

Investigating the Dark Figure of COVID-19 Cases in Austria: Borrowing From the deCODE Genetics Study in Iceland

Rainer Hirk
WU Vienna University of
Economics and Business

Gregor Kastner
WU Vienna University of
Economics and Business

Laura Vana
WU Vienna University of
Economics and Business

Abstract

The number of undetected cases of SARS-CoV-2 infections is expected to be a multiple of the reported figures mainly due to the assumed high proportion of asymptomatic infections and to limited availability of trustworthy testing resources. Relying on the deCODE genetics study in Iceland, which offers large scale testing among the general population, we investigate the magnitude and uncertainty of the number of undetected cases COVID-19 cases in Austria.

We formulate several scenarios relying on data on the number of COVID-19 cases which have been hospitalized, in intensive care, as well as on the number of deaths and positive tests in Iceland and Austria. We employ frequentist and Bayesian methods for estimating the dark figure in Austria based on the hypothesized scenarios and for accounting for the uncertainty surrounding this figure.

Using data available on April 01, 2020, our study contains two main findings: First, we find the estimated number of infections to be on average around 8.35 times higher than the recorded number of infections. Second, the width of the uncertainty bounds associated with this figure depends highly on the statistical method employed. At a 95% level, lower bounds range from 3.96 to 6.83 and upper bounds range from 9.82 to 12.61. Overall, our findings confirm the need for systematic tests in the general population of Austria.

Keywords: Comparative Study, SARS-CoV-2, Uncertainty Quantification, Unreported Infections.

1. Introduction

The number of confirmed infections with the SARS-CoV-2 is a central figure which many studies and models rely on for further analysis and also for evaluating whether the social distancing measures have proven effective. The same figure is being reported by media outlets and is the main information directed to the general public.

However, reports and scientific studies from different countries estimate the true number of infections to be a multiple of the reported figures (Li, Pei, Chen, Song, Zhang, Yang, and Shaman 2020; Maugeri, Barchitta, Battiato, and Agodi 2020; Zhao, Musa, Lin, Ran, Yang,

Wang, Lou, Yang, Gao, He *et al.* 2020). The number of undetected cases, the so-called dark figure, is expected to be large mainly due to two reasons: i) the characteristics of the disease and ii) the testing strategy (Czypionka and Reiss March 19, 2020). A dangerous characteristic of the COVID-19 disease is that the population is highly susceptible due to lack of antibodies and that transmission is likely to be carried out by infected individuals who exhibit light to no symptoms. Several studies report that roughly 50% of the infections are asymptomatic, which makes the early detection and isolation of the infected problematic (see, e.g., Nishiura, Kobayashi, Miyama, Suzuki, Jung, Hayashi, Kinoshita, Yang, Yuan, Akhmetzhanov *et al.* 2020; Day 2020; Shahan March 21, 2020; Castelfranco March 16, 2020). This means that even if all the symptomatic infections were recorded, the dark figure would be at least as high as the number of confirmed infections. Moreover, the infected individuals with mild symptoms are likely not going to get in contact with the health care system and will also not be recorded in official statistics. Therefore, extensive testing can be a viable strategy for accurately estimating for the prevalence of the disease in the general population. Nevertheless, countries around the world are struggling to set up such a strategy due to e.g., costs, testing capacity and availability of testing kits. Moreover, country-wide screenings with PCR (polymerase chain reaction) tests only provide a current snapshot of the active infections on a target date. As people who were infected but do not excrete the virus at this specific date are tested negative, these screenings do not offer information on the total number of individuals who have been infected with SARS-CoV-2.

Reports estimated the number of infected in Italy to be around 3.5 times higher than reported as of February 29, 2020 (Tuite, Ng, Rees, and Fisman 2020). Slightly lower estimates have been given for Germany (Kekulé March 14, 2020). In Austria, the limited test capacity has been an important set-back in the government's testing strategy. The capacity of PCR testing kits has been relatively low with the test capacity increasing slowly relatively to the number of infections (Czepel March 26, 2020). Secondly, until around March 22, 2020, only people with contact to confirmed infected people or coming from high risk areas such as Italy or China were tested. These issues point to a substantial number of unrecorded infections in Austria. Previous estimations place the number of infected in Austria between 16 000 to 55 000 as of March 18, 2020 (Czypionka and Reiss March 19, 2020; Sturn March 27, 2020; ORF March 27, 2020). In the four day period between March 31, 2020 and April 03, 2020 a field study commissioned by the Austrian government will be conducted by the research institute SORA, where 2000 individuals will be tested for COVID-19 in Austria. The results of this study will shed light on the number of active infections in Austria by ensuring that the sample of tested individuals is representative for the general population, but it does not provide information on the accumulated number of SARS-CoV-2 infections in Austria (ORF March 30, 2020). For this purpose reliable and accurate i.e., sensitive and specific antibody tests would be necessary.

In the absence of an extensive field study, the attempts to quantify the dark figure of infections can only produce rough estimates as they rely on several assumptions and/or on results from different countries. The difficulty in estimating undetected infections arises mainly due to data quality issues. The available data is highly dependent on the testing and reporting strategies of the different countries. Comparisons are not straightforward as most of the countries apply different testing approaches. Italy, for example, mostly focuses on testing in hospitals with symptoms (Onder, Rezza, and Brusaferro 2020). This results in a high proportion of positive tests and also the mortality rates are apparently higher in the tested sample due to more severe cases being tested. The roughly 50% asymptomatic are not covered by this approach so one can assume that the true mortality rate is lower than reported. Other countries have delayed or performed very limited testing until very recently. This was the case in the US, where less than 1000 cases were reported until March 10, 2020. Since then the number has increased by a factor of 196 times as of March 31, 2020 (Dong, Du, and Gardner 2020).

While the data quality issue is one most countries are battling with, one can make use of studies in other countries which have a better testing strategy. Iceland can be seen as a pioneer

in this respect as they launched a large scale testing of general population (see [Government of Iceland March 21, 2020](#)). The bio-pharmaceutical company deCODE genetics launched a testing program, where they test people on a voluntary basis in order to get a better understanding about the spread of the coronavirus across the country. As of April 01, 2020, Iceland has tested a higher proportion of its citizens than almost any other country in the world (except the Faroe Islands) and through the deCODE genetics study is in a position to obtain a rough approximation of the unrecorded infections in Iceland ([Nardelli and Ashton March 21, 2020](#)).

The goal of this article is to obtain rough estimates of the dark figure of infections in Austria by borrowing information from the screening of the general population in Iceland. The approach we propose in this article consists of defining several scenarios where we hypothesize on the relationship between the observed figures in Iceland versus the observed figures in Austria. While this study follows a rather naive approach and disregards most of the complexity of the problem by making simplifying assumptions, it is also, to the best of our knowledge, the only one making use of the Icelandic case study in order to extrapolate results to other countries and should provide rough approximations of the unrecorded infections in Austria while serving as a starting point for further research.

