# Estimates of the Undetected Rate among the SARS-CoV-2 Infected using Testing Data from Iceland

James H. Stock[a], Karl M. Aspelund[b], Michael Droste[a], Christopher D. Walker[a]

[a]Department of Economics, Harvard University
[b]Department of Economics, MIT

## Abstract

Testing for SARS-CoV-2 in the United States is currently targeted to individuals whose symptoms and/or jobs place them at a high presumed risk of infection. An open question is, what is the share of infections that are undetected under current testing guidelines? To answer this question, we turn to COVID-19 testing data from Iceland. The criteria for testing within the Icelandic medical system, processed by the National University Hospital of Iceland (NUHI), have also been targeted at high-risk individuals, but additionally most Icelanders qualify for voluntary testing through the biopharmaceutical company deCODE genetics. We use results from Iceland's two testing programs to estimate the share of infections that are undetected under standard (NUHI) testing guidelines. Because of complications in the deCODE testing regime, it is not possible to estimate a single value for this this undetected rate; however, a range can be estimated. Our primary estimates for the fraction of infections that are undetected range from 88.7% to 93.6%.

## COVID-19 Testing in Iceland

Iceland has conducted one of the most extensive testing regimes in the world on a per-capita basis. The first SARS-CoV-2 test was administered on January 31, and the first case was confirmed on February 27. Almost 21,000 Icelanders—about 6 percent of the island nation—have been tested for SARS-CoV-2 as of April 2.

Guðbjartsson et al (2020) describes the testing programs in Iceland in detail. In brief, Iceland has two ongoing testing programs. The first is through the regular healthcare system, which targets individuals exhibiting severe symptoms (cough, fever, muscle ache, and shortness of breath) and/or at high risk of infection because of close contact with a diagnosed individual or having returned from "high-risk areas" defined by the Directorate of Health. On March 19, the Directorate of Health broadened the criteria to include anyone returning from abroad and any symptomatic individual.[1] Samples are taken at local clinics or the National University Hospital of Iceland (NUHI) and analyzed by the hospital's microbiology department.

The second is a free voluntary program, with testing done by the biopharmaceutical company deCODE Genetics. The deCODE tests, which began on March 13, are open to individuals who have no or few symptoms, are not under quarantine, and have not been tested through NUHI. Individuals are placed under quarantine if they have been in close contact with someone diagnosed

[1] https://www.landlaeknir.is/um-embaettid/greinar/grein/item39194/Skilgreind-ahaettusvaedi---Defined-high-risk-areas

with COVID-19 or have been traveling, initially in high-risk areas—mostly ski areas in Italy and Austria—and later anywhere abroad. Quarantines last 14 days or until the individual develop symptoms and are tested positive for SARS-CoV-2 by NUHI.[2] At its peak, about 10,000 Icelanders were under quarantine.

The presence of these two testing programs makes it possible to draw inferences about the rate of infections in Iceland that are undetected under the NUHI testing guidelines. Because the NUHI testing guidelines are similar to those recommended by the Centers for Disease Control and Prevention (CDC) in the United States,[3] an estimate of the rate of undetected infections can inform epidemiological modeling in the United States. Estimates from the Icelandic population complement other estimates of the rate of undetected cases among the infected (Li et. al. (2020), Mizumoto et. al. (2020), Nishiura et. al. (2020), Russell et al. (2020); see Qui (2020)).

**Methods**

We are interested in the fraction of the infected who are not eligible for NUHI testing (the "undetected rate"). Expressed as a probability, this rate is $\Pr(NE = 0|I = 1)$, where $NE = 1$ if the individual is eligible for NUHI testing (NUHI-eligible) and 0 otherwise (NUHI-ineligible), and $I = 1$ if the individual is infected with SARS-CoV-2 and 0 otherwise. Denote this rate by $\theta$. Using Bayes Law, the undetected rate is related to the rates of infection among the NUHI-eligible and the NUHI-ineligible:

$$\theta = \Pr\left(NE = 0 \middle| I = 1\right) = \frac{f^{NI} \Pr\left(NE = 0\right)}{f^{NI} \Pr\left(NE = 0\right) + f^{NE} \Pr\left(NE = 1\right)}, \tag{1}$$

where $f^{NE} = \Pr\left(I = 1 \middle| NE = 1\right)$ is the positive testing probability among the NUHI-eligible and $f^{NI} = \Pr\left(I = 1 \middle| NE = 0\right)$ is the positive-testing probability among the NUHI-ineligible. Note that the denominator in (1) is the total infection rate.

