Folkhälsomyndigheten
PUBLIC HEALTH AGENCY OF SWEDEN

Estimates of the peak-day and the number of infected individuals during the covid-19 outbreak in the Stockholm region, Sweden February – April 2020

Article number: 20059

About the publication

The Public Health Agency of Sweden carried out a mathematical model of the outbreak of covid-19 in the Stockholm region to estimate the date with most new infections as well as the accumulated number of infected individuals at different dates. In the model, we divided infected individuals into “reported cases” and “unreported cases”. Reported cases are confirmed SARS-CoV-2 positive and reported to the Public Health Agency of Sweden between 17 February and 10 April 2020. Unreported cases are not included in the statistics and are assumed to have varying degrees of symptoms, from very mild to more severe. We used the result from a study that was carried out in the Stockholm region that showed that 2.5% of the population were SARS CoV-2 positive in the Stockholm region between 27 March and 3 April 2020.

Code for the modeling available at https://github.com/FohmAnalys/SEIR-model-Stockholm.

The work was carried out by staff from the Unit for Analysis during April 2020.

Public Health Agency of Sweden

Lisa Brouwers
Head of Unit for Analysis

Table of contents

About the publication 4
Glossary 6
Summary 7
Sammnfattning på svenska 8
Introduction 9
Objective 9
Transmission model 10
SEIR-model 10
Analysis 12
Fitting of observed data 12
Health report, Stockholm region 12
Method 13
Parameters of the model 14
Results 15
Peak day and number of infected 15
Forecast 3 and 7-days 20
Limitations 23
References 24
Sensitivity analyses 25

Glossary

Incidence	**Number of new cases (here: inflow to I_r and I_u)**
Observed incidence	Number of new reported cases (here: inflow to I_r)
Prevalence	Number of individuals in a population who simultaneously have a certain disease (here: all infectious individuals $I = I_r + I_u$)
Compartmental model	A mathematical model for infectious diseases where the population is divided into different compartments. All individuals within a compartment are assumed to have the same characteristics (here: we use a compartmental model of the type *SEIR*)
Susceptible	Individual who is not infected and can become infected (here: all individuals in *S*)
Latent	Individual who is infected but not yet infectious (here: all individuals in *E*)
Infected	Individual who is infectious (here: all individuals in I_r and I_u)
Recovered	Individual who has been infectious but is no longer and is assumed to be now immune (here: all individuals in *R*).
Latency phase	The time between the individual becoming infected until the same individual becomes infectious.
Unreported	Individuals who are infectious but have not been confirmed. These cases are not reported and therefore not included in the statistics (here: all individuals in I_u).

Summary

We modeled the spread of covid-19 in the Stockholm region with a mathematical model to estimate the total number of infected individuals and the date with the largest number of infectious individuals. According to our results, on 8 April 2020 approximately 70,500 individuals were infectious. This was the day with the largest number of simultaneously infectious individuals. From the model we also estimated that approximately 26% of the population in the Stockholm region will have been infected with covid-19 by 1 May 2020.

In the model we divided infectious individuals into “reported” and “unreported” cases. Reported cases in the model are those that were confirmed to be infected with SARS-CoV-2 by the healthcare system and were reported to the Public Health Agency of Sweden between 17 February and 10 April 2020; cases infected abroad are not considered. Unreported cases are not part of the reported statistics and could have from mild to more severe symptoms, but not so severe that they would be hospitalized. To calibrate the model, we also used data from a survey conducted in the Stockholm region which measured the current number of individuals infected with SARS-CoV-2 in the population. The survey showed that 2.5% of the population was positive for the virus between 27 March and 3 April.

Since the infectiousness of the unreported cases is currently unknown, we present three scenarios with different assumptions on how infectious an unreported case is relative to a reported case. In the first scenario, we assumed that an unreported case is as infectious as a reported case. In the other two scenarios we assume that the infectiousness of an unreported case is lower than that of a reported case. We assumed that the infectiousness correlates with symptoms and thus it is likely to assume that milder cases are less infectious than more severe cases. The scenario that best fit the number of reported cases between 17 February and 10 April was the scenario where the infectiousness of unreported cases was a tenth of that of reported cases; the difference in the fit of the other scenarios compared to the best scenario was however not large.

Sammanfattning på svenska

Vi har modellerat spridningen av covid-19 i Stockholms län med en matematisk modell för att skatta det totala antalet infekterade personer och tidpunkten då störst antal personer samtidigt är smittsamma. Enligt modelleringen var ungefär 70 500 personer samtidigt smittsamma den 8 april, dagen med flest samtidigt smittsamma personer. Enligt modelleringen kommer ungefär 26% av befolkningnen i Stockholms län att ha varit smittade med covid-19 den första maj.

I modellen har vi delat upp infekterade individer i ”rapporterade fall” och ”obekräftade fall”. De rapporterade fallen har av vården bekräftats vara infekterade med SARS-CoV-2 och utgör antal fall som rapporterats in till Folkhälsomyndigheten mellan 17 februari och 10 april 2020, fall som är smittade utomlands har exkluderats. De obekräftade fallen ingår inte i statistiken och utgör det så kallade mörkertalet. Dessa har olika grad av symptom, från mycket milda till mer allvarliga men inte så allvarliga att de läggs in på sjukhus. För skattning av modellen använder vi även resultat från den undersökning som genomförts i Stockholms län för att mäta aktuell förekomst av SARS-CoV-2 i samhället. Undersökning visade att 2.5% av befolkningen i Stockholms län var infekterade mellan 27 mars och 3 april.

Eftersom graden av smittsamhet bland de obekräftade fallen just nu är okänd har vi lagt in tre olika scenarier med olika antaganden om hur smittsamt ett obekräftat fall är i relation till ett rapporterat fall. I första fallet antar vi att de obekräftade fallen är lika smittsamma som de rapporterade fallen. I de andra fallen antar vi att de är mindre smittsamma än de rapporterade fallen. Vi antar att smittsamhet korrelerar med symptom och att det därför är troligt att anta att mildare fall är mindre smittsamma än mer allvarliga fall. Det scenario som ger bäst anpassning till antal rapporterade fall mellan 17 februari och 10 april 2020 är scenariot där obekräftade fall är cirka en tiondel så smittsamma som rapporterade fall, men skillnaden i anpassning är inte så stor.

Introduction

This report presents the model and results used to study the spread of covid-19 in the Stockholm region between 17 February and 10 April 2020. From surveillance data we know the number of reported cases. It is important however to also gain knowledge about the number of cases that are not reported since this will aide future forecasting, planning, and assessment of possible interventions.

