2020b]
f_a	Fraction of critical cases that are fatal*	Table 2	—
R_0	Reproductive number	3.5 - 6.0	[Imperial College COVID-19 Response Team, 2020a]
$M(t)$	Mitigation measures at time t	Table 4	—

Table 2: Age-specific clinical information

Age-group (years)	Age group population size*	Percentage of subclinical cases (m_a)	Percentage requiring critical care (c_a)	Percentage dying while in critical care (f_a)
≤ 9	11,890,343	99.90%	5%	30%
10-19	12,278,478	99.70%	5%	30%
20-29	17,087,151	98.80%	5%	30%
30-39	12,542,942	96.80%	5%	30%
40-49	8,937,230	95.10%	6.30%	40%
50-59	6,207,527	89.80%	12.20%	40%
60-69	3,206,638	83.40%	27.40%	50%
70-79	2,033,499	75.70%	43.20%	50%
≥ 80	919,539	72.70%	70.90%	50%

Table 3: Weekly passenger flux to countries included in the study

Date	Country	Passenger/week	Cases Reported
24 Feb	Oman	2660	2
22 Feb	UAE	13430	3
25 Feb	Kuwait	4025	3
24 Feb	Iraq	16254	1
21 Feb	Lebanon	800	1
6 March	China	4500	28

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070904; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Table 4: Timeline of public policies and mitigation strategies impacting the COVID-19 outbreak in Iran

Policies and main events	Description	Date effective
Social distancing encouraged	After the first few reported deaths, citizens are encouraged to avoid making unnecessary journeys and work from home. We assume this had an overall 20% impact on $M(t)$.	20/02/2020 [BBC Persian, 2020a]
Elderly encouraged to stay home	The elderly and people with underlying conditions are recommended to reduce their social contacts. With a population of 6,159,676, we assume this had an overall 20% impact on ζ_a for age groups > 60.	05/03/2020 [BBC Persian, 2020a]
School and university closure ordered	Nationwide school and university closure until further notice from the government. With a population of 41,255,972 aged under 30, we assume up to 50% effectiveness on ζ_a.	05/03/2020 [BBC Persian, 2020b]
New year's social and public events banned	During this period, social distancing was not strictly followed by citizens with reports of over 8.5 million travelling around the country during the 14-day holiday period. Nevertheless, since all shops and public gatherings were completely shut down, we assumed this had an extra 30% impact on mitigation measures $M(t)$.	20/03/2020 [BBC Persian, 2020c]
Majority of workers back to work ('smart social distancing')	The government announced that the majority of workers must return to work while they must observe some level of social distancing. We have assumed that the re-opening results in up to 20% drop in effectiveness of mitigation measures $M(t)$.	05/04/2020 [BBC Persian, 2020d]

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070904; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

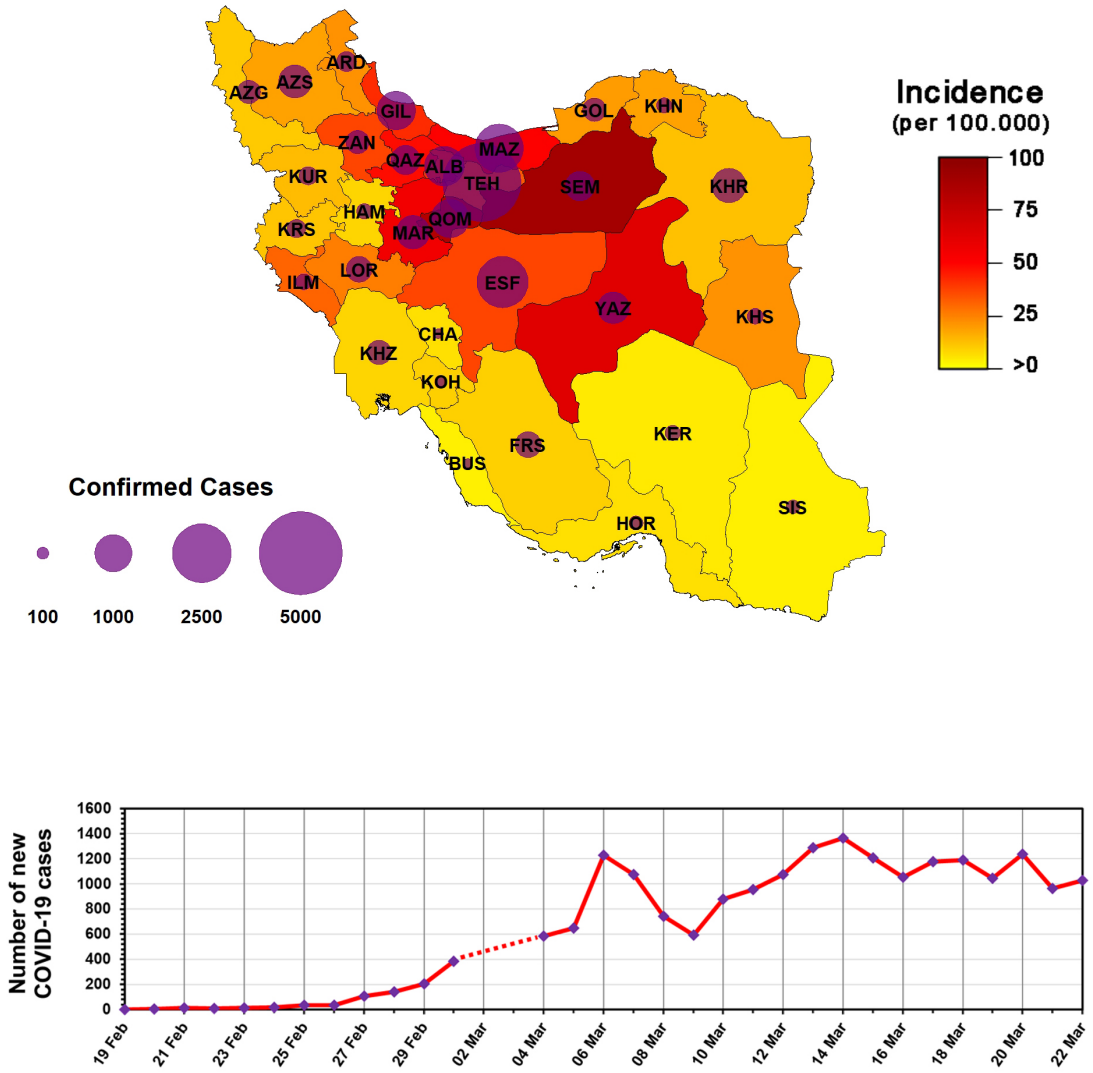

Figure 1: The COVID-19 Epidemic in Iran. (Top) Total incidence and confirmed cases by region, as of 22 March 2020. (Bottom) The early spread of the Iranian COVID-19 epidemic based on official reports. *Dotted line illustrates when no cases were reported officially and is for visual purposes only

