Folkhälsomyndigheten
PUBLIC HEALTH AGENCY OF SWEDEN

Estimates of the number of infected individuals during the covid-19 outbreak in the Dalarna region, Skåne region Stockholm region, and Västra Götaland region, Sweden

Article number: 20103

About the publication

This is an updated version of a previous report from the Public Health Agency of Sweden.

The Public Health Agency of Sweden developed a mathematical model to study the outbreak of covid-19 in four regions in Sweden. The aim was to estimate the date with most new infections as well as the accumulated number of infected individuals at different dates. In the model, infected individuals are divided into “reported cases” and “unreported cases”. Reported cases are confirmed SARS-CoV-2 positive and reported to the Public Health Agency of Sweden between February 17 and June 5, 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 results from two studies that were carried out in the Stockholm region. The first study showed that 2.5% of the population were SARS CoV-2 positive based on PCR-test in the Stockholm region between March 17 and April 3 2020, the second study showed that 2.3% of the population were positive in the Stockholm region between April 21 and April 24 2020.

The work was carried out by members of the Unit for Analysis during May and June 2020.

Public Health Agency of Sweden

Lisa Brouwers
Head of Unit for Analysis

Table of contents

Glossary

	Definition
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_1 *and* R_2).
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 $I = I_r + I_u$)
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).
Latency phase	The time between the individual becoming infected until the same individual becomes infectious.
PCR-test	Polymerase chain reaction-test for covid-19 infection.

Summary

With this updated report we present the mathematical model used to study the spread of covid-19 in four regions in Sweden: Dalarna, Skåne, Stockholm, and Västra Götaland. The main updates from the previous version consist of introducing a second compartment representing recovered individuals who still test positive on a PCR-test, capturing active or near-time infection. We have also added the two point-estimates of active infection in the Stockholm region in the fitting process.

With the mathematical model, we estimate the number of infected individuals at different time points and the date with the largest number of infectious individuals. According to our results, by July 1, 8.5% (5.9 – 12.9%) of the population in Dalarna will have been infected, 4% (2.4 – 9.9%) of the population in Skåne will have been infected, 19% (17.7 – 20.2%) of the population in Stockholm will have been infected, and 9% (6.3 – 12.2%) of the population in Västra Götaland will have been infected.

In the model, infectious individuals are grouped into “reported” and “unreported” cases. Reported cases 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 February 17 and June 5, 2020; cases infected abroad and cases identified through extended, screening-like testing, are not considered. Unreported cases are not part of the reported statistics and could have mild to more severe symptoms, but not severe enough to cause hospitalisation. To fit the model, additional data from two surveys conducted in the Stockholm region that measured the number of individuals currently infected with SARS-CoV-2 in the population at two time points are used. The surveys showed that 2.5% of the Stockholm population were positive for the virus between March 27 and April 3 and that 2.3% of the Stockholm population were positive for the virus between April 21 and April 24.

Sammanfattning på svenska

Denna uppdaterade rapport presenterar den matematiska modell som använts för att studera spridningen av covid-19 i fyra av Sveriges regioner: Dalarna, Skåne, Stockholm och Västra Götaland. De största uppdateringarna från föregående rapport består i att ett ytterligare fack introducerats i fackmodellen. Detta fack representerar individer som slutat vara smittsamma men som fortfarande kan testa positivt på ett PCR-test, det test som fångar aktiv eller nyligen aktiv infektion.

Med den matematiska modellen skattas antalet infekterade personer vid olika tidpunkter samt dagen med flest samtidigt smittsamma personer. Enligt modelleringen så kommer andelen som är eller har varit infekterade vid 1 juli vara: 8,5% (5,9 – 12,9%) i Dalarna, 4% (2,4 – 9,9%) i Skåne, 19% (17,7 – 20,2%) i Stockholm och 9% (6,3 – 12,2%) i Västra Götaland.

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 5 juni 2020, fall som är smittade utomlands och som ingår i den utökade testningen har exkluderats. Obekräftade fall 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 resultaten från de undersökningar som genomförts i Stockholms län för att mäta aktuell förekomst av SARS-CoV-2 i samhället. Undersökningarna visade att 2,5% av befolkningen i Stockholm var infekterade mellan 27 mars och 3 april och att 2,3% av befolkningen i Stockholm var infekterade mellan 21 april och 24 april.

Introduction

This report presents the updates of a previous model developed by the Public Health Agency of Sweden [1]. The model is used to study the spread of covid-19 in four regions of Sweden until June 5: Dalarna, Skåne, Stockholm, and Västra Götaland. 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 aid future forecasting, planning, and assessment of possible interventions.

Objective

The objective of this report is to estimate the total number of individuals infected with covid-19 in the regions Dalarna, Skåne, Stockholm, and Västra Götaland. We also investigate different scenarios based on different assumptions on contact intensity and the effect of varying the time period when infected individuals test positive on a PCR-test.

Major changes

Since this is an updated version of a previous report [1], we here state the modifications and extensions:

- Four regions are now studied instead of one.
- We have added a compartment of recovered individuals in order for infected individuals to be able to test positive on a PCR-test for longer than 5 days. In the previous report, the time of infectiousness and the assumed time-window for positive PCR-test coincided.
- We have fitted the model to two data-points on prevalence of active infection in Stockholm. Previously only one such point-prevalence was available and the model was calibrated to this value, generating a value on the fraction of unreported cases. Now we have included two point-prevalences at two different time points in the fitting of the model, allowing us to estimate the fraction of unreported cases.
- The model is seeded by the first (domestic) reported case and the corresponding number of unreported cases. The seeding is thereby included in the fitting.
- Since the last report, reporting of case data does now separate between cases identified in the healthcare system due to illness, and cases identified through screening-like testing. In order to get a stable series of case data, we have removed cases found trough screening-like testing.

Transmission model

SEIR-model

We will here describe the model, which is a brief extension to the model explained in [1].

We developed a compartmental model in which individuals are divided into different compartments depending on predetermined characteristics. 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 is 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 [2, 3].[1]

We divide infectious cases into two groups: reported cases and unreported cases. After the incubation period, an infected individual is either 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 $t_b =$ March 16 2020 (day 76 of the year) which is the day when people in Sweden, and particular in Stockholm, were recommended to work from home. The speed of the 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 who becomes infectious has an infectiousness that follows the time-dependent infectivity rate. We assume that reported and unreported cases are equally infectious. An individual is assumed to be infectious on average $\frac{1}{\gamma_1} = 5$ days [4].

[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 days (95% CI [4, 5.8]).

Data used in the fitting of the model include point prevalences found by PCR-testing in Stockholm at two different time points. Therefore, to fit the model we need to make assumptions on when individuals test positive on a PCR-test. Cases found in the random sample on which the Public Health Agency of Sweden base the estimates of ongoing or recent infection with covid-19 are generally mild cases. When infectious, we assume an individual can test positive on a PCR-test. In contrast to the previous modelling [1], we now split the recovered compartment into two: the first compartment R_1, in which an individual still can test positive on a PCR-test, and the second compartment R_2, in which an individual no longer test positive on a PCR-test. The median time an infected individual with mild symptoms is assumed to test positive on a PCR-test is 10 days, hence we assume the mean time spent in R_1 to be $\frac{1}{\gamma_2} = 5$ days [5].

We assume a closed population—no one enters and leaves the population. We denote the number of individuals in each compartment by S, E, I_r, I_u, R_1 and R_2. From this follows that the population size is $N = S + E + I_r + I_u + R_1 + R_2$. The transmission dynamics are described by the following equation system:

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

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

$$\frac{dI_o}{dt} = p_o \rho E - \gamma_1 I_u$$

$$\frac{dI_r}{dt} = (1 - p_o)\rho E - \gamma_1 I_r$$

$$\frac{dR_1}{dt} = \gamma_1(I_u + I_r) - \gamma_2 R_1$$

$$\frac{dR_2}{dt} = \gamma_2 R_1.$$

From this differential equation system, it is possible to calculate the numbers S, E, I_r, I_u, R_1 and R_2 at any time t, given the initial number of individuals in the different compartments at the start time t_0.

Note that the daily number of reported cases in the model at time t is given by $(1 - p_o)\rho E(t)$, which is fitted to the daily number of actual reported cases in the different studied regions.

Data

Observed number of daily reported cases

The observed data consists of the daily number of reported cases. Cases that were infected abroad are excluded, since they were not infected within the modelled population. We used the day of reported symptom onset (so-called *epi-date*) as the day when a case becomes infectious.[2] Note that the data the Public Health Agency of Sweden make publicly available is based on the reporting date. Data based on reporting date is more aggregated and do not adjust for the lag in reporting; therefore, the daily cases based on the epi-date and the reporting date will look different but the total numbers are the same. Further, we remove cases found based on extended testing of health care personnel, at primary care, and elderly care. Testing of health care personnel and patients at primary care are similar to a screening for covid-19, where the majority differ in severity from individuals not included in the extended testing. The effect of removing case data based on extended testing is studied in the sensitivity analysis.

Health report, Stockholm region

The Public Health Agency of Sweden conducted a study [6] in which 707 participants in a web panel of randomly recruited individuals in the Stockholm region, conducted self-sampling for active or recent covid-19 infection between March 27 and April 3, 2020. One additional study with the same panel as sampling frame was done between April 21 and April 24.[3] In the second study, 679 participants were recruited and tested.

In the first 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%). In the second study the weighted proportion positive individuals was 2.3%.

Studied regions

We studied four regions in Sweden, namely: Dalarna, Skåne, Stockholm, and Västra Götaland. See summary Table 1 for descriptive statistics for the four regions and Figure 1 for the daily number of reported cases by epi-date.

