Developing mathematical models to determine the COVID-19 incidence for different testing intensities

Rasmus Kristoffer Pedersen

Joint work with Christian Berrig and Viggo Andreasen
PandemiX Center, IMFUFA
Dept. Science and Environment,
Roskilde University, Denmark
Email: rakrpe@ruc.dk

IMFUFA-seminar
2^{nd} of November, 2022

Overview of this talk

- A problematic in working with COVID-19 data

Overview of this talk

- A problematic in working with COVID-19 data
- A mathematical model to investigate the problem

Overview of this talk

- A problematic in working with COVID-19 data
- A mathematical model to investigate the problem
- Model analysis

Overview of this talk

- A problematic in working with COVID-19 data
- A mathematical model to investigate the problem
- Model analysis
- Model simulations and “fitting” to data

Overview of this talk

- A problematic in working with COVID-19 data
- A mathematical model to investigate the problem
- Model analysis
- Model simulations and “fitting” to data
- Some lessons learned and discussion of results

Overview of this talk

- A problematic in working with COVID-19 data
- A mathematical model to investigate the problem
- Model analysis
- Model simulations and "fitting" to data
- Some lessons learned and discussion of results

Note: Work-in-progress

Introduction to the problem

- Different approaches to COVID-19 mitigation throughout the world.

Introduction to the problem

- Different approaches to COVID-19 mitigation throughout the world.
- One goal of retrospective research on COVID-19 is to evaluate the strategies.

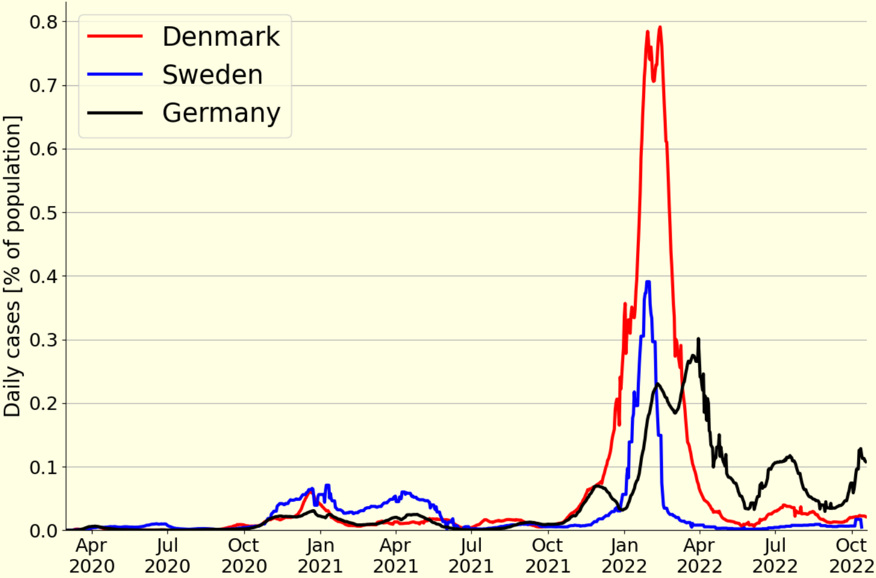
Introduction to the problem

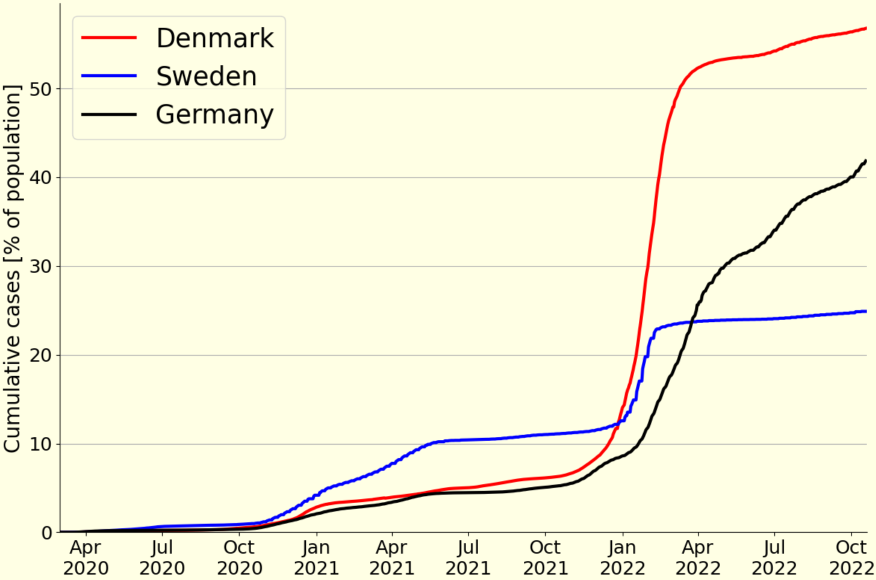
- Different approaches to COVID-19 mitigation throughout the world.
- One goal of retrospective research on COVID-19 is to evaluate the strategies.
- But how does one even begin?

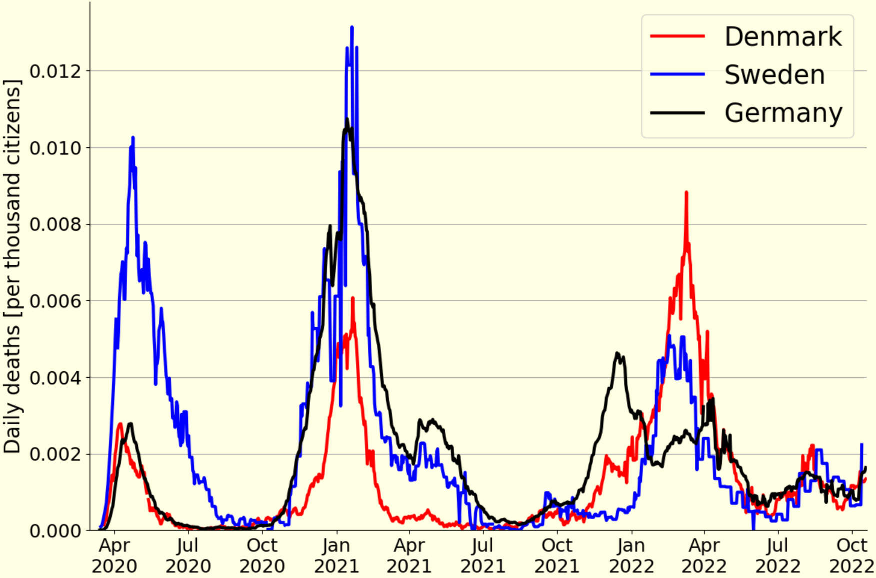
Introduction to the problem

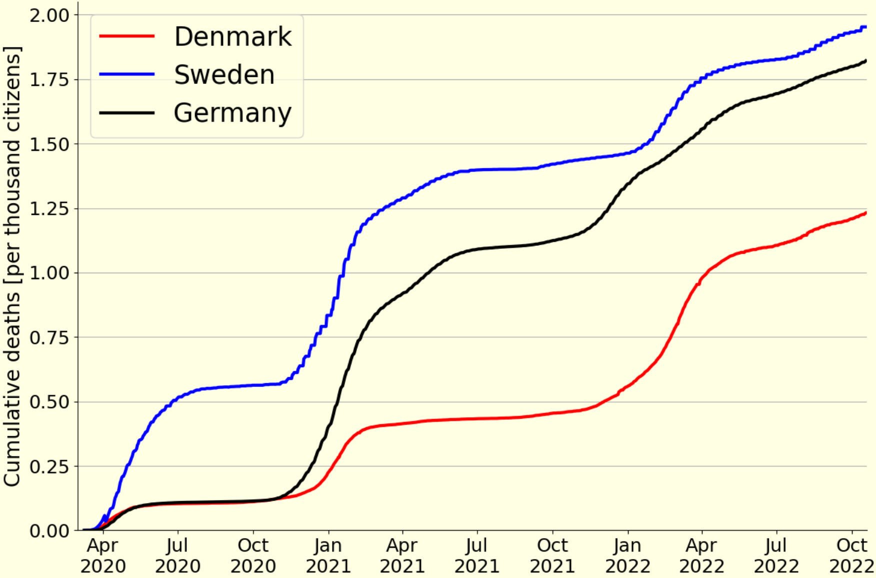
- Different approaches to COVID-19 mitigation throughout the world.
- One goal of retrospective research on COVID-19 is to evaluate the strategies.
- But how does one even begin?

Let's start by looking at some data...


Denmark
Sweden
Germany
Daily cases [% of population]
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Apr 2020
Jul 2020
Oct 2020
Jan 2021
Apr 2021
Jul 2021
Oct 2021
Jan 2022
Apr 2022
Jul 2022
Oct 2022


Denmark
Sweden
Germany
Cumulative cases [% of population]
0
10
20
30
40
50
Apr 2020
Jul 2020
Oct 2020
Jan 2021
Apr 2021
Jul 2021
Oct 2021
Jan 2022
Apr 2022
Jul 2022
Oct 2022

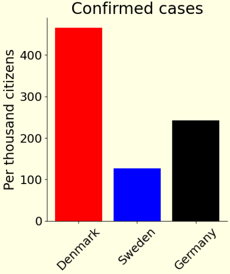

Denmark
Sweden
Germany
Daily deaths [per thousand citizens]
0.000
0.002
0.004
0.006
0.008
0.010
0.012
Apr 2020
Jul 2020
Oct 2020
Jan 2021
Apr 2021
Jul 2021
Oct 2021
Jan 2022
Apr 2022
Jul 2022
Oct 2022



Denmark
Sweden
Germany
Cumulative deaths [per thousand citizens]
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
Apr 2020
Jul 2020
Oct 2020
Jan 2021
Apr 2021
Jul 2021
Oct 2021
Jan 2022
Apr 2022
Jul 2022
Oct 2022

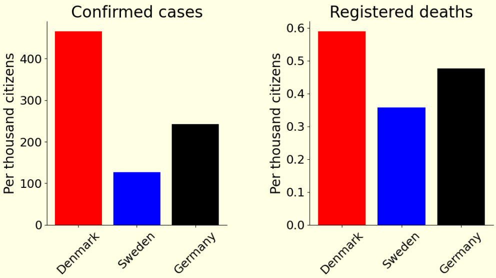
Let's make it a little simpler

Between October 1st, 2021 and May 1st, 2022:



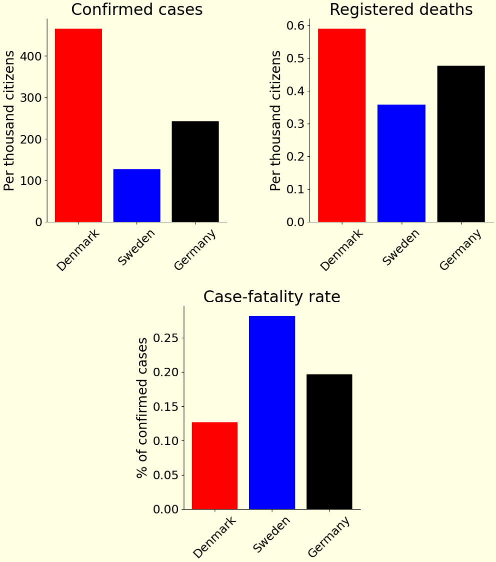
Let's make it a little simpler

Between October 1st, 2021 and May 1st, 2022:

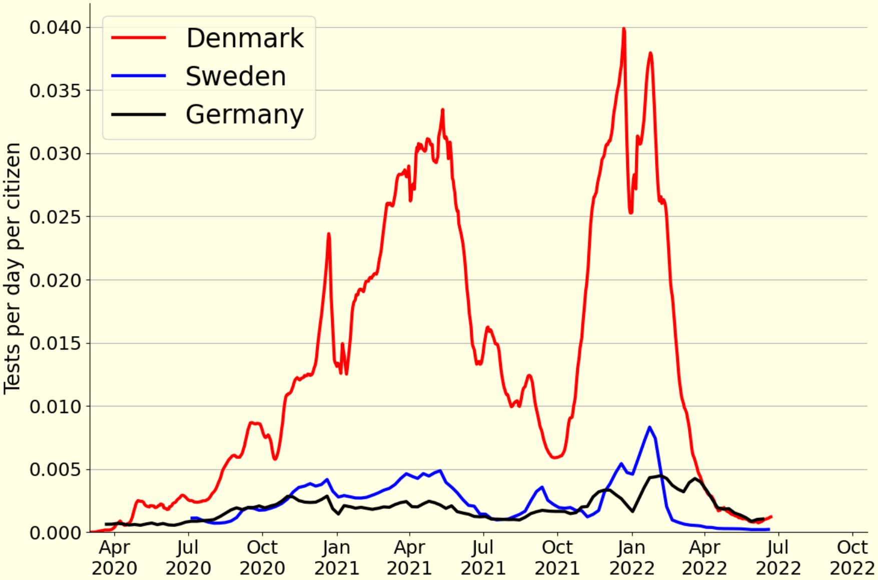


Let's make it a little simpler

Between October 1st, 2021 and May 1st, 2022:



Something must be different...