If the (a) fraction of the population that were NUHI-eligible on a given day were known and if (b) deCODE testing was a random sample of the NUHI-ineligible, then the first expression in (1) could be evaluated directly using data on the NUHI positive testing rate, $\Pr(I = 1|NE = 1)$, and the rate of infection among the NUHI-ineligible, $\Pr\left(I = 1 \middle| NE = 0\right)$, could be estimated from the deCODE testing data. Neither condition (a) nor (b) hold, however, and we address these issues in turn.

[2]Individuals with a positive test, whether quarantined or not at the time of testing, must enter a 10-day isolation with more stringent restrictions than those for quarantine. See https://www.landlaeknir.is/servlet/file/store93/item39065/GA.%20Instructions%20for%20the%20general%20public%20on%20home-based%20quarantine%2002.04.2020.pdf and https://www.landlaeknir.is/servlet/file/store93/item39066/KSJ%20Instructions%20for%20the%20general%20public%20on%20home-based%20isolation%2004042020.pdf.

[3] https://www.cdc.gov/coronavirus/2019-nCoV/hcp/clinical-criteria.html

The source of problem (a) is that, while the number of NUHI tests on a given day is known, the number of individuals who would qualify for a NUHI test is not. If all those who qualify for a NUHI test are referred to NUHI for testing, then they will get tested at some point, but on any given day only some of those eligible are tested.

Our first approach to addressing the NUHI stock/flow issue (a) is to estimate the probability of NUHI-eligibility, $\Pr(NE=1)$, as the fraction of the Icelandic population tested during an 8-day window, where the 8 days aligns with the data window reported in Guðbjartsson et al (2020). We refer to this as the "eligibility window" method.

Our second approach to addressing problem (a), which we refer to as the "odds ratio" method, is to use the laws of probability and algebraic manipulation to eliminate the probability of NUHI-eligibility from the expressions above. Doing so, however, introduces a new parameter, $\lambda_0 = \Pr(NE=1|I=0)$, which is the baseline rate of cough/cold/flu symptoms that are severe enough to qualify for a NUHI test, even though the individual turns out not to be infected. This baseline symptomatic rate can be estimated using historical data on flu cases in Iceland. It is shown in the appendix that the Bayes Law expression (1) and its counterpart for the NUHI-eligible imply,

$$\theta = \frac{\tau}{1+\tau}, \text{ where } \tau = \frac{(1-\lambda_0)(1-f^{NE})/f^{NE}}{\lambda_0(1-f^{NI})/f^{NI}}, \tag{2}$$

Equation (2) express the undetected rate, $\theta$, in terms of $f^{NE}$, $f^{NI}$, and $\lambda_0$, so if those three terms can be estimated, then $\theta$ can be estimated using (2).

The term $f^{NE}$ can estimated directly from NUHI testing data, and the term $\lambda_0$ can be estimated by the proportion of Icelanders who see a doctor about flu-like symptoms, see Appendix Section B for more details.

Both the eligibility window method (Equation (1) and the odds ratio method (Equation (2)) require an estimate of $f^{NI}$, the fraction of infections among the NUHI-ineligible. If the deCODE sample were a random sample of the NUHI-ineligible population, this would be estimated by the NUHI positive testing rate. However, the deCODE sample has two complications: it was voluntary, and those in quarantine were excluded. The next two sections discuss these issues turn, and show how the deCODE positive testing rate can be used to bound $f^{NI}$ that accounts for the complications in the deCODE testing sample. For the eligibility window method, we thereby estimate a lower bound (a conservative estimate of $\theta$), and for the odds ratio method we are able to estimate both a lower and upper bound.