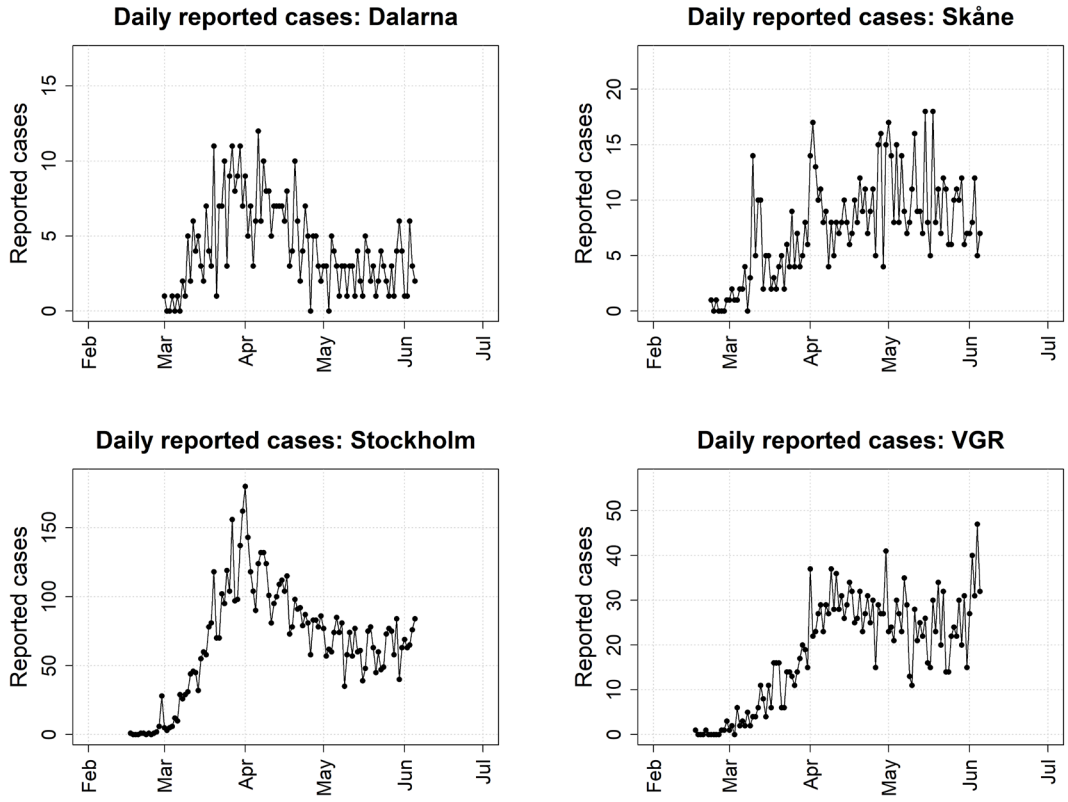
[2] If symptom onset day was not reported we used the day of testing.

[3] See: Folkhälsomyndigheten news (in Swedish)

Table 1: Summary of studied regions. Number of cases referrers to reported domestic cases, cases found in extended testing were removed.[5]

Region	Population[4]	Number of domestic reported cases[5] until June 5	% Domestic cases among population
Stockholm	2 374 550	7 378	0.311
Västra Götaland	1 724 529	2 034	0.118
Skåne	1 376 659	783	0.057
Dalarna	287 795	416	0.145

Figure 1: Daily reported cases[5] for the four studied regions.



[4] Population sizes from SCB, 2019

[5] Extended testing of health care personnel, at primary care, and elderly care excluded

Method

Fitting of the model to observed data

The parameters of the SEIR-model at time t were fitted to the reported number of domestic cases per day in the four regions region until June 5, 2020. More specifically, we fitted the model parameters so that the estimated incidence of reported cases each day ($(1-p_o)\rho E(t)$) were similar to the observed daily incidence of reported cases. We fixed the period when an infected individual is infectious and the duration of the latency period but estimated all parameters for the varying infectivity $\theta, \delta,$ and ε. We assume that at time t_0, for each region, there was one infectious reported individual and $p_u/p_r = p_u/(1-p_u)$ unreported cases.

When fitting the parameters to the case data, we minimised the residual squared error between the model-estimated number of new cases per day and the observed number of new cases per day. For the region of Stockholm, we additionally use the two data points on active infections of 2.5% between the dates March 27 and April 3 and of 2.3% between the dates April 21 and April 24. These are included in the fitting as binomial likelihoods. Since these point-estimates of the prevalence of active infection are only available for Stockholm, we can only estimate the proportion unreported cases for this region. The estimated proportion of unreported cases of Stockholm will be used as a known constant for the other regions.

To obtain parametric bootstrap confidence intervals (CIs), 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 result tables.

Simulating the outbreak beyond the date of cases—increased contacts

Since recommendations and restrictions were communicated by the Swedish government and the Public Health Agency of Sweden, physical distancing seem to have contributed to a lower infectivity rate. The infectivity rate is a combination of the probability of transmission and the number of contacts. From case data until June 5, we estimate the infectivity rate. However, in the near future, the physical distancing observed on June 5 may decrease when people start to ease up, leading to a higher infectivity rate again. This is the motivation to study the effect of increased contacts.

Table 2: Parameters of the model

Parameter	**Value**
Length of latency phase $\mathbf{1}/\boldsymbol{\rho}$	5.1 days
Length of infectiousness $\mathbf{1}/\boldsymbol{\gamma_1}$	5 days
Additional time of testing positive on a PCR-test after recovery $\mathbf{1}/\boldsymbol{\gamma_2}$	5 days
Start date $\boldsymbol{t_0}$	
Stockholm	February 17
Västra Götaland	February 17
Skåne	February 23
Dalarna	March 1
Turning point between $\boldsymbol{\theta}$ and $\boldsymbol{\delta\theta}$ occurs at time $\boldsymbol{t_b}$.	March 16
Infectivity rate parameters: $\boldsymbol{\theta}, \boldsymbol{\delta}$ och $\boldsymbol{\varepsilon}$	Estimated by the model
Infectivity rate at time $\boldsymbol{t}$: $\boldsymbol{b_t}$	Estimated by the model
Proportion unreported cases: $\boldsymbol{p_u}$	Estimated by the model

Results

Based on the reported cases in the four regions Dalarna, Skåne, Stockholm, and Västra Götaland until June 5, 2020, we estimated the spread of covid-19. Using the estimated infectivity rate b_t, we also make a forecast in which we assume no change in contact behaviour. We continue by showing the results of different scenarios of increased contacts, by increasing the infectivity by the same factor, from June 10.

Estimated model parameters and number of infected

In Table 2 the estimated parameters are shown. We could only estimate the proportion of cases that are unreported, p_u, for the region of Stockholm. For the other regions, the Stockholm region estimate is used as a known constant. The estimated unreported cases in Table 3 mean that there are 55 unreported cases for each reported case when extended testing and elderly care are excluded.

In Figure 2 to Figure 5 the estimated and observed number of reported cases, and the number of currently infectious individuals ($I_u + I_r$) are shown for the four regions. Note that the uncertainty for Skåne is quite large.

Table 3: Log-likelihood (LogL), estimated parameters, and confidence intervals (95% CI), by region.

Region	**LogL**	$\hat{\delta}$ **(95% CI)**	$\hat{\varepsilon}$ **(95% CI)**	$\hat{\theta}$ **(95% CI)**	$\hat{p}_u$ **(95% CI)**
Stockholm	-591.36	0.205	-0.258	0.944	0.982
		[0.201, 0.208]	[-0.283, -0.233]	[0.993, 1.012]	[0.981, 0.983]
Dalarna	-292.10	0.160	-0.188	1.020	-
		[0.143, 0.179]	[-0.243, -0.132]	[0.985, 1.056]	
Skåne	-368.51	0.286	-0.118	0.702	-
		[0.260, 0.313]	[-0.178, -0.057]	[0.677, 0.729]	
Västra Götaland	-453.94	0.290	-0.109	0.700	-
		[0.273, 0.308]	[-0.133, -0.084]	[0.685, 0.715]	

Figure 2: Stockholm. To the left: estimated daily incidence of newly reported cases, inflow to I_r (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 I_r and I_u, both reported and unreported cases (the red line) with a 95% confidence interval (the dashed lines).

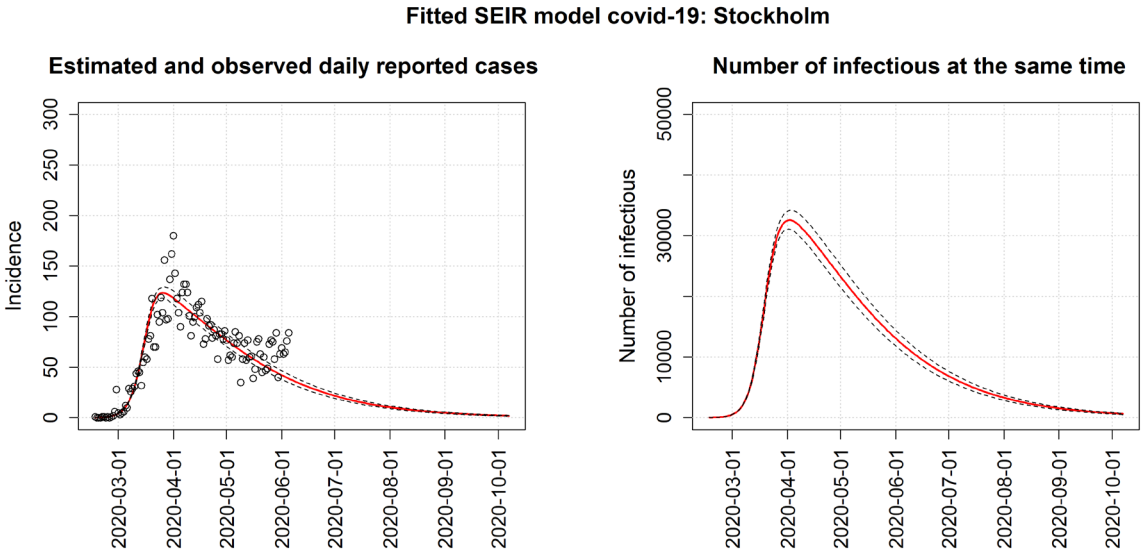


Figure 3: Dalarna. To the left: estimated daily incidence of newly reported cases, inflow to I_r (the red line), with a 95% confidence interval (the dashed lines) and observed data (circles). To the right: prevalence of infectious cases, all individuals in I_r and I_u, both reported and unreported cases (the red line) with a 95% confidence interval (the dashed lines).