Denmark
Sweden
Germany
Tests per day per citizen
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
Apr 2020
Jul 2020
Oct 2020
Jan 2021
Apr 2021
Jul 2021
Oct 2021
Jan 2022
Apr 2022
Jul 2022
Oct 2022

The core of the problem

- Different approaches to COVID-19 testing.

The core of the problem

- Different approaches to COVID-19 testing.
- The role of testing: Confirmation of symptoms, required for various activities or entirely voluntary?

The core of the problem

- Different approaches to COVID-19 testing.
- The role of testing: Confirmation of symptoms, required for various activities or entirely voluntary?
- The choice of testing-strategy also affects both the epidemic and the data-collection.

The core of the problem

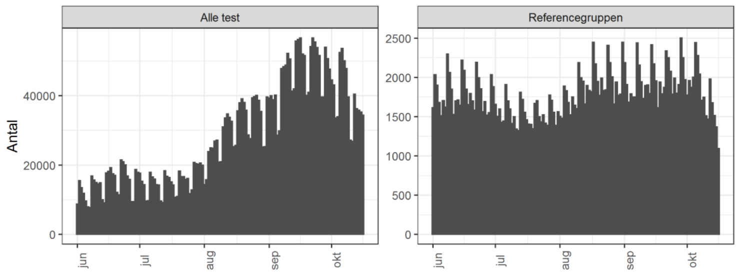
- Different approaches to COVID-19 testing.
- The role of testing: Confirmation of symptoms, required for various activities or entirely voluntary?
- The choice of testing-strategy also affects both the epidemic and the data-collection.
- The main problem: For each reported case of COVID-19, how many unidentified cases?

The core of the problem

- Different approaches to COVID-19 testing.
- The role of testing: Confirmation of symptoms, required for various activities or entirely voluntary?
- The choice of testing-strategy also affects both the epidemic and the data-collection.
- The main problem: For each reported case of COVID-19, how many unidentified cases?
- **How do we compare case-counts between periods and places where testing activity was different?**

Approach of Danish Health Agencies

Testing activity for non-COVID related hospitalizations in Denmark remained stable throughout 2020, despite large variation in testing activity in the population.



Approach of Danish Health Agencies

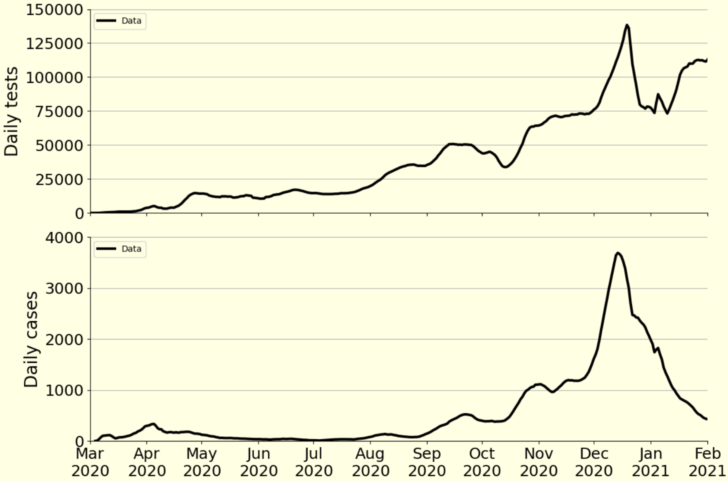
A method for correcting observed incidence was developed by Danish health agencies:

$$I_{corrected} = I_{observed} \left(\frac{50,000}{T} \right)^{0.7}$$

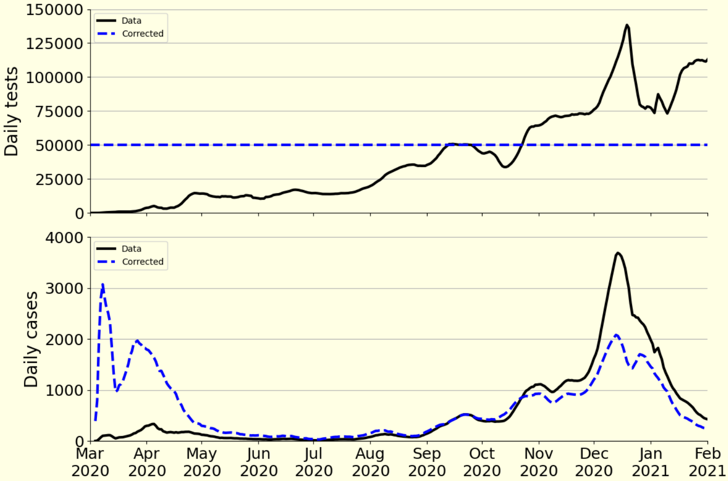
where T is the number of tests on the given day, $I_{observed}$ the observed cases, and $I_{corrected}$ is a measure for the expected number of positive tests if $50,000$ tests had been made. The 0.7 factor was estimated from data, and later changed to 0.3.

(SSI, 2020; *Ekspertrapport af d. 23. oktober 2020 - Incidens og fremskrivning af COVID-19 tilfælde*)

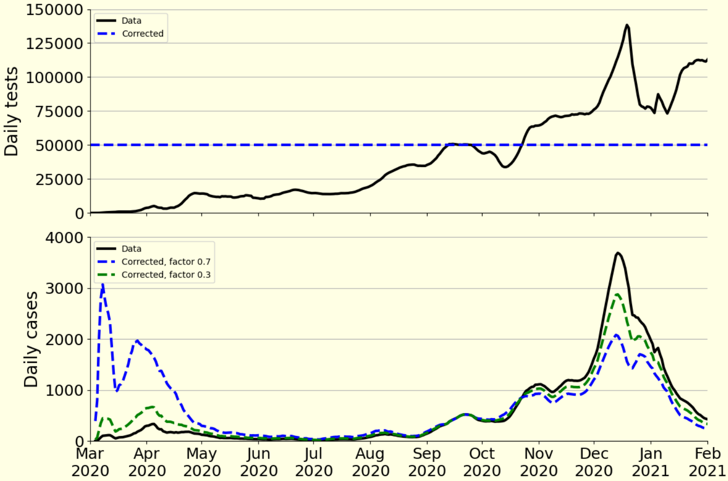
Approach of Danish Health Agencies



Approach of Danish Health Agencies



Approach of Danish Health Agencies



Approach of Danish Health Agencies

Allows us to compare different periods...

Approach of Danish Health Agencies

Allows us to compare different periods,
but does not estimate the *actual* number of infections.

Our approach

- We aim to determine the ratio between observed cases and the total number of COVID-19 cases.

Our approach

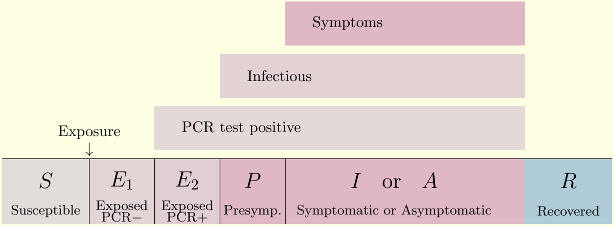
- We aim to determine the ratio between observed cases and the total number of COVID-19 cases.
- This ratio can be used as a correction-factor for observed data.

Our approach

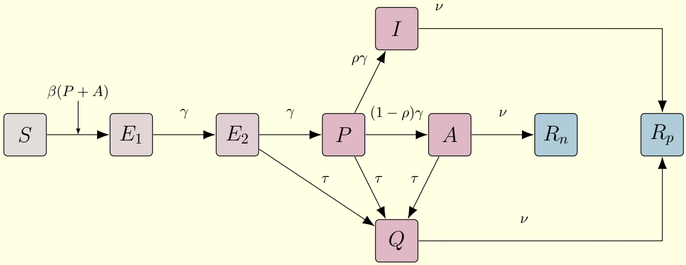
- We aim to determine the ratio between observed cases and the total number of COVID-19 cases.
- This ratio can be used as a correction-factor for observed data.
- We extend the classic SIR-model to include voluntary testing that identifies pre- and asymptomatic cases.

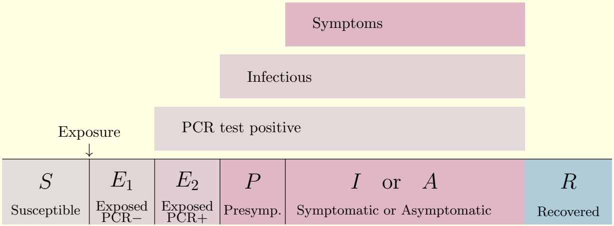
Model presentation

The model

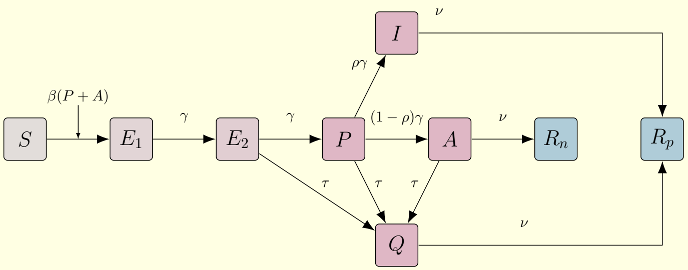

Symptoms
Infectious
PCR test positive
Exposure
S
Susceptible
E1
Exposed PCR−
E2
Exposed PCR+
P
Presymp.
I or A
Symptomatic or Asymptomatic
R
Recovered

The model


I
ν
$\rho\gamma$
$\beta(P+A)$
S
E_1
γ
E_2
γ
P
$(1-\rho)\gamma$
A
ν
R_n
R_p
τ
τ
τ
ν
Q


Symptoms
Infectious
PCR test positive
Exposure
S Susceptible
E_1 Exposed PCR−
E_2 Exposed PCR+
P Presymp.
I or A Symptomatic or Asymptomatic
R Recovered

The model



$$\dot{S} = -\beta S(P+A)$$
$$\dot{E}_1 = \beta S(P+A) - \gamma E_1$$
$$\dot{E}_2 = \gamma E_1 - \gamma E_2 - \tau E_2$$
$$\dot{P} = \gamma E_2 - \gamma P - \tau P$$
$$\dot{I} = \gamma \rho P - \nu I$$
$$\dot{A} = \gamma(1-\rho)P - \nu A - \tau A$$
$$\dot{Q} = \tau(E_2 + P + A) - \nu Q$$
$$\dot{R}_p = \nu Q + \nu I$$
$$\dot{R}_n = \nu A$$

The model

$$\dot{S} = -\beta S(P + A) \qquad \dot{A} = \gamma(1-\rho)P - \nu A - \tau A$$

$$\dot{E}_1 = \beta S(P + A) - \gamma E_1 \qquad \dot{Q} = \tau(E_2 + P + A) - \nu Q$$

$$\dot{E}_2 = \gamma E_1 - \gamma E_2 - \tau E_2 \qquad \dot{R}_p = \nu Q + \nu I$$

$$\dot{P} = \gamma E_2 - \gamma P - \tau P \qquad \dot{R}_n = \nu A$$

$$\dot{I} = \gamma \rho P - \nu I$$

Symbol	*Description*	*Default value*
β	Infectivity	$2/3$
ν	Rate of recovery	$1/3$
γ	Rate of disease progression	$1/3$
ρ	Fraction of symptomatic cases	$1/2$
τ	Rate of testing	0 to 0.5

All rates units of day^{-1}. Approximate R_0 of 1.4 initially.