*Selection into deCODE testing*

Because deCODE testing was voluntary during the time period analyzed here, it could overrepresent those who suspect they have the virus but are NUHI-ineligible. Approximately 44% of individuals tested by deCODE exhibit cold/flu symptoms (Guðbjartsson et al, 2020, Table 1),

much higher than the historical prevalence of cold/flu symptoms in Iceland. We will refer to these 44% as "mildly symptomatic." As a result, the deCODE positive testing rate could overestimate the true fraction of infected among the NUHI-ineligible.

To address this problem, note that $f^{NI}$ is a weighted average of the infection rates among the mildly symptomatic and the asymptomatic. Under the assumption that the asymptomatic and mildly symptomatic deCODE subsamples are randomly drawn from the asymptomatic and mildly symptomatic populations, respectively, then $f^{NI}$ can be estimated as the weighted average of these two (known) deCODE positive testing rates. Although the rate of mild symptoms among the NUHI-ineligible is not observed, we bound it below by $\lambda_0$ (the rate of severe flu symptoms in a normal March) and bound it above by the 44% of mild symptoms among the deCODE testing volunteers. See Appendix Section B for details.

*Exclusion of quarantined individuals*

Individuals in quarantine were excluded from deCODE testing and while some of the quarantined qualified for NUHI testing, some did not. Infections among those quarantined but NUHI-ineligible are therefore not detected by either the NUHI or deCODE testing. Under the assumption that the infection rate among the NUHI-ineligible quarantined does not exceed the infection rate of among NUHI-eligible, it is possible to use the laws of probability to bound the overall rate of infection among the NUHI-ineligible in terms of the infection rate among the NUHI-eligible and the deCODE rate (infections among the NUHI-ineligible who are not in quarantine). See Appendix Section B for details.

**Data**

We use data released by Icelandic public health authorities on the COVID-19 pandemic and on normal flu seasons.

The Icelandic Directorate of Health releases daily counts of new confirmed cases and tests conducted on its COVID-19 information website.[4] We use these data where possible. However, infection rates for individuals with no vs. mild symptoms are only available in the results by Guðbjartsson et al (2020), who report infection and test counts by symptom status for deCODE testing between March 13 and 19. We therefore focus on this time period when calculating parameters for the bounds. Only the most updated quarantine counts are available on the COVID-19 website, so we use daily announcements by the Icelandic Department of Civil Protection and Emergency Management to find past quarantine totals in the bound analysis.

To estimate the underlying prevalence of mild and severe flu-like symptoms, we use data from the Directorate of Health's website on influenza containing weekly counts of individuals who come

[4]English: https://www.covid.is/data Icelandic: https://www.covid.is/tolulegar-upplysingar

to clinics or emergency rooms with flu-like symptoms for every flu season from the 2009-10 through the 2017-18 flu season (the most recent season available).[5]

Appendix Section B contains the details on the data and construction of estimates.

**Results**

The results are summarized in Table 2. The primary estimate of the lower bound using the eligibility window method uses the infection rate among the asymptomatic from the deCODE tests and assumes that none of the NUHI-ineligible quarantined are infected; this yields an undetected rate of 88.7% (95% confidence interval 83.9% to 93.5%). The second line re-estimates the lower bound using the overall deCODE positive testing rate, which includes the mildly symptomatic; because the mildly symptomatic have a somewhat higher deCODE positive testing rate than the asymptomatic, the estimate of the population infection rate is somewhat larger and the undetected fraction is somewhat larger, 92.5%. As discussed, the NUHI testing eligibility guidelines were broadened after March 19, and using data postdating this broadening results in slightly lower estimates of the undetected rate, approximately 85%.

For the odds ratio estimate, we estimate $\lambda_0$ by the maximum weekly caseload for flu-like symptoms in the Icelandic medical system for all weeks from 2009 to 2017. This highest-caseload week occurred in November 2009 at the peak of the H1N1 pandemic and was 0.63%. The resulting estimated range is consistent with the estimates from the eligibility window method, with a range of 88.9% to 93.6%.