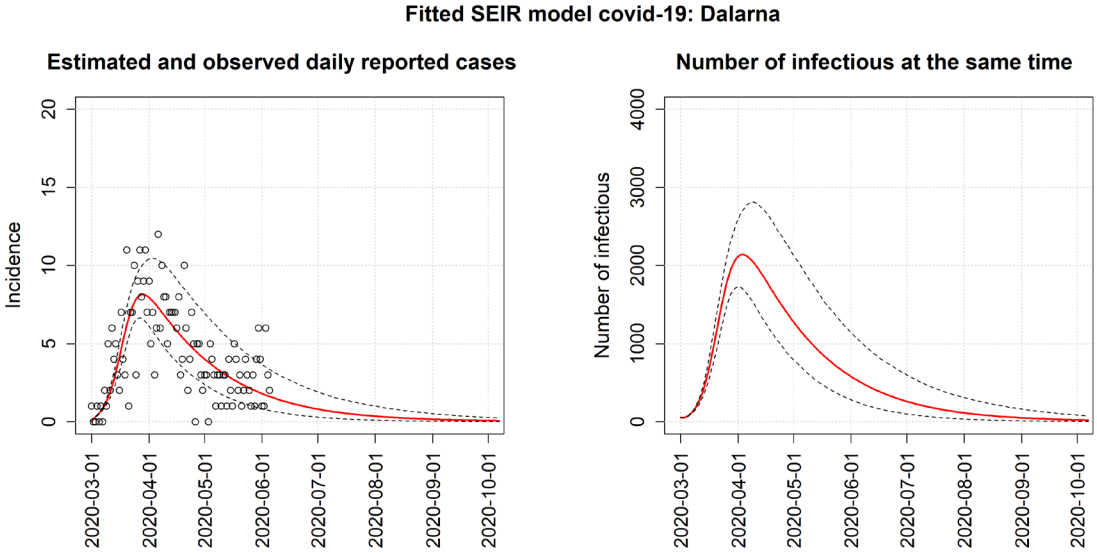


Figure 4: Skåne. To the left: estimated daily incidence of newly reported cases, inflow to I_r (the red line), with a 95% confidence interval (the dashed lines) and observed data (circles). To the right: prevalence of infectious cases, all individuals in I_r and I_u, both reported and unreported cases (the red line) with a 95% confidence interval (the dashed lines).

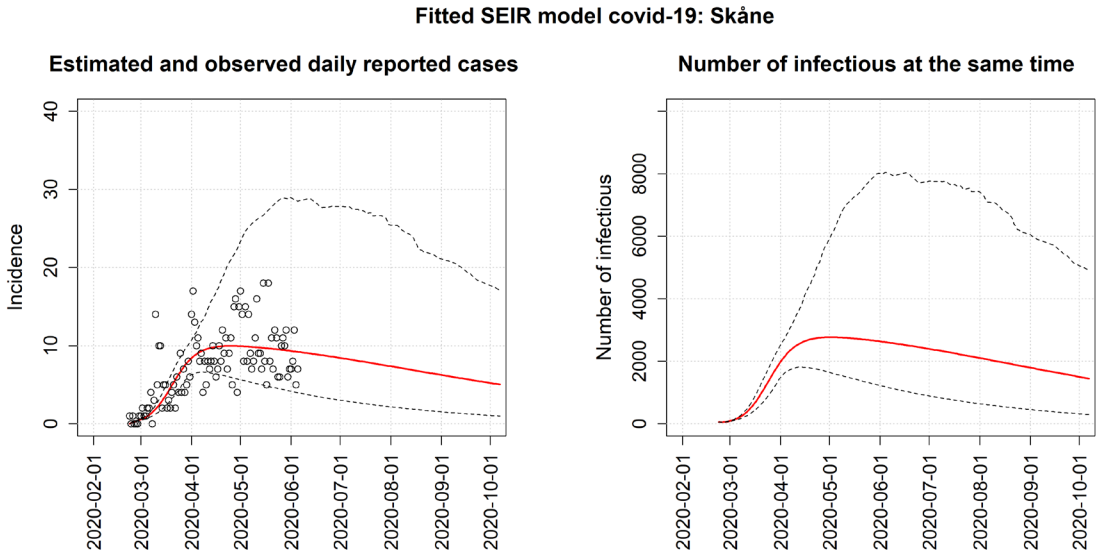
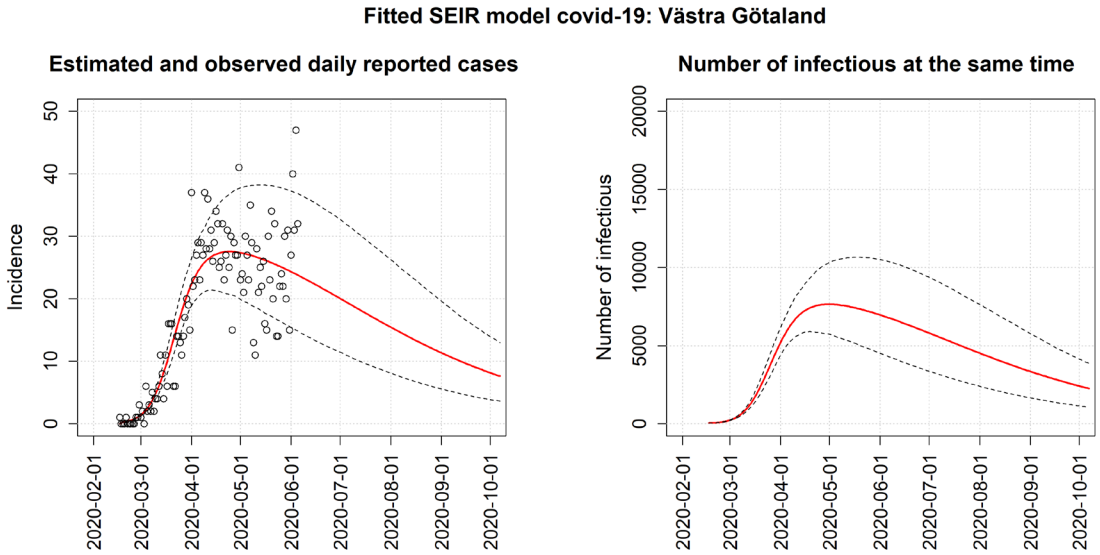


Figure 5: Västra Götaland. To the left: estimated daily incidence of newly reported cases, inflow to I_r (the red line), with a 95% confidence interval (the dashed lines) and observed data (circles). To the right: prevalence of infectious cases, all individuals in I_r and I_u, both reported and unreported cases (the red line) with a 95% confidence interval (the dashed lines).



In Table 4 we show the estimated number of accumulated infected individuals ($E + I + R$) at May 1, July 1, and September 1. For different dates until October 1, see Figure 6 to Figure 9. Note that these results assume the infectivity rate based on the estimated parameters in Table 3 for each region. In Figure 6 to Figure 9, we show the fraction of the population having had active infection ($I + R$). Given that

we assume that all who have had active infection develop antibodies, this would be the fraction positive in a serological test for antibodies. Furthermore, if we assume that it takes two weeks to develop antibodies after symptom onset, in order to answer what proportion has antibodies in e.g. week 20, we would need to inspect the graphs for week 18.

In Table 6 the estimated infectivity rate at the day of the first reported case and at June 5 are shown. The infectivity rate decreased by a factor of 5.9 for Dalarna, 4.9 for Stockholm, 3.3 for Skåne and 3.3 for Västra Götaland.

Table 4: Estimated number and proportion of accumulated infected individuals in each studied region. For values forward in time we assume individual's behaviour, and thereby the estimated infectivity, will not change. For the estimated proportions a 95% confidence interval (95% CI) are given.

Accumulated number of infected ($E + I + R$) and proportion of the population by region.

	2020-05-01		2020-07-01		2020-09-01	
Region	Number	Proportion (95% CI)	Number	Proportion (95% CI)	Number	Proportion (95% CI)
Stockholm	311 718	0.13	449 045	0.19	483 971	0.20
		[0.125, 0.138]		[0.177, 0.202]		[0.190, 0.218]
Dalarna	18 612	0.07	24 568	0.085	25 737	0.089
		[0.049, 0.088]		[0.059, 0.129]		[0.061, 0.139]
Skåne	26 074	0.02	57 245	0.04	82 072	0.06
		[0.013, 0.030]		[0.024, 0.099]		[0.029, 0.158]
Västra Götaland	70 562	0.04	150 864	0.09	202 204	0.12
		[0.033, 0.05]		[0.063, 0.122]		[0.079, 0.172]

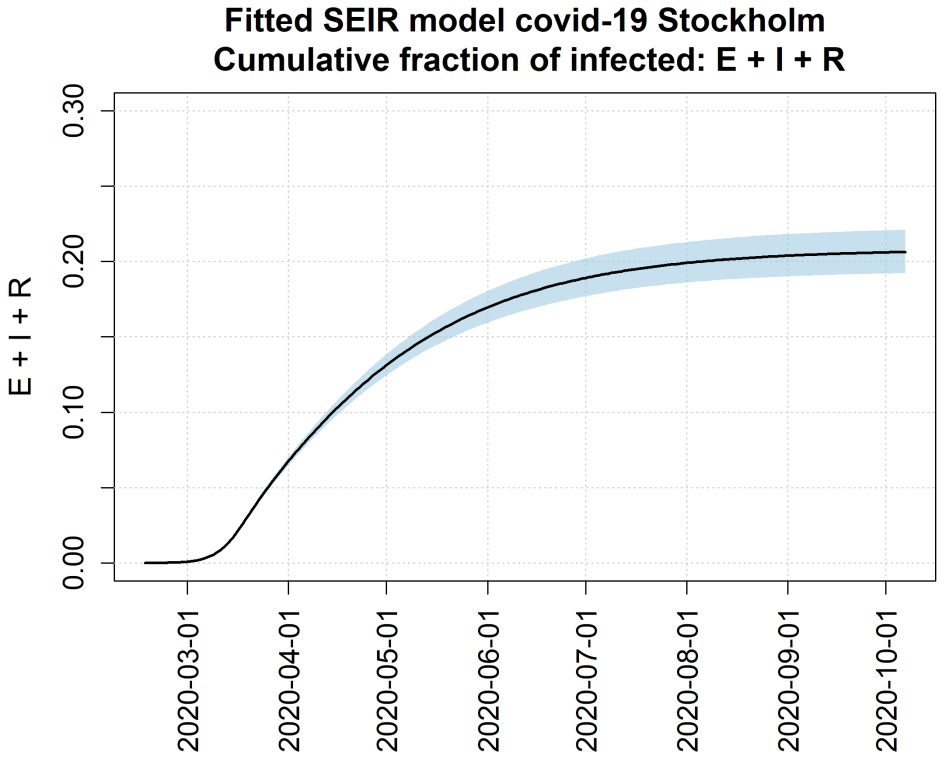
Table 5: Estimated peak-day and number of infectious individuals, $I_r + I_u$ (prevalence), and the estimated peak-day of the daily number of new cases (incidence), i.e. the daily inflow to I_r and I_u, by region. Below the estimated values, 95% confidence intervals are given.