The model

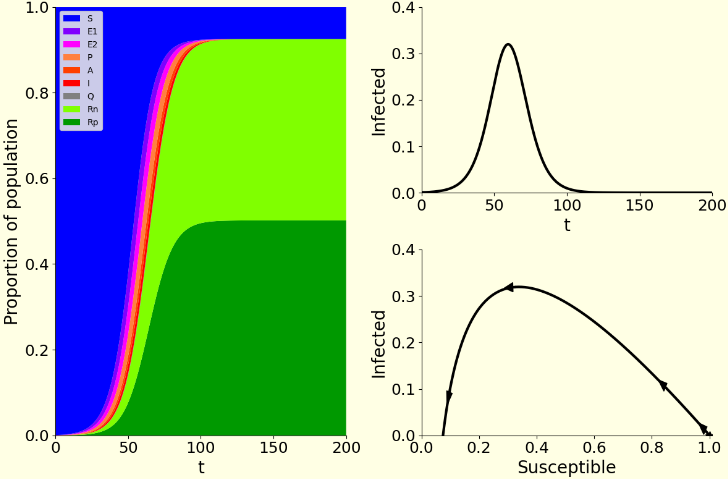
$$\dot{S} = -\beta S(P + A)$$
$$\dot{E}_1 = \beta S(P + A) - \gamma E_1$$
$$\dot{E}_2 = \gamma E_1 - \gamma E_2 - \tau E_2$$
$$\dot{P} = \gamma E_2 - \gamma P - \tau P$$
$$\dot{I} = \gamma \rho P - \nu I$$
$$\dot{A} = \gamma(1 - \rho)P - \nu A - \tau A$$
$$\dot{Q} = \tau(E_2 + P + A) - \nu Q$$
$$\dot{R}_p = \nu Q + \nu I$$
$$\dot{R}_n = \nu A$$

Symbol	*Description*	*Default value*
β	Infectivity	2/3
ν	Rate of recovery	1/3
γ	Rate of disease progression	1/3
ρ	Fraction of symptomatic cases	**1/2**
τ	Rate of testing	0 to 0.5

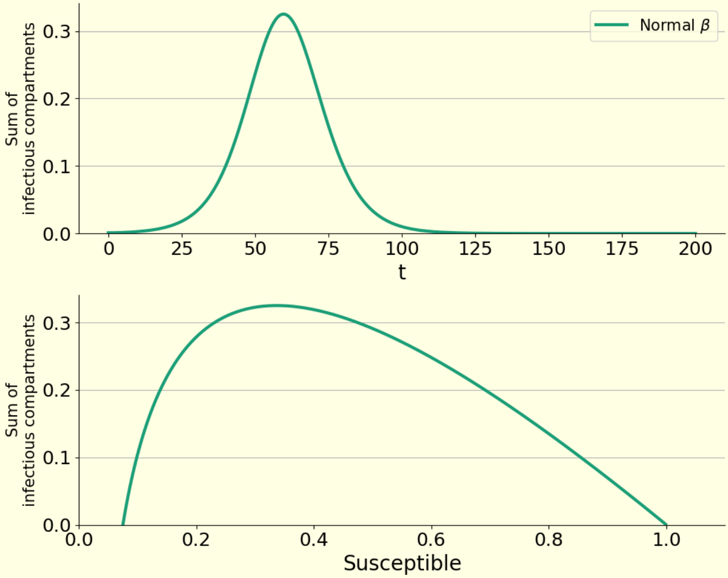
All rates units of day^{-1}. Approximate R_0 of 1.4 initially.

General model dynamics

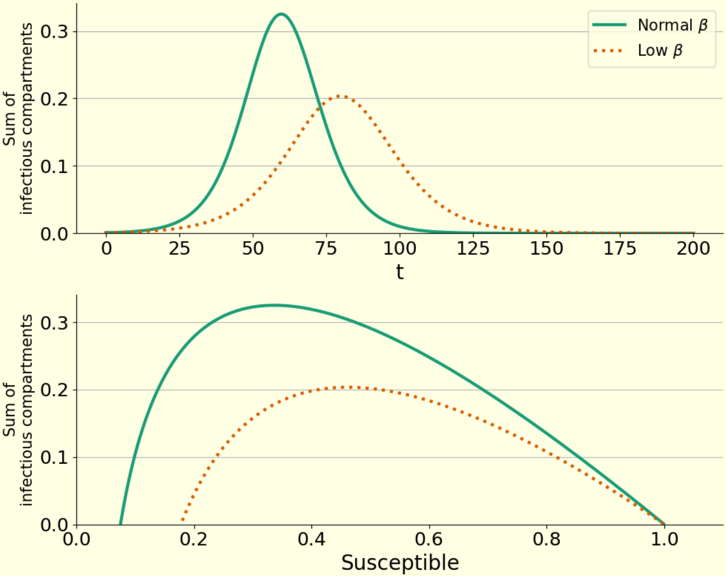
General dynamics like classic SIR-model.