Objective

The objective of the model is to estimate the total number of individuals infected with covid-19 in the Stockholm region, including unreported cases, and to estimate the peak day, i.e. the day with the largest number of infectious individuals, as well as with the largest number of new cases.

Transmission model

SEIR-model

We developed a so-called compartmental model in which individuals are divided into different compartments depending on predetermined properties. Within each compartment, individuals are assumed to have the same characteristics and act in the same way. The compartments are denoted *S* as in susceptible, *E* as in exposed, *I* as in infected, and *R* as in recovered.

When a healthy individual gets infected, he or she does not become infectious at once but enters the symptom-free phase *E* and remains in that compartment for an average of $\frac{1}{\rho} = 5.1$ days (1, 2).[1]

We divide infectious cases into two groups: reported cases and unreported cases. After the incubation period, an infected individual either is tested in the health care sector and becomes a confirmed and reported case or remains unconfirmed, i.e. an unreported case. If covid-19 is confirmed and reported, the individual is transferred from compartment E to compartment $I_r = I_{reported}$. If the individual is not tested and remains unreported, he or she is transferred from compartment E to the compartment $I_u = I_{unreported}$. The probability that a case remains unreported is denoted p_u, and the probability that a case becomes reported is $p_r = 1 - p_u$. The infectiousness, i.e. the infectivity rate, is assumed to vary between the value θ and the value $\delta\theta$, where the midpoint between θ and $\delta\theta$ occurs at the time t_b. We assume that the turning point occurs at $t_b = 16$ March 2020 (day 76 of the year) which is the day when people in Stockholm were recommended to work from home. The speed of this change in infectivity is determined by the parameter ε. Whether this is an increase or decrease is determined by the combination of ε and δ. The infectiousness at a time t is described as follows:

Time-dependent infectivity rate $b_t = b(t, t_b, \theta, \delta, \varepsilon) = \theta\left(\delta + \frac{1-\delta}{1+e^{-\varepsilon(t-t_b)}}\right)$.

The special case $\varepsilon = 0$ results in a constant infectivity rate.

We assume that an individual that becomes a reported case has an infectiousness that follows the time-dependent infectivity rate and that the infectiousness among unreported cases is a factor $q_o \in [0,1]$ of the infectiousness among reported cases.

An individual is assumed to be infectious on average $\frac{1}{\gamma} = 5$ days (3) and after that moves on to the compartment R.[2]

[1] Li et al. (1) has latency phase/incubation time of 5.2 days (95% CI [4.1, 7]), (2) Linton et al. estimates incubation time to 5 daysr (95% CI [4, 5.8]).

[2] . Wölfel et al. (3) studied a small sample of patients where they found an active viruse during day 3 to 8.

We assume a closed population, no one enters or leaves the population. We denote the number of individuals in each compartment by $S, E, I_r,\ I_u$ and R. This indicates that the population size is $N = S + E + I_r + I_u + R$. The transmission dynamics are described by the following equation system:

$$\frac{dS}{dt} = -S\frac{b_t\, I_r}{N} - Sq_o\frac{b_t\, I_u}{N}$$

$$\frac{dE}{dt} = S\frac{b_t\, I_r}{N} + Sq_o\frac{b_t\, I_u}{N} - \ \rho E$$

$$\frac{dI_o}{dt} = p_o\rho E - \gamma I_u$$

$$\frac{dI_r}{dt} = (1 - p_o)\rho E - \ \gamma I_r$$

$$\frac{dR}{dt} = \gamma(I_u + I_r).$$

From this differential equation system, one can calculate the numbers S, E, I_r, I_u and R at any time t, given the initial number of individuals in the different compartments at the desired start time t_0.

Note that the daily reported cases in the model at time t is given by $(1 - p_o)\rho E(t)$, which is calibrated to fit the daily number of actual reported cases in the Stockholm region.

The basic reproduction number R_0 is defined as the expected number of individuals infected by an initial infected individual in a fully susceptible population. In our model, R_0 varies in time t since the infectivity rate is assumed to depend on time t, and is given by the infectivity rate multiplied by the expected time the initially infected individual is infectious; the infectivity rate is the number of individuals an infectious individual infects per unit of time. In our model, an infected individual is either a reported case with infectiousness of b_t or an unreported case with a lower infectiousness of $q_u b_t$. Since the unreported proportion is denoted p_u, we get

$$R_0(t) = (1 - p_u)b_t/\gamma + p_0 q_u b_t/\gamma.$$

To obtain the effective reproduction number $R_e(t)$, we multiply $R_0(t)$ by the proportion of susceptible individuals at time t, i.e.

$$R_e(t) = R_0(t)\, S(t)/N.$$

Analysis

Fitting of observed data

We fitted the parameters of the SEIR-model at time t to the reported number of domestic cases per day in the Stockholm region[3] until 10 April 2020. More specifically, we fitted the model parameters so that the estimated incidence of reported cases each day ($(1-p_o)\rho E(t)$) was similar to the observed daily incidence of reported cases. There is some delay in the reporting, and since the last few days are uncertain, we excluded the last four days from the analysis (original analysis conducted 14 April 2020). The last day used for reported cases was therefore 10 April 2020. We fixed the length of time an infected individual is infectious and the duration of the latency period but estimated all parameters for the varying infectivity $\theta, \delta,$ and ε. Cases that had been infected abroad were excluded as they had not been infected within the modeled population. We used the day of reported symptom onset (so-called *epi-date*) as the day when a case becomes infectious.[4] Since we had contact tracing around the first early import cases in the Stockholm region, we assumed that we had no unreported cases until $t_0 =$ February 17. We, therefore, assume that at time t_0 = 17 February there was one infectious reported individual but no unreported cases. An alternative approach would be to set the number of reported cases at t_0 to one and the number of unreported cases to $p_u/p_r = p_u/(1-p_u)$.

When fitting the parameters to the data, we minimized the residual squared error between the model-estimated number of new cases per day and the observed number of new cases per day.

Health report, Stockholm region

The Public Health Agency of Sweden conducted a study[5] in which 707 participants in a web panel of randomly recruited individuals in the Stockholm region, conducted self-sampling for covid-19 between 27 March and 3 April 2020. The results from this study were used to calibrate our model and to estimate the number of unreported cases.