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070904; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

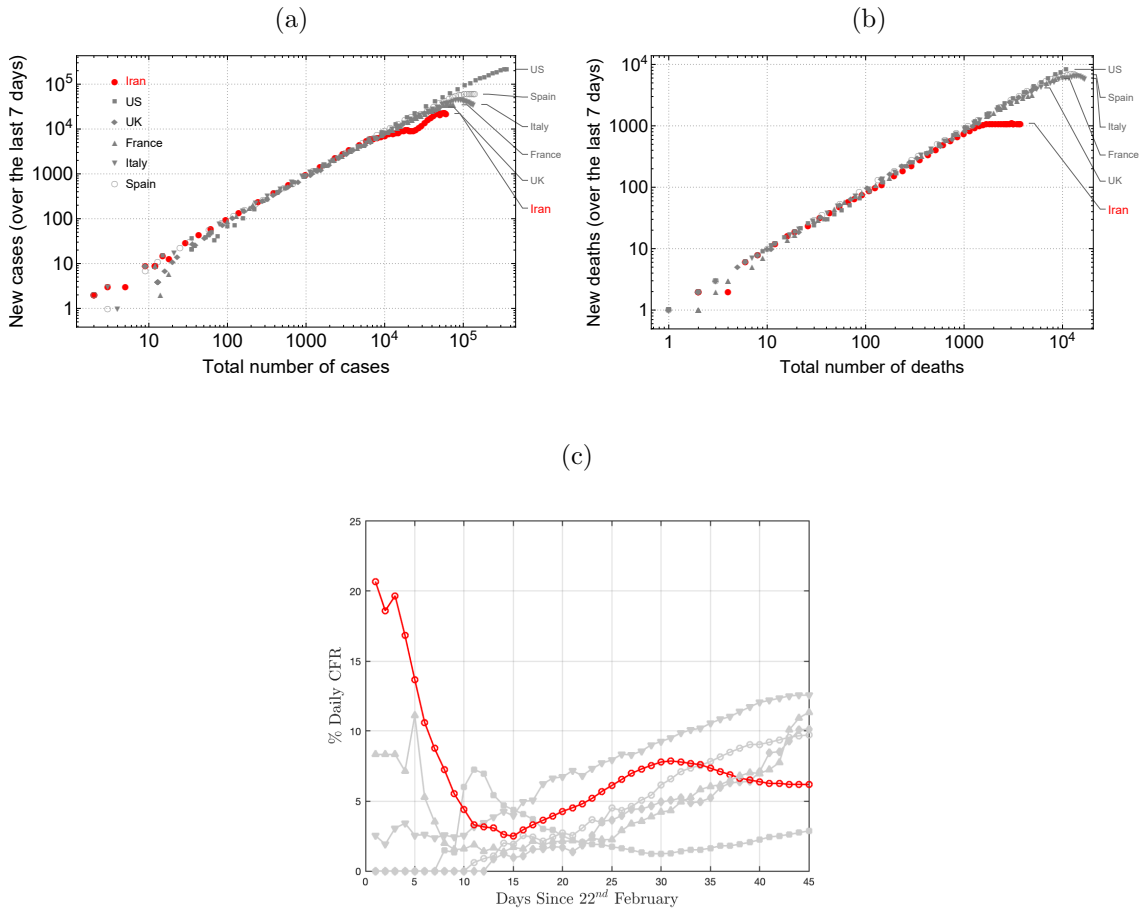

Figure 2: Reported number of cases and deaths from several countries including Iran (red). (a), (b) The sum of new cases/deaths over the last seven days since the announcement of reported numbers on that day as a function of the total number of cases/deaths to date. (c) The daily Case Fatality Ratio for all six countries since 22 February.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070904; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

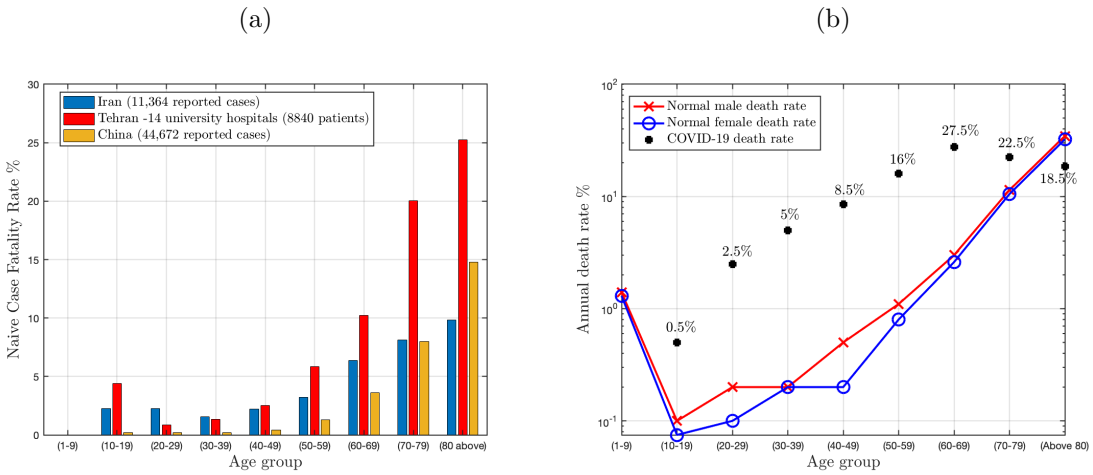

Figure 3: (a) Comparison of the age specific naïve Case Fatality Ratio reported by Iran's Ministry of Health, Tehran University of Medical Sciences hospitals, and CDC China. (b) Probability of death due to COVID-19 compared to the annual death rate in males and females based on their age in Iran [WHO DEFINITION].