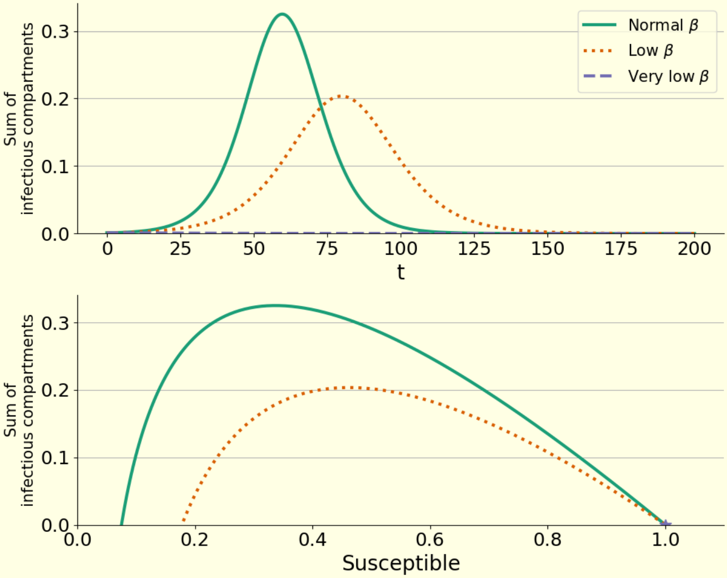
General model dynamics



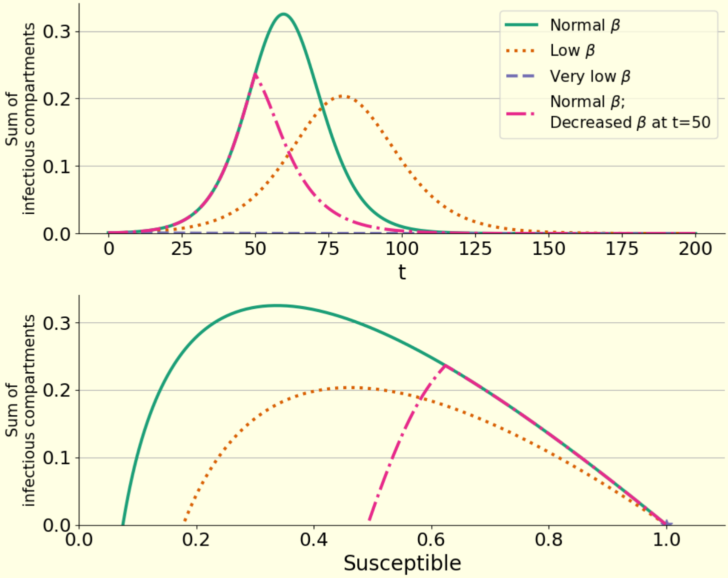
General model dynamics



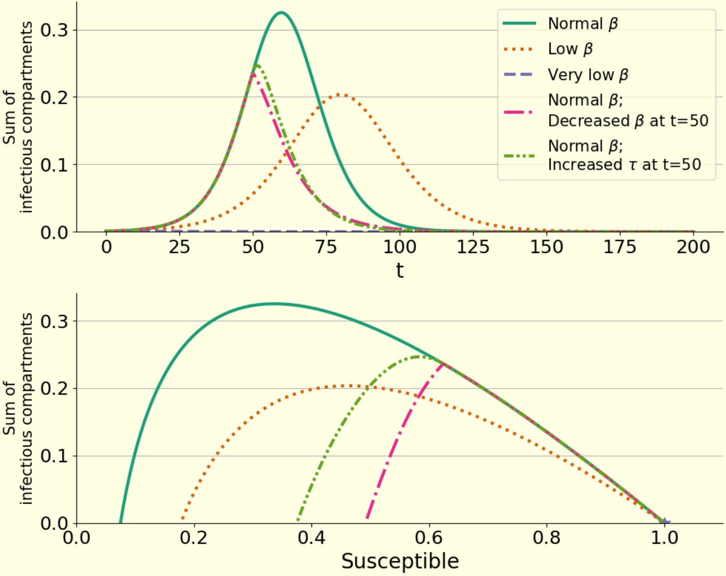
General model dynamics



General model dynamics



General model dynamics

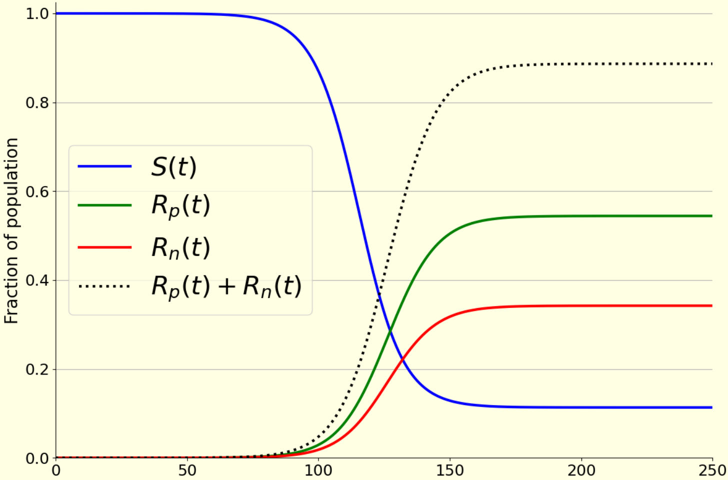


Final size calculation

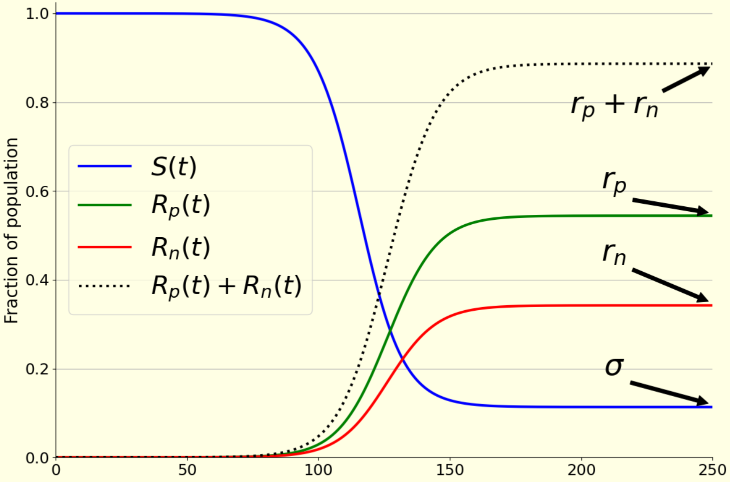
For parameters where an epidemic occurs, we define σ, r_p and r_n such that:

$$S(t) \underset{t\to\infty}{\longrightarrow} \sigma$$
$$R_p(t) \underset{t\to\infty}{\longrightarrow} r_p$$
$$R_n(t) \underset{t\to\infty}{\longrightarrow} r_n$$

Final size calculation



Final size calculation



Final size calculation

Similar to previous work (Andreasen, 2018), we define

$$T_X = \int_{t=0}^{t=\infty} X(t)dt$$

Final size calculation

Similar to previous work (Andreasen, 2018), we define

$$T_X = \int_{t=0}^{t=\infty} X(t)dt$$

From the model definition...

$$\dot{R}_p(t) = \nu Q(t) + \nu I(t)$$

Final size calculation

Similar to previous work (Andreasen, 2018), we define

$$T_X = \int_{t=0}^{t=\infty} X(t)dt$$

From the model definition, we integrate

$$\int_{t=0}^{t=\infty} \dot{R}_p(t)dt = \int_{t=0}^{t=\infty} (\nu Q(t) + \nu I(t))dt$$

Final size calculation

Similar to previous work (Andreasen, 2018), we define

$$T_X = \int_{t=0}^{t=\infty} X(t)dt$$

From the model definition, we integrate

$$r_p - R_{p,0} = \nu T_Q + \nu T_I$$

Final size calculation

Similar to previous work (Andreasen, 2018), we define

$$T_X = \int_{t=0}^{t=\infty} X(t)dt$$

From the model definition, we integrate

$$r_p - R_{p,0} = \nu T_Q + \nu T_I$$

and similarly for $\dot{R}_n$

$$r_n - R_{n,0} = \nu T_A$$

Final size calculation

Similar to previous work (Andreasen, 2018), we define

$$T_X = \int_{t=0}^{t=\infty} X(t)dt$$

From the model definition, we integrate

$$r_p - R_{p,0} = \nu T_Q + \nu T_I$$

and similarly for $\dot{R}_n$

$$r_n - R_{n,0} = \nu T_A$$

Or, for an initially susceptible population:

$$r_p = \nu T_Q + \nu T_I$$
$$r_n = \nu T_A$$

Final size calculation

We consider the following quantities:

$$\dot{S}/S = -\beta(P + A) \tag{1a}$$

$$\dot{S} + \dot{E}_1 + \dot{E}_2 = -(\gamma + \tau)E_2 \tag{1b}$$

$$\dot{S} + \dot{E}_1 + \dot{E}_2 + \dot{P} = -(\gamma + \tau)P - \tau E_2 \tag{1c}$$

$$\dot{S} + \dot{E}_1 + \dot{E}_2 + \dot{P} + \dot{A} = -(\nu + \tau)A - (\gamma\rho + \tau)P - \tau E_2 \tag{1d}$$

Final size calculation

In the limit where the entire population is initially susceptible, $S(0) = 1$, we can rewrite:

$$\beta = \frac{-\log\sigma}{T_P + T_A} \tag{1a}$$

$$T_{E_2} = \frac{1}{\gamma + \tau}(1 - \sigma) \tag{1b}$$

$$T_P = \frac{1}{\gamma + \tau}(1 - \sigma - \tau T_{E_2}) \tag{1c}$$

$$T_A = \frac{1}{\nu + \tau}(1 - \sigma - (\gamma\rho + \tau) T_P - \tau T_{E_2}) \tag{1d}$$

Final size calculation

In the limit where $S(0) = 1$, $R_p(0) = 0$, $R_n(0) = 0$, we can calculate the ratio of observed cases as

$$
\begin{aligned}
K_F &= \frac{r_p}{r_p + r_n} \\
&= 1 - \frac{r_n}{1-\sigma} \\
&= 1 - \frac{\nu}{1-\sigma} T_A \\
&= 1 - \frac{\nu\gamma^2(1-\rho)}{(\nu+\tau)(\gamma+\tau)^2}
\end{aligned}
$$

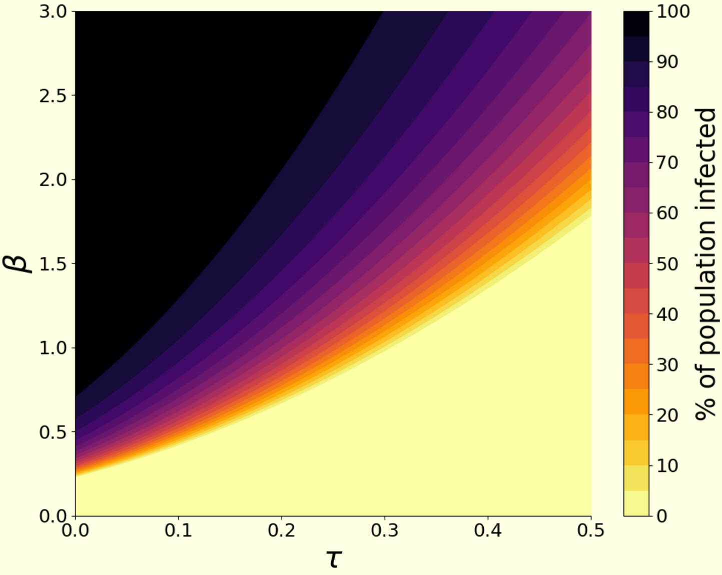
Note that K_F is independent of σ as well as β.

Fixed final size and (τ, β)

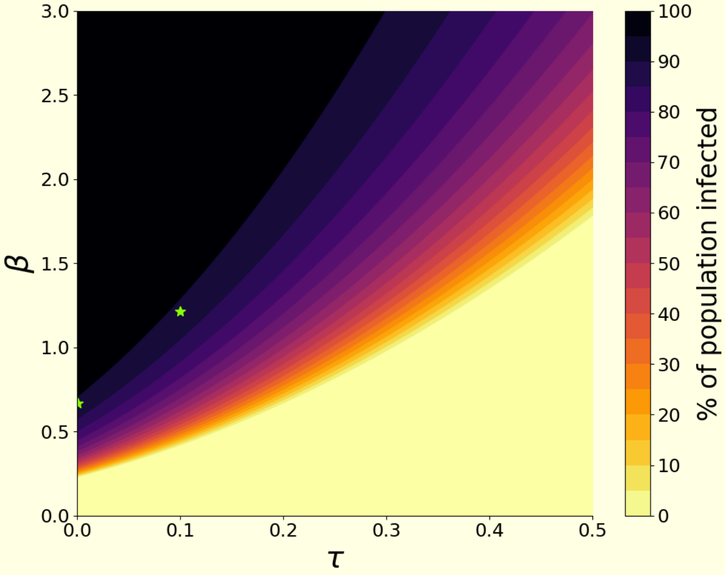
Although the fraction of cases identified, K, is independent of β, the epidemic final size, i.e. $r_n + r_p$, is not.

Let us take a look at the final size as a function of τ and β.

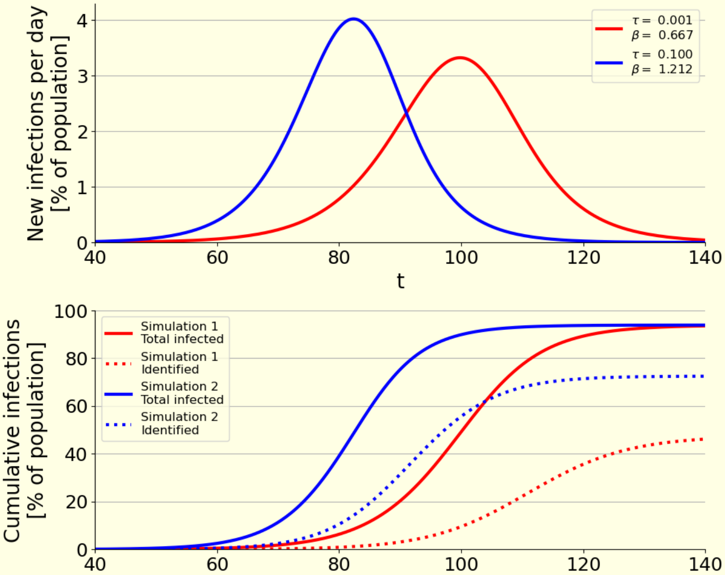
Fixed final size and (τ, β)



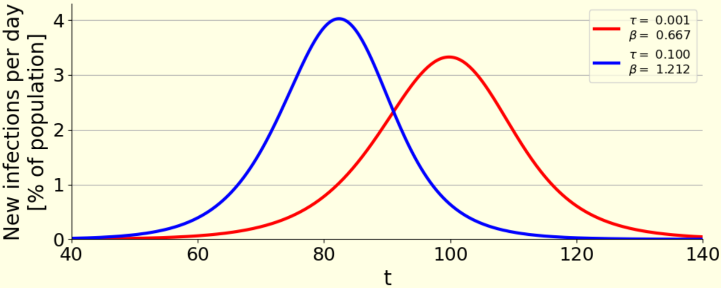
Fixed final size and (τ, β)

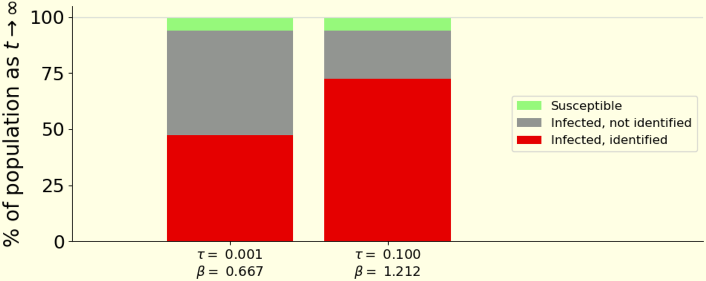


Fixed final size and (τ, β)