This remainder of the article is organized as follows: Section 2 introduces the deCODE genetics study in Iceland. Section 3 provides a comparison of Austria and Iceland in terms of COVID-19 figures. The different scenarios and our approach for estimating the dark figure in Austria are presented in Section 4. Section 5 discusses the findings and Section 6 provides concluding remarks.

2. Study performed by deCODE genetics

From March 15, 2020 until March 31, 2020, 10401 individuals were tested by the medical research company deCODE genetics, who joined forces with the Icelandic authorities in performing a large scale SARS-CoV-2 testing in Iceland. The company offers screening on a voluntary basis among the general population, by offering a free coronavirus test to anyone who fills in an online form. This approach to testing leads to a higher representation in the sample of non-symptomatic non-quarantined individuals and the sample of tested people can be considered to be more representative of the whole Icelandic population than the sample of individuals tested by the authorities, who are mainly symptomatic or considered high risk. Therefore, assuming that the deCODE genetics study covers a representative sample would allow conclusions to be drawn about the prevalence of the disease in the general population. The tests performed by deCODE genetics, as well as the ones performed by the Department of Microbiology of the National University Hospital of Iceland (NUHI), use the technology of polymerase chain reaction (PCR), which is considered to date to be the most accurate for COVID-19 diagnosis.

The number of infections discovered by the deCODE genetics study in the above mentioned time frame lies at 84. This constitutes a prevalence rate in the tested sample of 0.81%. Assuming that the sample of deCODE genetics is representative for the whole Icelandic population, 0.81% can be used as an estimate for the prevalence of the COVID-19 disease in the general population with a 95% frequentist bootstrap confidence interval (CI) of [0.64%, 0.98%]. This would correspond to a number of 2941 infected people in Iceland (with a 95% CI of [2314, 3567]). When extrapolating from the deCODE genetics sample to the whole population, these numbers would imply that, as of March 31, 2020, the official statistic of 1136 infections reported by the NUHI is underestimating the number of cases by 1805 (95% CI: [1178, 2431]), with only 38.63% of the cases being recorded officially (95% CI: [31.85%, 49.09%]). This indicates that the number of infections is 2.59 times larger than officially reported (95% CI: [2.04, 3.14]).

Table 1: Statistics related to COVID-19 disease available as of April 01, 2020. Source: [Austrian Social Ministry](#) and [Government of Iceland](#).

	Austria	Iceland	Austria (PMP)	Iceland (PMP)
Deaths	146	4	16.40	10.98
Hospitalized	1071	35	120.30	96.12
Intensive care	215	11	24.15	30.21
Confirmed infections	10482	1136	1177.41	3119.73
Tests	55863	9115	6274.91	25031.99
Percentage confirmed	18.76%	12.46%		

3. Comparison between Austria and Iceland

Iceland has a population of 364134 (reported by <https://statice.is/> on January 1, 2020), while Austria’s population of 8902600 (according to [Statistik Austria](#)) is 24.45 times larger. In terms of health care systems, the two countries are in the top 15 health care systems in the world based on a standardized set of metrics on health system performance, with Austria being few places ahead of Iceland (around place 9).¹

Table 1 shows the number of deaths, number of hospitalized patients, among the hospitalized the number of patients in intensive care, the number of confirmed infections and the number of tests performed in each country both in absolute value and per million of population (PMP). Note that the reported statistics for Iceland include only the results of the NUHI testing for comparison purposes, as both the Austrian and the NUHI testing focus mostly on symptomatic cases.

While in Iceland – as of April 01, 2020 – 4 people diagnosed with COVID-19 have died, at the same time in Austria 146 fatalities have been recorded. Among the active cases, 35 are currently hospitalized in Iceland and 1071 in Austria and 11 are reported to need intensive care in Iceland, compared to 215 in Austria. The per million of population figures can be compared among the two countries, with Iceland having a relatively higher number of intensive care patients, higher confirmed infections and roughly 4 times more tests compared to Austria. In Austria 18.76% of the tests are positive, in Iceland we observe a percentage of positive tests of 12.46%. This might be due to the relatively higher number of conducted tests in Iceland. Time series of the tests and the confirmed infections for the two countries are shown in the Appendix, Figure 7 and Figure 8.

Figure 1 shows the age distribution of the population in the two countries. While Iceland has relatively more people in the younger age groups, the percentage of people above 45 is relatively higher in Austria with the largest different being in the older adult population above 65.

When comparing the age distribution of the infected people presented in Figure 2, we find similar prevalence rates among most age groups. Only for the age groups from 35–44 and 65+ clear differences are observed. Within the 35–44 interval the proportion of infected people is higher in Iceland, while in Austria relatively more people aged 65 and above are infected among the tested individuals. Partly, this can be explained by Austria having a larger proportion of older adult population but could also highlight the different testing strategies of the two countries, where in Austria detection in non-risk groups might be less likely. Note that the number of infections in Iceland is given by both the deCODE genetics and the NUHI samples.

All data sources were updated on April 01, 2020 at 3:00 pm UTC+2.

¹Source: [Tandon, Murray, Lauer, and Evans \(2000\)](#) and more recently <https://worldpopulationreview.com/countries/best-healthcare-in-the-world/>

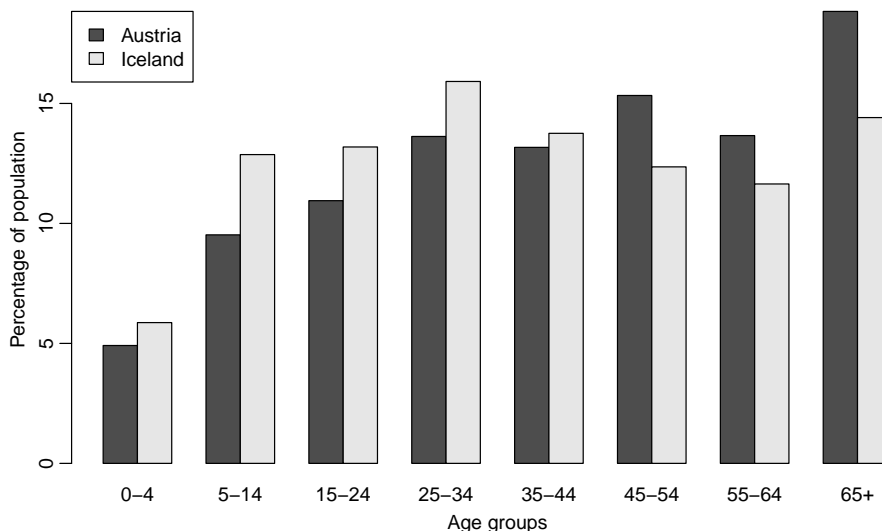


Figure 1: Age distribution in the population.