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070904; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

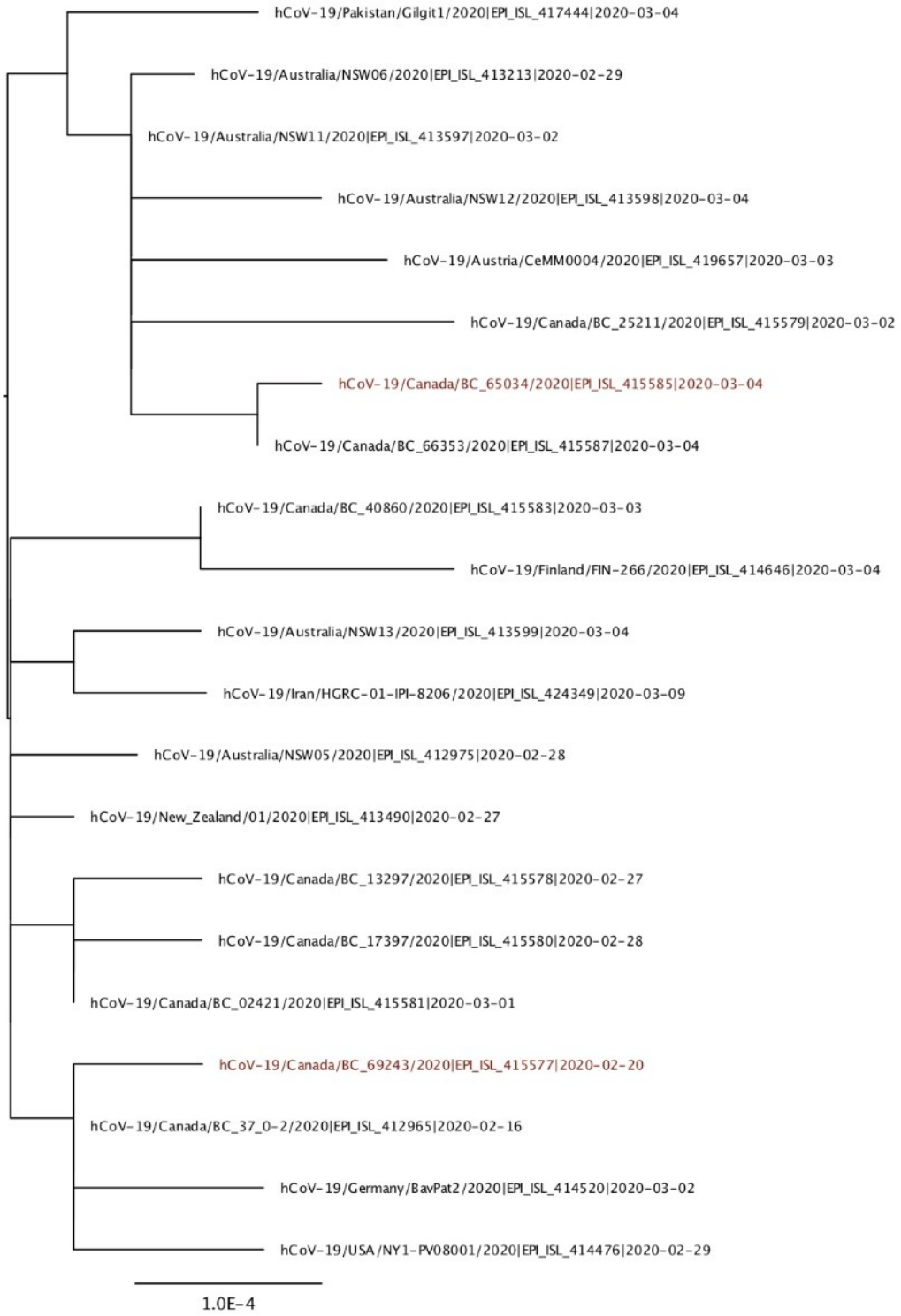

Figure 4: Maximum likelihood phylogenetic tree of samples linked to Iran. The labels include the sampling times used in the BEAST analysis. Red labels indicate the two epidemiologically linked samples that were excluded from the subsequent analysis.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070904; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

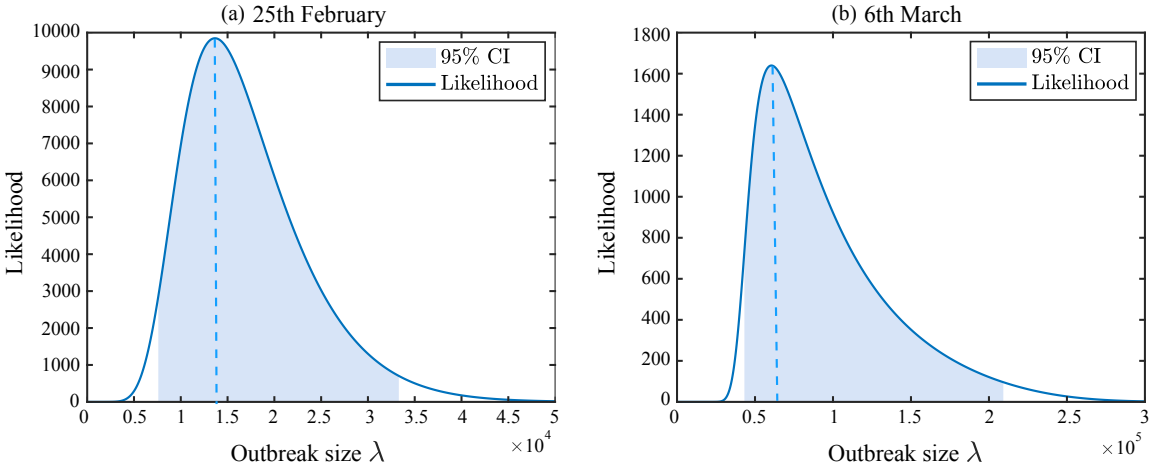

Figure 5: Likelihood function for (a) 6 exported cases to Oman, Lebanon, and Kuwait on 25 February and (b) 28 cases to China on 6 March. Red dashed line represents the maximum likelihood value and the gray area is the 95% confidence interval.