Region	**Prevalence**		**Incidence**	
	Peak day (95% CI)	Prevalence on peak-day (95% CI)	Peak day (95% CI)	Incidence on peak-day (95% CI)
Stockholm	2020-04-02	32 584	2020-03-26	6 902
	[2020-04-02, 2020-04-03]	[31 101, 34 206]	[2020-03-26, 2020-03-27]	[6 607, 7 212]
Dalarna	2020-04-03	2 139	2020-03-29	455
	[2020-04-02, 2020-04-09]	[1 731, 2 815]	[2020-03-27, 2020-04-03]	[372, 586]
Skåne	2020-05-02	2 768	2020-04-25	556
	[2020-04-13, 2020-06-05]	[1 817, 8 049]	[2020-04-07, 2020-05-31]	[370, 1 621]
Västra Götaland	2020-05-01	7 648	2020-04-24	1 538
	[2020-04-19, 2020-05-18]	[5 901, 10 657]	[2020-04-13, 2020-05-13]	[1 195, 2 136]

Table 6: Estimated infectivity on the day with the first reported case and the last day of reported cases in our analysis, June 5, by region.

Region	**Infectivity first day**	**Infectivity 2020-06-05**	**Reduction in infectivity**
Stockholm	0.944	0.193	4.885
Dalarna	0.971	0.163	5.947
Skåne	0.667	0.201	3.323
Västra Götaland	0.677	0.203	3.331

Figure 6: Stockholm. Upper graph: estimated cumulative fraction of infected ($E + I + R$). Lower graph: estimated cumulative fraction ever having had active infection ($I + R$). Both with 95% confidence interval in blue.



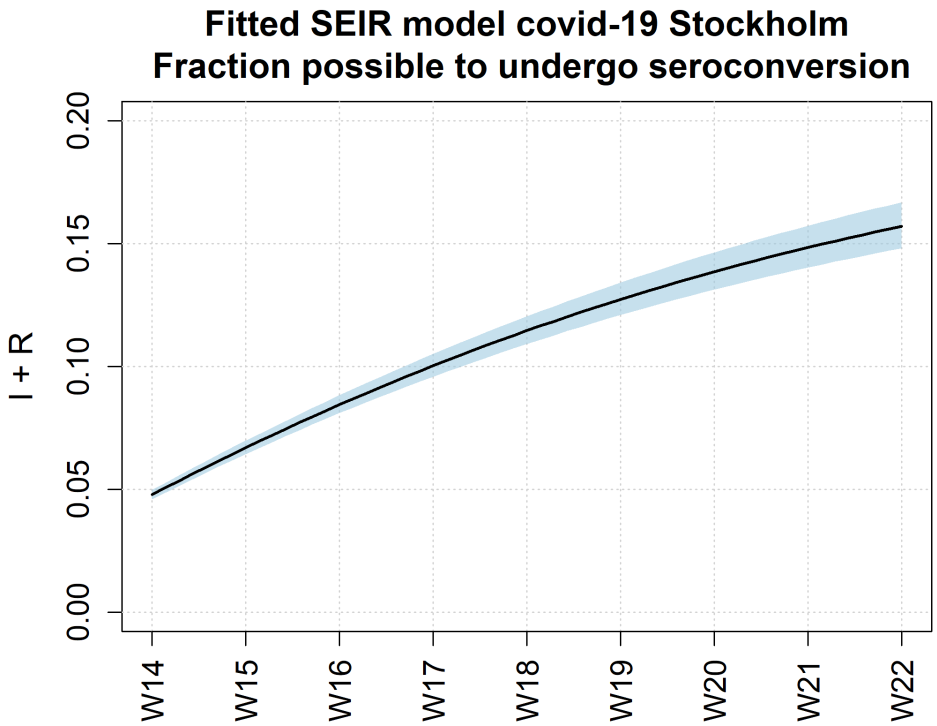
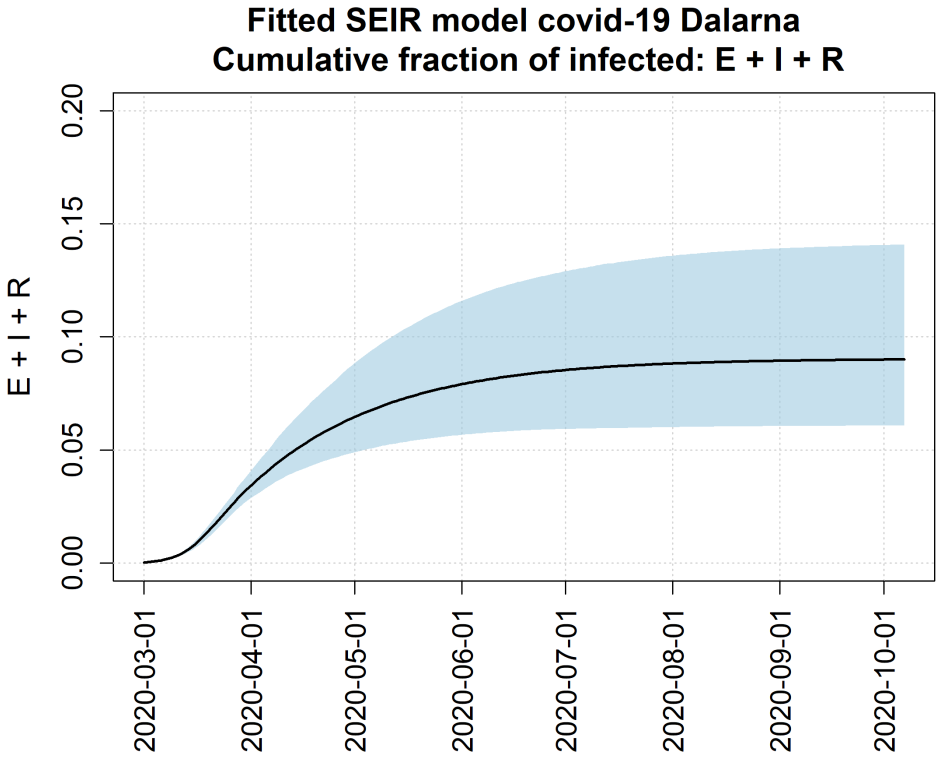


Figure 7: Dalarna. Upper graph: estimated cumulative fraction of infected ($E + I + R$). Lower graph: estimated cumulative fraction ever having had active infection ($I + R$). Both with 95% confidence interval in blue.



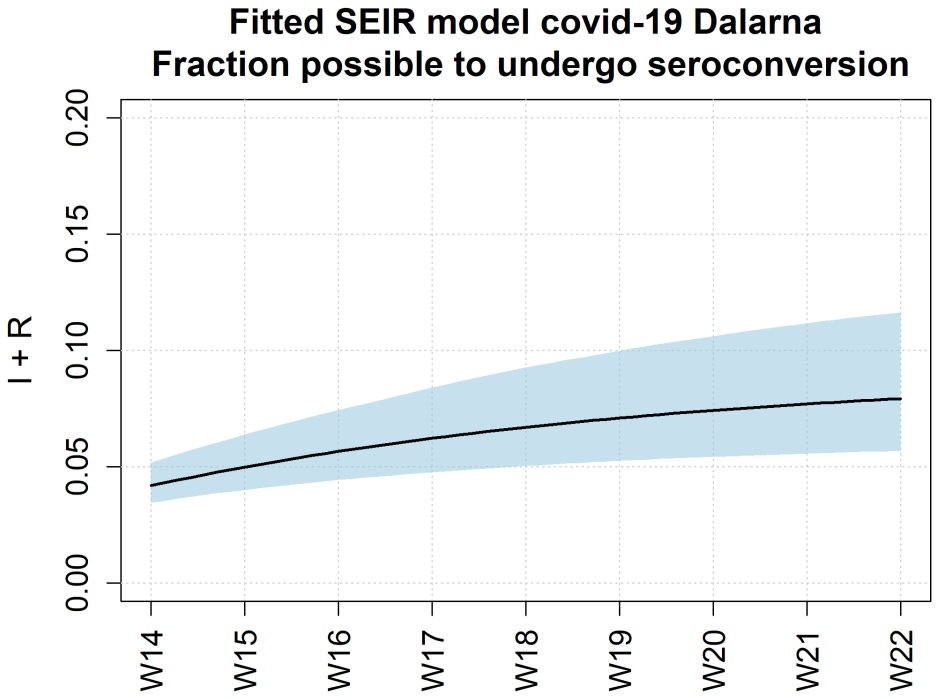
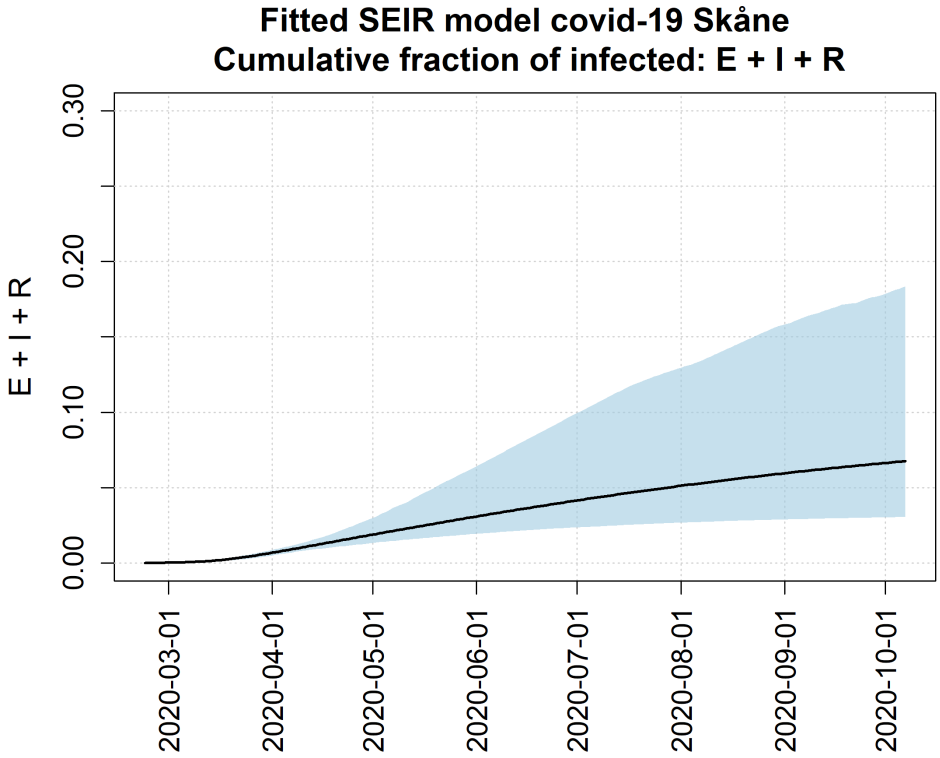


