Ongoing outbreak of COVID-19 in Iran: challenges and signs of concern with under-reporting of prevalence and deaths

M. Ghafari[1*], B. Hejazi[2], A. Karshenas[3], S. Dascalu[1,4], L. Ferretti[5], A. Ledda[6], M. A. Khosravi[7], M. Abbasalipour[7], S. Zeinali[7,8], 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. [7] Kawsar Human Genetic Research Center, Kawsar Biotech Company, Tehran, Iran. [8] Department of Molecular Medicine, Biotechnology Research Center, Pasteur Institute of Iran, Tehran, Iran.

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

Abstract

Iran was among the first group of countries with a major outbreak of COVID-19 in Asia. With nearly 100 exported cases to various other countries by Feb 28, 2020 it has since been the epicentre of the outbreak in the Middle East. By examining the age- and gender-stratified national data taken from PCR-confirmed cases and deaths related to COVID-19 on Mar 13 (reported by the Iranian ministry of health) and those taken from hospitalised patients in 14 university hospitals across Tehran on Apr 4 (reported by Tehran University of Medical Sciences), we find that the crude case fatality ratio of the two reports in those aged 60 and younger are similar to each other and almost 10 times higher than those reported from China, Italy, Spain and several other European countries (reported from government or ministry of health websites). Assuming a constant attack rate across all age-groups, we adjust for demography, delay from confirmation to death, and under-ascertainment of cases, to estimate the infection fatality ratio based on the reports from Mar 13. We find that our estimates are aligned with reports from China and the UK for those aged 60 and above [n=4609], but are 2-3 times higher in younger age-groups [n=6756] suggesting that only less than %10 of symptomatic cases were detected across the country at the time. Using inbound travel data (from China to Iran) and

matching the dates of the flights with prevalence of cases in China from Jan to Mar 2020, we assess the risk of importation of active cases into the country. Further, using outbound travel data, based on detected cases exported from Iran to several other countries, we estimate the size of the outbreak in the country on Feb 25 and Mar 6 to be 13,700 (95% CI: 7,600 - 33,300) and 60,500 (43,200 - 209,200), respectively. We next estimate the start of the outbreak using 18 whole-genome sequences obtained from cases with a travel history to Iran and the first sequence obtained from inside the country. Finally, we use a mathematical model to predict the evolution of the epidemic and assess its burden on the healthcare system. Our modelling analysis suggests the first peak of the epidemic was on Apr 5 and the next one likely follows within the next 6-10 weeks with approximately 30,000 ICU beds required (IQR: 12K - 60K) and over 1 million active cases (IQR: 740K - 3.7M) during the peak weeks. We caution that relaxed, stringent intervention measures, during a period of highly under-reported spread, would result in misinformed public health decisions and a significant burden on the hospitals in the coming weeks.

Introduction

The current pandemic of Coronavirus Disease 2019 (COVID-19) has had a major impact on societies around the world. Since the start of the first outbreak in Wuhan, China, back in January 2020, the SARS-CoV-2, virus responsible for the disease, spread rapidly to other parts of China and around the world. Iran was among the first group of countries 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 [1]. Following the reports from two deceased cases of COVID-19 in Qom on 19 Feb, in a matter of a few days, several other provinces around the country started to report cases of COVID-19 (see Figure 1a and the time lapse video of the outbreak across Iran during the first 32 days of the epidemic in the Supplementary Information). Given that Qom is the religious capital of the country with millions of pilgrims visiting its shrines every year, transmission chains likely grew much more rapidly compared to other parts of the country. As a result, we see the highest percentage of excess mortality during winter 2019-20 in Qom compared to all the other 30 provinces of Iran (see Figure 1b) with Qom, Golestan, Guilan, Mazandaran, and Alborz being the top five provinces with the highest levels of mortality compared to their 5-year average. Some of the early reports suggest the first two cases could have been infected by a merchant who had reportedly travelled to China [2]. However, with COVID-19, identifying the so-called 'patient zero' is problematic, due to the high rates of asymptomatic and pauci-symptomatic infections [3]. The most likely route of virus spread from China to Iran was via (asymptomatic) air travellers. Infected individuals typically show no symptoms for about 5 days [4] or sometimes no symptoms at all [5], while silently spreading the virus and therefore hindering control efforts [6].

In response to the first two confirmed cases, a number of policies and guidelines were implemented to slow the spread of the disease (see Table A3 in the appendix). The impact of these interventions along with increased physical distancing and public awareness gradually resulted in slowing down of the spread by mid-April. Despite these efforts, fears over economic uncertainty have prompted some employers and business owners affected by the outbreak not to follow recommendations from health officials which might have impacted them financially such as working from home or temporary closure of their businesses [7–9]. As a consequence, some activities followed as usual and the impact of some early public health measures were not as effective. These shortcomings are partly due to the economic sanctions imposed on Iran which have crippled the country's economy. While essential medicines and medical supplies are exempt from the sanctions, restrictions on banking, shipping, trade, and energy have complicated many of Iran's commercial ties and funding for essential equipment necessary to combat COVID-19. As a consequence, Iran has not been able to take the same measures as some other countries to prevent the spread [10; 11]. Furthermore, the recent announcement from the government to reopen parts of the economy in early April [12] has raised concerns over the possibility of a second wave of the epidemic across the country [13].

In this study, we investigate several aspects of the COVID-19 outbreak in Iran. By analysing the aggregate data on case and death counts, we find evidence of significant levels of under-reporting in the country. Further examination based on age-stratified data shows elevated levels of case and infection fatality ratios in those aged < 60 suggesting that less than %10 of symptomatic cases are being reported in those age-groups. Next, using genetic and epidemiological data from cases with travel history to Iran and the first genome collected from an individual inside the country, we estimate the start of the outbreak, its exponential growth rate, and the number of active cases during the early stages of the epidemic. Also using travel data from three major provinces in China with regular flights to Iran, we assess the risk of importation of infected individuals to the country from January to mid-March, 2020. We then examine the burden of the outbreak on the healthcare system under various intervention scenarios and assess the effectiveness of those measures to limit the surge in demand for hospital and ICU beds. In the end, we examine the possibility of data manipulation using Benford's law but find no evidence to support such a claim (see appendix C).

Results

Early signs of under-reporting based on aggregate data on mortality and confirmed cases

By comparing the reported prevalence and deaths in several countries with sizeable outbreaks up to early April, we can see that Iran was also initially on an exponential trajectory (see Figure 2a and Figure 2b). However, after reaching around 7,000 confirmed cases by mid-March, it significantly deviated from this trend which may partly be associated with the effect of intervention measures on Mar 5 to close schools and universities across the country (see Table A3). In particular, the constant accumulation of new deaths from Mar 21 to Apr 10 in Iran suggests that those reported numbers are likely bottlenecked by the fact that testing only took place at designated hospitals across the country while potentially many who became severely ill or died as a result of COVID-19 and were never tested at those hospitals have gone undetected. In addition, it is not customary in Iran to take tests from deceased individuals who were suspected of contracting COVID-19 unless the family requests for autopsy or further examination. Therefore, such issues with under-reporting may further influence the trends in prevalence and deaths. In Figure 2c we compare the daily Case Fatality Ratio (CFR) between several countries (daily CFR = daily reported deaths / daily reported cases). In Iran, we see a sharp initial decline of daily CFR from late February to early March which is due to having 100% CFR on Feb 19 that later dropped as new tests were taken and the number of confirmed cases started to grow. We also see a second peak in CFR on Mar 23 which, again, could be due to testing capacity not catching up with the rate at which the epidemic was growing across the country. Another possible explanation for the second peak in CFR is that changes in population behaviour (e.g. due to control or awareness of increasing deaths) can lead to a reduction in transmission, which, in turn, leads to a drop in reported infections while incidence in deaths follows this reduction with some delay.

Age- and gender-based analysis of fatality rates

We see a noticeable difference in naïve Case Fatality Ratio (nCFR) between Iran and other countries across almost all age groups (see Figure 3a). In particular, the estimated nCFR of Iran and those reported by the Tehran University of Medical Sciences (TUMS) for <60 age-groups are similar to each other, but are almost 10 times higher than those from other countries. This suggests that Iran is likely reporting only severe cases of infection. Figure 3 shows that our estimated infection fatality ratio (IFR) for Iran is 2-3 times higher for <60 age-groups than those from China [14] and the UK [15]. The elevated IFR estimates in those age-groups are partly a result of having a very short reported delay from hospitalisation to death in Iran (see Figure B.1) as well as significant under-reporting of confirmed cases. We also compare the percentage of deaths across the country based on gender (see Figures 3c and 3d) and find no significant difference between male and female mortality in different age-groups, again, highlighting potential under-reporting of prevalence and deaths based on gender. For details on the estimation of IFR, see "case and infection fatality ration analysis.xlsx" in the GitHub repository.

Phylogenetic analysis

We reconstruct a maximum likelihood phylogenetic tree of all sequences that were linked to Iran (see Figure 4) and find that, consistent with other studies [16; 17], our inferred substitution rate is 1.66×10^{-3} (HPD: 2.18×10^{-4} - 3.31×10^{-3}) and the exponential growth rate of 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 01/21/2020 (95% HPD: 12/05/2019 - 02/14/2020).

Risk assessment of case importation from China

Our analysis shows that the expected number of imported cases to Iran remained below one during the first two months of the outbreak in China (see appendix E). The fact that, despite the very low risk of importation, this event took place shows that it is possible that the infected individuals are from a high-risk sub-group of the population, such as businessmen, people working in the hospitality industry, or politicians who frequently travel in and out of the country.

Epidemiological analysis

Here, we estimate the outbreak size at two time points (Feb 25 and Mar 6). We initially carried out an independent likelihood-based analysis to estimate the outbreak size based on exported cases to each of the 6 countries with detected cases and direct flights from Iran and concluded that due to the very large passenger flow to Iraq and UAE there are likely numerous cases that have gone undetected (see appendix D). Therefore, we estimate the outbreak size on Feb 25 based on exported cases to Oman, Lebanon, and Kuwait. We use a similar approach to estimate the outbreak size on Mar 6 based on 28 exported cases to China (see Figure D.3). We find that the estimated number of cases in the country on Feb 25 and Mar 6 to be 13,700 (95% CI: 7,600 - 33,300) and 60,500 (95% CI: 43,200 - 209,200), respectively. We then use a range of doubling times 2.5 - 4 days to extrapolate the number of cases back in time to find the start of the epidemic in Iran (see Figure D.4a). Assuming an 18 days delay from the onset of illness to death [14; 18; 19], we also estimate the number of deaths across the country and compare those to the reported numbers from the ministry of health (see Figure D.4b). This results shows that by the time the first death is reported in Iran, hundreds to perhaps thousands of cases could have been present in the population [20].

Burden of the epidemic on the healthcare system

We consider a range of scenarios for the efficacy of non-pharmaceutical interventions from moderate to high based on of policy announcements by the government on set dates (Table A3). Our findings suggest that the current policy measures are likely not going to be enough to maintain the effective reproduction number below one (Figure 5a) and, as a result, the outbreak is likely going to spread across the country and return in the coming months (Figures 5b, 5d and 5c). The model shows two separate peaks in incidence, one in late-March and another 40-80 days following the first peak (see Figures 5c and 5f) with a significant surge of patients requiring intensive care through in early April across the country (Figure 5b) which is in agreement with the timing of the peak based on daily reports by the ministry of health. This first wave has very likely overwhelmed the healthcare system with approximately 30,000 ICU beds required (IQR: 12K - 60K) and over 1 million active cases (IQR: 740K - 3.7M) during the peak weeks given that there are approximately 3,800 ICU beds available in the country [21]. We predict that unless extreme intervention measures are in place to bring the epidemic under control, there can be up to 100,000 deaths due to COVID-19 (IQR: 88K - 163K) by the end of July 2020 (see Figure 5e). We note that 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 (see Figure 5f).