In that study, 18 of 707 tested positive for covid-19 and the estimated weighted proportion of positive individuals was 2.5% (95% CI [1.4% - 4.2%]). Given that estimate a total of 60 455 (95% KI [33,244; 99,731]) individuals in the Stockholm region were positive during this time. We calibrated the parameter p_u (the probability that a case remains unreported) to obtain an estimated average prevalence of approximately 2.5% during 27 March - 3 April.

---

[3] Stockholm region has a popluation of 2,374,550 (SCB, 2019).

[4] If symptom onset day was not reported we used the day of testing.

[5] See: https://www.folkhalsomyndigheten.se/nyheter-och-press/nyhetsarkiv/2020/april/resultat-fran-undersokning-av-forekomsten-av-covid-19-i-region-stockholm/ (in Swedish)

Method

We fitted the mathematical SEIR model to different assumptions of the parameter values of q_u, the factor that describes the infectiousness of unreported cases in relation to the infectiousness of reported cases. The three different scenarios were calibrated so that the estimated prevalence of infectious individuals, i.e. I_r and I_u, was 2.5% on average between 27 March and 3 April. For this prevalence to be 2.5% on the specified dates, the probability for a case to become unreported was $p_u = 0.987$ according to our calculations. That indicates that 98.7% of the infected individuals were unreported cases. We then studied three different degrees of infectiousness among the unreported cases relative to the infectiousness of a reported case: $q_u = \{1, 0.55, 0.11\}$. These values were arbitrarily chosen but were based on the basic assumption that infectiousness correlates with the degree of symptoms. It is thereby reasonable to assume that milder cases are less infectious that more severe cases. Based on the differential equation system above, we concluded that the scenarios will foremost affect the relative infectiousness of unreported to reported cases. We estimated the model parameters for the three scenarios and their standard errors. To obtain a parametric bootstrap confidence interval (CI), e.g. the estimated curve of the number of new daily reported cases, we drew 1,000 combinations of the model parameters, where each parameter was drawn from a normal distribution with the parameter estimate as mean and its standard error as the standard deviation and calculated the curve via the differential equations. The method was used to estimate the confidence intervals of all of the results presented in the tables.

In the next step, we studied how well the model succeeded in estimating the number of new cases over the next three or seven days. It was conducted by using data on the number of reported cases until 7 April. We then estimated the model based on these data, and investigated how well the estimated curve fitted the number of new cases in the next three days according to the reported number of cases. In the same way, we used data until 3 April, estimated the model, and then investigated how well the estimated curve fitted the number of new cases in the next seven days according to the reported number of cases.

In sensitivity analyses, we used vi $p_u \in \{0.5, 0.95\}$ and $q_u \in \{0.1, 0.5, 1.00\}$.

Parameters of the model

Parameter	**Värde**
Length of latency phase $\frac{1}{\rho}$	5.1 days
Length of infectiousness $\frac{1}{\gamma}$	5 days
Start date t_0	17 February
Turning point between θ and $\delta\theta$ occurs at time t_b.	16 March
Probability that a case remains unreported	0.987 (calibrated)
Reduction factor of the infectiousness of an unreported case compared to a reported case: q_u.	$[0,1]$ Scenarios with $\{1, 0.55, 0.11\}$
Infectivity rate parameters: θ, δ och ε	Estimated by the model
Infectivity rate at time t: b_t	Estimated by the model

Results

Based on the reported cases in the Stockholm region until 10 April 2020, we studied three scenarios. The results from the three scenarios were all consistent with the assumption of a covid-19 prevalence of approximately 2.5% between 27 March and 3 April, i.e. the prevalence estimated by the Stockholm study. Common to the three scenarios was that the unreported cases must account for 98.7% of the total number of cases to result in the observed prevalence. The scenarios differed regarding the infectiousness of unreported cases in relation to reported cases ($\boldsymbol{q_u}$).

	Infectiousness among unreported cases relative to reported cases (%)
Scenario 1	**100**
Scenario 2	**55**
Scenario 3	**11**

Peak day and number of infected

The peak day for prevalence, i.e. the date when the largest number of individuals were simultaneously infectious, occurred between 8 and 11 April, dependent on the scenario investigated, when approximately 70,500 were infectious. The date when the largest number of new cases occurred, i.e. the incidence or the inflow to I *(=* I_r + I_u*)*, occurred in the 2 or 3 April when approximately 14,500 individuals became infectious.

The results for the accumulated number and the proportion of individuals that were or had been infected ($E + I + R$) for the three different scenarios are presented in Table 1. On 11 April, 17% of the population in the Stockholm region were or had been infected. According to scenario 3, 26% will have been infected on 1 May. Simultaneously infectious (i.e. all individuals in $I = I_r + I_u$), and the peak-day for the different scenarios are presented in Table 2.

To the left in figures 1, 2, and 3, the estimated development of new reported cases, *daily incidence* or inflow to I_r, is shown, while the right graph shows the total number of infectious individuals, *prevalence,* or all individuals in I_r *and* I_u. The observed data is from 17 February to 10 April, and the curve forecasts beyond those dates to 25 April.

Based on reported cases during the period 17 February to 10 April, we calculated the value of the reproduction number at a certain time, $R_e(t)$. In all three scenarios, $R_e(t)$ takes a value just above or just below one on 10 April. See Table 3 for reproductions numbers with 95% confidence interval.

We present the estimated parameters and the residual square error (RSS) for the three scenarios in Table 4. It shows that scenario 3, where the unreported cases are 11% as infectious as the reported cases, provides the best fit to the data. According to scenario 3, the peak-day of new infectious cases occurred on 2 April (95% CI 27

March – 13 April), and the peak-day for prevalence of simultaneously infectious individuals occurred on 8 April (95% CI 1 April – 18 April). The number of new infectious cases, the incidence, was 14.749 on 2 April (95% CI 8,252 – 30,479) and the number of simultaneously infectious individuals, the prevalence, was 70,488 on 8 April (95% CI 38,006 – 148,485) in the Stockholm region.

Even though the infectiousness among reported and unreported cases is distributed differently among the scenarios, the outcome in the number of infectious and infected at different times does not differ much between the scenarios (see figures 1, 2, and 3 and Table 4).

Table 1. Accumulated number of and share of infected individuals (E + I + R) at two dates. Common for the three presented scenarios is that the probability for a case to become an unreported cases (p_u) is 0.987, but they differ in the assumed infectiousness of an unreported case in relation to the infectiousness of a reported case (q_u). All estimates are presented with a 95% CI (95% CI).