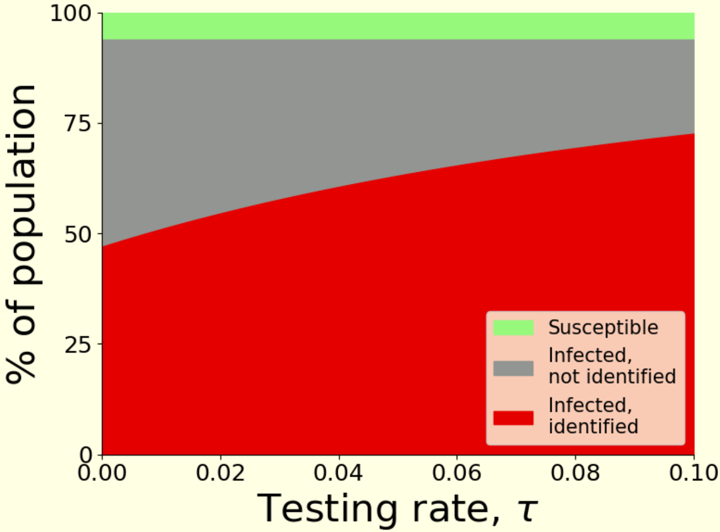


Fixed final size and (τ, β)



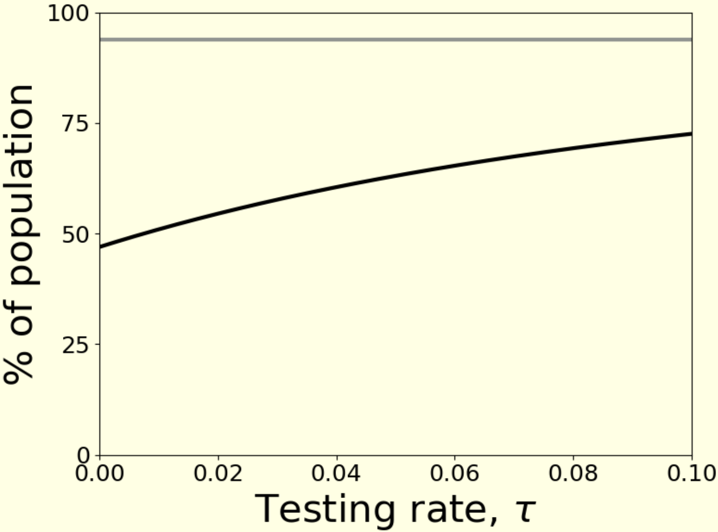


Fixed final size and (τ, β)



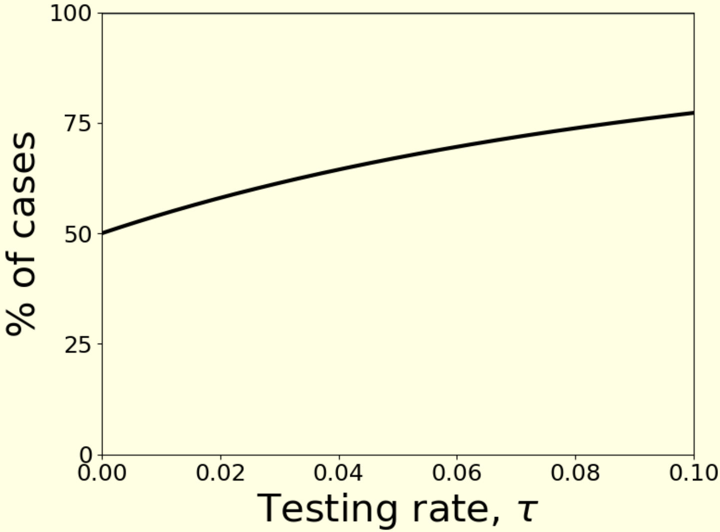
(β chosen such that final size is fixed)

Fixed final size and (τ, β)

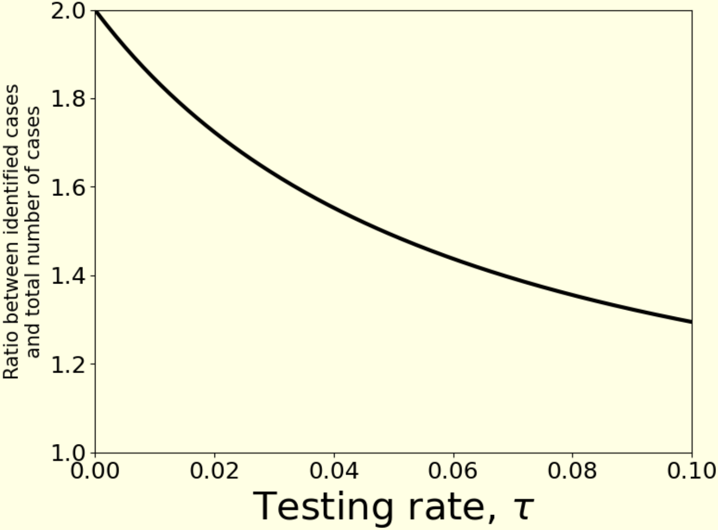


(β chosen such that final size is fixed)

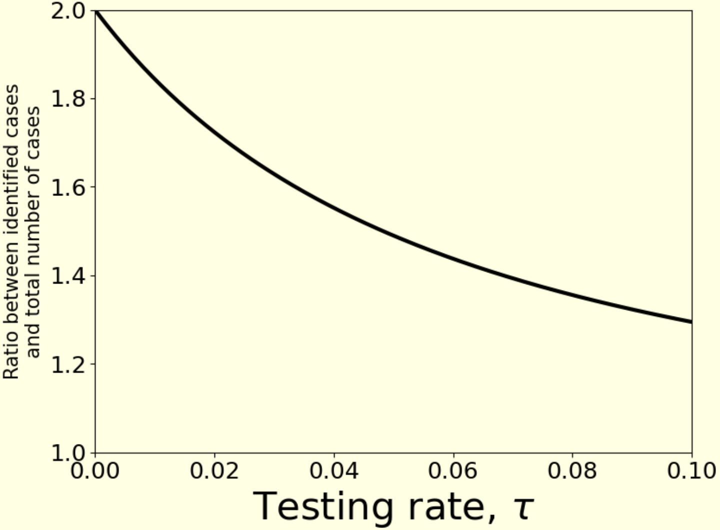
Fixed final size and (τ, β)



Fixed final size and (τ, β)



Fixed final size and (τ, β)



(This is actually just $1/K_F$)

Data and simulations

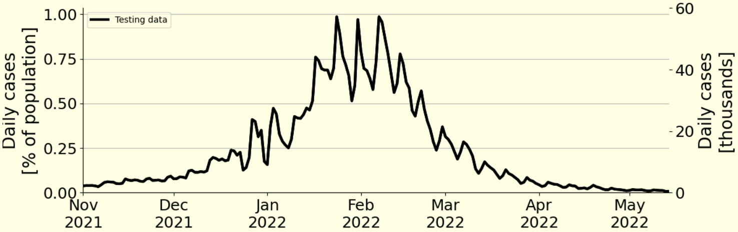
The Danish data

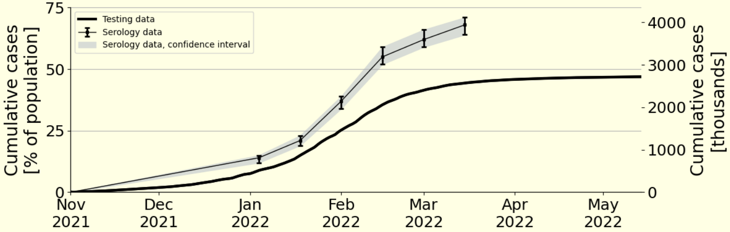
Let us take a look at a particular wave:

The Omicron-wave in Denmark,
Between December 2021 and March 2022.

The Danish data

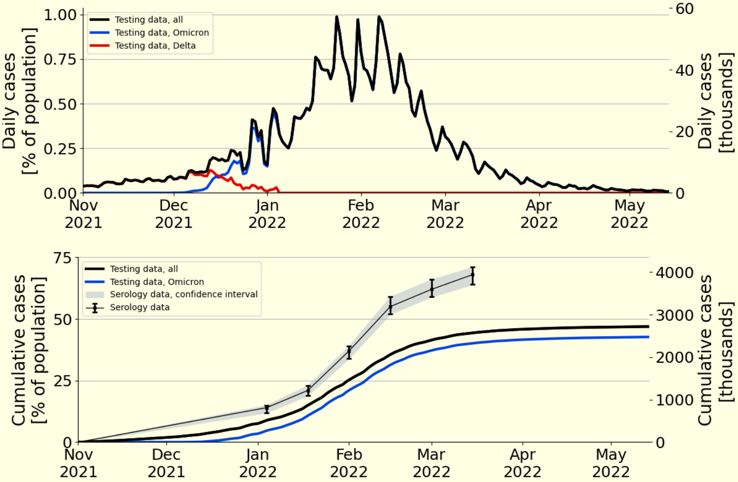
Daily cases and serology from blood-donors





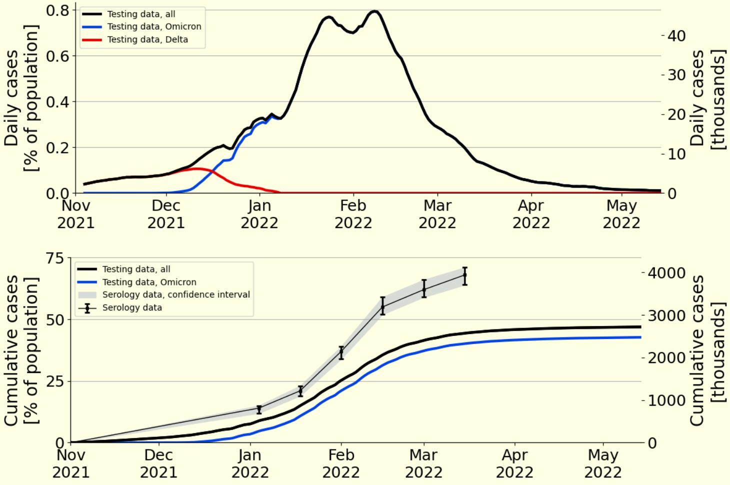
The Danish data

December 2021: Variant-sampling of 80 to 90% of cases.



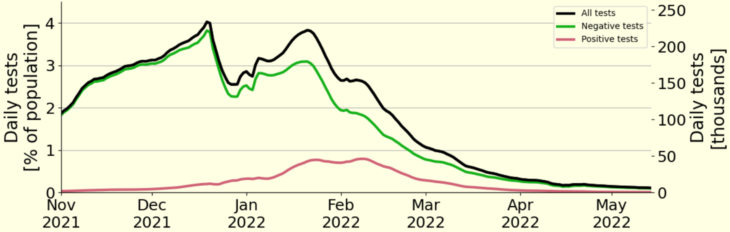
The Danish data

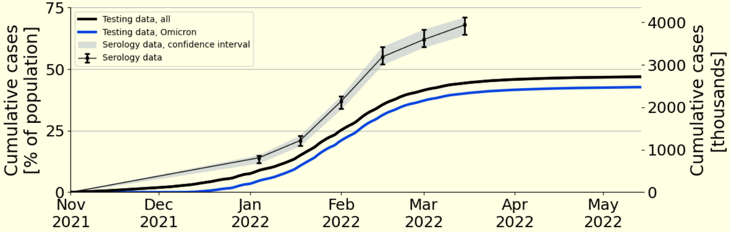
Smoothing: 7-day running mean.