Discussion

In this study we have provided a detailed analysis of the COVID-19 outbreak in Iran. We have examined how the epidemic evolved through time, how the epidemiological and genetic data enabled us to infer the dynamics of the outbreak in the presence of significant levels of under-reporting of prevalence and deaths, and assessed the effectiveness of the intervention measures in controlling the outbreak. This is also the first study that provides detailed analysis of cohort of COVID-19 patients in Iran using age- and gender-stratified data to estimate levels of under-reporting and infer case and infection fatality ratio in the country. We have also provided a detailed analysis on how the epidemic spread in all of the 31 provinces of Iran and investigated patterns of excess mortality across the country during winter 2019-20 and found five provinces that were likely affected the most by the outbreak. By Mar 19, Iran reported 1433 COVID-19-related deaths whereas there were an excess of 4565 deaths in winter compared to the same period last year with 3846 of deaths belonging to the five provinces we identified as having the highest percentage of excess mortality. Although one cannot directly attribute these extra registered deaths to COVID-19, this may suggest a potentially significant level of under-reporting of deaths related to COVID-19. In addition to the first genome-wide characterisation of the strains of SARS-CoV-2 in Iran, our study combined genomic sequence data from returning travellers to Australia, Austria, Canada, Finland, Germany, New Zealand, Pakistan, and the United States to estimate the start of outbreak in the country. In addition to improving previous methods of estimating outbreak size based on outbound flight information on Feb 25, our study also includes detected cases of COVID-19 from travellers returning to China from Iran on Mar 6 which not only allows us to estimate the likely start of the outbreak in the country but also suggests that the number of active cases could be up to hundreds of times higher than the official numbers at the time. We argue that this large discrepancy is partly due to limited detection of cases, particularly during the early stages of the epidemic, when the testing capacity of the country was not sufficient, and targeted testing of hospitalised patients at designated hospitals. Such discrepancies in reported numbers has been an issues in many other countries [22], but we find them to be particularly high for COVID-19 patients aged < 60 in Iran which is perhaps due to the fact that a large number of mild and subclinical cases in the younger age-groups are not getting tested. By comparing the nationwide age-stratified nCFR with those from hospitalised patients in Tehran, we see that the two are almost identical for $<$50 years-old individuals further supporting the observation of high levels of under-reporting for mild and asymptomatic cases. We also find the IFR in > 60 age-groups are close to those estimated for China and the UK but are 2-3 times higher in younger age-groups. Our phylogenetic and epidemiological analysis places the likely start of transmissions back to late January which suggests the virus had been circulating for, at least, a month prior to first official reports in Qom. We also estimate the growth rate of the epidemic from January to March 2020. Earlier analysis on the whole-genome data from cases with a history of travel to Iran (collected in the last week of February and early March) suggest the emergence of a new clade of SARS-CoV-2 [23; 24]. Similar analyses using genomic data from China and across the world have also revealed aspects of the timing and transmission of the virus [25; 26]. Sampled sequences from travellers with a link to Iran show signs of extreme genetic similarity which suggest that the outbreak across the country may have started from a single transmission event [27; 28]. Although our phylogenetic analysis based on these limited samples cannot strictly rule out the possibility of having multiple separate introductions from unsampled sources in the population, the fact that importation of COVID-19 cases from China occurred despite the low risk of importation suggests that it has likely been introduced to the country by a high-risk sub-group of asymptomatic travellers with frequent trips to Iran. This can also have major implications for screening of COVID-19 cases at the airports solely based on symptoms such as high temperature.

In the early days of the outbreak, only Pasteur Institute (PI) of Iran had the resources to test a few hundred of suspected COVID-19 cases. Therefore, most samples taken from hospitalised patients in Qom were sent directly to PI for analysis. As the demand for testing increased, several other clinics and pathology labs that routinely test for viral agents such as HIV and HBV and are equipped with real-time PCR machines received training at the PI to perform COVID-19 tests. By the end of the third week of the outbreak, the Chinese Embassy imported 100,000 tests and facilitated the purchase of another 200,000 tests which helped the healthcare system to manage the demand for patient testing. By mid-March, WHO delivered several shipments of emergency medical supplies and 100,000 additional test kits to Iran [29–31]. Although private labs also offer non-regulated tests to individuals with mild or no symptoms, their results are not counted in the official reports – only positive RT-PCR test results from designated hospitals across the country are reported in the daily announcements from the ministry of health. At the peak of the epidemic in early April, labs at PI tested 1500-2000 samples a day but as of early May these numbers are down to 200-400 samples a day partly because there are now over 100 labs across the country offering tests and there is no longer a high demand for testing at PI. An epidemiological report by TUMS on Apr 4, 2020 shows that of 8840 clinically diagnosed cases of COVID-19 only half of them received RT-PCR tests with a false negative rate of about 49% [32]. Given that most of these tests are taken from hospitalised patients, possibly with several days of delay since the onset of their symptoms, it is likely that the accuracy of the tests drop significantly over time [33]. Another report on Mar 13, 2020 by the ministry of health on reported deaths based on gender shows no significant difference in mortality between the two groups indicating that the RT-PCR tests are not sensitive enough to tease out the higher mortality reported in male individuals aged 20 and older and that there is likely under-reporting based on gender in those age groups [34–36]. In agreement with similar reports from hospitalised patients in France [37], we see that, in Iran, the fraction of hospitalised patients with intubation die much faster than those without need for intubation. However, the delay from hospitalisation to death is much shorter in the Iranian cohort likely suggesting that patients are under much more severe conditions when hospitalised. Even if we take the reports from TUMS at face value, it suggests that the correct number of hospitalised cases of COVID-19 could be up to four times higher than those officially reported. This is excluding the potentially much larger group of cases with mild or no symptoms who are never tested for COVID-19.

Our study also shows that the series of non-pharmaceutical interventions and the public compliance were effective in slowing down the speed of the spread of COVID-19 from mid-March to early-May, but they were likely not strong enough to prevent the re-emergence of a second peak. We caution that the recent policy to ease restrictions from early April, particularly during a period of highly under-reported spread which gives the appearance of a low number of active cases across the country, can lead to misinformed intervention measures which may, in turn, precipitate the likely emergence of the second peak. This quick return to normal life may help the SARS-CoV-2 become endemic in Iran and potentially create subsequent waves of infection across the country in autumn and winter. 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. This can potentially become an issue for any country attempting to control its epidemic by preventing the resurgence of new cases into its boarders [38]. 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, possibly because significant levels of under-reporting in some of those countries. The methods we discussed in this study can shed light on detecting such possible under-reporting of prevalence and deaths. Another explanation for the large outbreak in Iran compared to its neighbouring countries is the role of temperature and humidity in sustaining an epidemic in a country. Although the effects of temperature on COVID-19 outbreaks is still debated and may be small [39–41], 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 given that high susceptibility of the population is still the main driver of outbreaks [42].

Supporting information

We enclosed a supplementary video and an appendix to this manuscript.

Contributions and declaration of interests

MG, AL, and AK developed the idea and research. MG, BH, AKar, SD, and AK wrote the first draft of the manuscript, and all other authors discussed results and edited the manuscript. MG, BH, and AKar collected and analysed the epidemiological data. MAK, MA, and SZ collected and validated the genomic sample from Iran. Authors declare no competing interests.

Acknowledgement

We thank José Lourenço for helpful comments on the original draft of the manuscript. MG and SD are funded by the Biotechnology and Biological Science Research Council (BBSRC), grant number BB/M011224/1.

Methods

Data on aggregate number of cases and deaths

We obtain time series data for the number of confirmed cases and deaths in Iran, UK, US, France, Italy, and Spain from the John Hopkins University Centre for Systems Science and Engineering COVID-19 GitHub repository (accessed on 04/06/2020) [43]. We also obtain time series data on confirmed cases in all the 31 provinces of Iran from the ministry of health website (behdasht.gov.ir). We note that ministry stopped releasing province data from Mar 23 onward. They also did not release province data on Mar 2 and 3. We obtain mortality statistics during winter (from 2014 to 2020) from the National Organization for Civil Registration of Iran (www.sabteahval.ir). See appendix G and github repository for more details.

Age- and gender-stratified data from Iran

We use the only age- and gender-stratified nationwide data reported by the Iranian ministry of health on Mar 13 [44] along with the age-stratified data on patients from 14 hospitals associated with Tehran University of Medical Sciences (TUMS) on Apr 4 [32], the Chinese Center for Disease Control reported on Feb 11 [35], and several official reports from European countries with major outbreaks (see Tables A1 and A2 for the complete list of data and references) to estimate naïve Case Fatality Ratio (nCFR). After adjusting for Iran's demography [45] and under-ascertainment of cases using the method developed in [14], we use the known distribution of time between hospitalisation to hospital death (see Figure B.1b) to account for the delay from onset of symptoms to confirmation of cases [46]. Then, assuming a homogeneous attack rate across all age groups and that 50% of cases are asymptomatic, we calculate infection fatality ratio (IFR).

Genomics data of SARS-CoV-2 related to Iran and phylogenetic analysis

In this study, we collect the first whole-genome sequence of SARS-CoV-2 from inside Iran and include the additional 20 sequences from confirmed cases that flew out of Iran and had their viral genomes determined in other countries. The latter is made available through GISAID (see metadata on appendix I). We use these data to determine the time to the most recent common ancestor of the sampled sequences and characterise the local epidemic growth in Iran. Two of these sequences are epidemiologically linked and cannot be used for these purposes. We use the 21 whole-genome sequences to estimate parameters of the epidemic using phylodynamic models implemented in BEAST2 [47]. We reconstructed the maximum likelihood phylogenetic tree of all sequences using PhyML [48]. Two of these 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. We used the known sampling times, with a Continuous-Time Markov Chain reference prior [49] on the substitution 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, SD=2), and the growth rate implementation with a Laplace prior (scale=100), and an HKY+G model of substitution.

Air travel data on imported (exported) cases to (from) Iran

We collect a detailed list of air travel data to four of the largest international airports in Iran with the highest number of weekly international flights to countries that detect exported cases from Iran (see the summary Table 1 and detailed Table in appendix H). Flights from China to Iran mainly go from Beijing, Shanghai, and Guangdong to Tehran with a total of approximately 6,700 passengers per week. We use those flight information to estimate the outbreak size on Mar 6 (see appendix D) and assess the initial risk of importation of cases from those three provinces in China to Iran (see appendix E) We note that in early Mar, the Gulf Cooperation Council countries suspended their flights to Iran.

Excess seasonal mortality in 31 provinces of Iran

We use the publicly available data from the National Organisation for Civil Registration on the total number of registered deaths during winter (from Dec 22 to Mar 19) from 2015 to 2020 to calculate the percentage of excess mortality in winter 2019-20 (see appendix G for the dataset used in the analysis). To do this, we first find the proportion of deaths, $p(t)$, in winter by dividing the number of deaths during winter in each province by the total population of that province in that particular year. We then calculate the mean proportion of deaths, p^*, over the last five years (excluding winter 2019-20) for each province. Finally, we calculate the percentage of excess mortality, $EM(t)$ for each province in winter by

$$EM(t = \text{winter 2019-20}) = \frac{p(t) - p^*}{p^*} \times 100.$$

We note that the only census for province population of Iran is taken in 2016. To adjust for provincial population size changes each year, we assume the change in population size of each province is proportional to the change in national population size.

Modelling the outbreak using an SEIR model

To understand how the epidemic in Iran has unfolded, its burden on the healthcare system, and predict its likely trajectory in the coming months, we use an SEIR model [50]. We summarise all the model parameters and their corresponding values in Table A4 and Table A5. We assume that a few (1-2 individuals) initially infected sparked 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. Note that, as explained in appendix G, since there is no daily update on the number of deaths for each province, we only use our model to understand the dynamics at the country-level. 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). Deaths are assumed to occur only among severe cases (in the ICU). The fractions of demand for hospitalisation, ICU use, and deaths in different age-groups are estimated using data from China. Any difference based on Iran's demography and quality of healthcare could affect our estimates on health care demand. We also use the expected values for our clinically-informed parameters such as the incubation period (see Table A4) since we assume the uncertainties in the dynamics is dominated by non-pharmaceutical interventions to reduce effective reproduction number, $R(t)$ (see table A3). This parameter is defined as the product of reproduction number, R_0, strength of mitigation measures, $M(t)$, and effectiveness in isolation of individuals from certain age-groups, ζ_a.