4. Estimation of the dark figure in Austria

When trying to estimate the dark figure of SARS-CoV-2 infections in Austria, we consider five different scenarios based on available figures from the ministries of health of the two countries. In each of the scenarios we estimate the prevalence of COVID-19 in Austria by multiplying the estimated prevalence of the deCODE genetics study by a procedure-specific multiplier k . The scenarios focus on the number of hospitalized cases, the number cases in intensive care, the observed deaths due to COVID-19 and the ratio of positive SARS-CoV-2 tests.

4.1. Defining the scenarios

1. Scenario I: Ratio of hospitalized per million inhabitants

In the first scenario the multiplier is calculated by the number of hospitalized people in Austria divided by number of hospitalized people in Iceland (per million of population). As of April 01, 2020, 1071 (120.3 per million) COVID-19 cases are hospitalized in Austria, while 35 (96.12 per million) are hospitalized in Iceland. This gives a multiplier of 1.25 for this scenario.

In this scenario, we estimate a number of 89988 infected people as of April 01, 2020. This would mean that 79506 people were infected but not recorded and only 11.65% of the estimated infections were recorded. This would indicate that the number of infections is 8.59 times larger than officially reported.

2. Scenario II: Ratio of intensive care patients per million inhabitants

In the second scenario we calculate the multiplier by the number of COVID-19 cases in intensive care in Austria divided by number of COVID-19 cases in intensive care in Iceland. As of April 01, 2020, 11 (30.21 per million) persons are in intensive care in Iceland, while 215 (24.15 per million) are in intensive care in Austria. This gives a multiplier of 0.8 for this scenario. We do not account here for the difference in intensive care units (ICUs) among the two countries, as this stage in the pandemic the ICUs are not operating at full capacity.

This gives us an estimated number of infections of 57479. This would mean that 46997 people were infected but not recorded and only 18.24% of the estimated infections were

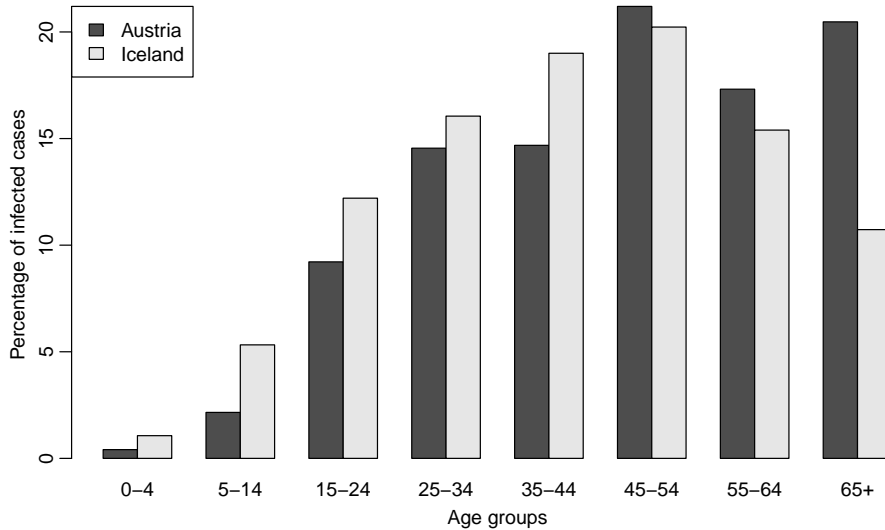


Figure 2: Distribution of infections among the age groups. For Iceland infections recorded by both deCODE genetics and NUHI are considered.

recorded as of April 01, 2020. This would indicate that the number of infections is 5.48 times larger than officially reported.

3. Scenario III: Ratio of deaths per million inhabitants

In this scenario the multiplier is given by the ratio of deaths in Austria and deaths in Iceland. In both countries a deceased person who has previously tested positive for the COVID-19 disease is counted as a death in the official statistics.² As of April 01, 2020, 4 (10.98 per million) persons have died in Iceland, while 146 (16.4 per million) deaths have been observed in Austria. This gives a multiplier of 1.49 for this scenario.

The estimated number of infections in Austria is 107339 in this scenario. Thus, as of April 01, 2020, 96857 people were infected but not recorded in the official statistics and only 18.24% of the estimated infections were recorded. This would indicate that the number of infections is 10.24 times larger than officially reported.

4. Scenario IV: Similar prevalence in the population as in Iceland

We assume the percentage of infected population in Austria is equal to the one in Iceland and use the prevalence of 0.81% estimated from deCODE genetics data to draw conclusions about the dark figure in Austria. This assumption implies that the disease had a similar development in the two countries.

When assuming a similar prevalence rate as in Iceland, we estimate a number of 71899 infected people as of April 01, 2020. This means that 61417 people were infected but not recorded and only 14.58% of the estimated infections were recorded. This indicates that the number of infections is 6.86 times larger than officially reported.

5. Scenario V: Ratio of the percentage positive SARS-CoV-2 tests

Throughout this scenario, we analyse and compare the testing behavior of both countries. A closer look at the testing results shows that Iceland has conducted roughly 3.99 times more tests per million people compared to Austria, but confirmed only 2.65 times as many infections per million people compared to Austria. Under the (strong) assumption that the probability of obtaining a positive test result does not decrease with

²Sources: <https://www.sozialministerium.at/Informationen-zum-Coronavirus/Coronavirus---Haeufig-gestellte-Fragen.html> and <https://www.covid.is/data>

Table 2: Estimated number of infections, the dark multiplier of confirmed infections (DM – the ratio of the estimated infections and the confirmed infections), 95% frequentist bootstrap confidence intervals for the number of infections and DM, and the multiplier k for each of the scenarios in Austria.

Scenario	Est. Infections	95% CI	DM	95% CI	k
Sc.I: hospitalized	89988	[70822, 109155]	8.59	[6.76, 10.41]	1.25
Sc.II: intensive care	57479	[45237, 69722]	5.48	[4.32, 6.65]	0.80
Sc.III: deaths	107339	[84477, 130201]	10.24	[8.06, 12.42]	1.49
Sc.IV: similar prevalence	71899	[56585, 87212]	6.86	[5.40, 8.32]	1.00
Sc.V: pos. test ratio	109926	[86513, 133339]	10.49	[8.25, 12.72]	1.53

increasing testing size, we take the ratio of the percentage of positive tests in Austria and the percentage of positive tests in Iceland of 1.53 as a multiplier of the prevalence rate.

As of April 01, 2020, we estimate a number of 109926 infected people for this scenario. This means that 99444 people were infected but not recorded and only 9.54% of the estimated infections were recorded. This would indicate that the number of infections 10.49 times larger than officially reported.

4.2. Assessing the uncertainty in the estimation of the dark figure in a simulation study

There are numerous sources of uncertainty in estimating the dark figure, some of which can be appropriately accounted for in a statistical modeling framework: i) the plausibility (likelihood) of the different scenarios, ii) the uncertainty related to the true value of the multiplier k , iii) uncertainty in the prevalence of the disease in Iceland.