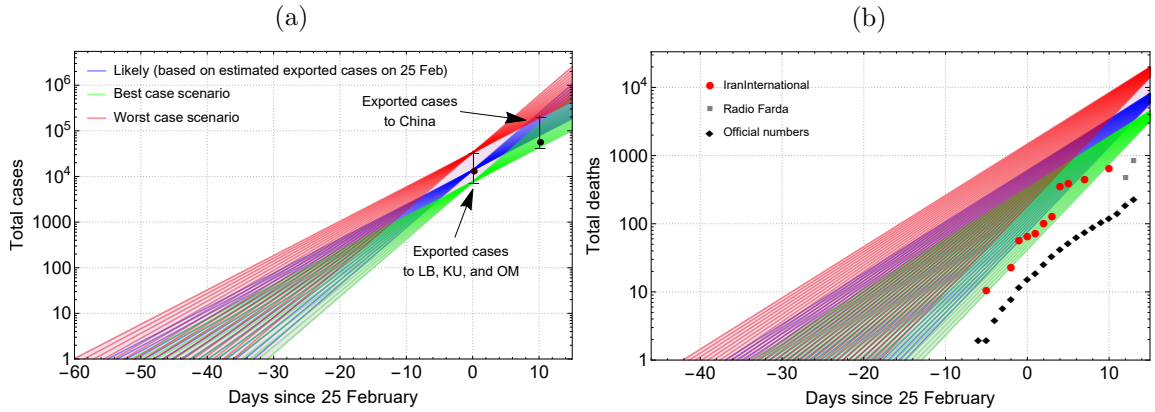

Figure 6: The likely (blue), worst case (red), and best case (green) scenario of the outbreak size in Iran (left) and total deaths since early January to 8 March. *Full list of all reported cases and deaths from Iran International, Radio Farda, and other news organisations can be found in the Supplementary material.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070904; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

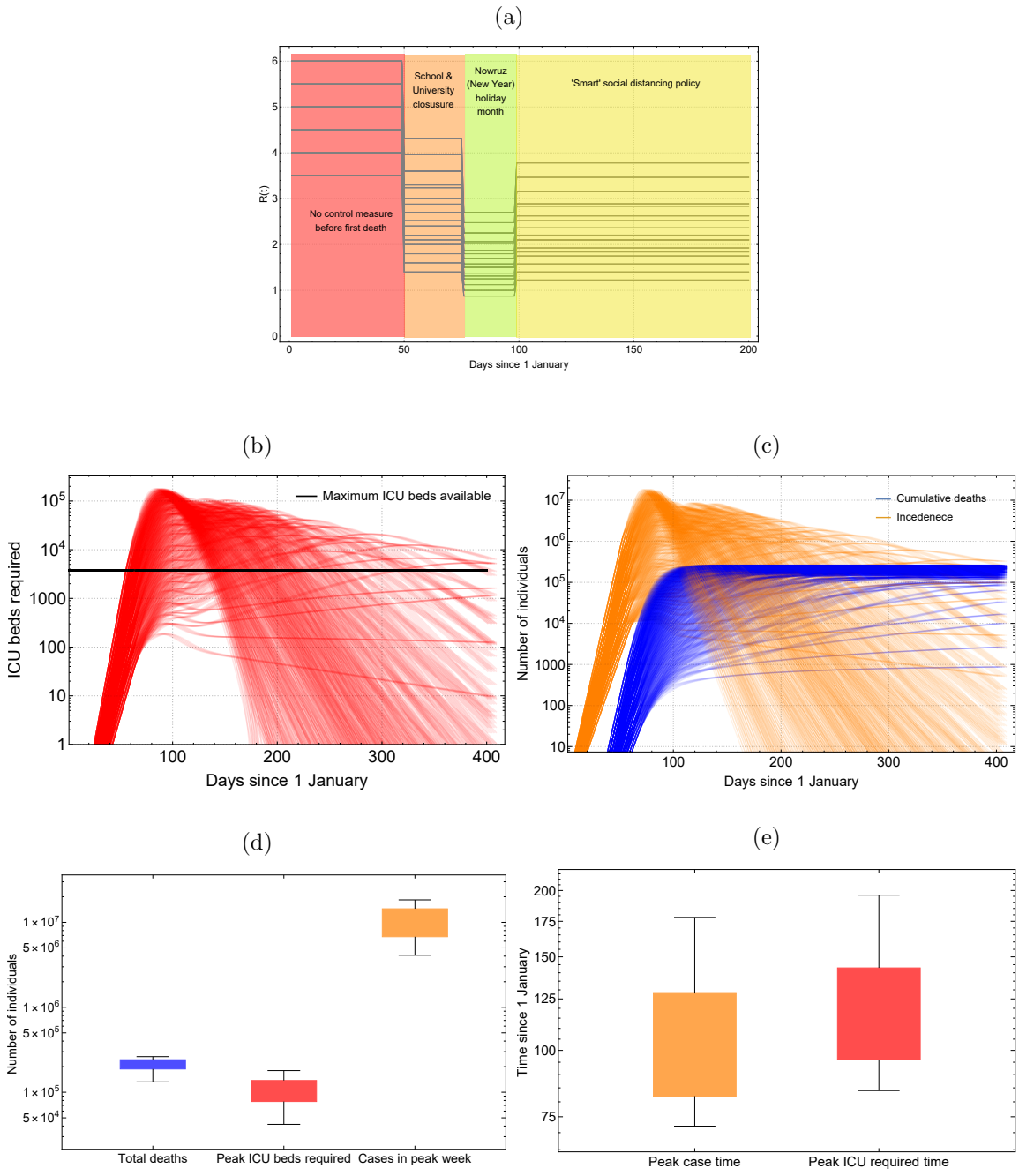

Figure 7: Projection of the impact of intervention measures on (a) effective reproduction number, $R(t)$, which we defined as the product of reproduction number, R0, strength of mitigation measures, $M(t)$, and isolation of individuals in specific age-categories, ζ_a, (b) number of ICU beds required (c) cumulative deaths and incidence over time. (d) and (e) show the summary of the analysis in total deaths, peak number of cases and ICU beds required, and the peak time since the seeding event for cases and ICU beds required times.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070904; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

References

Y. Bai, L. Yao, T. Wei, F. Tian, D.-Y. Jin, L. Chen, and M. Wang. Presumed asymptomatic carrier transmission of COVID-19. *JAMA*, 323(14):1406–1407, 2020.