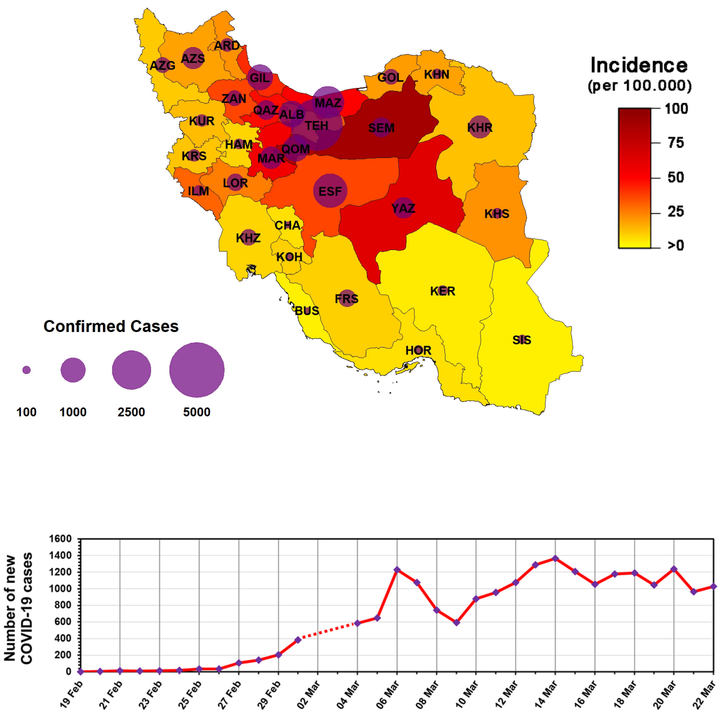
Code availability

The analysis on excess mortality and burden on hospitals are done in Mathematica 11.0 and estimating the outbreak size using air travel data and investigation into data manipulation using Benford's law are done in Matlab 2019a. We made all the codes and data used in this study available online on a GitHub repository (github.com/mg878/Iran_study).

Table 1: Weekly passenger flux to countries included in the study

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

(a)



(b)

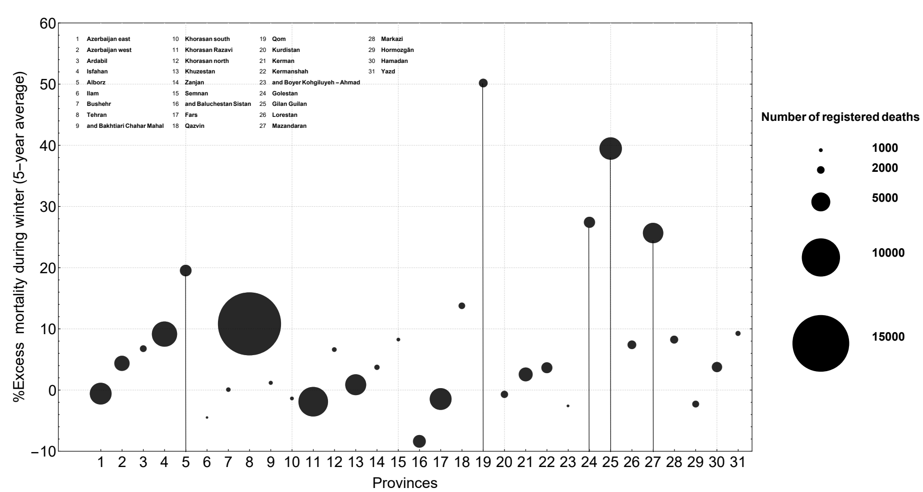


Figure 1: (a) The COVID-19 Epidemic in Iran. (Top) Total incidence and confirmed cases by region on Mar 22, 2020. (Bottom) The early spread of the COVID-19 epidemic based on official reports (dotted line illustrates when no cases were reported officially and is for visual purposes only). (b) Percentage of excess mortality during winter 2019-20 (from Dec 22, 2019 to Mar 19, 2020) in 31 provinces in Iran. Five provinces with the largest deviations from background mortality (averaged over five years) are highlighted by a vertical solid line. The size of the circles represents the total number of registered deaths in each province.

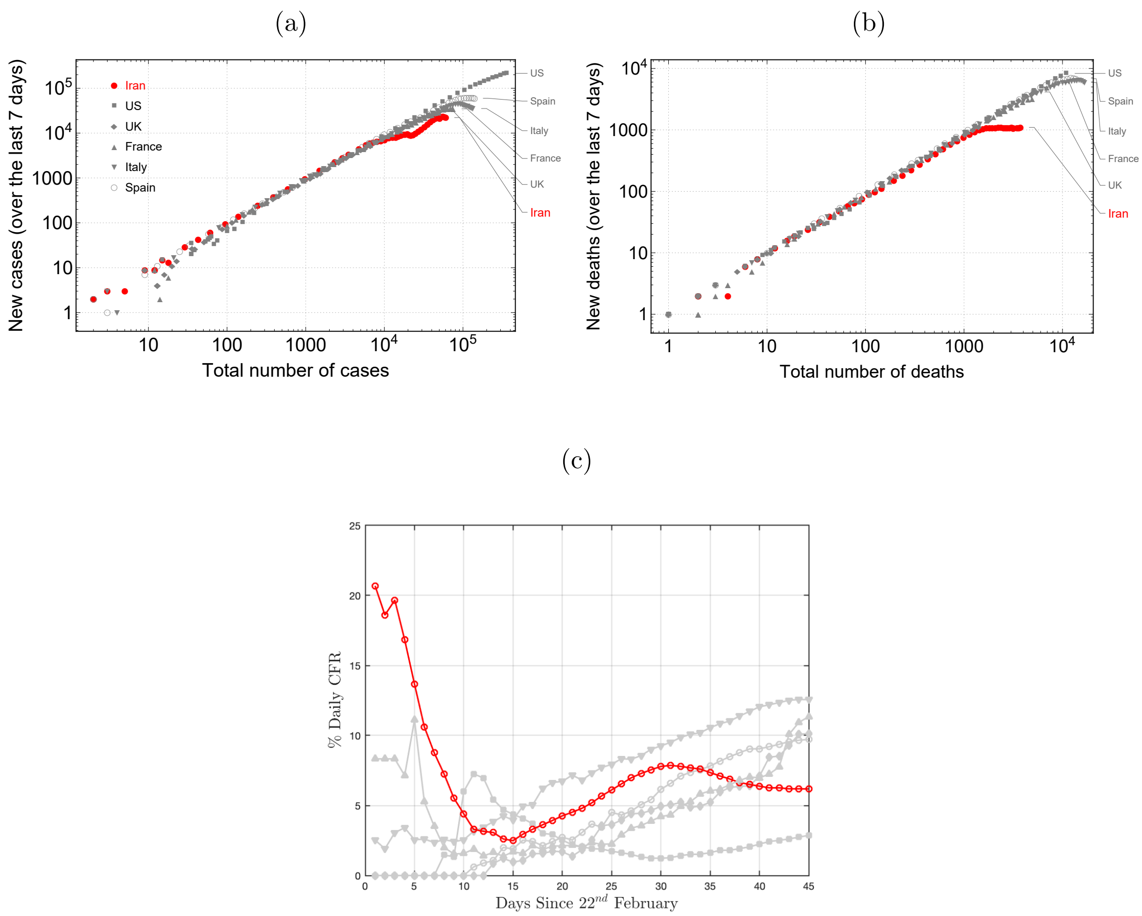


Figure 2: Reported number of cases and deaths from several countries including Iran (data shown in red circles). (a) and (b) show the sum of new cases and deaths (7-day rolling average) as a function of the total number of cases and deaths. (c) The daily Case Fatality Ratio (daily CFR) for all six countries from Feb 22 to Apr 7.

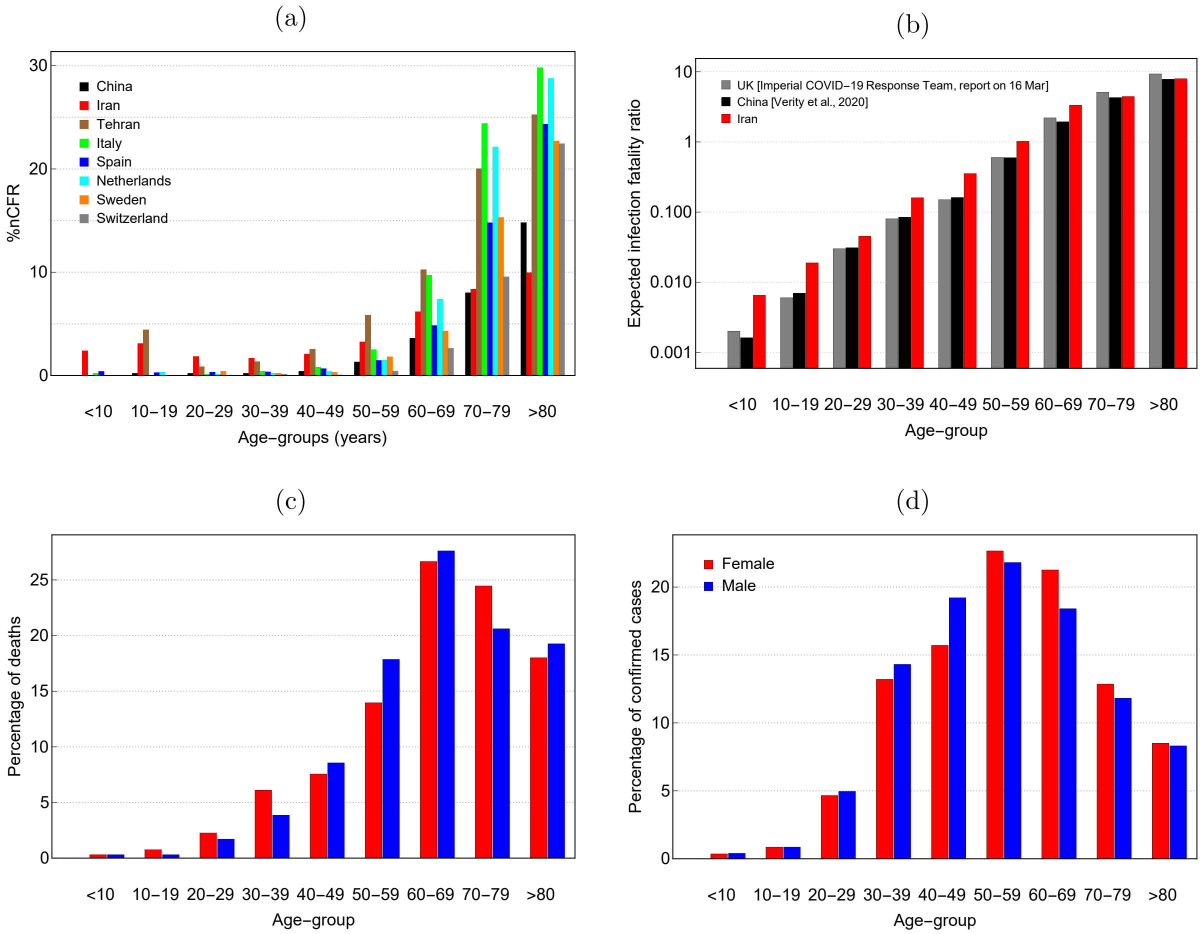


Figure 3: (a) Comparison of the age specific naïve Case Fatality Ratio reported by the Iranian ministry of health, Tehran University of Medical Sciences hospitals, several European countries, and China (see Tables A1 and A2 for more information). (b) Comparing the expected IFR in Iran, China, and the UK. Percentage of (c) deaths and (d) confirmed cases based on gender [44].