	Accumulated number of infected and proportion of the population in the Stockholm region by infectiousness of unreported cases relative to reported cases (q_u).			
	2020-04-11		2020-05-01	
q_u	Number (95% CI))	Proportion (95% CI)	Number (95% CI)	Proportion (95% CI)
Scenario 1: infecitousness 100%	409,723	0.17	648,557	0.27
	[266,654; 600,308]	[0.112, 0.253]	[360,847; 1,046,195]	[0.152, 0.441]
Scenario 2: infecitousness 55%	410,382	0.17	640,973	0.27
	[270,217; 613,828]	[0.114, 0.259]	[376,891; 1,067,624]	[0.159, 0.45]
Scenario 3: infecitousness 11%	412,674	0.17	616,655	0.26
	[222,824; 698,672]	[0.094, 0.294]	[273,34; 1,203,280]	[0.115, 0.507]

Table 2. Estimated peak-day and number of infectious individuals, Ir + Iu (prevalence), and the estimated peak-day for incidence, the daily number of new cases, i.e. the daily inflow to Ir and Iu, by the infectiousness of unreported cases relative to reported cases (q_u).

		Prevalence	**Incidence**	
q_u	Peak day (95% CI)	Prevalence on peak-day (95% CI)	Peak day (95% CI)	Incidence on peak-day (95% CI)
Scenario 1: infectiousness 100%	2020-04-11	70,581	2020-04-03	14,511
	[2020-04-02, 2020-04-16]	[44,701; 124,632]	[2020-03-28, 2020-04-09]	[9,717; 25,629]
Scenario 2: infectiousness 55%	2020-04-10	70,472	2020-04-03	14,565
	[2020-04-03, 2020-04-19]	[44,413; 128,138]	[2020-03-27, 2020-04-15]	[9,522; 26,291]
Scenario 3: infectiousness 11%	2020-04-08	70,488	2020-04-02	14,749
	[2020-04-01, 2020-04-18]	[38,006; 148,485]	[2020-03-27, 2020-04-13]	[8,252; 30,479]

Table 3. Estimated effective reproduction number (R_e) on the last day of reported cases in our analysis, 10 April 2020, with a 95% confidence interval, by the infectiousness of unreported cases relative to reported cases (q_u).

q_u	R_e **(2020-04-10) (95% CI))**
Scenario 1: infectiousness 100%	0.98
	[0.778, 1.167]
Scenario 2: infectiousness 55%	0.97
	[0.739, 1.172]
Scenario 3: infectiousness 11%	0.92
	[0.601, 1.222]

Table 4. Estimated prevalence as a proportion of the population in the Stockholm region, 27 March – 3 April 20202, residual sum of squares (RSS), and estimated parameters, standard error (se) and confidence interval (95% CI), by the infectiousness of unreported cases relative to reported cases (q_u).

q_u	**Prevalence 27 March – 3 April (proportion of the population)**	**RSS***	$\hat{\delta}$ **(se) 95% CI**	$\hat{\varepsilon}$ **(se) 95% CI**	$\hat{\theta}$ **(se) 95% CI**
Scenario 1: infectiousness 100%	2.5%	28,461	0.14 (0.018)	-0.24 (0.048)	1.68 (0.019)
			[0.101, 0.173]	[-0.336, -0.148]	[1.645, 1.719]
Scenario 2: infectiousness 55%	2.5%	27,576	0.14 (0.021)	-0.23 (0.046)	2.84 (0.038)
			[0.102, 0.185]	[-0.316, -0.138]	[2.763, 2.91]
Scenario 3: infectiousness 11%	2.5%	25,235	0.16 (0.032)	-0.19 (0.04)	10.9 (0.241)
			[0.094, 0.221]	[-0.264, -0.106]	[10.458, 11.402]

(*) **The model with the lowest RSS provides the best fit to the data**.

Figure 1. Scenario 1: we assume an infectiousness among unreported cases as 100% of that of reported cases. To the left: estimated daily incidence of newly reported cases, inflow to Ir (the red line), with a 95% confidence interval (the dashed lines) and observed data (circles). To the right: prevalence of infectious cases, all individuals in Ir och Iu, both reported and unreported cases (the red line) with a 95% confidence interval (the dashed lines).

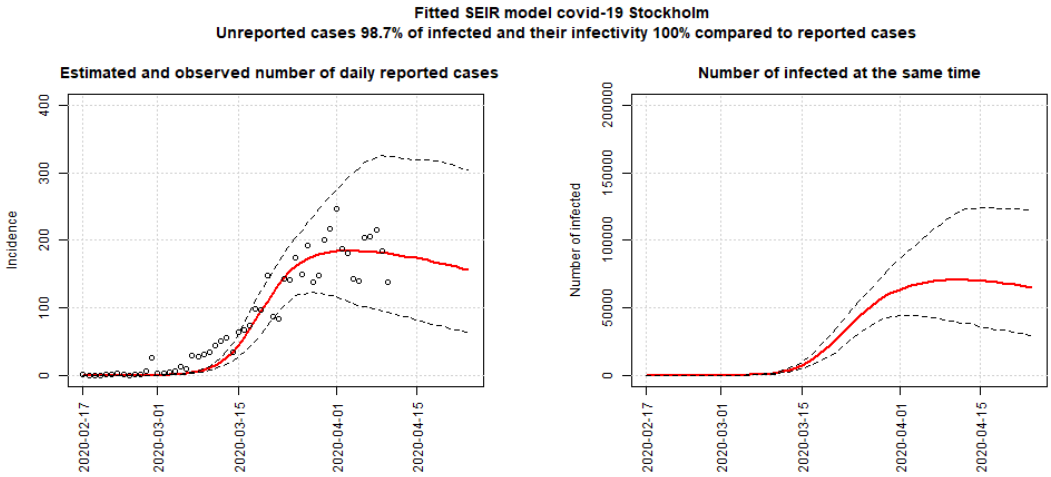


Figure 2. Scenario 2: we assume an infectiousness among unreported cases as 55% of that of reported cases. To the left: estimated daily incidence of newly reported cases, inflow to Ir (the red line), with a 95% confidence interval (the dashed lines) and observed data (circles). To the right: prevalence of infectious cases, all individuals in Ir och Iu, both reported and unreported cases (the red line) with a 95% confidence interval (the dashed lines).