The Danish data

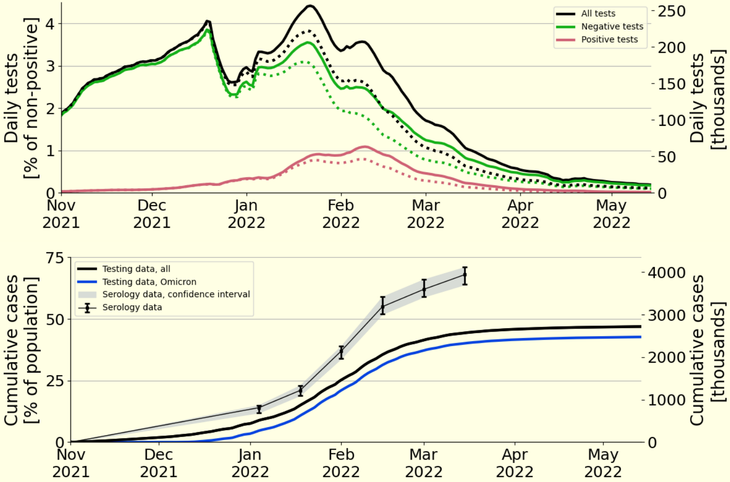
Testing data, by population.



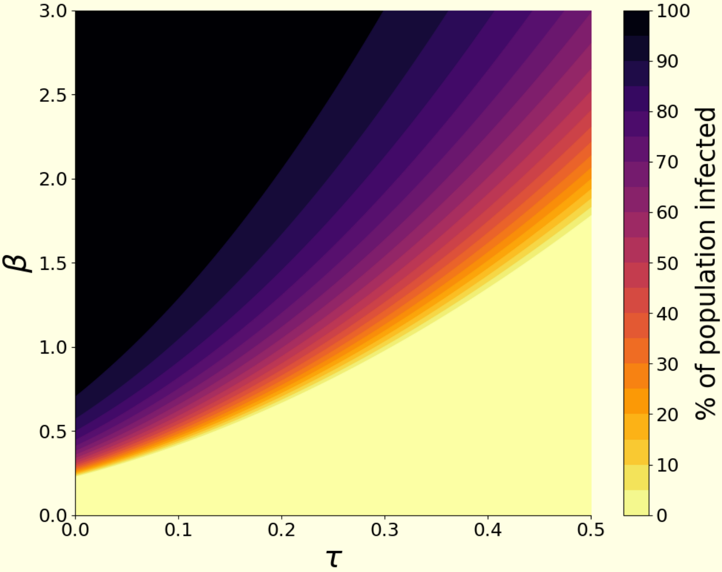


The Danish data

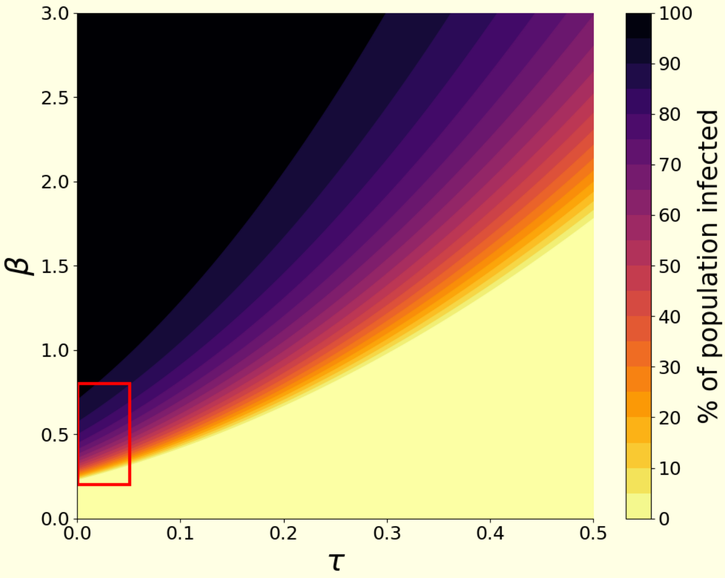
Testing data, by population that haven't tested positve.