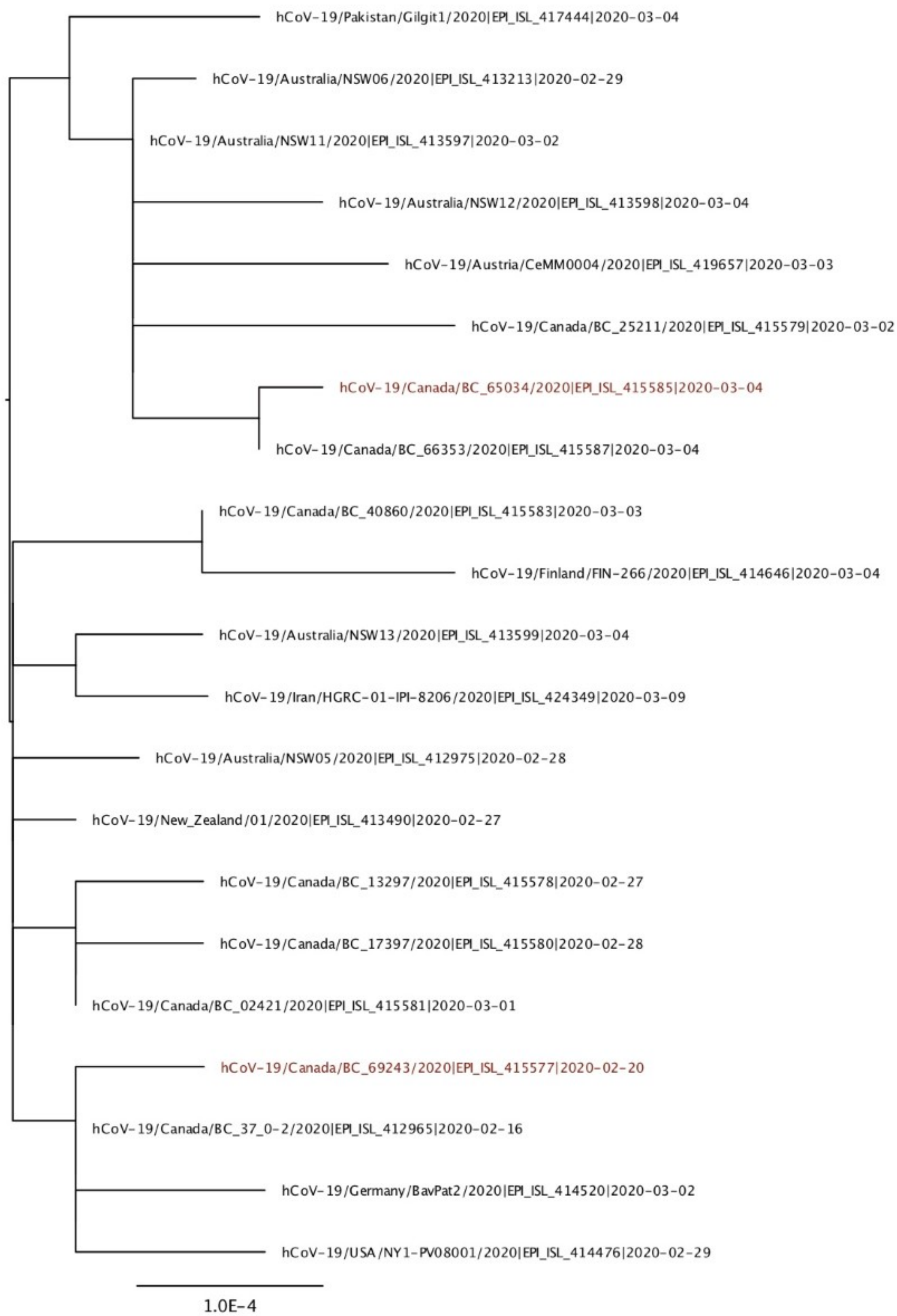


Figure 4: Maximum likelihood phylogenetic tree of samples linked to Iran. The labels include the sampling times used in the BEAST analysis. Red labels indicate the two epidemiologically linked samples that were excluded from the subsequent analysis.

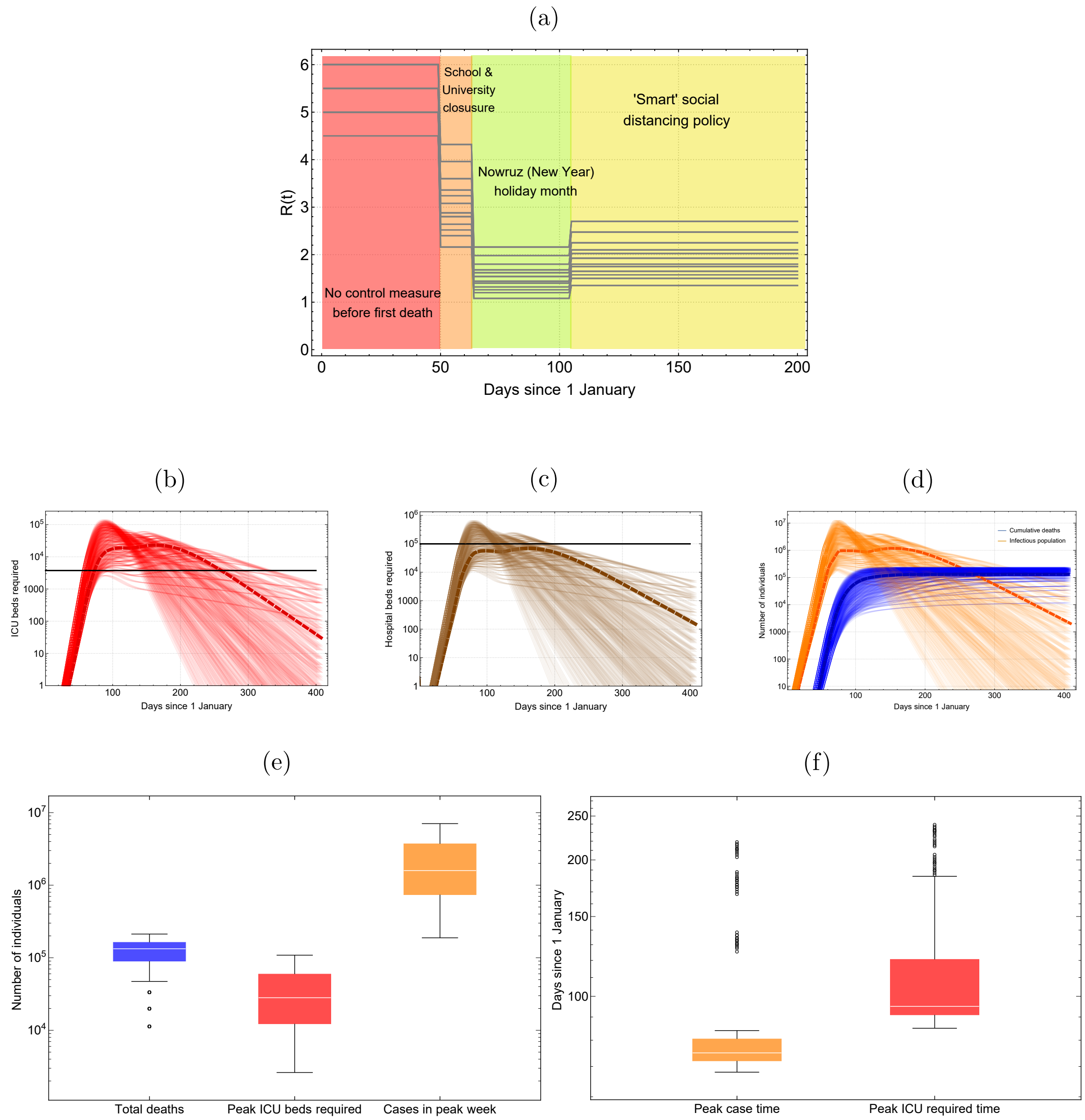


Figure 5: Projection of the impact of intervention measures on (a) effective reproduction number, $R(t)$, which we defined as the product of reproduction number, R_0, strength of mitigation measures, $M(t)$, and isolation of individuals in specific age-categories, ζ_a, (b) number of ICU and (c) hospital beds required (d) cumulative deaths and incidence over time. (e) and (f) 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. The horizontal solid lines in (b) and (c) show the total number of ICU and hospital beds available in Iran. In (e) and (f), coloured areas represent the interquartile range with median as a white horizontal line and outliers as open circles.

References

[1] World Health Organisation. WHO Director-General's opening remarks at the media briefing on COVID-19 - 28 February, 2020; https://www.who.int/dg/speeches/detail/who-director-general-s-remarks-at-the-media-briefing-on-covid-19---28-february-2020.

[2] 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.

[3] Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). Science. 2020;.

[4] Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. New England Journal of Medicine. 2020;382(13):1199–1207.

[5] Bai Y, Yao L, Wei T, Tian F, Jin DY, Chen L, et al. Presumed asymptomatic carrier transmission of COVID-19. JAMA. 2020;323(14):1406–1407.

[6] Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science. 2020;.

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

[8] Wintour P. 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.

[9] Basravi Z. 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.

[10] A Murphy IHMMEA Zhaleh Abdi. Economic sanctions and Iran's capacity to respond to COVID-19. The Lancet Public Health. 2020;.

[11] Danaei G, Farzadfar F, Kelishadi R, Rashidian A, Rouhani OM, Ahmadnia S, et al. Iran in transition. The Lancet. 2019;.

[12] IRNA. Daily COVID-19 reports. 2020;
https://www.irna.ir/news/83736401.

[13] IRNA. Daily COVID-19 reports. 2020;
https://www.irna.ir/news/83737152.

[14] Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. The Lancet Infectious Diseases. 2020;.

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

[16] Rambaut A. 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.

[17] Boni MF, Lemey P, Jiang X, Lam TTY, Perry B, Castoe T, et al. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. bioRxiv. 2020;.

[18] Linton NM, Kobayashi T, Yang Y, Hayashi K, Akhmetzhanov AR, Jung Sm, et al. 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. 2020 Feb;9:538.

[19] Wu JT, Leung K, Bushman M, Kishore N, Niehus R, de Salazar PM, et al. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. Nature Medicine. 2020;26(4):506–510.

[20] Jombart T, van Zandvoort K, Russell T, Jarvis C, Gimma A, Abbott S, et al. Inferring the number of COVID-19 cases from recently reported deaths;.

[21] Phua J, Faruq MO, Kulkarni AP, Redjeki IS, Detleuxay K, Mendsaikhan N, et al. Critical Care Bed Capacity in Asian Countries and Regions. Critical Care Medicine Society of Critical Care Medicine. 2020;48(5).

[22] Lavezzo E, Franchin E, Ciavarella C, Cuomo-Dannenburg G, Barzon L, Del Vecchio C, et al. Suppression of COVID-19 outbreak in the municipality of Vo, Italy. medRxiv. 2020;.

[23] Eden JS, Rockett R, Carter I, Rahman H, de Ligt J, Hadfield J, et al. An emergent clade of SARS-CoV-2 linked to returned travellers from Iran. Virus Evolution. 2020;6(1).

[24] Rambaut A, Holmes EC, Pybus OG. 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.

[25] Volz E, Baguelin M, Bhatia S, Boonyasiri A, Cori A, Cucunuba Perez Z, et al. Phylogenetic analysis of SARS-CoV-2. 2020;.

[26] Lu J, Plessis Ld, Liu Z, Hill V, Kang M, Lin H, et al. Genomic epidemiology of SARS-CoV-2 in Guangdong Province, China. medRxiv. 2020;.

[27] Hodcroft E, Müller N, Wagner C, Ilcisin M, Hadfield J, Bell SM, et al. 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.

[28] Bell SM, Müller N, Hodcroft E, adn James Hadfielda CW, Ilcisin M, Neher R, et al. 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.

[29] World Health Organisation. WHO ships emergency medical supplies to the Islamic Republic of Iran. 2020; https://www.who.int/news-room/feature-stories/detail/who-ships-emergency-medical-supplies-to-the-islamic-republic-of-iran.

[30] World Health Organisation. WHO delivers new shipment of medical supplies for COVID-19 response in Islamic Republic of Iran. 2020; http://www.emro.who.int/irn/iran-news/who-delivers-new-shipment-of-medical-supplies-for-covid-19-response-in-islamic-republic-of-irann.html.

[31] UNICEF. UNICEF delivers 3rd shipment of supplies to help Iran in fight against COVID-19. 2020; https://www.unicef.org/iran/en/press-releases/unicef-delivers-3rd-shipment-supplies-help-iran-fight-against-covid-19.

[32] Tehran University of Medical Sciences COVID-19 response team. Epidemiological report of COVID-19 in hospitals affiliated with Tehran University of Medical Sciences. Public Access Directory. 2020;.