Figure 8: Skåne. Upper graph: estimated cumulative fraction of infected ($E + I + R$). Lower graph: estimated cumulative fraction ever having had active infection ($I + R$). Both with 95% confidence interval in blue.



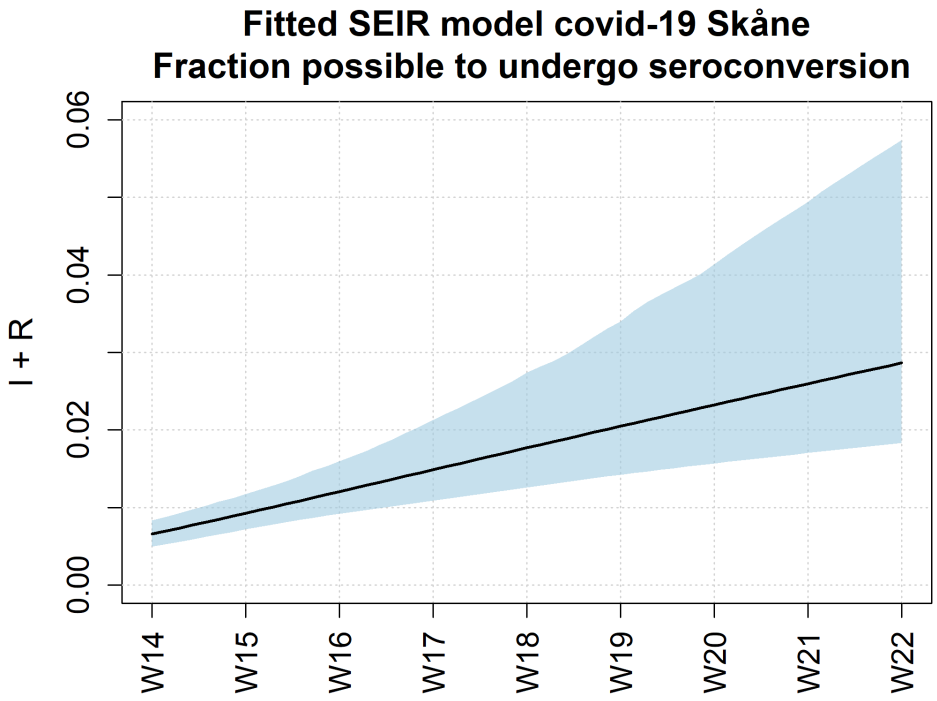
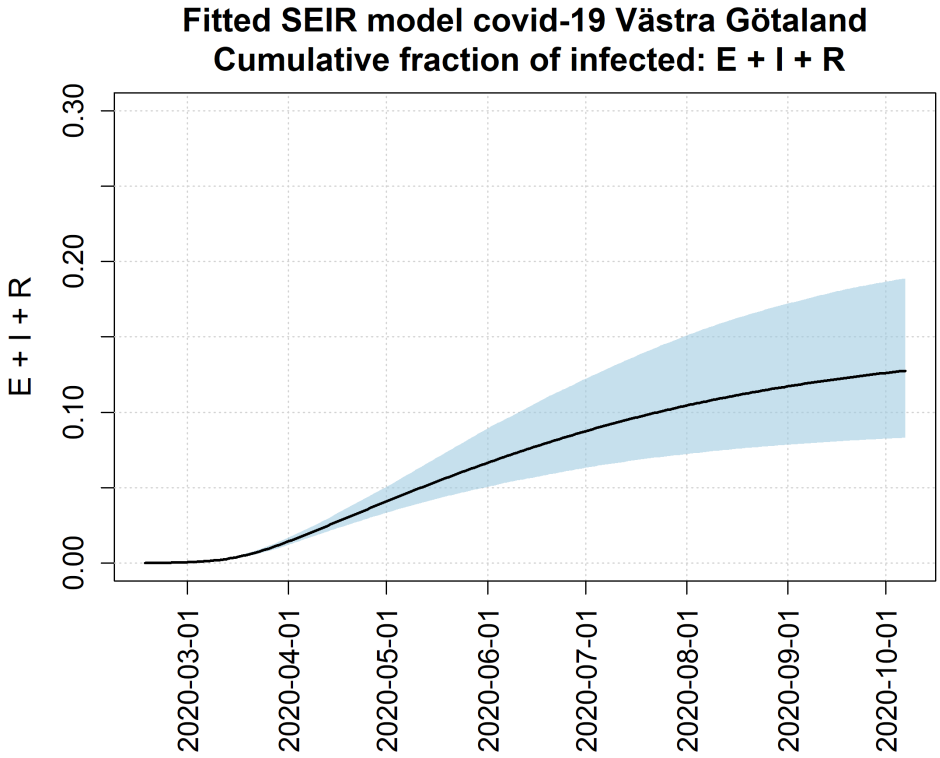
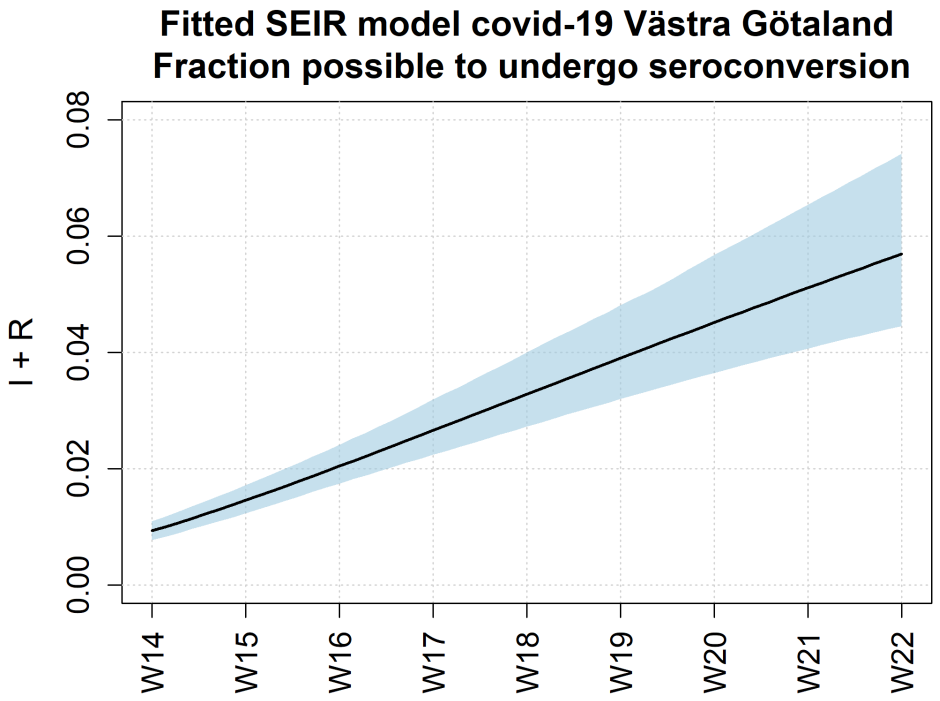


Figure 9: Västra Götaland. Upper graph: Estimated cumulative fraction of infected ($E + I + R$). Lower graph: Estimated cumulative fraction ever having had active infection ($I + R$). Both with 95% confidence interval in blue.





Increased contacts after June 10

The results in the previous sub-section assume that the estimated infectivity rates based on data until June 5 continue. The infectivity rates for all regions did decrease, to varying degrees (Table 6), from the first reported case until June 5. In Dalarna and Stockholm the estimated infectivity decreased the most, while the decrease was less in Skåne and Västra Götaland.

The infectivity rate is a combination of the contact rate and the probability of infection in one contact. Increasing the number of contacts by a factor 2 is equivalent to increasing the infectivity by a factor 2. In this section we investigate the effect of increasing the level of contacts from June 10. The level of contacts, and thereby the infectivity, is continuously increased from the estimated level at June 10 to the assumed highest level in the end of summer, August 31. The level on August 31 is then kept throughout the year.

In the analyses, the level of contacts are increased from June 10 by 20%, 40%, 60%, 80%, and 100% for Dalarna and Stockholm. In Dalarna, the estimated infectivity rate decreased from 0.97 in February to 0.16 in the beginning of June, a 100% increase of the June infectivity means that the infectivity is doubled to 0.32. Hence, the increased infectivity is still at a lower level than in the early beginning of the covid-19 pandemic in Sweden. For Skåne and Västra Götaland the contacts are increased by 10%, 20%, 30%, 40%, and 50%. The infectivity rate for Skåne decreased from 0.67 in February to 0.2 in the beginning of June, meaning that a 50% increase of the value in June would yield an infectivity rate of 0.3.