**Discussion**

The fraction of infections that are undetected under standard current testing guidelines is differs from the asymptomatic rate, which is the fraction of infected who are asymptomatic. Because the NUHI-ineligible include those who are mildly symptomatic, it is not surprising that the estimated nondetected rate exceeds estimates of the asymptomatic rate in the literature (Mizumoto et. al. (2020), Nishiura et. al. (2020)). Our estimate of the undetected rate is comparable to Li et. al (2020)'s 86% estimate for Wuhan.

One threat to the validity of these estimates is that they assume that the deCODE testing of the asymptomatic is a reliable estimate of the rate of infection among the asymptomatic in the Icelandic population. Because deCODE test recipients did not meet the quarantine requirements (returning from high-risk countries or close contact with a confirmed COVID-19 individual), this assumption seems plausible; it cannot be evaluated with existing data, however. Data to provide unbiased estimates of the undetected rate, and the overall rate of infection in the population, need to come from random testing. Such trials are commencing in Iceland, Germany, and Norway. If a random testing design excludes those eligible for tests under medically-based guidelines and/or those in quarantine, the combined data sets could be analyzed using the methods laid out here.

[5]Only in Icelandic:
https://www.landlaeknir.is/servlet/file/store93/item32967/influensulik.einkenni.uppfaersla.a.vef.okt2017.AG.xls_%C3%A1n%20myndar%20og%20aldursdreifingu.xls

## References

Guðbjartsson, Daníel F., et al. (2020). "Early Spread of SARS-Cov-2 in the Icelandic Population." Medrxiv Preprint. Posted online March 26, 2020, at https://www.medrxiv.org/content/10.1101/2020.03.26.20044446v1.full.pdf?fbclid=IwAR15VmgS2CU37oOkblJB-AUfd-EdE0H6v6ciaoY2Fwam5G7spa-ovOswUyI.

Li, R. et. al. (2020). "Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2)," *Science*, published online March 16, 2020 DOI: 10.1126/science.abb3221 at https://science.sciencemag.org/content/early/2020/03/13/science.abb3221.

Mizumoto, K. et. al. (2020). "Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020," *Eurosurveillance* 25(10), March 12, 2020 at https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2020.25.10.2000180.

Nishiura, H. et. al. (2020). "Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19)," forthcoming, *International Journal of Infectious Disease*, posted online February 13, 2020 at https://doi.org/10.1016/j.ijid.2020.03.020.

Qui, J. (2020). "Covert coronavirus infections could be seeding new outbreaks," *Nature* News, posted online March 20, 2020 at https://www.nature.com/articles/d41586-020-00822-x.

Russell, T. et al. (2020). "Using a delay-adjusted case fatality ratio to estimate under-reporting," Centre for Mathematical Modeling of Infectious Diseases Repository, posted online March 22, 2020, at https://cmmid.github.io/topics/covid19/severity/global_cfr_estimates.html

## Table 1: Parameter Values and Data Sources

| Parameter | Description | Estimate | Source |
|---|---|---|---|
| N | Population of Iceland | 364,134 | Statistics Iceland, 1/1/2020 |
| $\lambda_0$ | Share of non-infected pop. with severe flu-like symptoms | 0.0017, 0.0063 | Health Directorate, 2010-17 flu seasons |
| f^N | Share of NUHI-eligible pop. infected | 0.099-0.170 | NUHI infection rate, many date ranges, covid.is |
| P(Q) | Share of pop. quarantined | 0.0114 | Civil Defense Department, 3/19 |
| P(NE) | Share of pop. NUHI-eligible | 0.0100 | Total NUHI test count, covid.is, 3/19 |
| P(I\|NI, C=1) | Share of NUHI-ineligible asymptomatic pop. infected | 0.0135 | Guðbjartsson et al 2020, Table 1 |
| P(I\|NI, C=0) | Share of NUH-ineligible asymptomatic pop. Infected | 0.0056 | Guðbjartsson et al 2020, Table 1 |
| P(C=1\|decode) | Share of deCODE test sample who were symptomatic | 0.4437 | Guðbjartsson et al 2020, Table 1 |