[33] Wikramaratna P, Paton RS, Ghafari M, Lourenco J. Estimating false-negative detection rate of SARS-CoV-2 by RT-PCR. medRxiv. 2020;.

[34] Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, et al. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. JAMA. 2020;.

[35] 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. 2020;2:113.

[36] Jin JM, Bai P, He W, Wu F, Liu XF, Han DM, et al. Gender differences in patients with COVID-19: Focus on severity and mortality. medRxiv. 2020;.

[37] Salje H, Kiem CT, Lefrancq N, Courtejoie N, Bosetti P, Paireau J, et al. Estimating the burden of SARS-CoV-2 in France. medRxiv. 2020;.

[38] Kraemer MUG, Yang CH, Gutierrez B, Wu CH, Klein B, Pigott DM, et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. Science. 2020;.

[39] Ficetola GF, Rubolini D. Climate affects global patterns of COVID-19 early outbreak dynamics. medRxiv. 2020;.

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

[41] Notari A. Temperature dependence of COVID-19 transmission. medRxiv. 2020;.

[42] Baker RE, Yang W, Vecchi GA, Metcalf CJE, Grenfell BT. Susceptible supply limits the role of climate in the COVID-19 pandemic. medRxiv. 2020;.

[43] Novel Coronavirus (COVID-19) Cases, provided by Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. 2020; `https://github.com/CSSEGISandData/COVID-19`.

[44] National Committee on COVID-19 Epidemiology, Mar 13. Daily Situation Report on Coronavirus disease (COVID-19) in Iran. 2020; `http://corona.behdasht.gov.ir/files/site1/files/Factsheet-12.23-En.pdf`.

[45] Statistical Center of Iran. Selected findings of the 2011 national population and housing census. 2011; `https://web.archive.org/web/20151223181433/Iran`.

[46] Russell TW, Hellewell J, Jarvis CI, van Zandvoort K, Abbott S, Ratnayake R, et al. 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. 2020;25(12).

[47] Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evolution. 2018;4(1).

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

[49] Ferreira MA, Suchard MA. Bayesian analysis of elapsed times in continuous-time Markov chains. Canadian Journal of Statistics. 2008;36(3):355–368.

[50] Neher R, Aksamentov I, Noll N. About COVID-19 Scenarios. 2020; https://covid19-scenarios.org/about.

[51] Carlos III Health Institute. Informe sobre la situación de COVID-19 en España. 2020; https://www.isciii.es.

[52] National Institute for Public Health and the Environment. Ontwikkeling COVID-19 in grafieken. 2020;
https://www.rivm.nl/coronavirus-covid-19/grafieken.

[53] The COVID-19 Task force of the Department of Infectious Diseases and the IT Service Istituto Superiore di Sanità. Integrated surveillance of COVID-19 in Italy. 2020; https://www.epicentro.iss.it/en/coronavirus/bollettino/Infografica_20aprile%20ENG.pdf.

[54] The Public Health Agency of Sweden. Antal fall av covid-19 i Sverige. 2020; https://experience.arcgis.com/experience/09f821667ce64bf7be6f9f87457ed9aa.

[55] Federal Office of Public Health of the Swiss Confederation. COVID-19 in Switzerland. 2020;
https://covid-19-schweiz.bagapps.ch/fr-1.html.

[56] National Committee on COVID-19 Epidemiology, Mar 14. Daily Situation Report on Coronavirus disease (COVID-19) in Iran. 2020; http://corona.behdasht.gov.ir/files/site1/files/Factsheet-12.24-En1.pdf.

[57] Benford F. The Law of Anomalous Numbers. Proceedings of the American Philosophical Society. 1938;78(4):551–572.

[58] Durtschi C, Hillison WA, Pacini C. The effective use of Benford's Law to assist in detecting fraud in accounting data; 2004. .

[59] Brown RJC. Benford's Law and the screening of analytical data: the case of pollutant concentrations in ambient air. The Analyst. 2005;130(9):1280—1285.

[60] Idrovo AJ, Fernandez-Nino JA, Bojorquez-Chapela I, Moreno-Montoya J. 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. 2011;139(12):1827–1834.

[61] Crocetti E, Randi G. Using the Benford's Law as a First Step to Assess the Quality of the Cancer Registry Data. Frontiers in Public Health. 2016;4:225.

[62] Goodman W. The promises and pitfalls of Benford's law. Significance. 2016;13(3):38–41.

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

[64] Zhuang Z, Zhao S, Lin Q, Cao P, Lou Y, Yang L, et al. 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;.

[65] Fraser C, Donnelly CA, Cauchemez S, Hanage WP, Van Kerkhove MD, Hollingsworth TD, et al. Pandemic potential of a strain of influenza A (H1N1): early findings. Science. 2009;324(5934):1557–1561.

[66] 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`.

[67] World Tourism Organisation. Iran Islamic Republic Of Country-specific Basic indicators (Compendium) 2014 - 2018 (12.2019). Tourism Statistics. 2020;.

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

[69] VariFlight. Flight database; 2020.

[70] Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, et al. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. The Lancet Infectious Diseases. 2020;.

[71] McAloon CG, Collins A, Hunt K, Barber A, Byrne A, Butler F, et al. The incubation period of COVID-19: A rapid systematic review and meta-analysis of observational research. medRxiv. 2020;.

[72] IRNA. Daily COVID-19 reports. 2020;
https://www.irna.ir/news/83682584.

[73] IRNA. Daily COVID-19 reports. 2020;
https://www.irna.ir/news/83688246.

[74] IRNA. Daily COVID-19 reports. 2020;
https://www.irna.ir/news/83688090.

[75] IRNA. Daily COVID-19 reports. 2020;
https://www.irna.ir/news/83687927.

[76] IRNA. Daily COVID-19 reports. 2020;
https://www.irna.ir/news/83689130.

[77] IRNA. Daily COVID-19 reports. 2020;
https://www.irna.ir/news/83688131.

[78] IRNA. Daily COVID-19 reports. 2020;
https://www.irna.ir/news/83687367.

[79] IRNA. Daily COVID-19 reports. 2020;
https://www.irna.ir/news/83688153.

[80] IRNA. Daily COVID-19 reports. 2020;
https://www.irna.ir/news/83688367.

[81] IRNA. Daily COVID-19 reports. 2020;
https://www.irna.ir/news/83705027.

[82] IRNA. Daily COVID-19 reports. 2020;
https://www.irna.ir/news/83702881.

[83] IRNA. Daily COVID-19 reports. 2020;
https://www.irna.ir/news/83717334.

[84] IRNA. Daily COVID-19 reports. 2020;
https://www.irna.ir/news/83705278.

[85] IRNA. Daily COVID-19 reports. 2020;
https://www.irna.ir/news/83692780.

[86] Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. International Journal of Infectious Diseases. 2020 2020/04/15;93:284–286.

[87] Davies NG, Kucharski AJ, Eggo RM, Gimma A, , Edmunds WJ. The effect of non-pharmaceutical interventions on COVID-19 cases, deaths and demand for hospital services in the UK: a modelling study. medRxiv. 2020;.

[88] NHS. Hospital Admitted Patient Care Activity 2018-19. 2018-19;.

[89] Cao B, Wang Y, Wen D, Liu W, Wang J, Fan G, et al. A trial of lopinavir–ritonavir in adults hospitalized with severe COVID-19. New England Journal of Medicine. 2020;.

[90] 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. 2020;.

Supplementary appendix

A Age-stratified data on confirmed cases and deaths

Table A1: Age-specific data for Iran [44] and Tehran [32]

Age-group	National Confirmed Cases, N (%)	National Death, N (%)	National nCFR %	Tehran Confirmed Cases, N (%)	Tehran Death, N (%)	Tehran nCFR %
0-9	42 (0.37)	1* (0.19)	2.38	63 (0.71)	0 (0.00)	0.00
10-19	97 (0.85)	3 (0.58)	3.09	136 (1.54)	6 (0.92)	4.41
20-29	545 (4.80)	10 (1.95)	1.83	710 (8.03)	6 (0.92)	0.85
30-39	1,563 (13.75)	26 (5.06)	1.66	1,643 (18.59)	22 (3.36)	1.34
40-49	1,983 (17.44)	41 (7.98)	2.07	1,699 (19.22)	43 (6.57)	2.53
50-59	2,526 (22.23)	82 (15.95)	3.25	1,610 (18.21)	94 (14.37)	5.84
60-69	2,252 (19.82)	139 (27.04)	6.17	1,475 (16.69)	151 (23.09)	10.24
70-79	1,401 (12.33)	117 (22.76)	8.35	914 (10.34)	183 (27.98)	20.02
>80	955 (8.40)	95 (18.84)	9.95	590 (6.67)	149 (22.78)	25.25
Total	11,364	514		8,840	654	

*The only COVID-19 death in under-10 children was observed in a 3-year old child who had leukemia.

Table A2: Age-specific data for Spain [51], Netherlands [52], Italy [53], China [35], Sweden [54], and Switzerland [55].

Age-group	China (11 February)			Italy (20 April)			Netherlands (21 April)		
	Confirmed Cases, N (%)	Death, N (%)	nCFR %	Confirmed Cases, N (%)	Death, N (%)	nCFR %	Confirmed Cases, N (%)	Death, N (%)	nCFR %
0-9	416 (0.93)	0 (0.00)	0	1,000 (0.60)	2 (0.01)	0.2	80 (0.23)	0 (0.00)	0.00
10-19	549 (1.23)	1 (0.10)	0.2	0 (0.00)	0 (0.00)	0	320 (0.94)	1 (0.03)	0.31
20-29	3,619 (8.10)	7 (0.68)	0.2	7,000 (4.25)	7 (0.03)	0.1	2,868 (8.41)	3 (0.08)	0.10
30-39	7,600 (17.01)	18 (1.76)	0.2	11,250 (6.83)	45 (0.21)	0.4	2,929 (8.58)	6 (0.15)	0.20
40-49	8,571 (19.19)	38 (3.71)	0.4	23,000 (13.96)	184 (0.85)	0.8	3,802 (11.14)	15 (0.38)	0.39
50-59	10,008 (22.40)	130 (12.71)	1.3	31,960 (19.40)	799 (3.71)	2.5	6,267 (18.37)	92 (2.35)	1.47
60-69	8,583 (19.21)	309 (30.21)	3.6	24,928 (15.13)	2,418 (11.22)	9.7	4,641 (13.60)	343 (8.76)	7.39
70-79	3,918 (8.77)	312 (30.50)	8	26,770 (16.25)	6,532 (30.31)	24.4	5,207 (15.26)	1,152 (29.43)	22.12
>80	1,408 (3.15)	208 (20.33)	14.8	38,802 (23.56)	11,563 (53.66)	29.8	8,005 (23.46)	2,303 (58.83)	28.77
Total	44,672	1,023		164,710	21,550		34,119	3,915	
Age-group	Spain (21 April)			Sweden (21 April)			Switzerland (22 April)		
	Confirmed Cases, N (%)	Death, N (%)	nCFR %	Confirmed Cases, N (%)	Death, N (%)	nCFR	Confirmed Cases, N (%)	Death, N (%)	nCFR %
0-9	512 (0.35)	2 (0.02)	0.39	83 (0.52)	0 (0.00)	0	110 (0.39)	0 (0.00)	0.00
10-19	2,882 (1.99)	8 (0.06)	0.28	199 (1.24)	0 (0.00)	0	749 (2.67)	0 (0.00)	0.00
20-29	5,371 (3.71)	17 (0.14)	0.32	1,200 (7.46)	0 (0.00)	0.4	3,400 (12.10)	1 (0.08)	0.03
30-39	13,580 (9.37)	46 (0.38)	0.34	1,500 (9.33)	7 (0.36)	0.2	3,742 (13.31)	4 (0.33)	0.11
40-49	21,221 (14.65)	140 (1.14)	0.66	2,000 (12.44)	20 (1.04)	0.3	4,432 (15.77)	2 (0.16)	0.05
50-59	26,461 (18.26)	384 (3.12)	1.45	2,800 (17.41)	66 (3.42)	1.8	5,784 (20.58)	24 (1.98)	0.41
60-69	22,721 (15.68)	1,099 (8.92)	4.84	2,200 (13.68)	148 (7.66)	4.3	3,434 (12.22)	90 (7.41)	2.62
70-79	21,739 (15.00)	3,215 (26.11)	14.79	2,200 (13.68)	459 (23.76)	15.3	2,745 (9.77)	262 (21.56)	9.54
>80	30,415 (20.99)	7,403 (60.12)	24.34	3,900 (24.25)	1,232 (63.77)	22.7	3,708 (13.19)	832 (68.48)	22.44
Total	144,902	12,314		16,082	1,932		28,104	1,215	