The results of increasing the contact rate are shown in Figure 10 to Figure 13. In Figure 10 and Figure 11, we see that if the increase in contacts can be kept lower than 60% for Stockholm and Dalarna, a second wave would not become worse than the first wave. For Skåne, Figure 12, the increase in contacts would need to be within 10% of the value estimated for June 10 to obtain a level of cases lower than reported cases already observed. In Figure 13 we see that if the contact increase in Västra Götaland is less than 20%, the second wave will not become worse than the first observed wave.

Figure 10: Stockholm. Simulated number of reported cases when the contacts are allowed to increase by different levels.

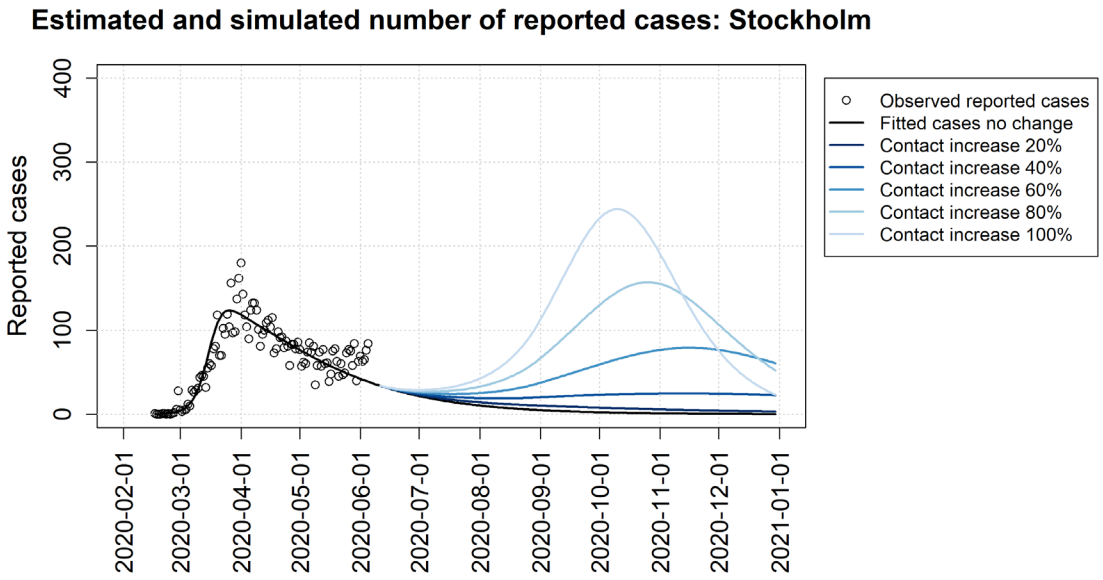


Figure 11: Dalarna. Simulated number of reported cases when the contacts are allowed to increase by different levels.

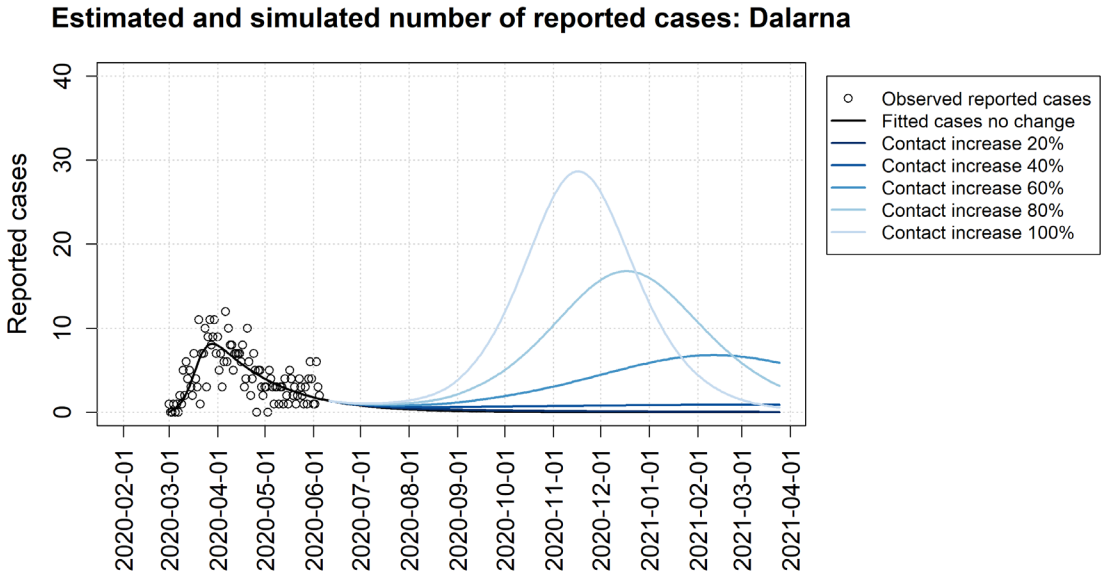


Figure 12: Skåne. Simulated number of reported cases when the contacts are allowed to increase by different levels.

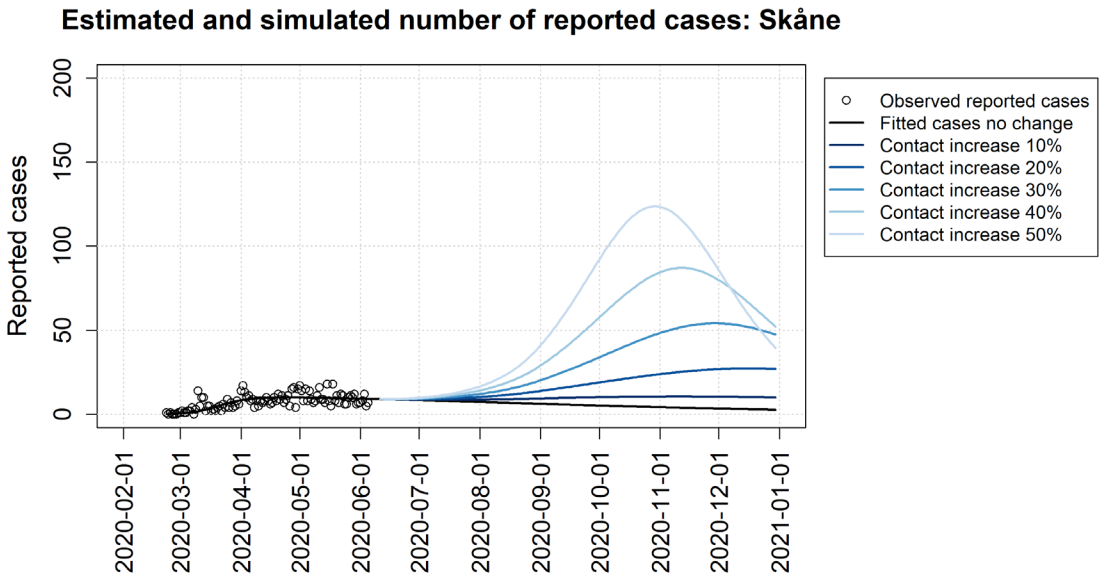
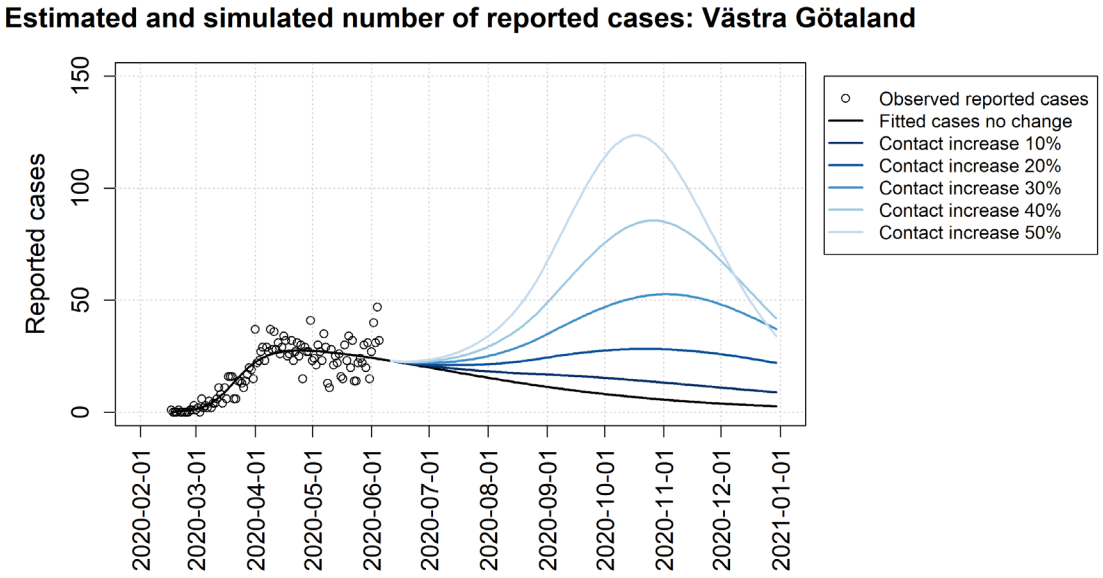


Figure 13: Västra Götaland. Simulated number of reported cases when the contacts are allowed to increase by different levels.