In this section we address these points in the following way. Regarding i), the assignment of probabilities to each of the scenarios is a challenging task due to disagreement of experts and the lack of data availability. Therefore we resort to simulation and in a first experiment, simulate weights for the different scenarios from a prior distribution and assume the prevalence of the disease in Iceland to be fixed. This allows us to obtain a distribution of the expected value for the number of infections (and DM) over all scenarios. In a second experiment, we propose a Bayesian statistical model for the prevalence rate in Iceland. Moreover, we shift focus from the different scenarios to the multiplier k and propose a prior distribution for this quantity.

Due to the limited testing availability in Austria, the number of recorded infections is an unrealistic indicator when comparing the prevalence rate of Iceland and Austria. Testing data in Austria cannot be assumed to represent a picture of the whole population and but rather it displays the infections in a subgroup of the population. Currently, the most comparable, trustworthy and informative indicators are observed figures in the hospitals like the number of hospitalized people, the number of people in intensive care and the observed number of deaths. Other meaningful scenarios are: assuming a similar prevalence rate like in Iceland as the first recorded infections took place in both countries at a similar point in time; using the ratio of the percentage of positive tests as a reasonable estimation/guess of a multiplier of the prevalence rate. Therefore, we focus on these five scenarios for our statistical modeling purposes.

Simulating the probability of the five scenarios

We choose a Dirichlet prior on the probability of the five scenarios:

$$p_1, p_2, p_3, p_4, p_5 \sim Dir(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5),$$

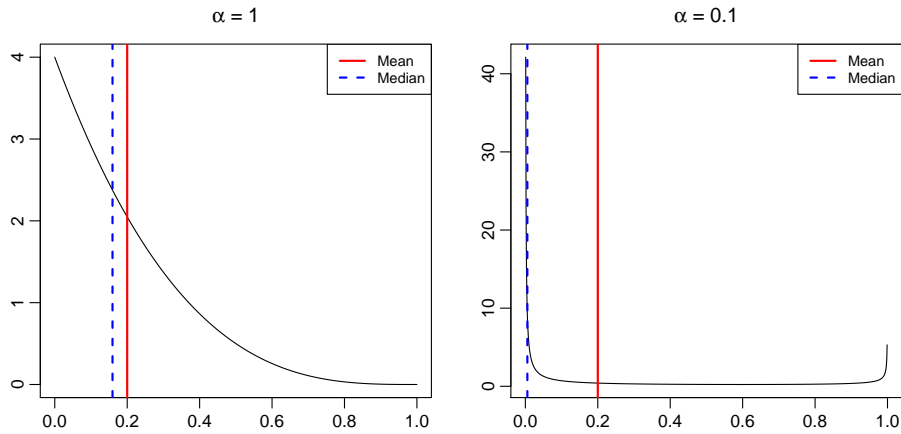


Figure 3: Marginal prior densities $p_i \sim \mathcal{B}(\alpha_i, \sum_{j \neq i} \alpha_j)$, $i = 1, \dots, 5$, for $\alpha = 1$ (left) and $\alpha = 0.1$ (right).

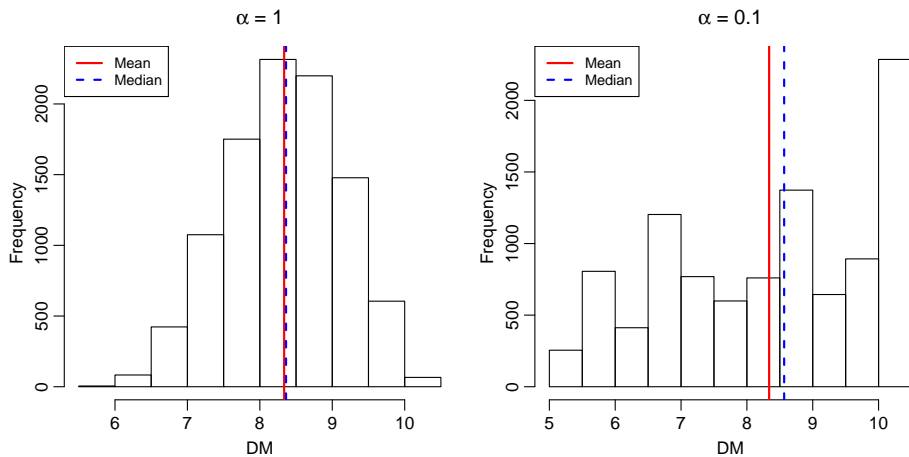


Figure 4: Histograms of the estimated DM for 10000 simulated scenario probabilities from the Dirichlet distribution with hyper-parameters $\alpha = 1$ and $\alpha = 0.1$.

where $\alpha_1 = \alpha_2 = \dots = \alpha_5 \equiv \alpha$ is a hyper-parameter to be chosen. Setting all the parameters equal implies that in the long run we believe the scenarios to be equally likely. In the simulation we quantify the variability of the DM with respect to the variability in the probability of the scenarios. In this subsection we consider the prevalence in Iceland to be fixed to 0.81% in this experiment.

We choose two values of hyper-parameters of the Dirichlet distribution: $\alpha = 1$ which corresponds to assigning evenly distributed probabilities to the scenarios, and $\alpha = 0.1$ which corresponds to assigning most mass to one of the scenarios. For a visualization of the implied marginal prior distributions of p_i , $i = 1, \dots, 5$, see Figure 3.

In our simulation, we draw 10000 samples from the Dirichlet distribution and weigh the estimated number of infections in each scenario by the sampled weights in order to obtain the expected value of the DM. For both values of the hyper-parameters we obtain the same mean of 8.34 for the DM. Figure 4 shows the distribution of the DM under variability of the scenarios. More mass is assigned to the tails of the distribution under the case $\alpha = 0.1$. This is due the fact that for this set of hyper-parameters the more extreme scenarios get more weight because the probability of one of the scenarios is close to one and the others close to zero. The modes correspond to the estimated multipliers calculated in the scenarios.

Statistical modeling of the prevalence rate

The prevalence of the disease in Iceland's general population has been considered fixed in the previous experiment. In order to account for the uncertainty surrounding this quantity, we make use of the beta-binomial model. In this Bayesian model, a Beta prior is assumed on the prevalence rate in Iceland p^{ISL} :

$$p^{\text{ISL}} \sim \mathcal{B}(a, b),$$

where a and b are hyper-parameters to be chosen. The prior distribution on the prevalence of the disease should assign high mass on small values of p^{ISL} . Suitable hyper-parameters could be e.g., $a = 1$ and $b = 50$ which implies a prior mean and a prior standard deviation of roughly 0.02.