B Delay from hospitalisation to death

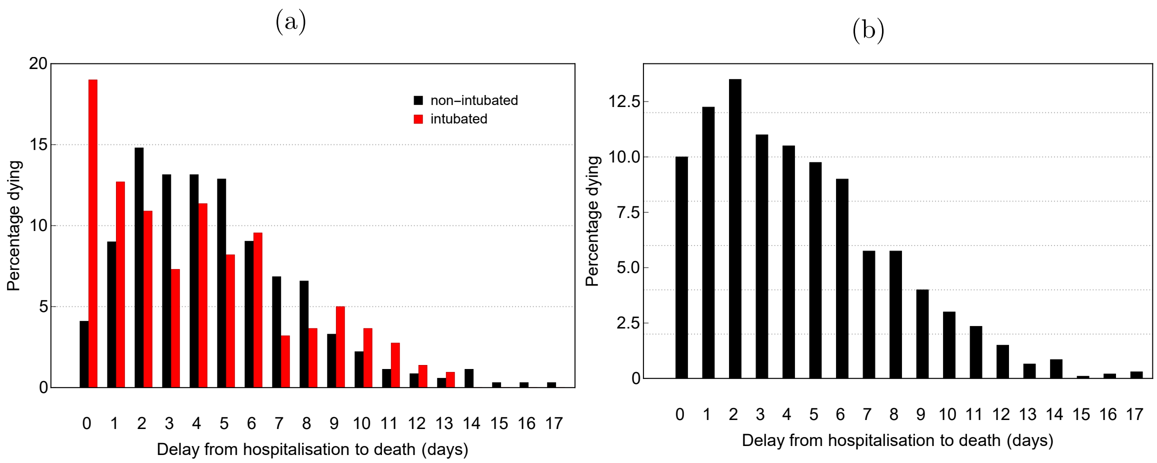


Figure B.1: (a) Comparison of delay from hospitalisation to hospital death for patients with (red) and without tracheal intubation (black) in Iran. (b) Delay from hospital admission to hospital death in Iran [56].

C Investigating irregularities in reported numbers using Benford's law

Benford's law (BL) gives the probability distribution of leading digits in a determined set of numbers [57]. The probability that digit $d = 1, 2, ..., 9$ is a leading number is given by $P(d) = log_{10}(1 + (1/d))$ where numbers with leading digit 1 have the highest probability of appearance and this probability steadily decreases as the starting digit becomes larger. BL is used in a variety of different areas to study irregularities in data [58; 59] and is also frequently used to assess the quality of epimeiological and clinical data [60; 61].

An exponential function in the form $2^{t/d_t}$ obeys BL perfectly, where this exponential function can represent the doubling rate of an outbreak with t being the time in days and d_t the doubling time in days. If we combine the data from two different exponential functions into one larger data set then that larger data set will also obey BL. We can do this in the current context of reported numbers for COVID-19 during the exponential growth phase and create a larger data set that contains the total number of cases and total number of deaths, both of which are exponential functions but with different growth rates. We note that if the numbers grow linearly over time, then they do not obey BL. Also, a data set that contains a mix of both exponential and linear growth will also not necessarily obey BL.

Figure C.2 shows the distribution of leading digits in data during the exponential phase (see Figure 2) from Iran, the USA, and the UK compared to the Benford distribution. The data for Iran includes the total cases reported from the start of the outbreak to Apr 1, 2020 and total deaths from the first reported death to Apr 10, 2020. Note that we take the reported cases up to Apr 1 since all three countries follow the exponential growth up to that day. We sample 40 random numbers 5 times from this data set and count the number of occurrences for each leading digit. This sampling of data allows us to generate an average distribution with error bars representing the standard deviation shown in Figure C.2. The probability distribution does not show any conclusive evidence to suggest a manipulation of data in any of the three countries. From this, we conclude that the likely low or inaccurate number of reported cases in Iran are due to other issues mentioned in this study and not manipulation of data [62]. We note that while this method can be used to test if data manipulation has occurred, it does not give any information about deliberate absence of data by, for instance, not reporting deaths from specific hospitals.

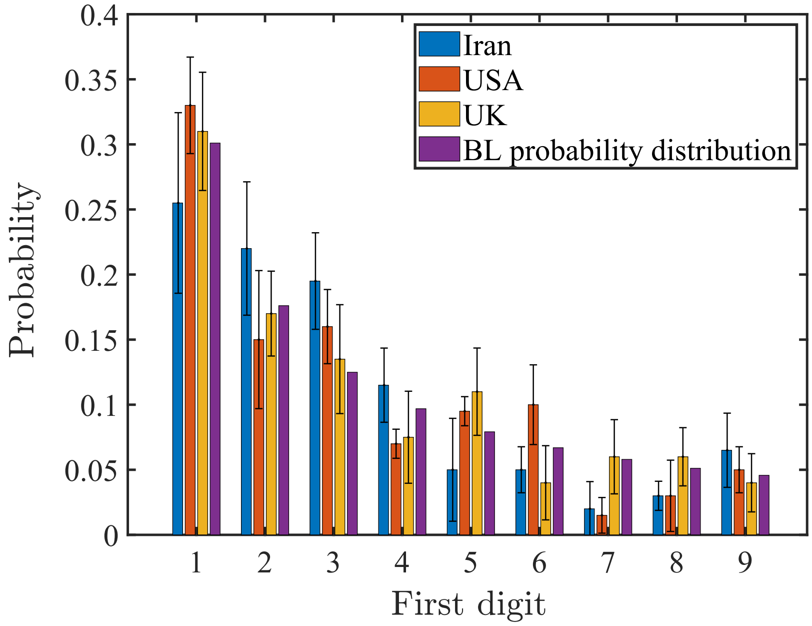


Figure C.2: The probability distribution of leading digits in the data of total confirmed cases from the start of the outbreak to Apr 1 and total deaths from the first reported death in Iran to Apr 10 and are compared to the distribution given by Benford's law (BL). Error bars represent 1 standard deviation unit from the mean.

D Estimating the outbreak size based on air travel data

Two recent studies [63; 64] estimate the size of the outbreak in Iran on 25 February to be approximately 17,000 (4,800 - 44,500). The study in [63] uses the method developed by [65] 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 [66]), proportion of international travels who are residents of Iran [67], and the 'infectious disease vulnerability index', a proxy for measuring the country's ability to detect cases. Apart from Imam Khomeini airport (the airport 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 [64] uses a Binomial sampling method [68] to estimate the outbreak size based on 5 exported cases to Oman, UAE, and Lebanon. They use the VariFlight platform [69] to estimate the weekly number of passengers to Iran. Although we cannot independently verify those numbers, our estimates are fairly similar (see Table 1). The method used in [64] demands fewer parameters and produces similar results as [63], 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.

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 1 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.

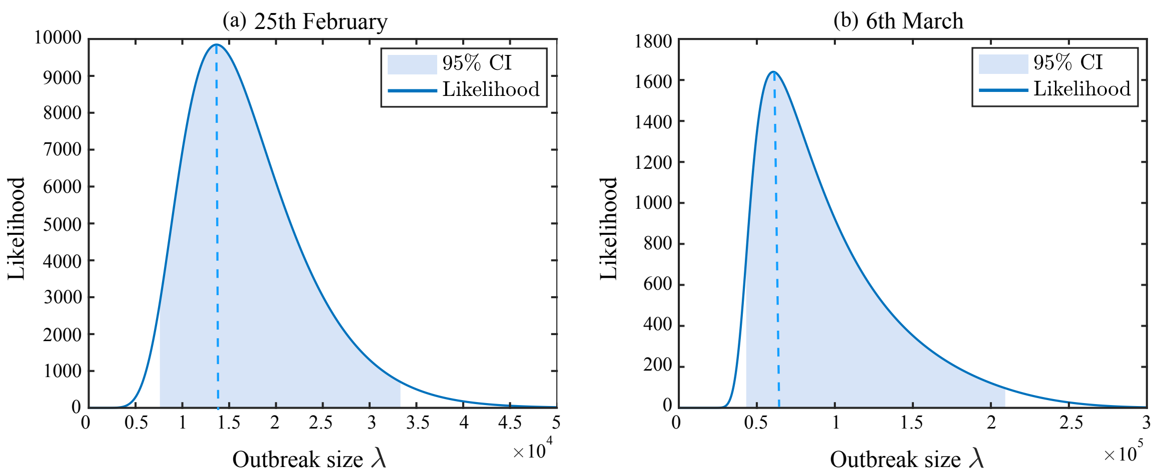


Figure D.3: Likelihood function for (a) 6 exported cases to Oman, Lebanon, and Kuwait on Feb 25 and (b) 28 cases to China on Mar 6. Blue dashed line represents the maximum likelihood value and the shaded area is the 95% confidence interval.

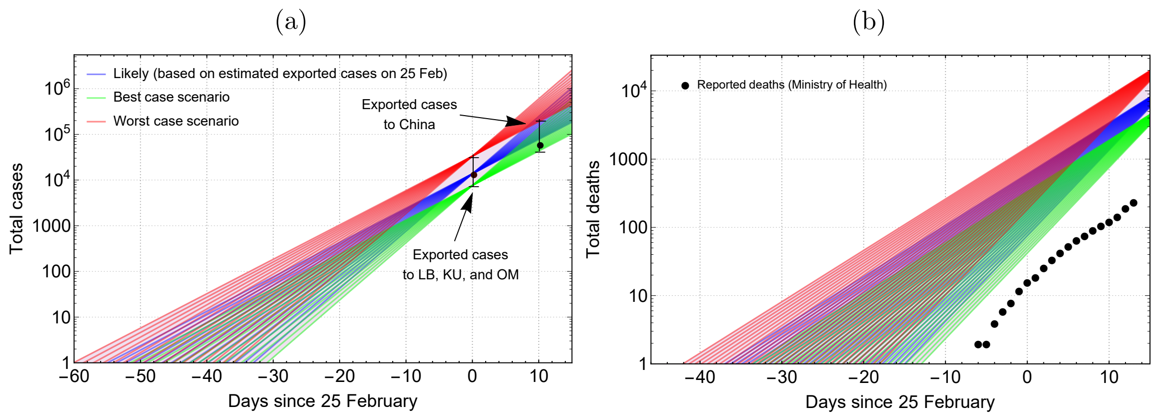


Figure D.4: The likely (blue), worst case (red), and best case (green) scenario of the outbreak size in Iran (left) and total deaths since early Jan to Mar 8.

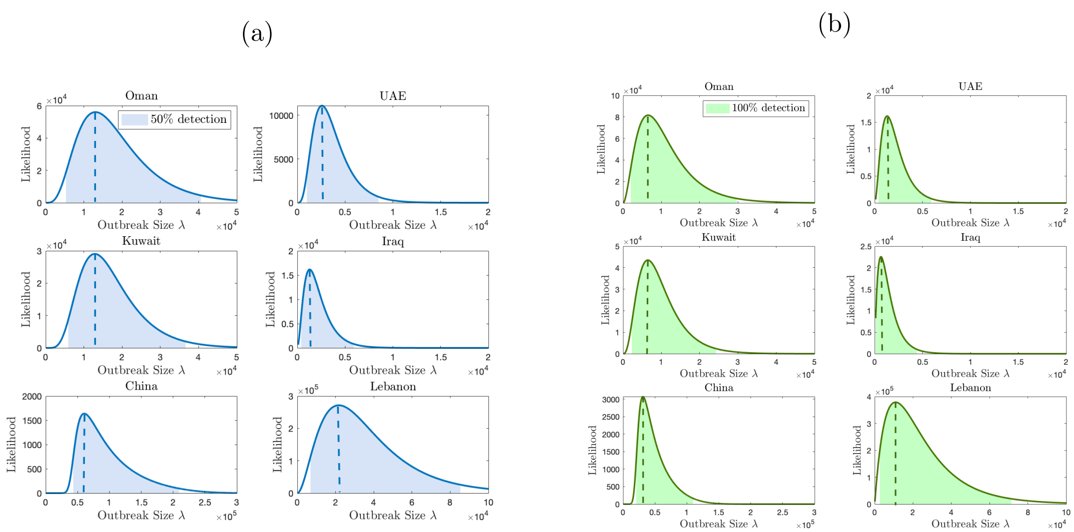


Figure D.5: Likelihood estimation of outbreak size in Iran based on flight travel data assuming (a) 50% (b) 100% of cases detected at the airports in Oman, UAE, Kuwait, Iraq, and Lebanon on Feb 25 and China on Mar 6. Dotted lines represent the maximum likelihood values and shaded areas show the 95% confidence intervals.