Sensitivity analysis

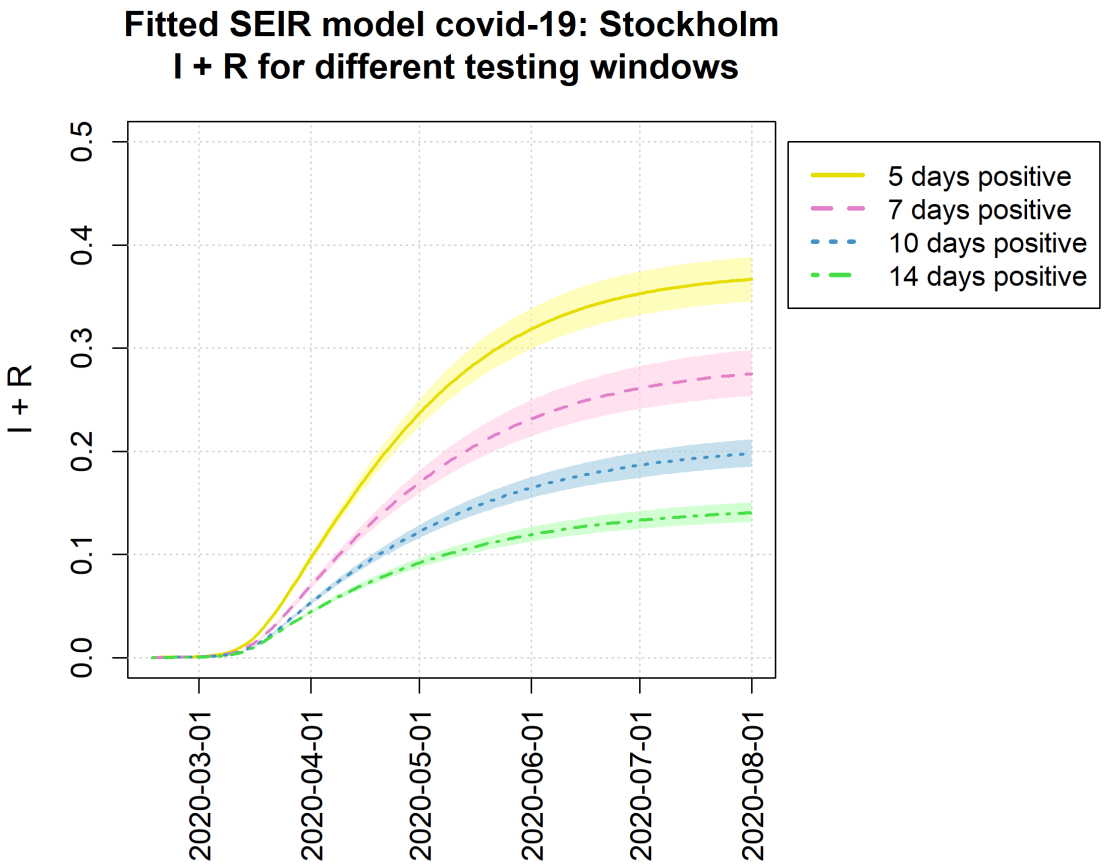
Median time testing positive on a PCR-test when having mild infection

As explained in the section named SEIR-model, in the main analysis we assume that individuals not requiring hospital care test positive on a PCR-test on average 10 days. The assumption of a testing window of 10 days is incorporated in the model by the two data points on point-prevalences at two different dates provided by the Health report Stockholm. The model estimates how many the unreported cases must be in order to fit the observed point-prevalences, given a 10 day testing window. In Table 7 we show the estimated fraction of unreported cases, when extended testing and elderly care are excluded, given the assumption of a testing window of either 5, 7, 10, or 14 days. A shorter testing window means that a larger proportion must be infected at the given dates in order to reach the observed prevalences of 2.5% and 2.3%. In Figure 14, we show the fraction infectious and recovered $(I + R)$ for the different testing windows.

Table 7: Estimated fraction of unreported cases and their 95% CI for different testing windows on a PCR-test for covid-19.

Days positive	**5**	**7**	**10**	**14**
Fraction unreported ($\hat{p}_u$)	0.9908	0.9872	0.9821	0.9760
95% CI	[0.9904, 0.9911]	[0.9868, 0.9876]	[0.9815, 0.9827]	[0.975, 0.9769]
Number of unreported cases per reported case	107.3	77.2	54.9	40.7
95% CI	[103.17, 111.36]	[74.76, 79.65]	[53.05, 56.80]	[39.00, 42.29]

Figure 14: Estimated cumulative fraction of infectious individuals given different testing windows on a PCR-test for covid-19. The lines show the estimated values and the shaded areas the 95% confidence intervals.



Removing case data based on extended testing

To obtain a stable series of case data, in which the numbers would not vary due to current testing strategy, we excluded the following groups of cases: health care personnel, primary care cases, and cases from elderly care centres. We investigate the effect of varying exclusion of groups in the Stockholm region. The analysis is performed with

- no data excluded,
- health care personnel excluded
- health care personnel and primary care cases excluded
- health care personnel, primary care cases, and cases from the elderly care excluded, as in the main analysis.

The fraction of unreported cases varies most with different exclusions but the other results are similar. This shows that the model seems to adjust for the removed data and still estimate the spread in a similar way; removing more data on cases increases the fraction unreported cases (Table 8) but still the estimated fraction infected by May 1, July 1, and September 1 is very similar (Table 9).

Figure 15: Number of confirmed domestic cases in Stockholm.

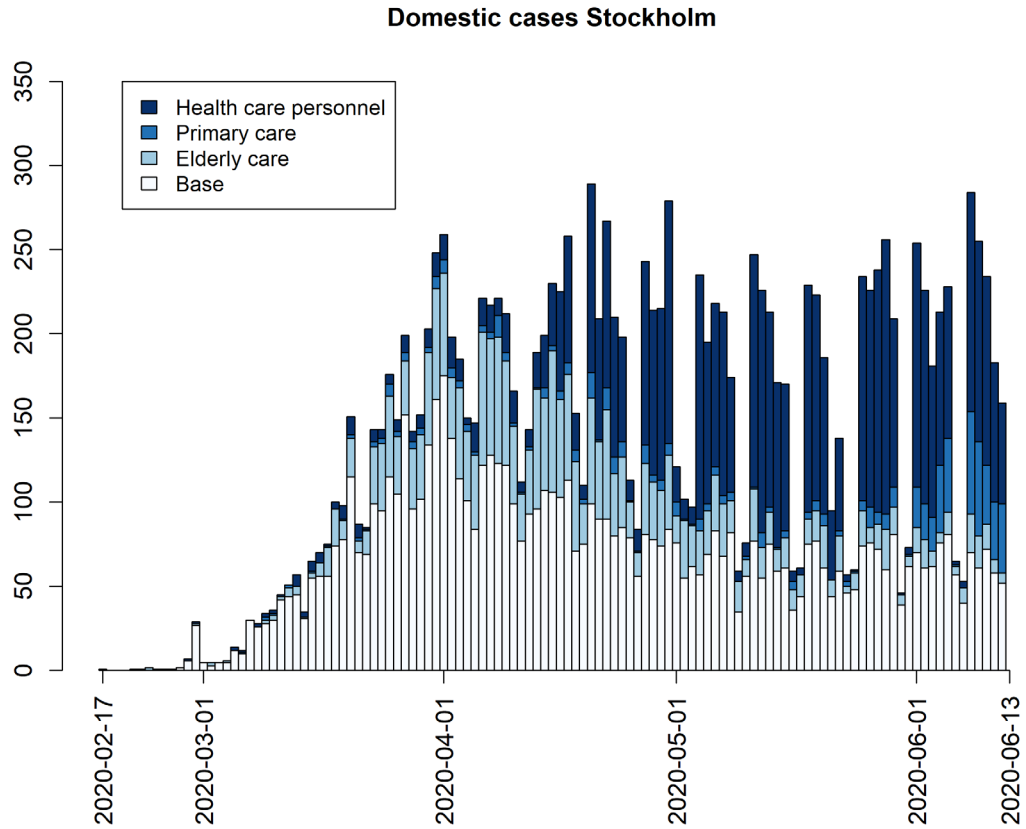


Table 8: Estimated fraction of unreported cases with 95% CI for different exclusions of case data.

Case data removed	**Health care personnel, primary care cases, elderly care cases**	**Health care personnel, primary care cases**	**Health care personnel**	**None**
Fraction unreported ($\hat{p}_u$)	0.9821	0.9752	0.9744	0.967
95% CI	[0.9815, 0.9827]	[0.9743, 0.976]	[0.9734, 0.9752]	[0.9642, 0.9697]
Number of unreported cases per reported case	54.8	39.29	38.06	29.4
95% CI	[53.05, 56.80]	[37.91, 40.67]	[36.59, 39.32]	[26.9, 32.0]

Table 9: Estimated number and proportion of accumulated infected individuals with 95% CI in Stockholm for different exclusions of case data.

Accumulated number of infected ($E + I + R$) and proportion of the population.

	2020-05-01		**2020-07-01**		**2020-09-01**	
Case data removed	**Number**	**Proportion (95% CI)**	**Number**	**Proportion (95% CI)**	**Number**	**Proportion (95% CI)**
Health care personnel, primary care cases, elderly care cases	311 718	0.13	449 045	0.19	483 971	0.20
		[0.125, 0.138]		[0.177, 0.202]		[0.190, 0.218]
Health care personnel, primary care cases	307 616	0.13	427 591	0.18	454 113	0.19
		[0.121, 0.138]		[0.165, 0.196]		[0.174, 0.209]
Health care personnel	308 592	0.13	435 671	0.18	465 541	0.20
		[0.123, 0.138]		[0.170, 0.197]		[0.181, 0.212]
None	315 507	0.13	477 900	0.20	526 248	0.22
		[0.118, 0.150]		[0.180, 0.226]		[0.199, 0.247]

Table 10: Estimated peak-day and number of infectious individuals, $I_r + I_u$ (prevalence), and the estimated peak-day the daily number of new cases (incidence), i.e. the daily inflow to I_r and I_u, for different exclusions of case data. Below the estimated values 95% confidence intervals are given.