R. E. Baker, W. Yang, G. A. Vecchi, C. J. E. Metcalf, and B. T. Grenfell. Susceptible supply limits the role of climate in the covid-19 pandemic. *medRxiv*, 2020.

Z. Basravi. Iran to reopen businesses as COVID-19 lockdown eased. 2020. `https://www.aljazeera.com/news/2020/04/iran-reopen-businesses-covid-19-lockdown-eased-200417132846820.html`.

D. Baud, X. Qi, K. Nielsen-Saines, D. Musso, L. Pomar, and G. Favre. Real estimates of mortality following COVID-19 infection. *The Lancet Infectious Diseases*.

BBC Persian. People are asked not to leave their houses. 2020a. `https://www.bbc.com/persian/iran-51969772`.

BBC Persian. School closure. 2020b. `https://www.bbc.com/persian/iran-51761418`.

BBC Persian. 8.5 millions have gone on trips amid COVID-19 outbreak. 2020c. `https://www.bbc.com/persian/iran-51942853`.

BBC Persian. Majority of people return to work. 2020d. `https://www.bbc.com/persian/business-52219694`.

BBC Persian. At least 210 people have died with COVID-19. 2020e. `https://www.bbc.com/persian/iran-51676966`.

S. M. Bell, N. Müller, E. Hodcroft, C. W. adn James Hadfielda, M. Ilcisin, R. Neher, and T. Bedford. Genomic analysis of COVID-19 spread. Situation report 2020-03-27. *Nextstrain*, 2020. `https://nextstrain.org/narratives/ncov/sit-rep/2020-03-27?f_division_exposure=Iran&n=14`.

F. Benford. The law of anomalous numbers. *Proceedings of the American Philosophical Society*, 78(4):551–572, 1938.

Q. Bi, Y. Wu, S. Mei, C. Ye, X. Zou, Z. Zhang, X. Liu, L. Wei, S. A. Truelove, T. Zhang, W. Gao, C. Cheng, X. Tang, X. Wu, Y. Wu, B. Sun, S. Huang, Y. Sun, J. Zhang, T. Ma, J. Lessler, and T. Feng. Epidemiology and transmission of COVID-19 in Shenzhen China: analysis of 391 cases and 1,286 of their close contacts. *medRxiv*, 2020.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070904; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

M. F. Boni, P. Lemey, X. Jiang, T. T.-Y. Lam, B. Perry, T. Castoe, A. Rambaut, and D. L. Robertson. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *bioRxiv*, 2020.

R. J. C. Brown. Benford's law and the screening of analytical data: the case of pollutant concentrations in ambient air. *The Analyst*, 130(9):1280—1285, 2005. doi: 10.1039/b504462f.

B. Cao, Y. Wang, D. Wen, W. Liu, J. Wang, G. Fan, L. Ruan, B. Song, Y. Cai, M. Wei, X. Li, J. Xia, N. Chen, J. Xiang, T. Yu, T. Bai, X. Xie, L. Zhang, C. Li, Y. Yuan, H. Chen, H. Li, H. Huang, S. Tu, F. Gong, Y. Liu, Y. Wei, C. Dong, F. Zhou, X. Gu, J. Xu, Z. Liu, Y. Zhang, H. Li, L. Shang, K. Wang, K. Li, X. Zhou, X. Dong, Z. Qu, S. Lu, X. Hu, S. Ruan, S. Luo, J. Wu, L. Peng, F. Cheng, L. Pan, J. Zou, C. Jia, J. Wang, X. Liu, S. Wang, X. Wu, Q. Ge, J. He, H. Zhan, F. Qiu, L. Guo, C. Huang, T. Jaki, F. G. Hayden, P. W. Horby, D. Zhang, and C. Wang. A trial of lopinavir–ritonavir in adults hospitalized with severe COVID-19. *New England Journal of Medicine*, 2020.

E. Crocetti and G. Randi. Using the Benford's Law as a First Step to Assess the Quality of the Cancer Registry Data. *Frontiers in Public Health*, 4:225, 2016. ISSN 2296-2565.

N. G. Davies, A. J. Kucharski, R. M. Eggo, A. Gimma, , and W. J. Edmunds. The effect of non-pharmaceutical interventions on COVID-19 cases, deaths and demand for hospital services in the UK: a modelling study. *medRxiv*, 2020.

C. Durtschi, W. A. Hillison, and C. Pacini. The effective use of benford's law to assist in detecting fraud in accounting data. 2004.

J.-S. Eden, R. Rockett, I. Carter, H. Rahman, J. de Ligt, J. Hadfield, M. Storey, X. Ren, R. Tulloch, K. Basile, J. Wells, R. Byun, N. Gilroy, M. V. O'Sullivan, V. Sintchenko, S. C. Chen, S. Maddocks, T. C. Sorrell, E. C. Holmes, D. E. Dwyer, J. Kok, and for the 2019-nCoV Study Group. An emergent clade of SARS-CoV-2 linked to returned travellers from Iran. *Virus Evolution*, 6(1), 2020.

Encyclopedia of Nations. Average length of stay of visitors - Tourism indicators - UNCTAD Handbook of Statistics - Country Comparison. *UNCTAD Handbook of Statistics*, 2003. `https://www.nationsencyclopedia.com/WorldStats/UNCTAD-average-length-stay-visitors.html`.

L. Ferretti, C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner, M. Parker, D. Bonsall, and C. Fraser. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science*, 2020.

G. F. Ficetola and D. Rubolini. Climate affects global patterns of COVID-19 early outbreak dynamics. *medRxiv*, 2020.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070904; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

C. Fraser, C. A. Donnelly, S. Cauchemez, W. P. Hanage, M. D. Van Kerkhove, T. D. Hollingsworth, J. Griffin, R. F. Baggaley, H. E. Jenkins, E. J. Lyons, T. Jombart, W. R. Hinsley, N. C. Grassly, F. Balloux, A. C. Ghani, N. M. Ferguson, A. Rambaut, O. G. Pybus, H. Lopez-Gatell, C. M. Alpuche-Aranda, I. B. Chapela, E. P. Zavala, D. M. E. Guevara, F. Checchi, E. Garcia, S. Hugonnet, and C. Roth. Pandemic potential of a strain of influenza A (H1N1): early findings. *Science*, 324(5934):1557–1561, 2009.