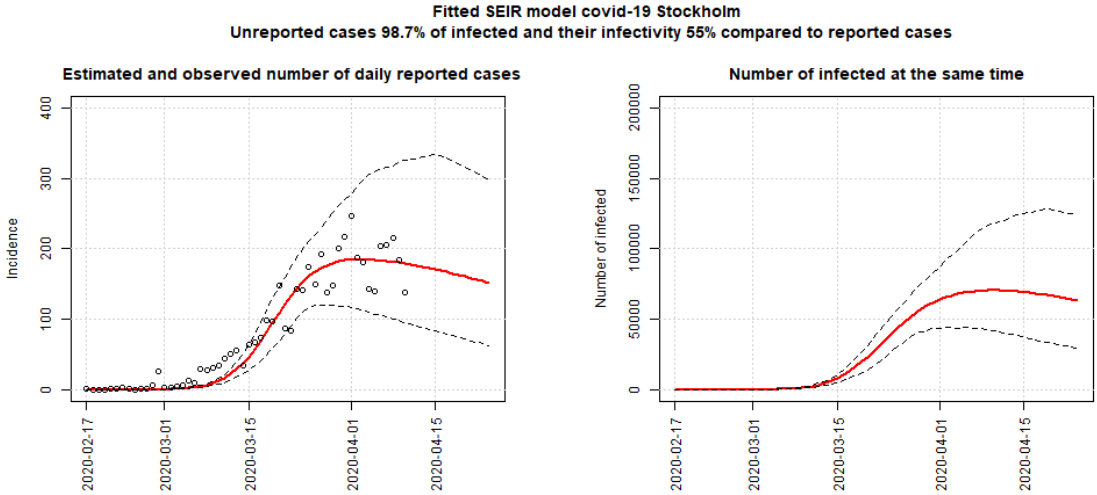
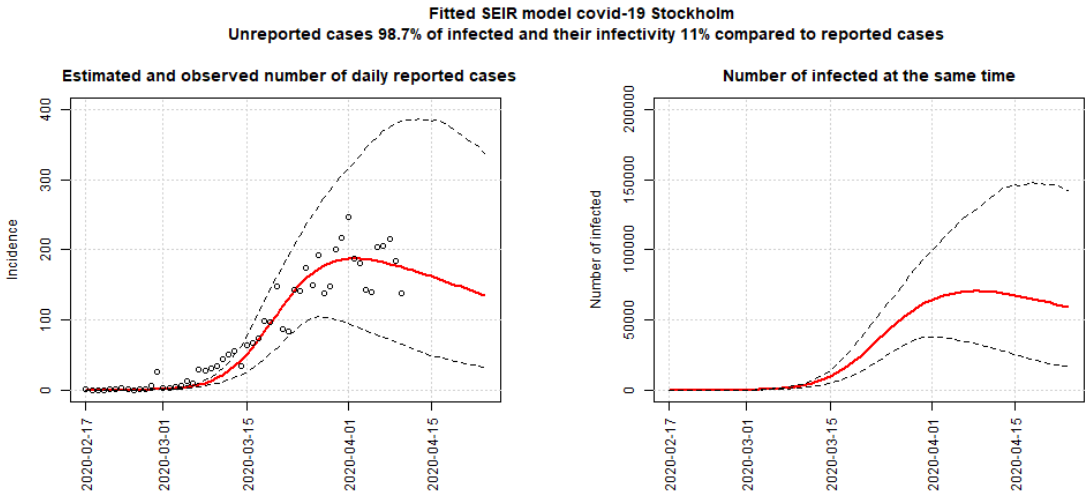


Figure 3. Scenario 3: we assume an infectiousness among unreported cases as 11% of that of reported cases. To the left: estimated daily incidence of newly reported cases, inflow to Ir (the red line), with a 95% confidence interval (the dashed lines) and observed data (circles). To the right: prevalence of infectious cases, all individuals in Ir och Iu, both reported and unreported cases (the red line) with a 95% confidence interval (the dashed lines).



Forecast 3 and 7-days

We studied how well our estimated models based on the number of cases up to April 7 predicted new cases 3 days ahead, and how well our estimated models based on the number of cases up to 3 April predicted new cases 7 days ahead. In figure 4, we see that all three models performed equally well in predicting the number of new cases 3 days ahead. In figure 5, we see that the scenario 3, where the infectiousness among unreported cases was 11% of that among reported cases, was best at predicting the 7 days following 3 April. All three scenarios overestimated the number of new cases, but the assumption of 11% infectiousness provided a better prediction 7 days ahead than the other two scenarios.

Figure 4. The figure presents how well the estimated curves for the three different scenarios on data until April 7 predict the daily number of newly reported cases 3 days ahead. The top graph is scenario 1, the middle graph is scenario 2, and the graph at the bottom scenario 3. The verticle dashed line in each graph indicates the date up to which point data was used in the estimation of the curve. The circles to the left of the dashed line were used to estimate the curves, and the filled circles to the right of the dashed line were used to determine how well the curves predicted new cases.