The number infected people in the deCODE genetics sample $N_{\text{inf}}^{\text{ISL}}$ is then modeled as a binomial distribution with size equal to the number of tested individuals by deCODE genetics $N_{\text{test}}^{\text{ISL}}$ and probability p^{ISL} :

$$N_{\text{inf}}^{\text{ISL}} | p^{\text{ISL}} \sim \text{Bin}(N_{\text{test}}^{\text{ISL}}, p^{\text{ISL}}).$$

We use the deCODE genetics results in order to estimate the true number of infected in Austria. We assume that $N_{\text{inf}}^{\text{AUT}}$, the true number of infected individuals in Austria is a binomial random variable with size equal to the number of inhabitants of Austria $N_{\text{pop}}^{\text{AUT}}$ and probability which is a multiple of the probability in Iceland $k \cdot p^{\text{ISL}}$:

$$N_{\text{inf}}^{\text{AUT}} | p^{\text{ISL}} \sim \text{Bin}(N_{\text{pop}}^{\text{AUT}}, k \cdot p^{\text{ISL}}), \quad (1)$$

where k is a multiplier. For k fixed, this would represent the posterior beta-binomial predictive distribution, where the posterior probability of p^{ISL} would be given by:

$$p^{\text{ISL}} | N_{\text{test}}^{\text{ISL}}, N_{\text{inf}}^{\text{ISL}} \sim \mathcal{B}(a + N_{\text{inf}}^{\text{ISL}}, b + N_{\text{test}}^{\text{ISL}} - N_{\text{inf}}^{\text{ISL}}). \quad (2)$$

However, there is also uncertainty coming from the value of k . We consider here two prior probability distributions for the multiplier k .

Prior 1: Discrete probability We consider each of the five values of k presented in Table 2 equally likely.

$$P(k = 1.25) = P(k = 0.8) = P(k = 1.49) = P(k = 1) = P(k = 1.53) = 1/5.$$

Prior 2: Mixture of gamma distributions It is unrealistic to assume that k only has 5 possible values. A more realistic assumption is that k is a continuous variable coming from a mixture of five distributions which have their mean at the values computed in the different scenarios. Moreover, the variance of the different components should differ among the scenarios, due to the few observations observed in some of the scenarios. For example, the value of k for the third scenario is computed by using data on only 4 deaths in Iceland, so this component can be expected to have higher variance as one additional death can change the multiplier significantly. Given that the multipliers are positive values, we assume a gamma distribution for the components:

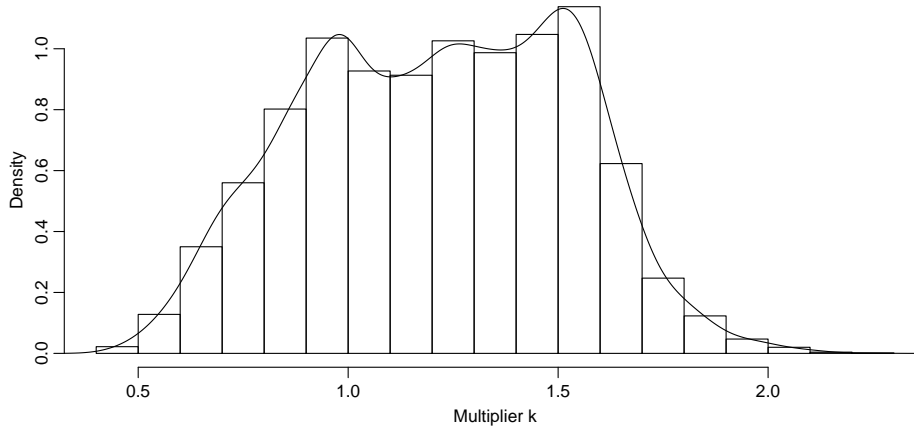
$$p(k) = \sum_{s=1}^S \phi_s \mathcal{G}(\alpha_s, \beta_s),$$

where S denotes the number of scenarios, ϕ_s denote the probability of scenario s , $\alpha_s > 0$ is the shape and $\beta_s > 0$ is the rate parameter of the gamma distribution for scenario s . Table 3 presents the chosen hyper-parameters each of the gamma distributions corresponding to the scenarios. We choose the hyper-parameters α_s and β_s such that the mean of the distribution equal the value of the multiplier estimated in each of the scenarios and the variance equals predetermined values.

The mixture of gamma prior for the multiplier k is displayed in Figure 5.

Table 3: Parameter choices for the gamma distributions in the mixture model.

Scenario	ϕ_s	α_s	β_s	$\mu_s = \alpha_s/\beta_s$	$\sigma_s = \sqrt{\alpha_s}/\beta_s$
Sc.I: hospitalized	0.20	156.65	125.16	1.25	0.10
Sc.II: intensive care	0.20	31.96	39.97	0.80	0.14
Sc.III: deaths	0.20	44.58	29.86	1.49	0.22
Sc.IV: similar prevalence	0.20	100.00	100.00	1.00	0.10
Sc.V: pos. test ratio	0.20	233.76	152.89	1.53	0.10

Figure 5: Mixture of gamma prior for the multiplier k .

Sampling After simulating 10000 values for k from the priors introduced above, we sample p^{ISL} from the distribution in Equation 2. Finally, $N_{\text{inf}}^{\text{AUT}}$ is simulated from the binomial distribution given in Equation 1 where we replace k and p^{ISL} with the sampled values. Figure 6 provides the predictive distributions for the DM with the two prior specification. The average DM lies at 8.39 for prior 1 and at 8.38 for prior 2.

4.3. Results

Table 4 displays a summary of the ratio of estimated infections and confirmed infections with corresponding intervals for quantifying the uncertainty for all five modeling approaches. We find that the mean estimates of the DM are between 8.3 and 8.4 for all five approaches. In the frequentist approach, where we average over the equally weighted scenarios, we find a mean estimate of 8.33 with a 95% frequentist bootstrap confidence interval of [6.56, 10.11]. When simulating the probabilities of the five scenarios with a Dirichlet distribution with $\alpha = 1$, we find similar results like in the frequentist approach, while a hyper-parameter of $\alpha = 0.1$ increases the uncertainty bounds as more mass is assigned to one of the scenarios in each of the 10000 simulations. When accounting for both the uncertainty in the prevalence in Iceland and the value of the multiplier in a beta-binomial model, the mean estimates are 8.39 for prior 1 and 8.38 for the mixture of gamma prior on the multiplier k . As expected, the range of the credible intervals increases as more uncertainty is being accounted for. Especially for the mixture of gamma prior, where a plausible continuous range for the multiplier k is being accounted for, the uncertainty bounds are the widest. This final approach also incorporates uncertainty regarding the calculation of the multipliers in the scenarios.