E Assessing the importation risk from China

Assuming that only pre/asymptomatic cases travel, the probability that an individual on the plane is infected is $p(t) = \sum_{k=0}^{\infty} 2(1 - S(k))i(t + k)/N$ where $i(t)$ is the incidence in a given province at time t with a total population size of N and $S(k)$ is the cumulative of the incubation time period and is assumed to be a log-normally distributed number with mean 4.8 days and standard deviation 1.9 days [70; 71]. The factor of 2 is for correcting the true incidence having an additional %50 asymptomatic cases [22]. We further allow a fraction, f, of the passenger flying to Tehran to come from Hubei province where f is the fraction of passengers from airports in Hubei (compared to the 100 busiest airports in China in 2018 ordered by total passenger traffic, according to Civil Aviation Administration of China statistics). To calculate the expected number of imported cases of COVID-19 to Iran from mid-January to mid-March, we find

$$\sum_{i=(\text{Beijing, Shanghai, and Guangdong})} \sum_{t=0}^{60} (1 - f)c_i p_i(t) + f c_{\text{Hubei}} p_{\text{Hubei}}(t) \approx 0.2 \ll 1$$

where $t = 0$ corresponds to Jan 19 and c_i is the number of passengers flying from province i per day.

F Intervention policy announcements and model parameters for the SEIR model

Table A3: Timeline of public policies and mitigation strategies in Iran

Policies and main events	Description	Date effective
School closure in Tehran and Qom ordered	After the first few reported deaths, the government ordered to close the universities and schools in the affected cities. We assume this had an overall 20% impact on $M(t)$.	02/20/2020 [72; 73]
Disinfecting and cleaning of public transport and schools	Disinfection of schools and public transport in multiple cities.	02/24/2020 [74–76]
Closure of sports complexes and gyms	Sporting facilities closed and cancelling or postponing sporting events in Zanjan.	02/24/2020 [77]
School closures in multiple cities	Schools closed in multiple other cities until further notice	02/24/2020 [78–80]
Social distancing encouraged and elderly advised to stay home	The elderly and people with underlying conditions are recommended to reduce their social contacts. We assume this had an overall 30% and 20% impact on ζ_a for > 60 and between 30 and 60 age-groups, respectively.	03/05/2020 [81]
Nationwide 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.	03/05/2020 [82]
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 assume this had an extra 60% impact on mitigation measures $M(t)$.	03/20/2020 [83–85]
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 $M(t)$.	04/05/2020 [12; 13]

Table A4: Model parameters

Parameter	Description	Expected value	Reference
t_l	Latency from infection to infectiousness	4 days	[70],[86]
t_i	Time to recover/falling severely ill	5 days	[87]
t_h	Time to recover/move to a critical state for a severe patient	8 days	[88]
t_c	Time to stabilise/die for a critical patient	10 days	[89]
m_a	Fraction of sub-clinical cases in age-group a	Table A5	[15]
c_a	Fraction of severe cases that turn critical in age-group a	Table A5	[15]
f_a	Fraction of critical cases that are fatal in age-group a	Table A5	—
R_0	Reproduction number	4.5 - 6.5	[90]
$M(t)$	Mitigation measures at time t	Table A3	—

Table A5: 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%

*National census in 2011 [45].

G Data from 31 provinces in Iran

Deaths during winter 2019-20 in 31 provinces in Iran

Annual winter deaths report (2013-2020)						
Province\year	1398(2019-2020)	1397(2018-2019)	1396(2017-2018)	1395(2016-2017)	1394 (2015-2016)	1393(2014-2015)
East Azerbaijan	5,762	5,880	5,725	5,648	5,858	5,768
West Azerbaijan	4,023	4,069	3,913	3,721	3,678	3,987
Ardabil	1,770	1,825	1,625	1,592	1,581	1,971
Isfahan	6,681	6,414	6,168	5,808	6,012	5,907
Alborz	3,062	2,972	2,571	2,266	2,417	2,149
Ilam	629	671	702	604	656	929
Bushehr	1,174	1,206	1,175	1,134	1,172	1,150
Tehran	16,676	15,644	15,528	14,244	14,302	14,756
Chahar Mahaal and Bakhtiari	1,054	992	1,087	1,014	1,069	1,963
Khorasan South	978	967	1,048	942	1,007	1,041
Khorasan Razavi	7,825	7,854	8,330	7,899	7,693	7,359
Khorasan North	1,229	1,183	1,210	1,114	1,100	1,111
Khuzestan	5,546	5,522	5,750	5,426	5,223	5,348
Zanjan	1,381	1,416	1,346	1,221	1,339	1,827
Semnan	946	927	917	803	847	815
Sistan and Baluchestan	3,374	3,627	3,888	3,931	3,249	3,054
Fars	5,776	5,841	6,101	5,624	5,798	5,641
Qazvin	1,720	1,617	1,515	1,491	1,417	1,466
Qom	2,301	1,558	1,546	1,519	1,498	1,389
Kurdistan	1,910	1,970	1,843	1,893	1,982	1,943
Kerman	3,671	3,903	3,768	3,356	3,267	3,721
Kermanshah	2,896	2,837	2,948	2,670	2,699	2,745
Kohgiluyeh and Boyer-Ahmad	707	748	797	680	676	768
Golestan	2,941	2,336	2,326	2,329	2,232	2,652
Gilan	5,936	4,461	4,310	4,085	4,125	4,048
Lorestan	2,285	2,305	2,158	2,023	2,013	2,227
Mazandaran	5,261	4,328	4,253	4,155	3,973	3,933
Markazi	2,101	1,973	2,046	1,830	1,904	2,153
Hormozgan	1,796	1,807	1,923	1,737	1,874	1,778
Hamadan	2,708	2,754	2,631	2,539	2,494	2,850
Yazd	1,328	1,275	1,250	1,174	1,156	1,204
Total	105,447	100,882	100,398	94,472	94,311	97,653

Number of confirmed cases in 31 provinces. *Data collected from the ministry of health's website (behdasht.gov.ir).**Ministry's website stopped sharing province data from Mar 23 onward and the data on Mar 2 and 3 was not available.

Data source*

Date	Qom	Teh	Maz	Alb	Sem	Gol	Qaz
19-Feb-20	2						
20-Feb-20	2						
21-Feb-20	7	4					
22-Feb-20	8	2					
23-Feb-20	7	4	1				
24-Feb-20	8	3					
25-Feb-20	16	8	2	2			
26-Feb-20	15	4	1		1		
27-Feb-20	7	38	7	3	3		
28-Feb-20	16	64	9	3	2	2	2
29-Feb-20	21	52	12	8	4	22	6
01-Mar-20	30	170	11	31	3		
02-Mar-20	N/A**						
03-Mar-20	N/A**						
04-Mar-20	101	253	9	19	8	9	25
05-Mar-20	32	56	50	61	22	4	31
06-Mar-20	137	61	180	129	54	67	65
07-Mar-20	145	126	305	3	24	58	2
08-Mar-20	17	266	14	2	37	13	29
09-Mar-20	27	140	13		45		40
10-Mar-20	39	169	253	32	1	4	10
11-Mar-20	53	256	32	45	63	9	27
12-Mar-20	42	303	79	74	40	25	42
13-Mar-20	42	303	192	6	38	21	12
14-Mar-20	32	347	17	134	47		47
15-Mar-20	84	251	72	67	41	25	22
16-Mar-20	19	200	96	49	34	42	40
17-Mar-20	29	273	59	116	33	30	26
18-Mar-20	22	213	61	76	60	20	53
19-Mar-20	31	137	58	61	16	9	47
20-Mar-20	36	220	84	95	52	15	8
21-Mar-20	20	232	28	55		6	59
22-Mar-20	17	249	36	60		10	29
Total	**1178**	**5098**	**1700**	**1177**	**645**	**391**	**669**

Data source*

Date	Esf	Frs	Hor	Koh	Cha	Bus	Gil
19-Feb-20							
20-Feb-20							
21-Feb-20							2
22-Feb-20							
23-Feb-20							2
24-Feb-20	2						2
25-Feb-20		1	1				2
26-Feb-20		2	1	2			1
27-Feb-20	8						23
28-Feb-20	10						25
29-Feb-20	12	8					17
01-Mar-20	13	8	2				28
02-Mar-20							
03-Mar-20							
04-Mar-20		14			1		35
05-Mar-20	118	19	2		8	3	80
06-Mar-20	150	13	6	4	5	2	91
07-Mar-20	96	14	11		2	3	70
08-Mar-20	80	9	5	3	5	2	2
09-Mar-20	37	29	17		5	1	28
10-Mar-20	17	18	17	2	11	7	
11-Mar-20	170	19	10		1		5
12-Mar-20	4	29		15	11	3	84
13-Mar-20	110	33	4				71
14-Mar-20	155	11				7	113
15-Mar-20	126	9	7	1		6	43
16-Mar-20	118	43	16	5	3	4	18
17-Mar-20	75	31	6		2	4	45
18-Mar-20	162	60	12	13	4	4	21
19-Mar-20	108	21	5	15	3	8	73
20-Mar-20	145	50	17	1	3	1	99
21-Mar-20	101	22		12	2		57
22-Mar-20	87	26	2		2		38
Total	**1979**	**505**	**148**	**73**	**68**	**55**	**1191**

Data source*

Date	Ard	Azs	Azg	Kur	Zan	Mar	Ham
19-Feb-20							
20-Feb-20						1	
21-Feb-20							
22-Feb-20							
23-Feb-20						1	
24-Feb-20						2	1
25-Feb-20							
26-Feb-20						1	
27-Feb-20	5	2	1	1			1
28-Feb-20	1	4		2			
29-Feb-20	3	6		5		18	
01-Mar-20						44	3
02-Mar-20							
03-Mar-20							
04-Mar-20		9		4	6	8	
05-Mar-20	17	10	5	6	7	31	6
06-Mar-20	11	10	1	24	35	48	7
07-Mar-20	1	2	22	2	2	37	11
08-Mar-20	9	30	7	19	1	70	26
09-Mar-20	22	22	29	3	15	54	
10-Mar-20	19	17	4	12	11	27	17
11-Mar-20	1	29	27	6	22	31	7
12-Mar-20	14	7	19	9	17	88	12
13-Mar-20	15	97	25	9	20	48	10
14-Mar-20	33	59	31	5	31	115	7
15-Mar-20		36	18	27	32		11
16-Mar-20	30	35	31	31	15	14	12
17-Mar-20	19	78	34	2	25	47	6
18-Mar-20	10	84	42	22	20	23	13
19-Mar-20	16	58	26	11	26	47	3
20-Mar-20	35	55	27	15	44	17	5
21-Mar-20	6	72	14	7	35		5
22-Mar-20	19	57	28	16	28	36	7
Total	**289**	**813**	**395**	**238**	**394**	**882**	**175**