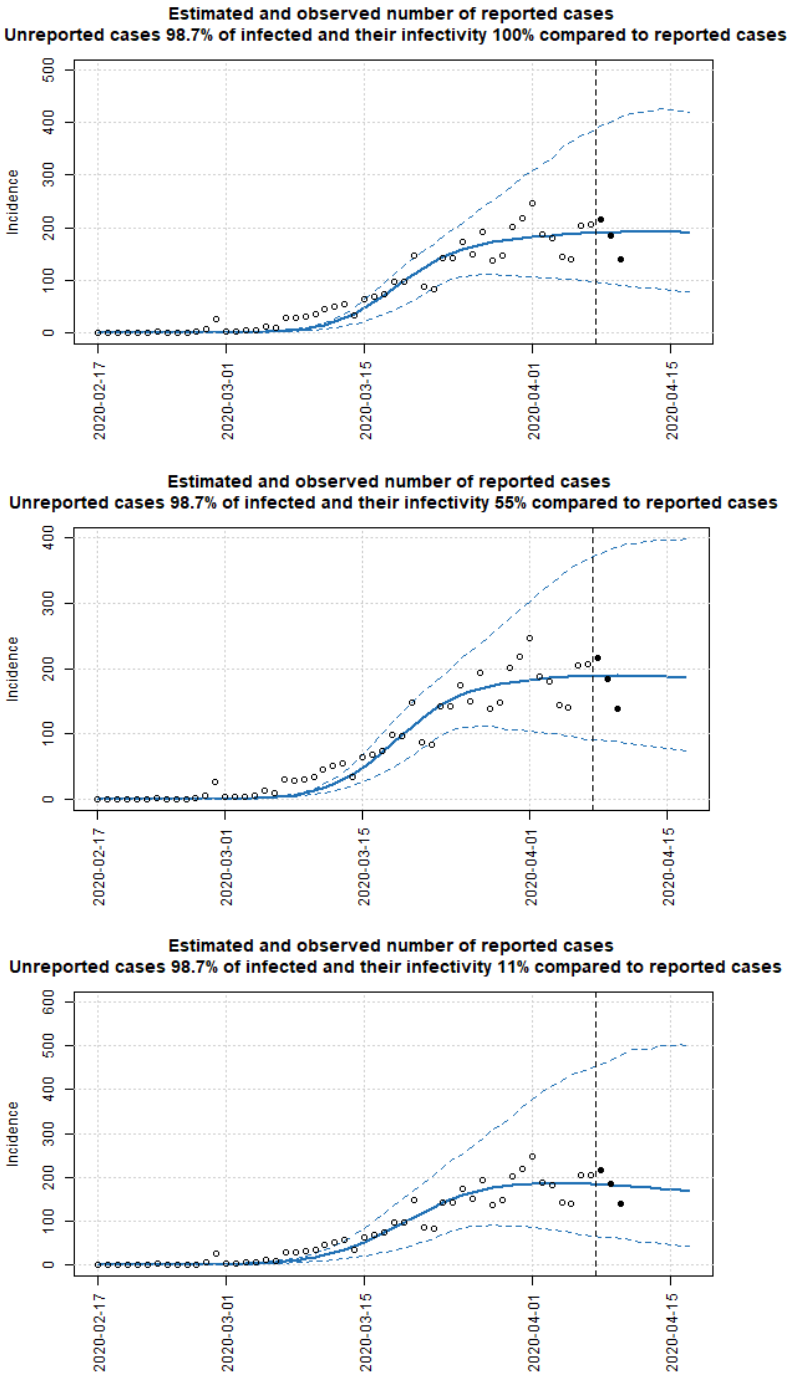
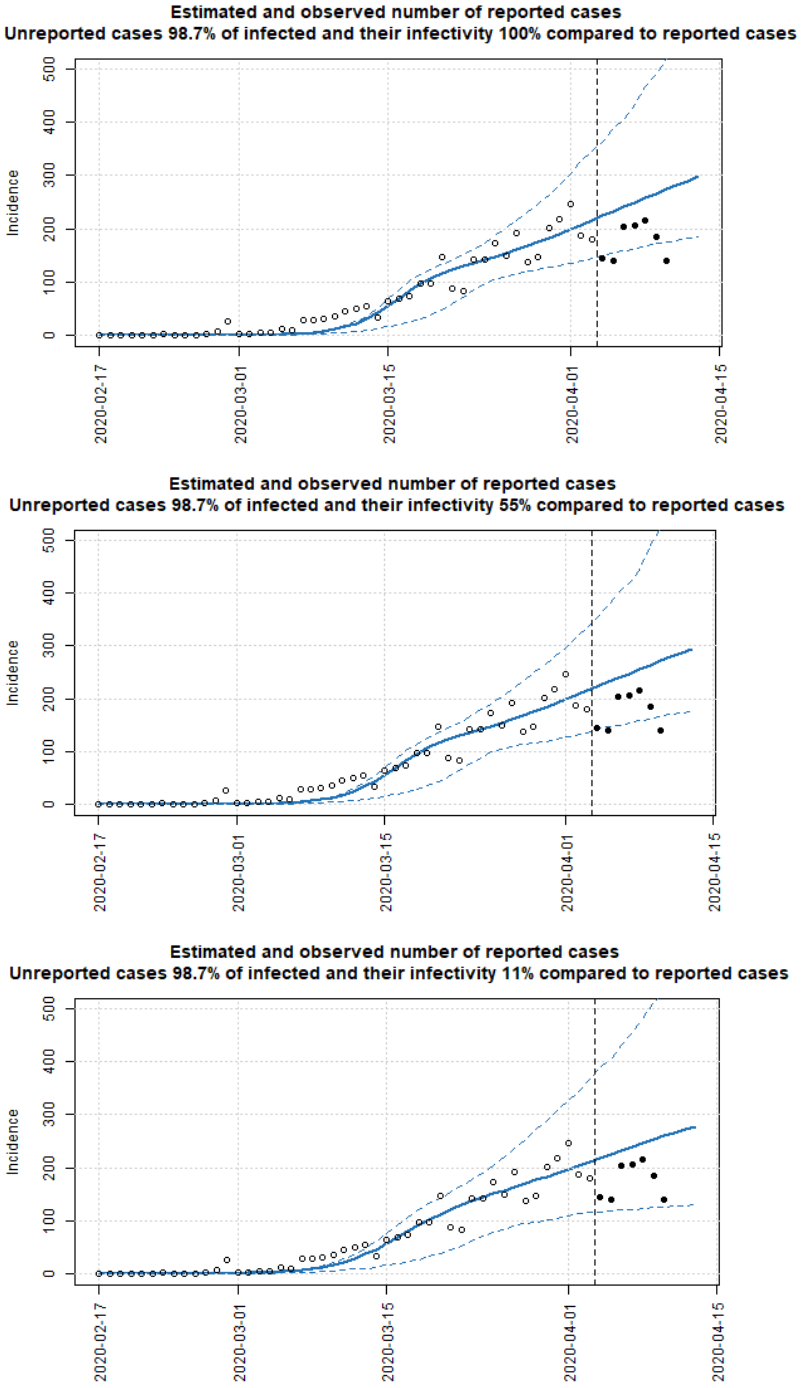


Figure 5. The figure presents how well the estimated curves for the three different scenarios based on data until April 3 predict the daily number of newly reported cases 7 days ahead. The top graph is scenario 1, the middle graph is scenario 2, and the graph at the bottomscenario 3. The verticle dashed line in each graph indicates the date up to which point data was used in the estimation of the curve. The circles to the left of the dashed line were used to estimate the curves, and the filled circles to the right of the dashed line were used to determine how well the curves predicted new cases.



Limitations

Covid-19 is primarily transmitted through droplet infection, which indicates that the social contact structure in the population is important for the dynamics of infection. The compartmental model does not take into account variation in contacts between people, which occur in a society where few individuals could have many contacts and the majority have fewer contacts. This simplification in the model, i.e. a homogenous contact structure, usually results in a somewhat faster growth of an epidemic than if heterogeneity is included in the model. The model, therefore, runs the risk of overestimating the speed of the outbreak in the Stockholm region. This is not included in the specified confidence intervals, as a confidence interval cannot report such uncertainties.