5. Discussion

The figures used in computing the multipliers between Austria and Iceland in the different

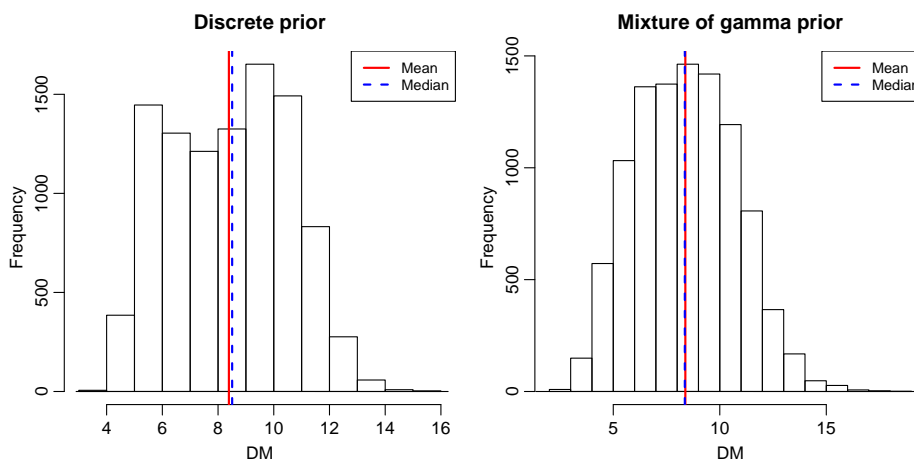


Figure 6: Histograms of the estimated DM for 10000 simulated scenarios with discrete probabilities and the mixture of gamma prior for k .

Table 4: DM with uncertainty bounds (95% bootstrap confidence intervals for the frequentist approach and 95% highest density intervals for the Bayesian approaches).

Approach	DM	95% Intervals
avg. frequentist approach	8.33	[6.56, 10.11]
Dirichlet $\alpha = 1$	8.34	[6.83, 9.82]
Dirichlet $\alpha = 0.1$	8.34	[5.57, 10.49]
Beta-Binomial – discrete probability	8.39	[4.71, 12.01]
Beta-Binomial – mixture of gamma	8.38	[3.96, 12.61]

scenarios are compared on the same date, April 01, 2020, which assumes the same temporal development of the disease in the two countries. Following the suggestion of the reviewers, we investigated the robustness of the results when comparing the figures at different time points, rather than contemporaneously. For this purpose, we assumed a three day lag of the Icelandic figures, as the first infection in Iceland was recorded on February 28, 2020 while the first confirmed infection in Austria happened on February 25, 2020. This is also confirmed by the modes of active infections in the two countries. We computed the multipliers by taking the ratios of the Austrian figures on March 29, 2020 and the Icelandic figures on April 01, 2020 and obtain an average multiplier of 8.75. For Austrian figures on April 01, 2020 and the Icelandic figures on April 04, 2020 the average multiplier lies at 8.17.

At the time of the revision of the paper, the aforementioned field study conducted by the research institute SORA has been completed on 1544 Austrian individuals (this corresponds to 173 tests per million inhabitants). Among these, 0.33% active infections were confirmed (Ogris and Hofinger April 10, 2020). This figure provides a snapshot for active SARS-CoV-2 infections for the period April 01, 2020 to April 06, 2020. Note, however, that the SORA study is not directly comparable with the deCODE genetics study for several reasons. First, the sample size is much larger in the deCODE genetics study, with 16886 tests (46373 tests per million inhabitants) performed by April 06, 2020. Second, the sample selection is voluntary in the deCODE study but more complex in the SORA study (Ogris and Hofinger April 10, 2020). Third, the studies are conducted over different time-periods, with the deCODE genetics study being still ongoing since March 15, 2020. This is important because infected who recovered before the beginning of each testing period are not accounted for. With the population being widely compliant with the social-distancing measures imposed by the Austrian government on March 13, it is plausible that the number of newly infected started to decrease immediately.

Hence, we suspect a higher prevalence rate in Austria one or two weeks before April 06.³

6. Conclusion

Our findings suggest that the number of undetected infections of SARS-CoV-2 in Austria, both active and recovered, is a multiple of the reported figures, with an average ratio of estimated infections and confirmed infections of around 8.36. However, the uncertainty surrounding this estimate is significant. We employ several frequentist and Bayesian methods to appropriately account for the uncertainty and find that plausible estimates may well lie in the interval [3.96, 12.61]. The analysis relies on the deCODE genetics study in Iceland, which offers large scale testing among the general population. We investigate the magnitude and uncertainty of the dark figure of the SARS-CoV-2 infections in Austria by formulating several scenarios relying on data on the number of COVID-19 cases which have been hospitalized, in intensive care, as well as on the number of deaths and positive tests in Iceland and Austria.

We provide simulation experiments in a statistical framework for accounting for different sources of uncertainty such as the likelihood of the different scenarios, the prevalence of COVID-19 in Iceland and the uncertainty regarding the multiplier between the prevalence in Austria and the prevalence in Iceland.

One of the primary limitations of the proposed framework is the assumption that the deCODE genetics study in Iceland consists of a representative sample of the entire population. It is to be noted that the application for testing occurs on a voluntary basis through an online form, which can be a source of self-selection bias leading to the underrepresentation of some age groups or to an over-proportional percentage of infections in the sample. A second limitation is the reliance of the study on key figures reported by the two governments where a lag in reporting is likely. Moreover, while the data is mostly comparable, there might still be differences in the definitions of the reported figures. Third, we assume a static setting for modeling the prevalence rate in both countries, which does not control for time span over which the deCODE genetics study has recorded the observations. Last, other factors such as the handling of the COVID-19 disease might to some extent vary among the countries. While some of the performed experiments are able to account at least partially for these issues, we stress that our results remain dependent on the assumptions made throughout the analysis.

Acknowledgments

We thank Matthias Templ (editor) as well as Georg Heinze (reviewer) and Peter Filzmoser (reviewer) for their insightful and timely comments and suggestions which helped to improve the paper and in particular sharpen the discussion about its limitations.

The authors acknowledge funding from the Austrian Science Fund (FWF) for the project “High-dimensional statistical learning: New methods to advance economic and sustainability policies” (ZK 35), jointly carried out by WU Vienna University of Economics and Business, Paris Lodron University Salzburg, TU Wien, and the Austrian Institute of Economic Research (WIFO).

References

Castelfranco S (March 16, 2020). “The Hard Lessons of Italy’s Devastating Coronavirus Outbreak.” <http://www.rfi.fr/en/europe/20200316-the-hard-lessons-of-italy-s-devastating-coronavirus-outbreak/> [Accessed March 31, 2020].

³A similar pattern can be observed for Iceland, where five-day rolling window averages of the infections recorded by deCODE genetics started to decrease around the beginning of April.