W. Goodman. The promises and pitfalls of benford's law. *Significance*, 13(3):38–41, 2016.

S. Guindon, J. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, and O. Gascuel. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*, 59(3), 2010.

E. Hodcroft, N. Müller, C. Wagner, M. Ilcisin, J. Hadfield, S. M. Bell, R. Neher, and T. Bedford. Genomic analysis of COVID-19 spread. Situation report 2020-03-13. *Nextstrain*, 2020.
`https://nextstrain.org/narratives/ncov/sit-rep/2020-03-13?n=6`.

A. J. Idrovo, J. A. Fernandez-Nino, I. Bojorquez-Chapela, and J. Moreno-Montoya. Performance of public health surveillance systems during the influenza A(H1N1) pandemic in the Americas: testing a new method based on Benford's Law. *Epidemiology and Infection*, 139(12):1827–1834, 2011.

N. Imai, I. Dorigatti, A. Cori, C. Donnelly, S. Riley, and N. M. Ferguson. Estimating the potential total number of novel Coronavirus cases in Wuhan City, China. *Imperial College London COVID-19 Response Team*, 2020.

Imperial College COVID-19 Response Team. Estimating the number of infections and the impact of non- pharmaceutical interventions on COVID-19 in 11 European countries. 2020a.

Imperial College COVID-19 Response Team. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID- 19 mortality and healthcare demand. 2020b.

Iran International. Daily COVID-19 reports. 2020.
`https://iranintl.com/en`.

M. U. G. Kraemer, C.-H. Yang, B. Gutierrez, C.-H. Wu, B. Klein, D. M. Pigott, , L. du Plessis, N. R. Faria, R. Li, W. P. Hanage, J. S. Brownstein, M. Layan, A. Vespignani, H. Tian, C. Dye, O. G. Pybus, and S. V. Scarpino. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*, 2020.

Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. Leung, E. H. Lau, J. Y. Wong, X. Xing, N. Xiang, Y. Wu, C. Li, Q. Chen, D. Li, T. Liu, J. Zhao, M. Liu, W. Tu,

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070904; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

C. Chen, L. Jin, R. Yang, Q. Wang, S. Zhou, R. Wang, H. Liu, Y. Luo, Y. Liu, G. Shao, H. Li, Z. Tao, Y. Yang, Z. Deng, B. Liu, Z. Ma, Y. Zhang, G. Shi, T. T. Lam, J. T. Wu, G. F. Gao, B. J. Cowling, B. Yang, G. M. Leung, and Z. Feng. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *New England Journal of Medicine*, 382(13):1199–1207, 2020a.

R. Li, S. Pei, B. Chen, Y. Song, T. Zhang, W. Yang, and J. Shaman. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science*, 2020b.

N. M. Linton, T. Kobayashi, Y. Yang, K. Hayashi, A. R. Akhmetzhanov, S.-m. Jung, B. Yuan, R. Kinoshita, and H. Nishiura. Incubation Period and Other Epidemiological Characteristics of 2019 Novel Coronavirus Infections with Right Truncation: A Statistical Analysis of Publicly Available Case Data. *Journal of Clinical Medicine*, 9:538, Feb 2020.

J. Lourenço, R. Paton, M. Ghafari, M. Kraemer, C. Thompson, P. Simmonds, P. Klenerman, and S. Gupta. Fundamental principles of epidemic spread highlight the immediate need for large-scale serological surveys to assess the stage of the sars-cov-2 epidemic. *medRxiv*, 2020.

J. Lu, L. d. Plessis, Z. Liu, V. Hill, M. Kang, H. Lin, J. Sun, S. Francois, M. U. G. Kraemer, N. R. Faria, J. T. McCrone, J. Peng, Q. Xiong, R. Yuan, L. Zeng, P. Zhou, C. Liang, L. Yi, J. Liu, J. Xiao, J. Hu, T. Liu, W. Ma, W. Li, J. Su, H. Zheng, B. Peng, S. Fang, W. Su, K. Li, R. Sun, R. Bai, X. Tang, M. Liang, J. Quick, T. Song, A. Rambaut, N. Loman, J. Raghwani, O. Pybus, and C. Ke. genomic epidemiology of sars-cov-2 in guangdong province, china.

R. Neher, I. Aksamentov, and N. Noll. About COVID-19 Scenarios. 2020. `https://covid19-scenarios.org/about`.

NHS. Hospital Admitted Patient Care Activity 2018-19. 2018-19.

H. Nishiura, N. M. Linton, and A. R. Akhmetzhanov. Serial interval of novel coronavirus (COVID-19) infections. *International Journal of Infectious Diseases*, 93:284–286, 2020/04/15 2020.

A. Notari. Temperature dependence of COVID-19 transmission. *medRxiv*, 2020.

J. Phua, M. O. Faruq, A. P. Kulkarni, I. S. Redjeki, K. Detleuxay, N. Mendsaikhan, K. K. Sann, B. R. Shrestha, M. Hashmi, J. E. M. Palo, R. Haniffa, C. Wang, S. M. R. Hashemian, A. Konkayev, M. B. Mat Nor, B. Patjanasoontorn, K. M. K. Nafees, L. Ling, M. Nishimura, M. J. Al Bahrani, Y. M. Arabi, C.-M. Lim, W.-F. Fang, for the Asian Analysis of Bed Capacity in Critical Care (ABC) Study Investigators, and the Asian

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070904; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Critical Care Clinical Trials Group. Critical care bed capacity in asian countries and regions. *Critical Care Medicine Society of Critical Care Medicine*, 48(5), 2020.

Radio Farda. Daily covid-19 reports. 2020a.
https://en.radiofarda.com.

Radio Farda. WHO chief says 97 cases of coronavirus in 11 countries originated from Iran. 2020b.
https://en.radiofarda.com/a/.