We assume that the infection rate (infectivity) varies over time, but that the parameters that control the length of the latency phase (ρ) and the length of infectiousness (γ) are constant. This indicates that they control exponentially distributed random variables. This is a (common) simplification in compartmental models in order to facilitate calculation but is worth highlighting. Exponentially distributed time is not always the best description of what we observe. For example, the variation in the latency phase may be lower than what an exponentially distributed time with an average of 5.1 results in.

Further, we assume that the relationship between reported and unreported cases is constant over time. This is a limitation since the routines for sampling were changed on 12 March 2020, when the focus was shifted from testing individuals with symptoms coming from known risk areas abroad or contact tracing to individuals in need of inpatient hospital care. We do not believe that this change significantly affected the reporting of cases of domestic infection but rather affected the reporting of the previous import cases.

References

1. Li Q, Guan X, Wu P, et al. "Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia". N Engl J Med (2020). DOI: 10.1056/NEJMoa2001316
2. Linton NM, Kobayashi T, Yang Y, 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). DOI: 10.3390/jcm9020538
3. Wölfel, R., Corman, V.M., Guggemos, W. et al. "Virological assessment of hospitalized patients with COVID-2019". Nature (2020). https://doi.org/10.1038/s41586-020-2196-x
4. S. Flaxman, S. Mishra, A. Gandy, et al., "Estimating the number of infections and the impact of non- pharmaceutical interventions on COVID-19 in 11 European countries," tech. rep., Imperial College London, 2020.

Sensitivity analyses

This section presents estimations to reported cases for more combinations of p_u and q_u, more precisely $p_u \in \{0.5, 0.95\}$ and $q_u \in \{0.1, 0.5, 1.00\}$. Figures 6 and 7 present the estimated daily incidence of reported cases and the prevalence of infectious individuals. Table 5 shows the accumulated number of cases until 11 April and until 1 May 2020, as well as the peak-day for these combinations of p_u and q_u.

For the lower share of reported cases of 50%, the prevalence was not close to that which was observed in the Stockholm study. In the different scenarios, we can see a peak around April 8-9. Figure 6 shows that if the share of unreported cases is high (95%), and if the infectiousness among unreported cases is high (100% and 50% of that among reported cases), then the peak does not occur with a fast decrease, but rather a longer plateau. In the scenario with lower infectiousness and a smaller share of unreported cases, the decrease in the number of cases is faster (but still rather slow).

Figure 6. The scenario where 5% of the cases are reported and 95% remain unreported. The infectiousness of unreported cases is 10% of the reported cases in the top graph, 50% in the middle, and 100% in the bottom graph. The red line in the left graph shows the estimated daily incidence of reported cases and the circles the observed data. The right graph shows the prevalence of infectious cases.

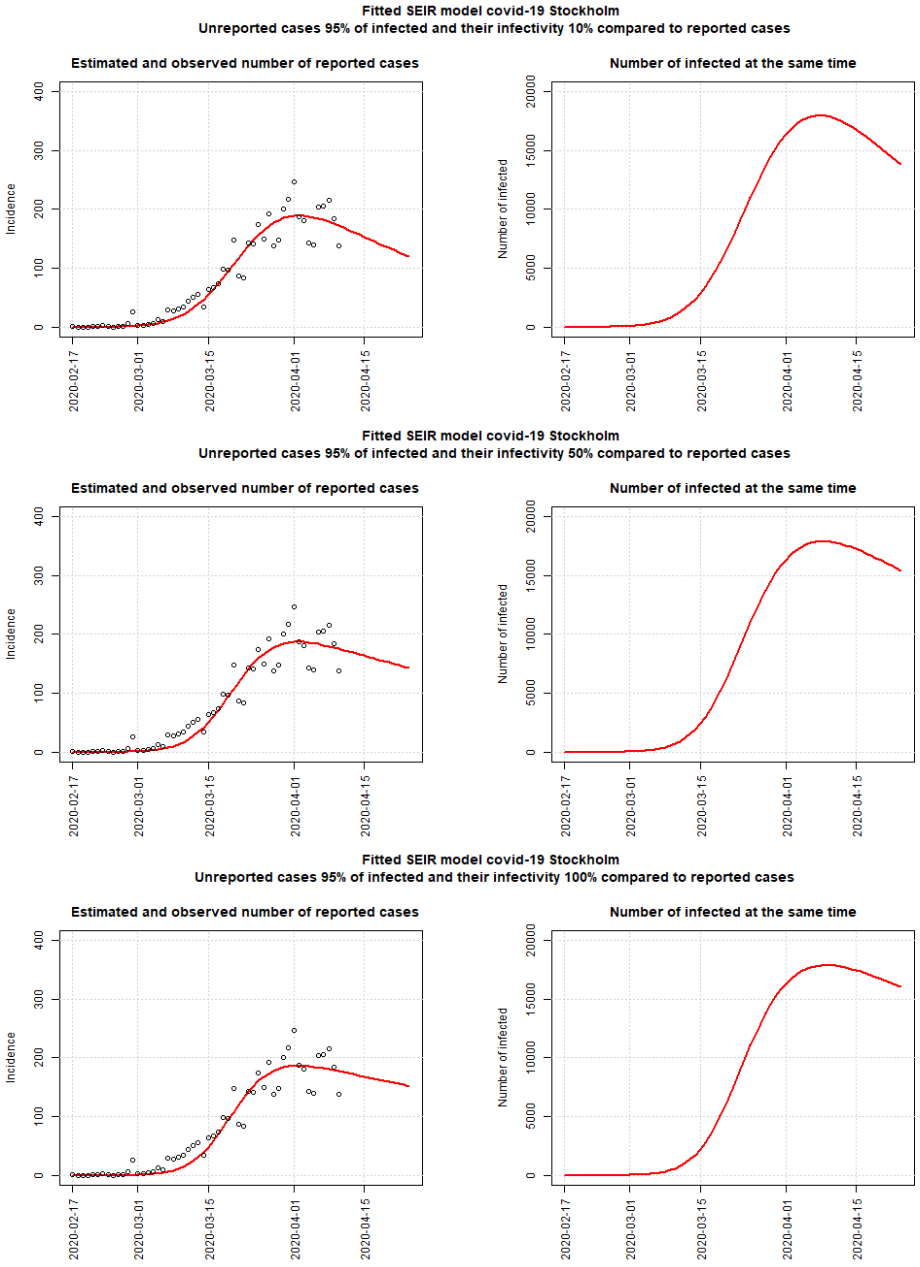


Figure 7. The scenario where 50% of the cases are reported and 50% remain unreported. The infectiousness of unreported cases is 10% of the reported cases in the top graph, 50% in the middle, and 100% in the bottom graph. . The red line in the left graph shows the estimated daily incidence of reported cases and the circles the observed data. The right graph shows the prevalence of infectious cases.