- Czepel R (March 26, 2020). “Neue Tests in zwei Wochen möglich.” <https://science.orf.at/stories/3200426/> [Accessed March 31, 2020].
- Czypionka T, Reiss M (March 19, 2020). “Ein Blick in die Glaskugel. Welche drei Faktoren es für das Verständnis der SARS-CoV-2 Fallzahlen braucht.” <https://www.ihs.ac.at/publications-hub/blog/beitraege/infizierte-coronavirus/> [Accessed March 31, 2020].
- Day M (2020). “COVID-19: Identifying and Isolating Asymptomatic People Helped Eliminate Virus in Italian village.” doi:10.1136/bmj.m1165.
- Dong E, Du H, Gardner L (2020). “An Interactive Web-Based Dashboard to Track COVID-19 in Real Time.” *The Lancet Infectious Diseases*. doi:10.1016/S1473-3099(20)30120-1.
- Government of Iceland (March 21, 2020). “Large Scale Testing of General Population in Iceland Underway.” <https://www.government.is/news/article/2020/03/15/Large-scale-testing-of-general-population-in-Iceland-underway/> (Press release) [Accessed March 31, 2020].
- Kekulé AS (March 14, 2020). “The Pandemic Reached Europe.” <https://www.kekule.com/> [Accessed March 31, 2020].
- Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, Shaman J (2020). “Substantial Undocumented Infection Facilitates the Rapid Dissemination of Novel Coronavirus (SARS-CoV-2).” *Science*. doi:10.1126/science.abb3221.
- Maugeri A, Barchitta M, Battiato S, Agodi A (2020). “Estimation of Unreported Novel Coronavirus (SARS-CoV-2) Infections from Reported Deaths: A Susceptible Exposed Infectious Recovered Dead Model.” *Preprints 2020*. doi:10.20944/preprints202004.0052.v1.
- Nardelli A, Ashton E (March 21, 2020). “Everyone In Iceland Can Get Tested For The Coronavirus. Here’s How The Results Could Help All Of Us.” <https://www.buzzfeed.com/albertonardelli/coronavirus-testing-Iceland> (Press release) [Accessed March 31, 2020].
- Nishiura H, Kobayashi T, Miyama T, Suzuki A, Jung S, Hayashi K, Kinoshita R, Yang Y, Yuan B, Akhmetzhanov AR, *et al.* (2020). “Estimation of the Asymptomatic Ratio of Novel Coronavirus Infections (COVID-19).” *medRxiv*.
- Ogris G, Hofinger C (April 10, 2020). “COVID-19 Prevalence (Media information).” https://www.sora.at/uploads/media/Austria_COVID-19_Prevalence_BMBWF_SORA_20200410_EN_Version_fuer_HP.pdf [Accessed April 17, 2020].
- Onder G, Rezza G, Brusaferro S (2020). “Case-Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy.” *JAMA*. ISSN 0098-7484. doi:10.1001/jama.2020.4683. https://jamanetwork.com/journals/jama/articlepdf/2763667/jama_ond er_2020_vp_200059.pdf, URL <https://doi.org/10.1001/jama.2020.4683>.
- ORF (March 27, 2020). “Neue Berechnung der Dunkelziffer.” <https://science.orf.at/stories/3200438/> [Accessed March 31, 2020].
- ORF (March 30, 2020). “Stichprobenstudie soll schon Dienstag starten.” <https://science.orf.at/stories/3200456/> [Accessed March 31, 2020].
- Shahan Z (March 21, 2020). “Iceland is Doing Science 50% of People with COVID-19 Not Showing Symptoms, 50% Have Very Moderate Cold Symptoms.” <https://cleantechnica.com/2020/03/21/iceland-is-doing-science-50-of-people-with-covid-19-not-showing-symptoms-50-have-very-moderate-cold-symptoms/> [Accessed March 31, 2020].

- Sturn S (March 27, 2020). “Wie hoch ist die Dunkelziffer von Covid-19 Infektionen in Österreich?” <https://www.wu.ac.at/other/zukunftsperspektiven-nach-der-coronakrise-1/corona-medien-details/detail/wie-hoch-ist-die-dunkelziffer-von-covid-19-infektionen-in-oesterreich> [Accessed March 31, 2020].
- Tandon A, Murray CJ, Lauer JA, Evans DB (2000). “Measuring Overall Health System Performance for 191 Countries.” *Geneva: World Health Organization*.
- Tuite A, Ng V, Rees E, Fisman D (2020). “Estimation of COVID-19 Outbreak Size in Italy Based on International Case Exportations.” *medRxiv*.
- Zhao S, Musa SS, Lin Q, Ran J, Yang G, Wang W, Lou Y, Yang L, Gao D, He D, *et al.* (2020). “Estimating the Unreported Number of Novel Coronavirus (2019-nCoV) Cases in China in the First Half of January 2020: A Data-Driven Modelling Analysis of the Early Outbreak.” *Journal of Clinical Medicine*, **9**(2), 388.

A. Figures

A.1. Number of tests per million inhabitants

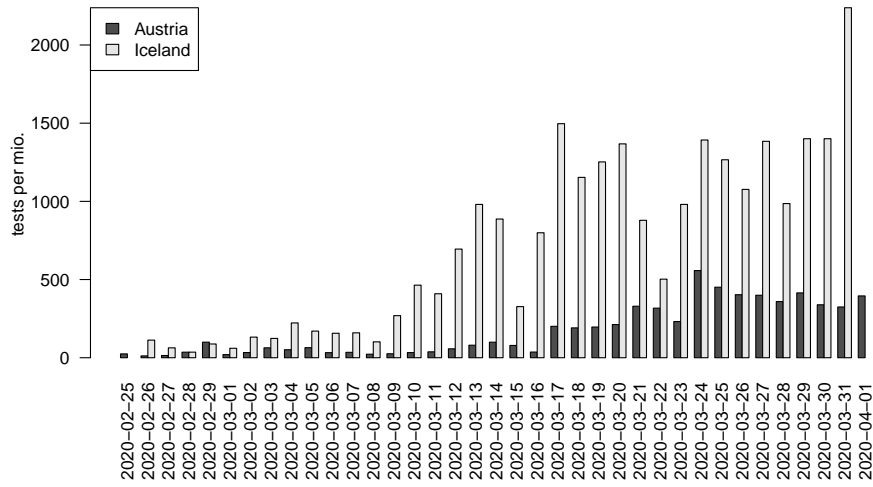


Figure 7: Times series of the number of tests per million inhabitants.

A.2. Number of confirmed infections per million inhabitants

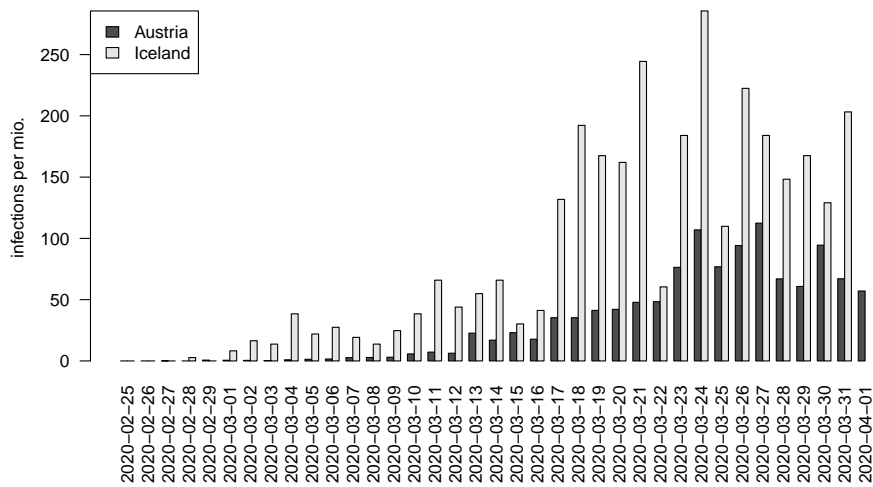


Figure 8: Times series of the number of confirmed infections per million inhabitants.