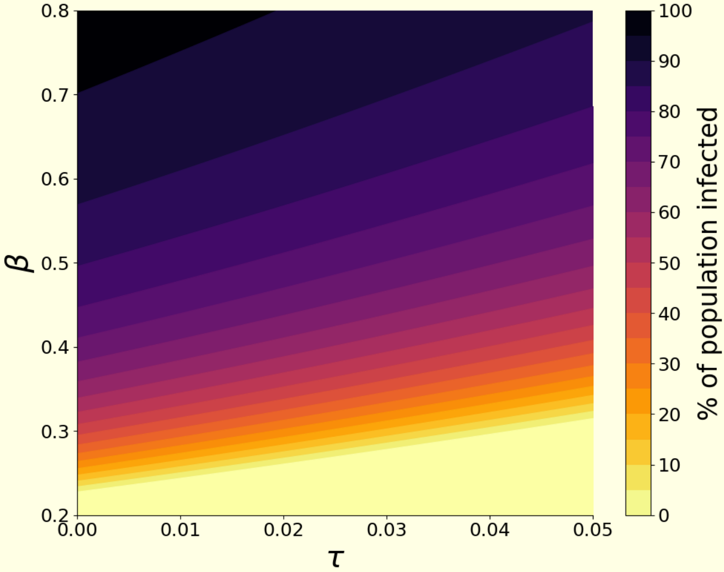
Model simulation



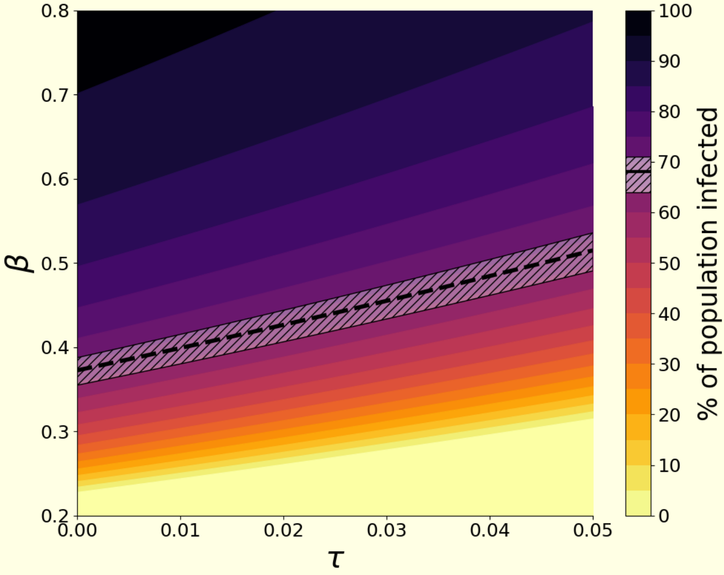
Model simulation



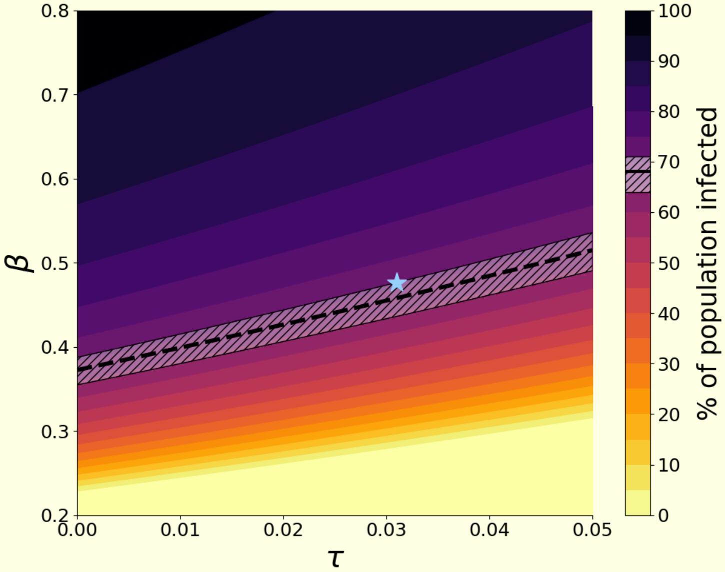
Model simulation



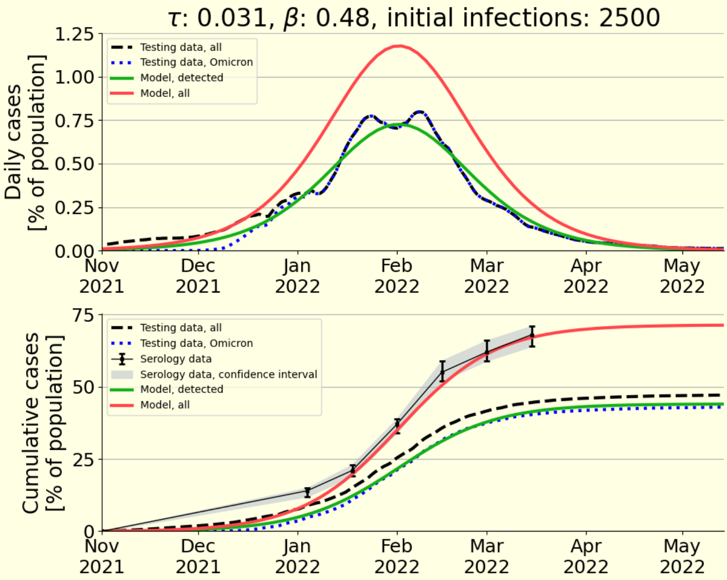
Model simulation



Model simulation

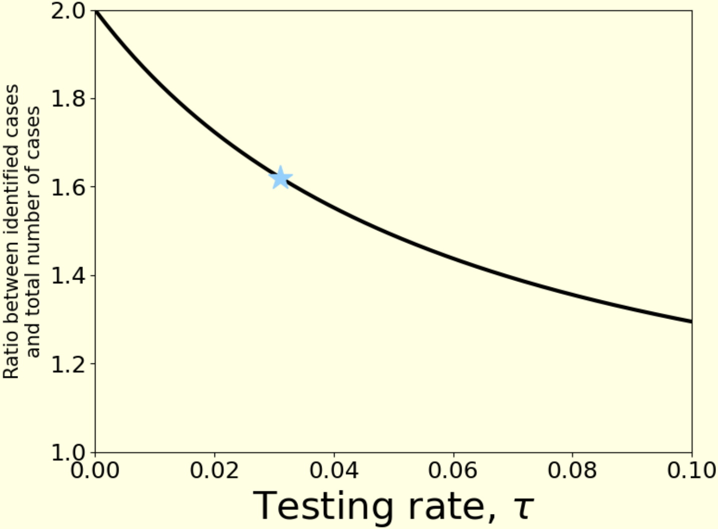


Model simulation

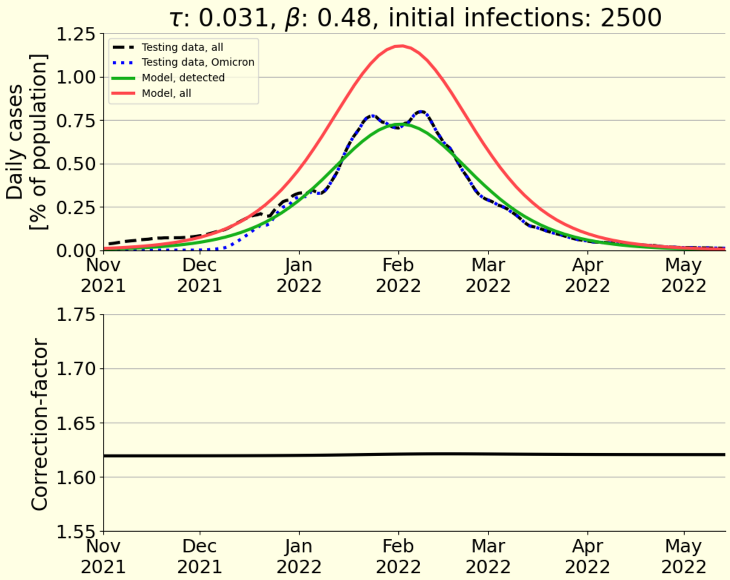


Model simulation

We chose $\tau = 0.031$

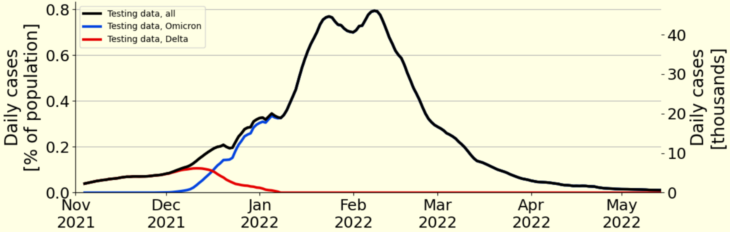


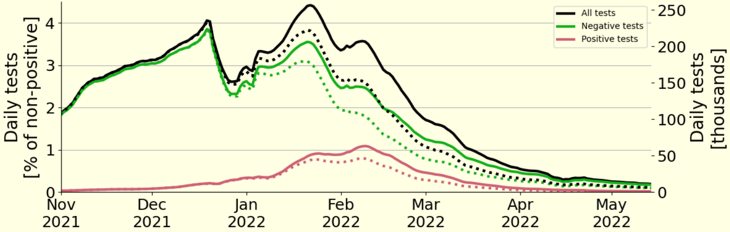
Model simulation



Simulating with data

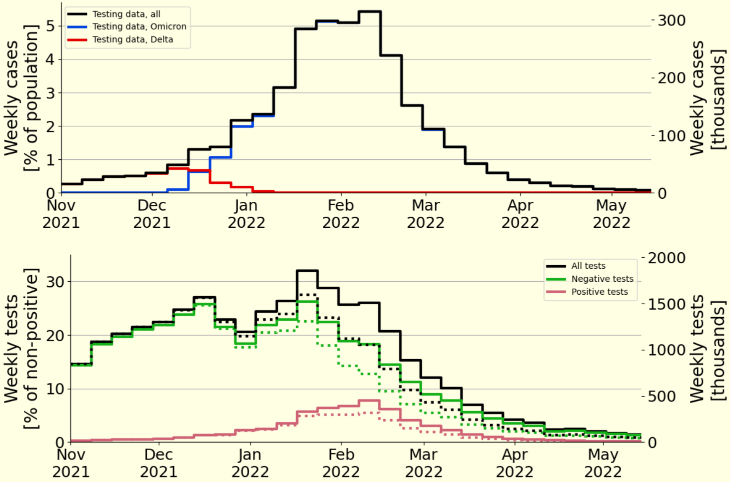
For practical reasons...



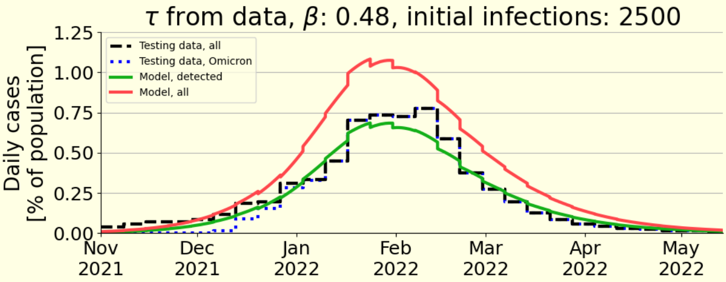


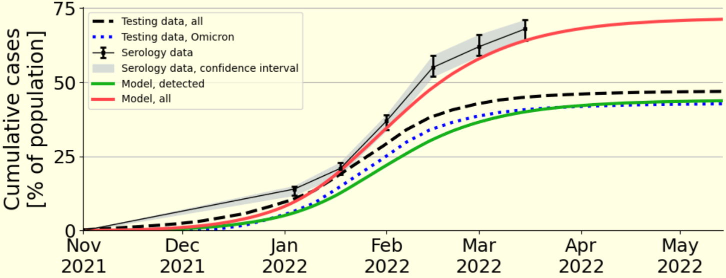
Simulating with data

For practical reasons, weekly sums.

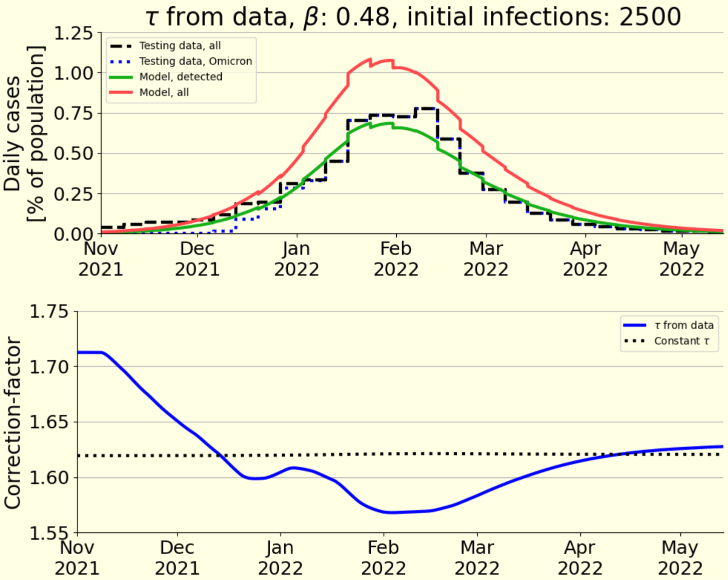


Simulating with data

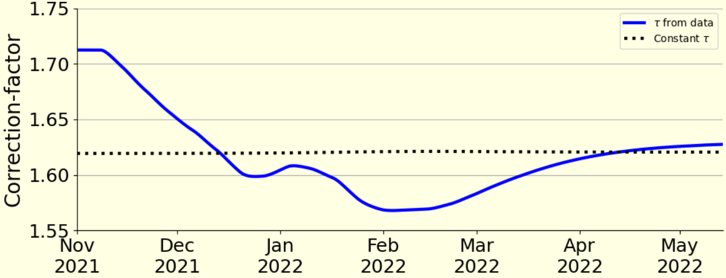




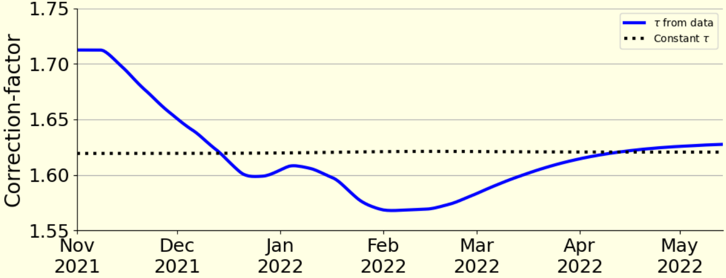
Simulating with data



But wait...

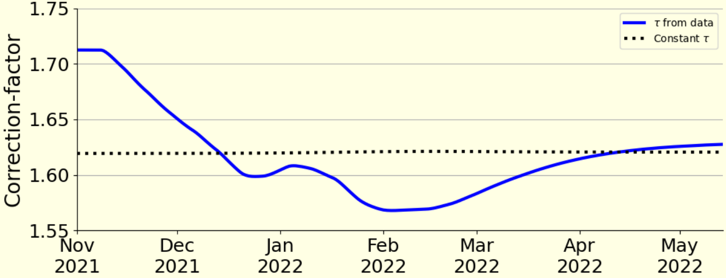

1.75
1.70
1.65
1.60
1.55
Correction-factor
τ from data
Constant τ
Nov 2021
Dec 2021
Jan 2022
Feb 2022
Mar 2022
Apr 2022
May 2022

But wait...



What is "Correction-factor" actually here?

But wait...

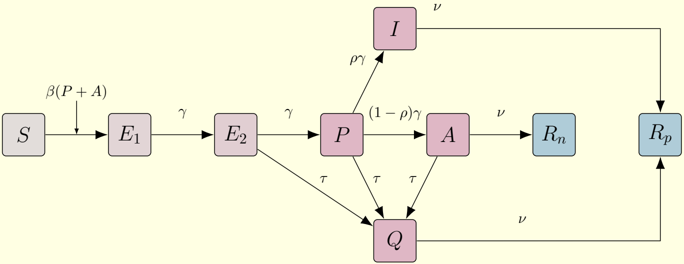


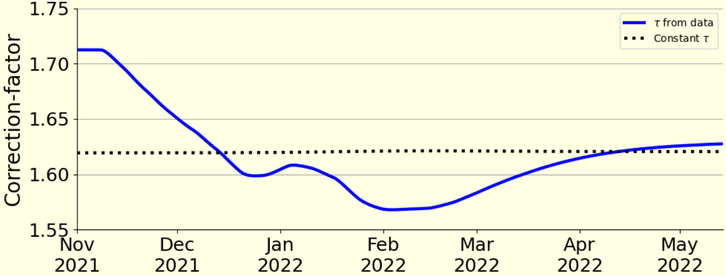
What is "Correction-factor" actually here?

Serology: 68% infected.
Testing: 42% infected.