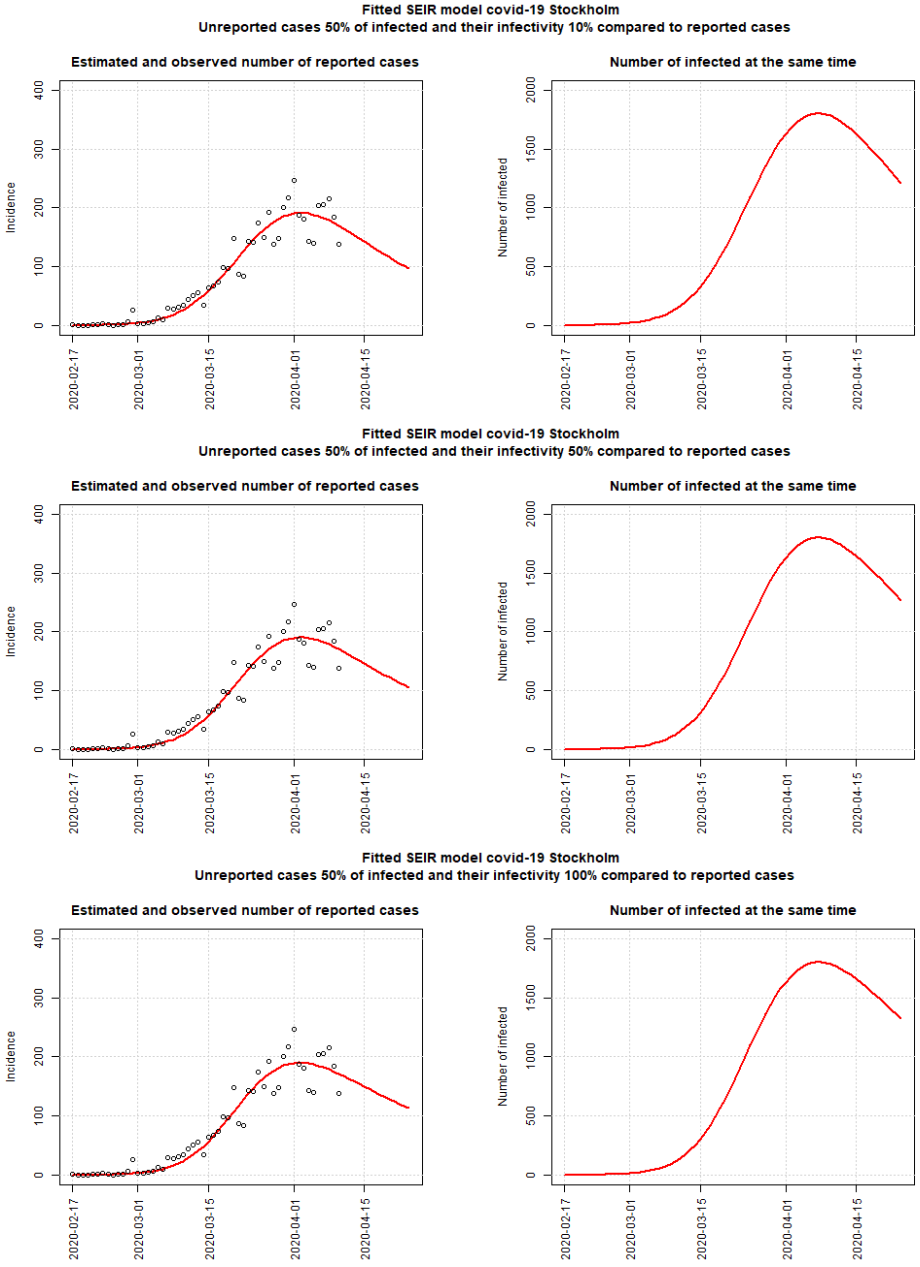


Table 5. Sensitivity analysis of the accumulated number of infected individuals (E+I+R) and the share infected at two different dates, and the estimated peak-day of infectious individuals.

		Accumulated number of cases and share of infected until				**Peak-day (95% CI)**	**Prevalence peak-day (95% CI)**
q_u	p_u	2020-04-11		2020-05-01			
		Accumulated number (CI 95%)	Accumulated share (CI 95%)	Accumulated number (CI 95%)	Accumulated share (CI 95%)		
Infectiousness of unreported 100%	unreported 95%	104,447	0.044	163,804	0.069	2020-04-09	17,881
		[62,439; 180,281]	[0.026, 0.076]	[78,671; 383,153]	[0.033, 0.161]	[2020-04-03, 2020-05-11]	[10,622; 53,906]
	unreported 50%	10,575	0.004	14,811	0.006	2020-04-08	1,801
		[3,811; 34,648]	[0.002, 0.015]	[4,306; 91,933]	[0.002, 0.039]	[2020-04-02, 2020-05-11]	[629; 19,898]
Infectiousness of unreported 50%	unreported 95%	104,733	0.044	160,211	0.067	2020-04-08	17,895
		[60,393; 191,661]	[0.025, 0.081]	[73,202; 408,747]	[0.031, 0.172]	[2020-04-02, 2020-05-11]	[10,179; 56,861]
	unreported 50%	10,597	0.004	14,515	0.006	2020-04-08	1,803
		[3,426; 37,383]	[0.001, 0.016]	[3,577; 102,755]	[0.002, 0.043]	[2020-04-03, 2020-05-11]	[538; 23,960]
Infectiousness of unreported 10%	unreported 95%	105,463	0.044	151,094	0.064	2020-04-08	17,976
		[45,766; 241,136]	[0.019, 0.102]	[51,057; 466,851]	[0.022, 0.197]	[2020-04-02, 2020-05-08]	[7,584; 61,467]
	unreported 50%	10,625	0.004	14,172	0.006	2020-04-08	1,806
		[2,887; 54,572]	[0.001, 0.023]	[3,048; 171,524]	[0.001, 0.072]	[2020-04-01, 2020-05-11]	[461; 45,666]

Folkhälsomyndigheten
PUBLIC HEALTH AGENCY OF SWEDEN

Solna Nobels väg 18, SE-171 82 Solna. **Östersund** Forskarens väg 3. Box 505, SE-831 26 Östersund.

www.folkhalsomyndigheten.se