B. Data

B.1. SARS-CoV-2 tests

Table 5: Statistics on the number of SARS CoV-2 tests performed in Austria and Iceland (tests performed by NUHI).

Date	Total Tests		Total Tests p. mio			Perc. Positive Tests	
	AUT	ISL	AUT	ISL	ISL/AUT	AUT	ISL
2020-02-25	218		24.49			0.92%	
2020-02-26	321	41	36.06	112.60	3.12	0.62%	0%
2020-02-27	447	64	50.21	175.76	3.50	0.67%	0%
2020-02-28	763	77	85.71	211.46	2.47	0.39%	1.3%
2020-02-29	1649	109	185.23	299.34	1.62	0.55%	0.92%
2020-03-01	1826	131	205.11	359.76	1.75	0.77%	3.05%
2020-03-02	2120	179	238.13	491.58	2.06	0.85%	5.59%
2020-03-03	2683	224	301.37	615.16	2.04	0.78%	6.7%
2020-03-04	3138	305	352.48	837.60	2.38	0.92%	9.51%
2020-03-05	3711	367	416.84	1007.87	2.42	1.1%	10.08%
2020-03-06	4000	424	449.31	1164.41	2.59	1.38%	11.08%
2020-03-07	4308	482	483.90	1323.69	2.74	1.83%	11.2%
2020-03-08	4509	519	506.48	1425.30	2.81	2.31%	11.37%
2020-03-09	4734	617	531.75	1694.43	3.19	2.77%	11.02%
2020-03-10	5026	786	564.55	2158.55	3.82	3.62%	10.43%
2020-03-11	5362	935	602.30	2567.74	4.26	4.59%	11.34%
2020-03-12	5869	1188	659.25	3262.54	4.95	5.15%	10.27%
2020-03-13	6582	1545	739.33	4242.94	5.74	7.66%	9.19%
2020-03-14	7467	1868	838.74	5129.98	6.12	8.77%	8.89%
2020-03-15	8167	1987	917.37	5456.78	5.95	10.53%	8.91%
2020-03-16	8490	2278	953.65	6255.94	6.56	11.99%	8.43%
2020-03-17	10278	2823	1154.49	7752.64	6.72	12.96%	8.5%
2020-03-18	11977	3243	1345.34	8906.06	6.62	13.74%	9.56%
2020-03-19	13724	3699	1541.57	10158.35	6.59	14.67%	10.03%
2020-03-20	15613	4197	1753.76	11525.98	6.57	15.29%	10.25%
2020-03-21	18545	4517	2083.10	12404.77	5.95	15.17%	11.49%
2020-03-22	21368	4700	2400.20	12907.34	5.38	15.18%	11.51%
2020-03-23	23429	5057	2631.70	13887.74	5.28	16.75%	12.02%
2020-03-24	28391	5564	3189.07	15280.09	4.79	17.17%	12.8%
2020-03-25	32407	6025	3640.17	16546.11	4.55	17.16%	12.48%
2020-03-26	35995	6417	4043.20	17622.63	4.36	17.77%	12.98%
2020-03-27	39552	6921	4442.75	19006.74	4.28	18.71%	13%
2020-03-28	42750	7280	4801.97	19992.64	4.16	18.7%	13.1%
2020-03-29	46441	7790	5216.57	21393.22	4.10	18.38%	13.03%
2020-03-30	49455	8300	5555.12	22793.81	4.10	18.96%	12.8%
2020-03-31	52344	9115	5879.63	25031.99	4.26	19.05%	12.46%
2020-04-01	55863		6274.91			18.76%	

B.2. SARS-CoV-2 confirmed infections

Table 6: Statistics on the number of SARS CoV-2 infections in Austria and Iceland (infections in Iceland are reported by NUHI).

Date	Infections		Infections p. mio		
	AUT	ISL	AUT	ISL	ISL/AUT
2020-02-25	2	0	0.22	0.00	0.00
2020-02-26	2	0	0.22	0.00	0.00
2020-02-27	3	0	0.34	0.00	0.00
2020-02-28	3	1	0.34	2.75	8.15
2020-02-29	9	1	1.01	2.75	2.72
2020-03-01	14	4	1.57	10.98	6.99
2020-03-02	18	10	2.02	27.46	13.58
2020-03-03	21	15	2.36	41.19	17.46
2020-03-04	29	29	3.26	79.64	24.45
2020-03-05	41	37	4.61	101.61	22.06
2020-03-06	55	47	6.18	129.07	20.89
2020-03-07	79	54	8.87	148.30	16.71
2020-03-08	104	59	11.68	162.03	13.87
2020-03-09	131	68	14.71	186.74	12.69
2020-03-10	182	82	20.44	225.19	11.02
2020-03-11	246	106	27.63	291.10	10.53
2020-03-12	302	122	33.92	335.04	9.88
2020-03-13	504	142	56.61	389.97	6.89
2020-03-14	655	166	73.57	455.88	6.20
2020-03-15	860	177	96.60	486.08	5.03
2020-03-16	1018	192	114.35	527.28	4.61
2020-03-17	1332	240	149.62	659.10	4.41
2020-03-18	1646	310	184.89	851.33	4.60
2020-03-19	2013	371	226.11	1018.86	4.51
2020-03-20	2388	430	268.24	1180.88	4.40
2020-03-21	2814	519	316.09	1425.30	4.51
2020-03-22	3244	541	364.39	1485.72	4.08
2020-03-23	3924	608	440.77	1669.71	3.79
2020-03-24	4876	712	547.71	1955.32	3.57
2020-03-25	5560	752	624.54	2065.17	3.31
2020-03-26	6398	833	718.67	2287.62	3.18
2020-03-27	7399	900	831.11	2471.62	2.97
2020-03-28	7995	954	898.05	2619.91	2.92
2020-03-29	8536	1015	958.82	2787.44	2.91
2020-03-30	9377	1062	1053.29	2916.51	2.77
2020-03-31	9974	1136	1120.35	3119.73	2.78
2020-04-01	10482		1177.41		

Affiliation:

Rainer Hirk, Gregor Kastner, Laura Vana
 Institute for Statistics and Mathematics
 WU Vienna University of Economics and Business
 Building D4, Level 4
 1020 Vienna, Austria

E-mail: Rainer.Hirk@wu.ac.at, Gregor.Kastner@wu.ac.at, Laura.Vana@wu.ac.at

Austrian Journal of Statistics

published by the Austrian Society of Statistics

<http://www.ajs.or.at/>

<http://www.osg.or.at/>

Volume 49

April 2020

Submitted: 2020-04-15

Accepted: 2020-04-18