Data source*

Date	Khz	Krs	Lor	Ilm	Khr	Sis	Yaz
19-Feb-20							
20-Feb-20							
21-Feb-20							
22-Feb-20							
23-Feb-20							
24-Feb-20							
25-Feb-20					1		
26-Feb-20	3	1	1			2	
27-Feb-20		2	3		1		1
28-Feb-20	3						
29-Feb-20	4			1		1	4
01-Mar-20	9	2	8	1	17	2	
02-Mar-20							
03-Mar-20							
04-Mar-20	27	15	5	7	8	8	
05-Mar-20	1		11		53	5	
06-Mar-20		5	50	2		2	50
07-Mar-20	1	4	27	2	46	13	24
08-Mar-20	5	2	37		19		6
09-Mar-20	4	11	7	16	29		3
10-Mar-20	37	7	25	25	39	6	37
11-Mar-20	17	8	9	6	34	4	23
12-Mar-20	31	20	25	7	29	6	21
13-Mar-20	15	15	25	8	110	2	46
14-Mar-20	48	8	2	6	30	2	73
15-Mar-20	32	13	52	18	143	6	36
16-Mar-20	53	16		2		11	51
17-Mar-20	25	14	35	7	63	11	49
18-Mar-20	28	7	39	10	30	6	45
19-Mar-20	22	23	37	8	50	18	69
20-Mar-20	10	22	9	20	48	8	49
21-Mar-20	31	27	34	18	57	5	52
22-Mar-20	22	19	33	17	42	15	84
Total	**444**	**243**	**476**	**183**	**858**	**134**	**725**

Data source*

Date	Khs	Ker	Khn
19-Feb-20			
20-Feb-20			
21-Feb-20			
22-Feb-20			
23-Feb-20			
24-Feb-20			
25-Feb-20			
26-Feb-20			
27-Feb-20			
28-Feb-20			
29-Feb-20	1		
01-Mar-20	2	2	
02-Mar-20			
03-Mar-20			
04-Mar-20		14	1
05-Mar-20	9		2
06-Mar-20	3	7	12
07-Mar-20		15	8
08-Mar-20	26	1	1
09-Mar-20	-7	-5	10
10-Mar-20	14	4	
11-Mar-20	15	18	11
12-Mar-20	12	7	
13-Mar-20		12	
14-Mar-20	4		1
15-Mar-20	18	7	6
16-Mar-20	7	13	45
17-Mar-20	12	19	3
18-Mar-20	21	11	
19-Mar-20	5	14	21
20-Mar-20	18	12	17
21-Mar-20		8	1
22-Mar-20	15	8	26
Total	**178**	**169**	**165**

H List of flights to/from Iran with estimated number of passengers per week per airline per airport

Approximate number of passengers/week - based on number of flights and aircraft capacity						
Oman						
Origin	**Tehran**	**Isfahan**	**Shiraz**	**Mashhad**	**Total**	
Weekly passengers	1,610	0	1,050	0	2,660	
Oman Air	WY432/Emr 175 - Daily	NA	NA	NA		
Salam Air	OV545/A320N - Daily	NA	OV556/A320N - Daily	NA		
Qatar						
Weekly passengers	5,600	250	1,350	1,050	8,250	
Qatar Airways	QR491/A350 - 9 - Daily	QR471/A330 - Once a week	QR477/A320 - Daily	QR493/A320 - Daily		Source for Qatar: Qatarairways.com
Qatar Airways	QR483/A350 -9 - Daily					
Qatar Airways	QR499A350 -1 - Daily					
Iran Air			IR683/ A320 - Twice a week			
UAE						
Weekly passengers	12,230	0	450	750	13,430	
Emirates	EK972/B777 300 - Daily					Source for emirates: emirates.com
Emirates	EK980/B777 300 - Daily					
Mahan Air	W561/A340 - Daily					
Mahan Air	W563/A310 - Daily					

Mahan Air	W565/A310 - Daily					
Iran Air	IR658/A321 - Daily					
Qeshm Air	QB2202/A319 - 6 times a week					
Flydubai			FZ272/B737-800 - 3 times a week	FZ254/B737-800 - 5 times a week		source for flydubai: https://www.flydubai.com/en/
Kuwait						
Weekly passengers	725	300	450	2,550	4,025	
Kuwait Airways	KU516/A320 - roughly 4 times a week			KU512/A320 - 3 times a week		Source for Kuwait airlines: http://www.kuwaitairways.com/en
Iran Air	IR601/A321 - once a week	IR667/A319 - twice a week	IR665/A320 - 3 times a week	IR669/A321 - 4 times a week		
Ata Air				I36605/A320 - 3 times a week		
Jazeera Air				J9152/A320 - Daily		

Iraq					
Weekly passengers	11,888	0	0	4,366	16,254
Iraqi Airways	IA112/B737-800 Daily			IA118/B737-800 twice a day	
Iraqi Airways	IA114/B737-800 once a week				
Iraqi Airways	IA188/B737-800 once a week				
Iran Air	IR5323/A300 - Daily			IR5331/A300 3 times a week	

Iran Air	IR5301/A300 roughly 3 times a week				
Qeshm Air	QB2217/A320 Daily			QB2297/A320 3 times a week	
Qeshm Air	QB2205/A320 - 3 times a week				
Qeshm Air	QB2289/A320 Once a week				
Qeshm Air	QB2299/A319 - Daily				
Caspian Air	IV7912 & IV7914/MD83 - Daily			IV7928/MD83 3 times a week	
Caspian Air	IV7906/MD83 3 times a week				
Caspian Air	IV7926/MD83 4 times a week				
Mahan Air	W55062/A310 - 4 times a week				
Mahan Air	W55060/A310 4 times a week				
Mahan Air	W55058/A310 4 times a week				
Lebanon					
Weekly passengers	600	0	0	200	800
Iran Air	IR661/A300 3 times a week			IR663/A300 once a week	
China					
Weekly passengers	6,700				6,700
Mahan Air	W577/A340 Daily				
Mahan Air	W579/A340 5 times a week				

Mahan Air	W581/ A340/ 3 times a week				
Mahan Air	W587/A340 4 times a week				
China Southern	CZ6026/B737 5 times a week				
Total number of passengers before flight suspension					**52,119**

Bahrain App number of passengers					
Weekly passengers	250	250	250	250	1000
Air Arabia	G9202/A320- 5 times a week	G9218/A320 5 times a week to Lar	G9214/A320 5 times a week	G9206/A320 5 times a week	
There are no direct flights between Iran and Bahrain but as BBC Persian reported on 25/02/2020, most of the cases were travelling with Air Arabia. The number of passengers for Bahrain has been derived based on the number of flights out of Sharjeh International airport into gulf countries. This is not at all close to the actual figure and should not be reported as credible in the paper					

Number of passengers per week 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	6700	28

Flight schedule prior to suspension from Flightradar24. The estimate number of passengers is based on the usual aircraft used and the seating capacity of the planes. Flight numbers provided are unique and could be googled for more information. There are many sources that the schedule could be found from using the information provided here.

Aircraft app seating capacity*	
Model	Capacity
A320/A321	150
B737-800	200
B777 - 300	270
Emr 175	80
A350-900	250
A350-1000	300
A310	230
A319	160
A340	300
A300	200

* Seating capacity varies between airlines and these are approximate values with a max 10% tolerance

I Genomic metadata

We gratefully acknowledge the authors, originating and submitting labs of the sequences from GISAID's Database (www.gisaid.org/) on which part of the genetic analysis is based. The list is detailed below.

Strain	GISAID ID	Collection Data	Country	Location	Originating Lab	Sex	Author	Submitting Lab	Exposure History
Canada/BC_40860/2020	EPI_ISL_415583	2020-03-03	Canada		BCCDC Public Health Laboratory	Female	Harrigan et al	BCCDC Public Health Laboratory	Travel to Iran
Australia/NSW13/2020	EPI_ISL_413599	2020-03-04	Australia	Sydney	Centre for Infectious Diseases and Microbiology - Public Health	Male	Timms et al	NSW Health Pathology - Institute of Clinical Pathology and Medical Research; Westmead Hospital; University of Sydney	Travel to Iran
Australia/NSW05/2020	EPI_ISL_412975	2020-02-28	Australia	Sydney	Centre for Infectious Diseases and Microbiology Laboratory Services	Male	Eden et al	NSW Health Pathology - Institute of Clinical Pathology and Medical Research; Westmead Hospital; University of Sydney	Travel to Iran
Australia/NSW06/2020	EPI_ISL_413213	2020-02-29	Australia	Sydney	Centre for Infectious Diseases and Microbiology Laboratory Services	Female	Eden et al	NSW Health Pathology - Institute of Clinical Pathology and Medical Research; Westmead Hospital; University of Sydney	Travel to Iran
Australia/NSW11/2020	EPI_ISL_413597	02/03/2020	Australia	Sydney	Centre for Infectious Diseases and Microbiology- Public Health	Male	Lam et al	NSW Health Pathology - Institute of Clinical Pathology and Medical Research; Westmead Hospital; University of Sydney	Travel to Iran
Australia/NSW12/2020	EPI_ISL_413598	04/03/2020	Australia	Sydney	Centre for Infectious Diseases and Microbiology - Public Health	Male	Gray et al	NSW Health Pathology - Institute of Clinical Pathology and Medical Research; Westmead Hospital; University of Sydney	Travel to Iran
Canada/BC_37_0-2/2020	EPI_ISL_412965	2020-02-16	Canada		BCCDC Public Health Laboratory	Female	Harrigan et al	BCCDC Public Health Laboratory	Travel to Iran
Canada/BC_69243/2020	EPI_ISL_415577	2020-02-20	Canada		BCCDC Public Health Laboratory	Male	Harrigan et al	BCCDC Public Health Laboratory	Contact of case (Iran), no travel history
Canada/BC_25211/2020	EPI_ISL_415579	02/03/2020	Canada		BCCDC Public Health Laboratory	Female	Harrigan et al	BCCDC Public Health Laboratory	Travel to Iran
Canada/BC_66353/2020	EPI_ISL_415587	04/03/2020	Canada		BCCDC Public Health Laboratory	Male	Harrigan et al	BCCDC Public Health Laboratory	Travel to Iran
Canada/BC_65034/2020	EPI_ISL_415585	04/03/2020	Canada		BCCDC Public Health Laboratory	Female	Harrigan et al	BCCDC Public Health Laboratory	Contact of case (Iran), no travel history
Canada/BC_17397/2020	EPI_ISL_415580	2020-02-28	Canada		BCCDC Public Health Laboratory	unknown	Harrigan et al	BCCDC Public Health Laboratory	Travel to Iran
Canada/BC_13297/2020	EPI_ISL_415578	2020-02-27	Canada		BCCDC Public Health Laboratory	Male	Harrigan et al	BCCDC Public Health Laboratory	Travel to Iran
Canada/BC_02421/2020	EPI_ISL_415581	01/03/2020	Canada		BCCDC Public Health Laboratory	Male	Harrigan et al	BCCDC Public Health Laboratory	Travel to Iran
NewZealand/01/2020	EPI_ISL_413490	2020-02-27	New Zealand	Auckland	Auckland Hospital	Female	Storey et al	Institute of Environmental Science and Research (ESR)	Travel to Iran
Germany/BavPat2/2020	EPI_ISL_414520	02/03/2020	Germany	Munich	Bundeswehr Institute of Microbiology	Female	Walter et al	Bundeswehr Institute of Microbiology	Travel to Iran
USA/NY1-PV08001/2020	EPI_ISL_414476	2020-02-29	USA		MSHS Clinical Microbiology Laboratories	Female	Patel et al	MSHS Pathogen Surveillance Program	Travel to Iran
Pakistan/Gilgit1/2020	EPI_ISL_417444	04/03/2020	Pakistan	Gilgit		Female	Javed et al	Department of Healthcare Biotechnology	Travel to Iran
Austria/CeMM0004/2020	EPI_ISL_419657	03/03/2020	Austria		Center for Virology, Medical University of Vienna	Male	Popa et al	Bergthaler laboratory, CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences	Travel to Iran
Finland/FIN-266/2020	EPI_ISL_414646	04/03/2020	Finland		Department of Virology and Immunology, University of Helsinki and Helsinki University Hospital, Huslab Finland	Female	Smura et al	Department of Virology, Faculty of Medicine, University of Helsinki, Helsinki, Finland	Travel to Iran
Iran/HGRC-01-IPI-8206/2020	EPI_ISL_424349	2020-09-03	Iran	Iran	Human Genetic Research Center	unknown	Zeinali,S et al		First sequence from inside Iran