A. Rambaut. Phylogenetic analysis of nCoV-2019 genomes. 2020.
http://virological.org/t/phylodynamic-analysis-176-genomes-6-mar-2020/356
phylodynamic-analysis-176-genomes-6-mar-2020/356.

A. Rambaut, E. C. Holmes, and O. G. Pybus. A dynamic nomenclature for SARS-CoV-2 to assist genomic epidemiology. 2020.
http://virological.org/t/
a-dynamic-nomenclature-for-sars-cov-2-to-assist-genomic-epidemiology/458.

Reality Check team. Coronavirus: Iran is facing a major challenge controlling the outbreak. 2020.
https://www.bbc.co.uk/news/world-middle-east-51642926.

Robin Wright. How Iran Became a New Epicenter of the Coronavirus Outbreak. *The New Yorker*, 2020.
https://www.newyorker.com/news/our-columnists/
how-iran-became-a-new-epicenter-of-the-coronavirus-outbreak.

T. W. Russell, J. Hellewell, C. I. Jarvis, K. van Zandvoort, S. Abbott, R. Ratnayake, C. C.-. working group, S. Flasche, R. M. Eggo, W. J. Edmunds, and A. J. Kucharski. Estimating the infection and case fatality ratio for coronavirus disease (COVID-19) using age-adjusted data from the outbreak on the Diamond Princess cruise ship, February 2020. *Eurosurveillance*, 25(12):2000256, 2020.

M. M. Sajadi, P. Habibzadeh, A. Vintzileos, S. Shokouhi, F. Miralles-Wilhelm, and A. Amoroso. Temperature, Humidity and Latitude Analysis to Predict Potential Spread and Seasonality for COVID-19. 2020.

P. Spychalski, A. Błażyńska-Spychalska, and J. Kobiela. Estimating case fatality rates of COVID-19. *The Lancet Infectious Diseases*.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070904; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

M. A. Suchard, P. Lemey, G. Baele, D. L. Ayres, A. J. Drummond, and A. Rambaut. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution*, 4(1), 2018.

Tehran university school of medicine COVID-19 response team. Epidemiological report of COVID-19 in hospitals affiliated with Tehran university school of medicine. *Public Access Directory*, 2020.

The Islamic Republic News Agency(IRNA). COVID-19 fatalities age distribution. 2020. `www.irna.ir/news/83712959/`.

The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. The Epidemiological Characteristics of an Outbreak of 2019 Novel Coronavirus Diseases (COVID-19) — China, 2020. *China CDC Weekly*, 2:113, 2020.

N. Triggle. Coronavirus: How to understand the death toll. 2020. `https://www.bbc.com/news/health-51979654`.

A. R. Tuite, I. I. Bogoch, R. Sherbo, A. Watts, D. Fisman, and K. Khan. Estimation of coronavirus disease 2019 (COVID-19) burden and potential for international dissemination of infection from Iran. *Annals of Internal Medicine*, 2020.

VariFlight. Flight database, 2020.

R. Verity, L. C. Okell, I. Dorigatti, P. Winskill, C. Whittaker, N. Imai, G. Cuomo-Dannenburg, H. Thompson, P. G. T. Walker, H. Fu, A. Dighe, J. T. Griffin, M. Baguelin, S. Bhatia, A. Boonyasiri, A. Cori, Z. Cucunubá, R. FitzJohn, K. Gaythorpe, W. Green, A. Hamlet, W. Hinsley, D. Laydon, G. Nedjati-Gilani, S. Riley, S. van Elsland, E. Volz, H. Wang, Y. Wang, X. Xi, C. A. Donnelly, A. C. Ghani, and N. M. Ferguson. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet Infectious Diseases*, 2020.

E. Volz, M. Baguelin, S. Bhatia, A. Boonyasiri, A. Cori, Z. Cucunuba Perez, G. Cuomo-Dannenburg, C. Donnelly, I. Dorigatti, R. Fitzjohn, H. Fu, K. Gaythorpe, A. Ghani, A. Hamlet, W. Hinsley, N. Imai, D. Laydon, G. Nedjati Gilani, L. Okell, S. Riley, S. Van Elsland, H. Wang, Y. Wang, X. Xi, and N. Ferguson. Phylogenetic analysis of SARS-CoV-2. 2020.

P. Wintour. Iran's president has left nation open to second Covid-19 wave. 2020. `https://www.theguardian.com/world/2020/apr/13/irans-president-has-left-nation-open-to-second-covid-19-wave-critics`.

World Tourism Organisation. Iran Islamic Republic Of Country-specific Basic indicators (Compendium) 2014 - 2018 (12.2019). *Tourism Statistics*, 2020.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070904; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

J. T. Wu, K. Leung, M. Bushman, N. Kishore, R. Niehus, P. M. de Salazar, B. J. Cowling, M. Lipsitch, and G. M. Leung. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nature Medicine*, 26(4):506–510, 2020.

Z. Zhuang, S. Zhao, Q. Lin, P. Cao, Y. Lou, L. Yang, and D. He. Preliminary estimation of the novel coronavirus disease (COVID-19) cases in Iran: A modelling analysis based on overseas cases and air travel data. *International Journal of Infectious Diseases*, 2020.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070904; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Appendix A: Investigating manipulation of reported numbers using Benford's law

Benford's law (BL) gives the probability distribution of leading digits in a determined set of numbers [Benford, 1938]. The probability that digit $d = 1, 2, ..., 9$ is a leading number is given by $P(d) = log10(1 + (1/d))$ where numbers with leading digit 1 have the highest probability of appearance and this probability steadily decreases as the starting digit becomes larger. Benford's law is used in a variety of different areas to study irregularities in data [Brown, 2005, Durtschi et al., 2004] and is also frequently used in medicine to assess the quality of data [Crocetti and Randi, 2016, Idrovo et al., 2011].