## Table 2: Estimates of the Proportion of NUHI-Ineligible among the Infected

| Method | $\lambda_0$ | Undetected rate θ | | Dates of NUHI Testing | Dates of deCODE Testing |
|---|---|---|---|---|---|
| | | Lower bound | Upper bound | | |
| eligibility window[a] | | 0.887 (0.839, 0.935) | | 3/13-3/19 | 3/13-3/19 |
| eligibility window[b] | | 0.925 (0.905, 0.946) | | 3/13-3/19 | 3/13-3/19 |
| eligibility window[b] | | 0.853 (0.786, 0.921) | | 3/20-3/26 | 3/20-3/26 |
| eligibility window[b] | | 0.847 (0.803, 0.891) | | 3/27-4/2 | 3/27-4/2 |
| odds ratio | 0.0063 | 0.889 | 0.936 | 3/13-3/19 | 3/13-3/19 |

*Notes*: nonparametric bootstrap 95% confidence intervals are given in parentheses under the lower bound estimates for the eligibility window method.

[a]Uses NUHI-ineligible positive testing rate on the asymptomatic from deCODE tests (Guðbjartsson et al, 2020, Table 1).

[b]Uses overall deCODE positive testing rate (Icelandic Directorate of Health at https://www.covid.is/data).

## Appendix

### *A. Derivations*

*Derivation of Equation* (2).

Let $P(NE = 1|I = 0) = \lambda_0$ and $\theta = P(NE = 0|I = 1)$. By Bayes Law,

$$f^{NE} = \frac{(1-\theta)P(I=1)}{(1-\theta)P(I=1) + \lambda_0\big(1-P(I=1)\big)}$$

and

$$f^{NI} = \frac{\theta P(I=1)}{\theta P(I=1) + (1-\lambda_0)\big(1-P(I=1)\big)}.$$

Taking the ratio of odds ratios, we obtain that

$$\frac{(1-f^{NE})/f^{NE}}{(1-f^{NI})/f^{NI}} = \left(\frac{\lambda_0}{1-\lambda_0}\right)\left(\frac{\theta}{1-\theta}\right)$$

which yields

$$\frac{\theta}{1-\theta} = \left(\frac{1-\lambda_0}{\lambda_0}\right)\frac{(1-f^{NE})/f^{NE}}{(1-f^{NI})/f^{NI}} \equiv \tau$$

and by rearranging,

$$\theta = \frac{\tau}{1+\tau}.$$

*Bounds for* $\theta$.

The bounds for $\theta$ arise because of the overrepresentation of mildly symptomatic in the (voluntary) deCODE sample and the exclusion of quarantined individuals from the deCODE sample, some of whom are not NUHI-eligible. The available information, however, provides bounds on $f^{NI}$. We make the following assumptions:

1. $\lambda_0 \leq P(C = 1|Q = 0, NE = 0) \leq P(C = 1|deCODE\ volunteer)$
2. $P(I = 1|NE = 0, Q = 1) \leq P(I = 1|NE = 1)$

where $C = 1$ indicates whether an individual has mild symptoms and $Q = 1$ indicates whether an individual is in quarantine. Note that ($Q = 0$, $NE = 0$) corresponds to deCODE eligibility.

By the law of total probability,

$$f^{NI} = P(I = 1|NE = 0, Q = 1)P(Q = 1|NE = 0) + P(I = 1|NE = 0, Q = 0)P(Q = 0|NE = 0)$$

and

$$P(I = 1|NE = 0, Q = 0) = (p^C - p^A)P(C = 1|Q = 0, NE = 0) + p^A.$$

Note that $p^C \equiv P(I = 1|NE = 0, Q = 0, C = 1)$ and $p^A \equiv P(I = 1|NE = 0, Q = 0, C = 0)$ are identified from Guðbjartsson et al. (2020) and are estimated by the deCODE positive testing rates for the symptomatic and asymptomatic, respectively.