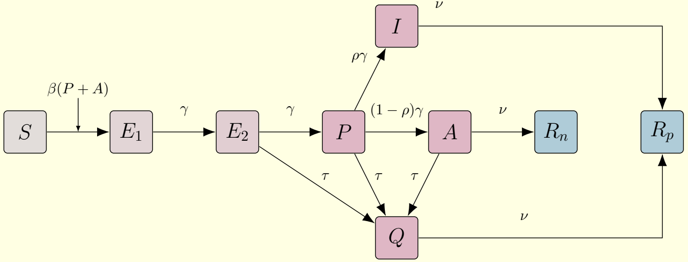
Ratio: $\dfrac{0.68}{0.42} = 1.62$

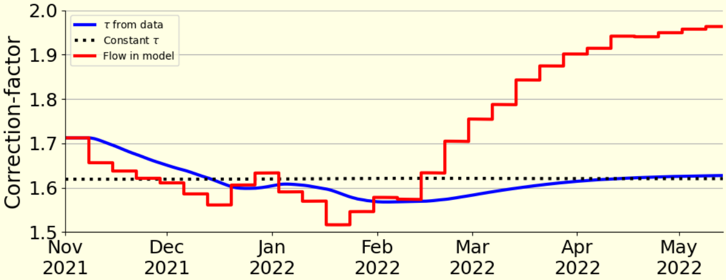
Getting the correction-factor right



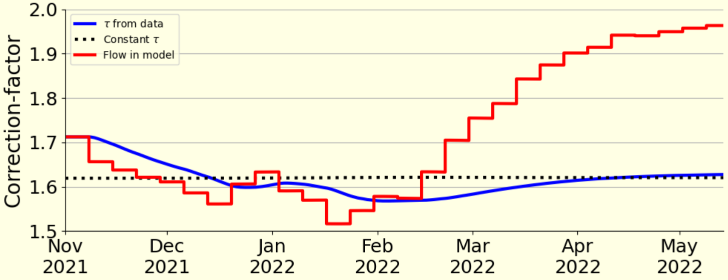


Getting the correction-factor right



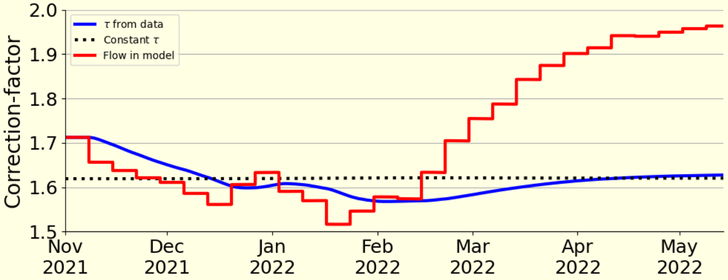


Getting the correction-factor right



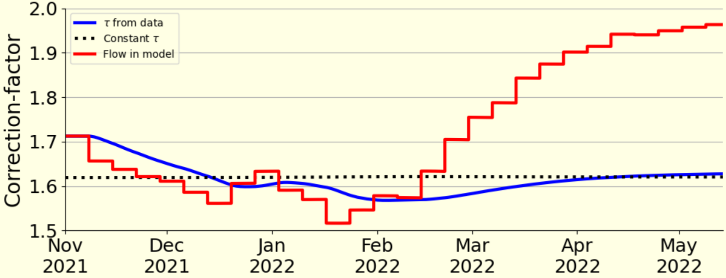
$$K_F = 1 - \frac{\nu\gamma^2(1-\rho)}{(\nu+\tau)(\gamma+\tau)^2}$$

Getting the correction-factor right



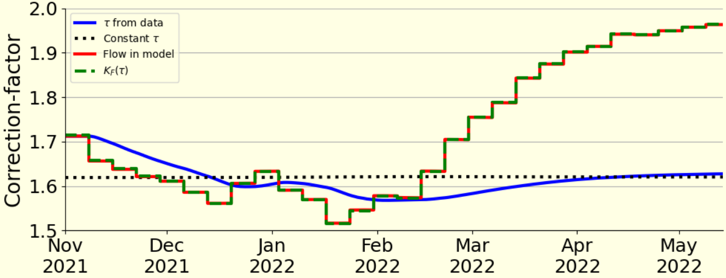
$$K_F(\tau) = 1 - \frac{\frac{1}{3}\cdot\frac{1}{3}\cdot\frac{1}{3}\left(1-\frac{1}{2}\right)}{(\frac{1}{3}+\tau)(\frac{1}{3}+\tau)(\frac{1}{3}+\tau)}$$

Getting the correction-factor right



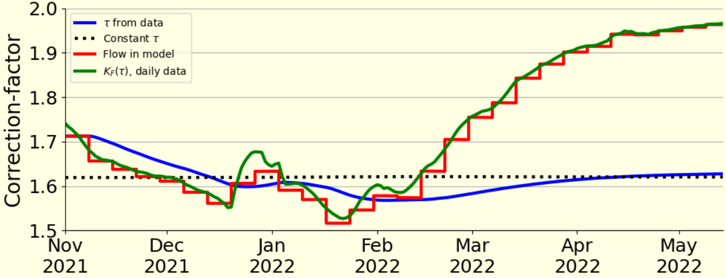
$$K_F(\tau) = 1 - \frac{1}{2\left(1+3\tau\right)^3}$$

Getting the correction-factor right



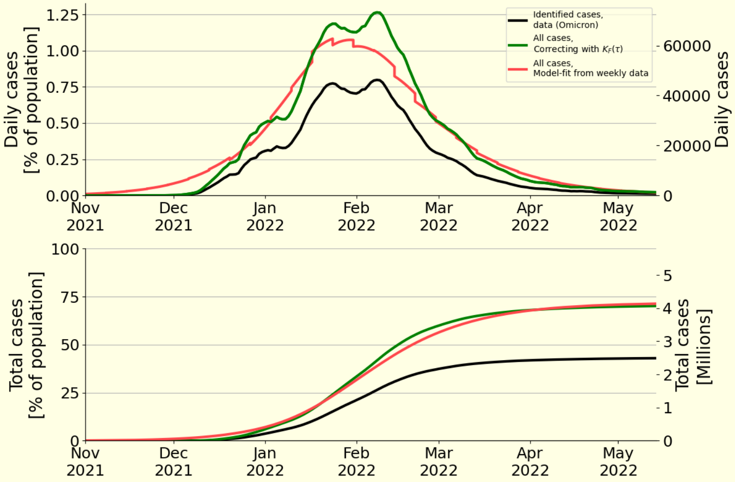
$$K_F(\tau) = 1 - \frac{1}{2\left(1 + 3\tau\right)^3}$$

Getting the correction-factor right

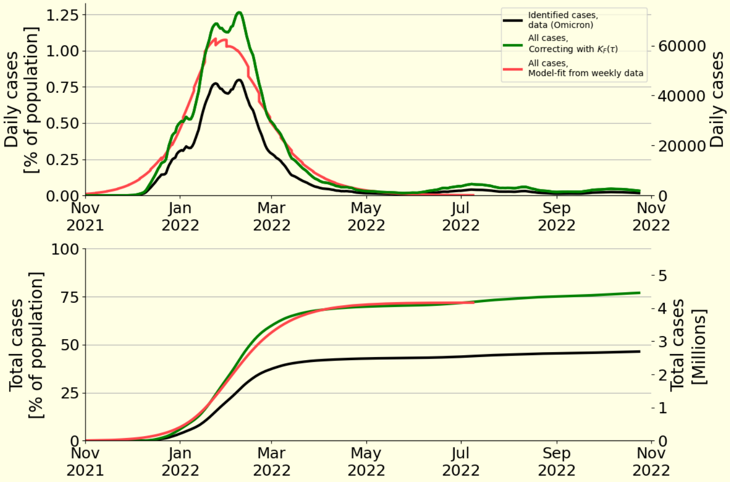


$$K_F(\tau) = 1 - \frac{1}{2\left(1 + 3\tau\right)^3}$$

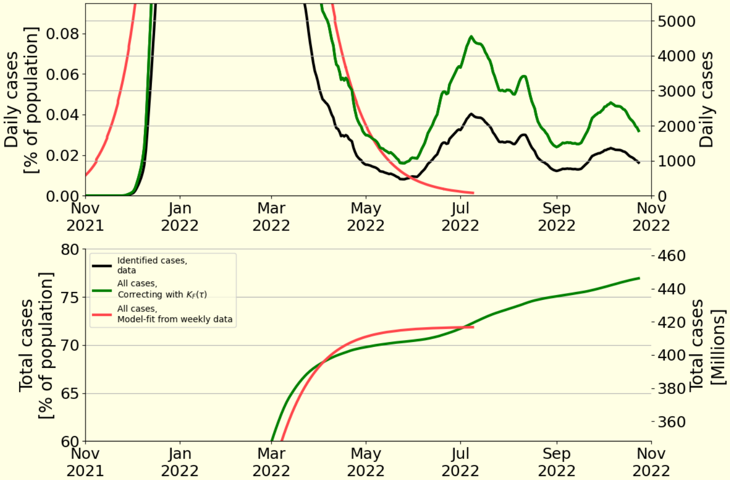
Correcting the Danish data



Correcting the Danish data



Correcting the Danish data



Summing up

- Using an extended SIR-model, we are able to estimate the fraction of COVID-19 cases identified in the Omicron wave of early 2022 in Denmark.

Summing up

- Using an extended SIR-model, we are able to estimate the fraction of COVID-19 cases identified in the Omicron wave of early 2022 in Denmark.
- Analytical results suggest a simple relation between testing intensity and fraction of cases identified.

Summing up

- Using an extended SIR-model, we are able to estimate the fraction of COVID-19 cases identified in the Omicron wave of early 2022 in Denmark.
- Analytical results suggest a simple relation between testing intensity and fraction of cases identified.
- Simulations with hand-fitted parameters agree well with data.

Summing up

- Using an extended SIR-model, we are able to estimate the fraction of COVID-19 cases identified in the Omicron wave of early 2022 in Denmark.
- Analytical results suggest a simple relation between testing intensity and fraction of cases identified.
- Simulations with hand-fitted parameters agree well with data.
- Results can easily be expanded to other COVID-19 waves and other countries, to allow for comparison.

Summing up

- Using an extended SIR-model, we are able to estimate the fraction of COVID-19 cases identified in the Omicron wave of early 2022 in Denmark.
- Analytical results suggest a simple relation between testing intensity and fraction of cases identified.
- Simulations with hand-fitted parameters agree well with data.
- Results can easily be expanded to other COVID-19 waves and other countries, to allow for comparison.
- Provides an estimate of the actual numbers of cases, in contrast to other methods used for test-correction.

Limitations and final comments

- **For low testing, parameter ρ decides everything.**
  For $\tau \to 0$, we simply have $K_F = \frac{1}{\rho}$.

Limitations and final comments

- **For low testing, parameter ρ decides everything.**
  For $\tau \to 0$, we simply have $K_F = \frac{1}{\rho}$.
- **Assumes voluntary testing is available.**
  Not the case in early 2020.

Limitations and final comments

- **For low testing, parameter ρ decides everything.**
  For $\tau \to 0$, we simply have $K_F = \frac{1}{\rho}$.
- **Assumes voluntary testing is available.**
  Not the case in early 2020.
- **Re-infections are counted as new infections.**
  Acceptable in early 2022, but not in summer 2022.

Limitations and final comments

- **For low testing, parameter ρ decides everything.**
  For $\tau \to 0$, we simply have $K_F = \frac{1}{\rho}$.
- **Assumes voluntary testing is available.**
  Not the case in early 2020.
- **Re-infections are counted as new infections.**
  Acceptable in early 2022, but not in summer 2022.
- **The many changes in restrictions makes it unfeasible to reproduce a longer period in the model.**
  December lockdown, January school re-opening, February general re-opening.

Limitations and final comments

- **For low testing, parameter ρ decides everything.**
  For $\tau \to 0$, we simply have $K_F = \frac{1}{\rho}$.
- **Assumes voluntary testing is available.**
  Not the case in early 2020.
- **Re-infections are counted as new infections.**
  Acceptable in early 2022, but not in summer 2022.
- **The many changes in restrictions makes it unfeasible to reproduce a longer period in the model.**
  December lockdown, January school re-opening, February general re-opening.
- **Definition of "a test" is not that well-defined.**
  International data (from OurWorldInData) longer updated.

Limitations and final comments

- **For low testing, parameter ρ decides everything.**
  For $\tau \to 0$, we simply have $K_F = \frac{1}{\rho}$.
- **Assumes voluntary testing is available.**
  Not the case in early 2020.
- **Re-infections are counted as new infections.**
  Acceptable in early 2022, but not in summer 2022.
- **The many changes in restrictions makes it unfeasible to reproduce a longer period in the model.**
  December lockdown, January school re-opening, February general re-opening.
- **Definition of "a test" is not that well-defined.**
  International data (from OurWorldInData) longer updated.
- **Parameter-values**
  Proper parameter-estimation should be carried out.

Thank you for your attention.

Any questions?

Website: rasmuspedersen.com
Email: rakrpe@ruc.dk