Figure 8a shows the distribution of leading digits in data acquired from Iran, the USA, and the UK compared to the Benford distribution. The observed data in Iran is the cases reported by each province from 18/02/2020 to 20/03/2020. We find numbers that are larger than 9 in the data to take out any bias that may come into the distribution from single digit numbers. We find a total of 397 numbers larger than 9 from which we then sample 300 random numbers 8 times to generate an average distribution with error bars representing the standard deviation. Similarly, we find 19494 numbers larger than 9 from the data of US states from 22/01/2020 to 11/04/2020 and 619 numbers from the data of English provinces from 09/03/2020 to 11/04/2020. We sample the numbers obtained from the US and the UK in a similar way. We then normalise the frequency distribution by the sample size of 300 to obtain the probability distribution shown in Fig. 8a. The probability distribution does not show any conclusive evidence to suggest a manipulation of data in any of the three countries. The low or inaccurate number of reported cases in Iran are therefore likely due to limitations in testing and failure to detect infected individuals. While this method can be used to test if data manipulation has occurred it does not give any information about the deliberate absence of data, for example not reporting deaths from specific hospitals [Goodman, 2016].

Figure 8b shows the distribution of leading numbers for the total daily reported deaths across the whole country for Iran, the US, and the UK until 11/04/2020. There are a total of 42 reported numbers larger than 9 in Iran, 28 in the US, and 29 in the UK. We randomly sample a section of numbers larger than 9 equal to the smallest available data which is 28 numbers larger than 9 from the US. The frequency distribution of the leading number is then normalised by 28 to obtain a probability distribution. There are no error bars since the available data is small and sampling was only done once. The daily reported number of deaths in the entire country is perhaps not a reliable statistic to study for comparison to Benford's law since the available data is too small. But since Iran does not report deaths for individual provinces and cities, the total number of daily deaths in the country is the only statistic available of deaths due to COVID-19 form Iran and the only means of comparison to data from other countries. We still observe a significant distribution of numbers leading with 1 which can be an explanation for Fig. 2c and why Iran has a fatality

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070904; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

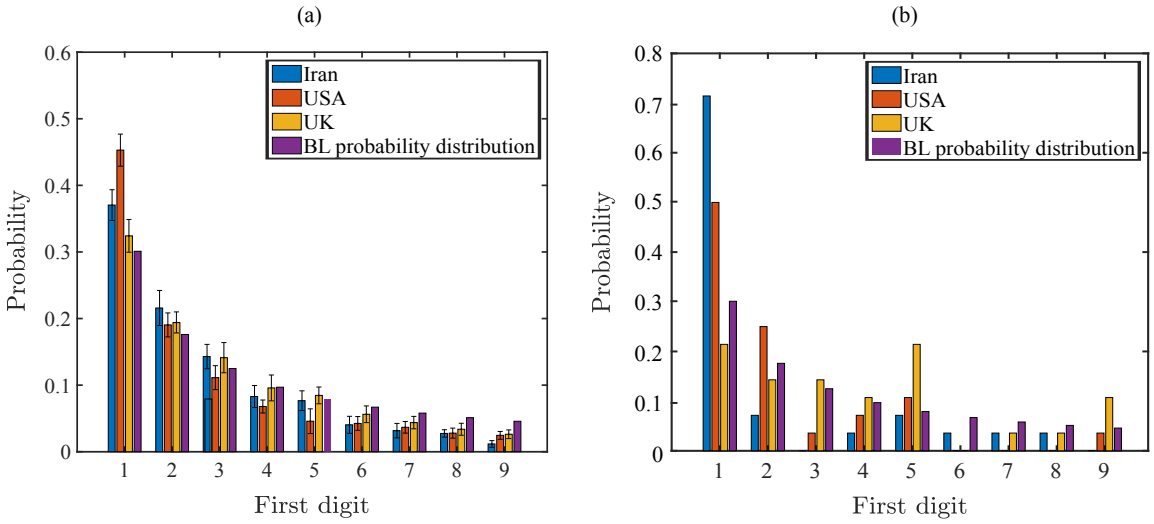

Figure 8: The probability distribution of leading digits (a) in daily reported cases of infected individuals and (b) in daily reported deaths due to COVID-19. The data from provinces in Iran, states in the US, and counties in the UK are compared to the distribution predicted by Benford's law.

rate that is almost constant from 40 days after Feb 22. This can mean that the constant death rate is due to reported deaths that all lead with the number 1 which may be an indication of the under-reporting of deaths due to COVID-19 since we would expect the death rate to gradually increase from what we observe in other countries.

Appendix B: Air travel analysis

medRxiv preprint doi: https://doi.org/10.1101/2020.04.18.20070904; this version posted April 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

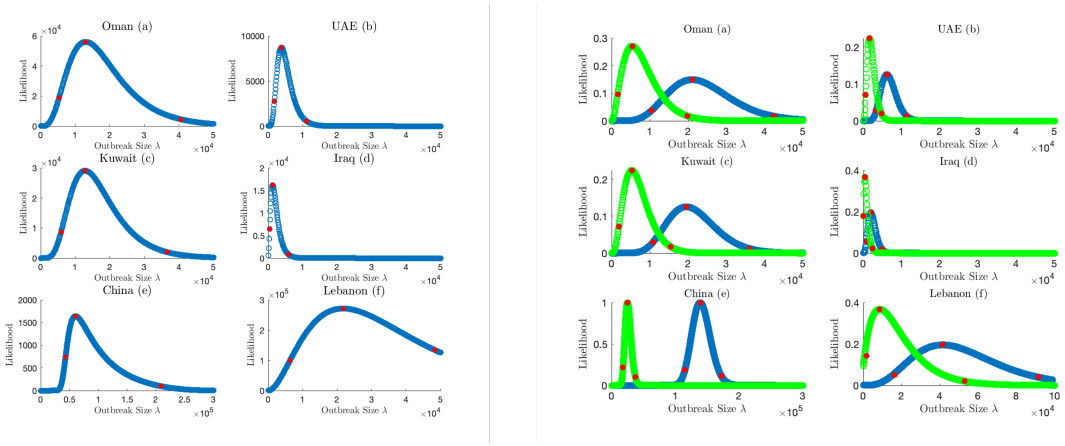

Figure 9: Likelihood estimation of outbreak size in Iran based on travel data from Oman, UAE, Kuwait, Iraq, and Lebanon on 25 February and China on 6 March. (Left) Beta-binomial likelihood analysis of outbreak size for each country assuming 50% detection. (Right) Binomial likelihood analysis for each country assuming 100% detection (green) and 50% detection (blue). Red dots represent the maximum likelihood and 95% confidence interval.