We consider each term individually. Bayes Law and the fact that $0 \leq P(NE = 0|Q = 1) \leq 1$ yields that

$$0 \leq P(Q = 1|NE = 0) = \frac{P(NE = 0|Q = 1)P(Q = 1)}{1 - P(NE = 1)} \leq \frac{P(Q = 1)}{1 - P(NE = 1)}$$

and

$$1 - \frac{P(Q = 1)}{1 - P(NE = 1)} \leq P(Q = 0|NE = 0) \leq 1.$$

by the complement rule. Next, Assumption 1 generates

$$(p^C - p^A)\lambda_0 + p^A \leq P(I = 1|NE = 0, Q = 0)$$

and

$$P(I = 1|NE = 0, Q = 0) \leq (p^C - p^A)P(C = 1|deCODE\ volunteer) + p^A.$$

Combining these results with assumption 2, the implied bounds for $f^{NI}$ are

$$f^{NI} \leq (p^C - p^A)P(C = 1|deCODE\ volunteer) + p^A + \left(\frac{P(Q = 1)}{1 - P(NE = 1)}\right)P(I = 1|NE = 1) \equiv \overline{f^{NI}}$$

and

$$f^{NI} \geq \left((p^C - p^A)\lambda_0 + p^A\right)\left(1 - \frac{P(Q = 1)}{1 - P(NE = 1)}\right) \equiv \underline{f^{NI}}.$$

Evaluating (2) at the extremes of the range for $f^{NI}$ and taking the resulting extremes for $\theta$ yields the identified set for $\theta$.

### *B. Data Construction*

See the text and Appendix A for definitions of the various terms.

- Population: Statistics Iceland estimates that the population of Iceland was 364,134 on January 1, 2020.[6]
- $f^{NE}$: From the Directorate of Health's counts, we calculate the total count of NUHI infected cases and divide by the total count of NUHI tests from March 13 through March 19, to match the deCODE testing dates.
- $\lambda_0$: We use Directorate of Health weekly influenza count data to estimate $\lambda_0$ by the greatest one-week fraction of Icelanders reporting flu symptoms.[7] This period occurred in November 2009 during the H1N1 pandemic and was 0.63%. To bound the proportion mildly symptomatic, we use a lower value of $\lambda_0$ which is the average rate of reported flu symptoms in the two-week period around March 15 from 2011-2018, which is 0.17%.
- $P(I|C=1,NE=0,Q=0), P(I|C=0,NE=0,Q=0)$: Under the assumption that deCODE randomly tests *within* the deCODE-eligible symptomatic and asymptomatic groups, then these terms are estimated by the deCODE positive testing rates within these two groups as reported in Table 1 of Guðbjartsson et al (2020) for deCODE testing from March 13 through March 19.
- $P(C=1|NE=0,Q=0)$: As a lower bound, we use the lower bound of $\lambda_0$ of 0.17%. As an upper bound, we use the proportion symptomatic from Guðbjartsson et al (2020), 44%, which reflects the assumption that mildly symptomatic individuals will be oversampled in the voluntary deCODE testing.
- $P(Q=1)$: The COVID-19 website only reports the most updated count of individuals under quarantine, but we require the proportion not quarantined $P(Q=0)$ at the time of testing. We use the number of quarantined individuals on March 19, the last day of deCODE testing reported in Guðbjartsson et al (2020), from the report of the Department of Civil Defense on March 20.[8]
- $P(NE=1)$ (for use in bounds calculation, odds ratio approach): We sum the number of NUHI tests from March 13 through March 19 from the Health Directorate count data to match the timing of the deCODE testing and the quarantine count figure in the bound analysis, then divide this sum by the population of Iceland.

---

6 https://px.hagstofa.is/pxen/pxweb/en/Ibuar/Ibuar__mannfjoldi__1_yfirlit__yfirlit_mannfjolda/MAN00000.px/table/tableViewLayout1/?rxid=cdf68733-56cd-41c8-8801-d73ca7ded1cd

[7] In Icelandic: https://www.landlaeknir.is/servlet/file/store93/item32967/influensulik.einkenni.uppfaersla.a.vef.okt2017.AG.xls_%C3%A1n%20myndar%20og%20aldursdreifingu.xls

[8] In Icelandic: https://www.almannavarnir.is/utgefid-efni/stoduskyrsla-koronaveira-covid-19-20032020/?wpdmdl=24791