Case data removed	**Prevalence**		**Incidence**	
	Peak day (95% CI)	Prevalence on peak-day (95% CI)	Peak day (95% CI)	Incidence on peak-day (95% CI)
Health care personnel, primary care cases, elderly care cases	2020-04-02	32 584	2020-03-26	6 902
	[2020-04-02, 2020-04-03]	[31 101, 34 206]	[2020-03-26, 2020-03-27]	[6 607, 7 212]
Health care personnel, primary care cases	2020-04-03	33 440	2020-03-28	7 078
	[2020-04-02, 2020-04-04]	[31 646, 35 415]	[2020-03-27, 2020-03-28]	[6 723, 7 463]
Health care personnel	2020-04-03	33 105	2020-03-27	7 017
	[2020-04-02, 2020-04-04]	[31 575, 34 872]	[2020-03-26, 2020-03-28]	[6 719, 7 341]
None	2020-03-31	31 462	2020-03-22	6 834
	[2020-03-30, 2020-04-02]	[27 965, 35 739]	[2020-03-22, 2020-03-22]	[6 124, 7 607]

Table 11: Estimated infectivity on the day with the first reported case and the last day of reported cases, June 5, for different exclusions of case data.

Case data removed	Infectivity first day	Infectivity 2020-06-05	Reduction in infectivity
Health care personnel, primary care cases, elderly care cases	0.944	0.193	4.885
Health care personnel, primary care cases	0.993	0.186	5.339
Health care personnel	0.997	0.189	5.272
None	1.033	0.202	5.114

Discussion and limitations

Discussion

We have modelled the spread of covid-19 in four regions in Sweden by fitting an SEIR model to reported cases and two point-prevalences at two different dates in Stockholm. We find that by July 1, 8.5% (5.9 – 12.9%) of the population in Dalarna will have been infected, 19% (17.7 – 20.2%) of the population in Stockholm will have been infected, 4% (2.4 – 9.9%) of the population in Skåne will have been infected, and 9% (6.3 – 12.2%) of the population in Västra Götaland will have been infected. For Stockholm it is possible to estimate the fraction unreported cases and we find it to be 0.9821, i.e. per one reported case there are 55 unreported cases.

We also investigated the effect of increased contacts during the summer that stabilises in autumn. We found that if the contacts in Stockholm and Dalarna increase by less than 60% in comparison to the contact rate in the beginning of June, the second wave will not exceed the observed first wave. For Skåne the contacts cannot increase with more than 10% in comparison to the contact rate observed in the beginning of June to achieve a level of cases lower than already observed. In Table 1 we saw that only 0.057% of the Skåne population were confirmed cases, when excluding extended testing. This is the region with the lowest percentage among the four regions. In Västra Götaland, if the contact increase is less than 20% the second wave will not become worse than the first observed wave.

In choosing the window period for the PCR-test we need to consider the sample of individuals tested. Naturally, most studies investigating the testing window are carried out with a sample of hospitalised patients. Our chosen duration of 10 days for the testing window is based on a study [5] on mild and asymptomatic cases found by contact tracing. We believe that this sample is more similar to the cases found by Health report than hospitalised cases in other studies are. In the sensitivity analysis we studied the effect of assuming different durations an infected individual could be detected positive with a PCR-test. We found that this assumption has a major impact on the estimated number that have been infected at different time points. There are still many unknowns concerning the covid-19 infection, especially regarding mild infections. Recently, information has been presented that mild cases never seem to develop antibodies against SARS-CoV-2, but only T-cell-mediated immunity. If it is also the case that infected individuals with very mild symptoms have short PCR test windows, perhaps being missed entirely, the assumed test window of 10 days could be an overestimate and we would thereby underestimate the number of infected.

In Sweden, the testing strategy has varied with time. To maintain a stable series of case data that does not vary with the testing strategy, we excluded cases found by the extended testing, in which a larger proportion of mild cases are found. In the analysis studying the effect of different exclusions of case data, it was shown that

the estimated number of infections was insensitive to different exclusion scenarios. The parameter being sensitive to different exclusions was the proportion of unreported cases.

Limitations

As stated in [1], 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 used in this analysis does not account for variation in contacts, where few individuals may have many contacts while the majority have fewer. This simplification, i.e. a homogenous contact structure, usually results in a somewhat faster growth of an epidemic. There is therefore a risk of overestimating the speed of the outbreaks.

The model used in this study is not age-stratified. The disease affects different age groups differently; e.g., young people seem to get milder infections. In this modelling study we assume that each infected individual has the same infectivity and the same risk of becoming a reported case, disregarded his or her age. Additionally, different age groups normally have varied degrees of contacts and have changed their behaviour differently during the covid-19 pandemic. This is not captured in the model and could for example affect our analysis on increased contacts. If, in reality, an increase in contacts in the population mainly is attributed to young people but those above the age of 70 maintain their physical distancing, then our results on increased contacts could be misleading. Since then the increased spread due to increased infectivity would mainly occur within a group commonly known not to develop severe enough symptoms requiring hospital care and therefore not becoming a reported case within the non-extended testing. While in the model, the increase occurs for all individuals to the same degree, and the model would overestimate the cases requiring hospital care.

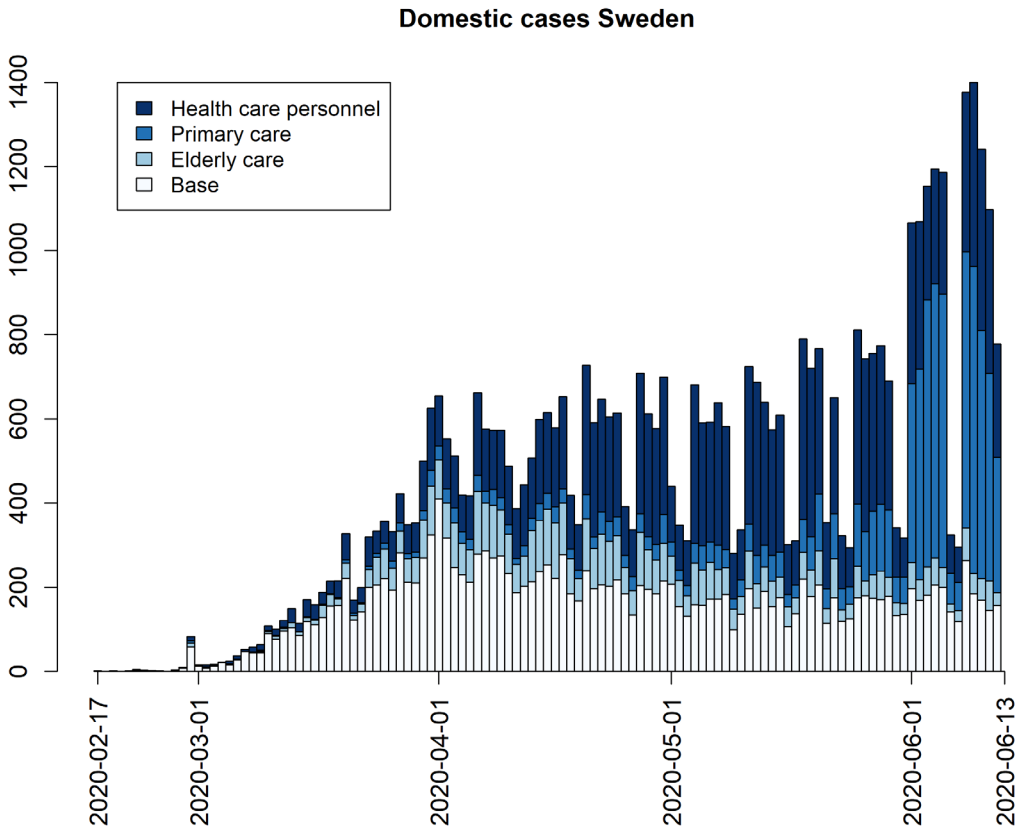
We assume a constant relationship between reported and unreported cases over time. This is a limitation since the routines and strategies for testing were changed on three occasions the studied period. The first change took place on March 12, 2020, when the focus was shifted from testing individuals with symptoms coming from known risk areas abroad (and contact tracing of these individuals) to individuals in need of inpatient hospital care. The second change was in April when more health care workers began to be tested. This testing can be seen as a screening and finds much more mild infections than those identified when seeking care. The third change happened in late May and the beginning of June when even more screening was put into place, larger groups of patients were tested in the primary care. See Figure 15 for the number of cases in Sweden by test type. To avoid the problem of unstable case data we excluded certain groups of cases from the analysis. However, this is a loss of information even though the estimated process seems to be rather unaffected.

The results on point-prevalences in Stockholm are incorporated in the model as two data points. However, we do not incorporate the uncertainty of these two estimates.

Including these uncertainties would most likely increase the estimated confidence intervals for the Stockholm estimates.

The estimated fraction of unreported cases, based on data from Stockholm, is applied for all regions. It is possible that this fraction in reality differs by regions.

Figure 16: Number of confirmed domestic cases in Sweden.



References

[1] 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,” Public health agency of Sweden, 2020.

[2] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong and al., “Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia,” *N Engl J Med,* vol. 382, no. 13, pp. 1199-1207, 2020.

[3] N. Linton, T. Kobayashi, Y. Yang, K. Hayashi, A. Akhmetzhanov and S.-M. Jung, "Incubation Period and Other Epidemiological Characteristics of 2019 Novel Coronavirus Infections with Right Truncation: A Statistical Analysis of Publicly Available Case Data," *J. Clin. Med.,* vol. 9, no. 2, p. 538, 2020.

[4] R. Wölfel, V. Corman, W. Guggemos, M. Seilmaier, S. Zange, M. Müller, D. Niemeyer, T. Jones and P. Vollmar, "Virological assessment of hospitalized patients with COVID-2019," *Nature,* vol. 581, pp. 465 - 469, 2020.

[5] Z. Hu, S. C, C. Xu, G. Jin, Y. Chen, X. Xu, M. H, W. Chen, Y. Lin, Y. Zheng, J. Wang, Y. Yi and H. Shen, "Clinical characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in Nanjing, China," *Sci. China Life Sci.,* vol. 63, pp. 706-711, 2020.

This report presents the mathematical model used to model the spread of covid-19 in four regions of Sweden: Dalarna, Skåne, Stockholm and Västra Götaland.

This is an updated version of the report 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.

The Public Health Agency of Sweden is an expert authority with responsibility for public health issues at a national level. The Agency develops and supports activities to promote health, prevent illness and improve preparedness for health threats. Our vision statement: a public health that strengthens the positive development of society.

Folkhälsomyndigheten

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

www.folkhalsomyndigheten.se