

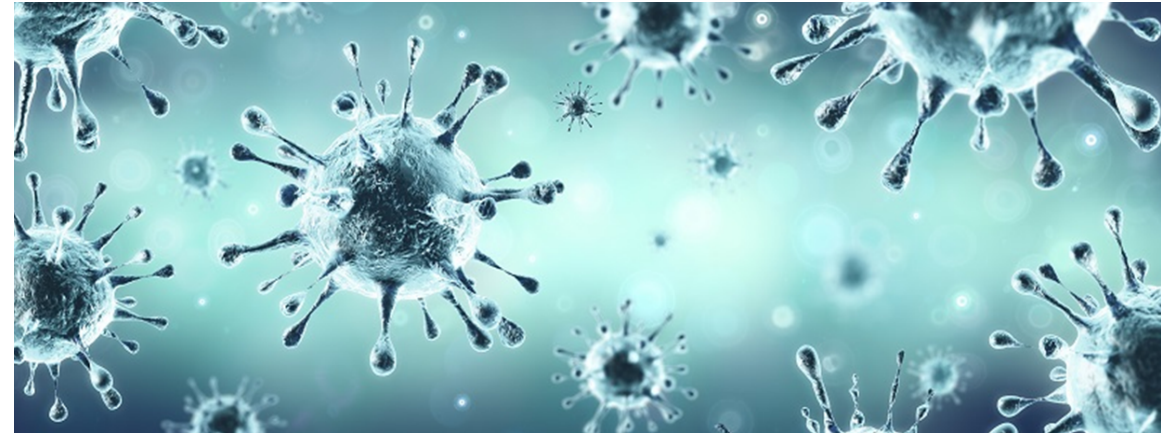


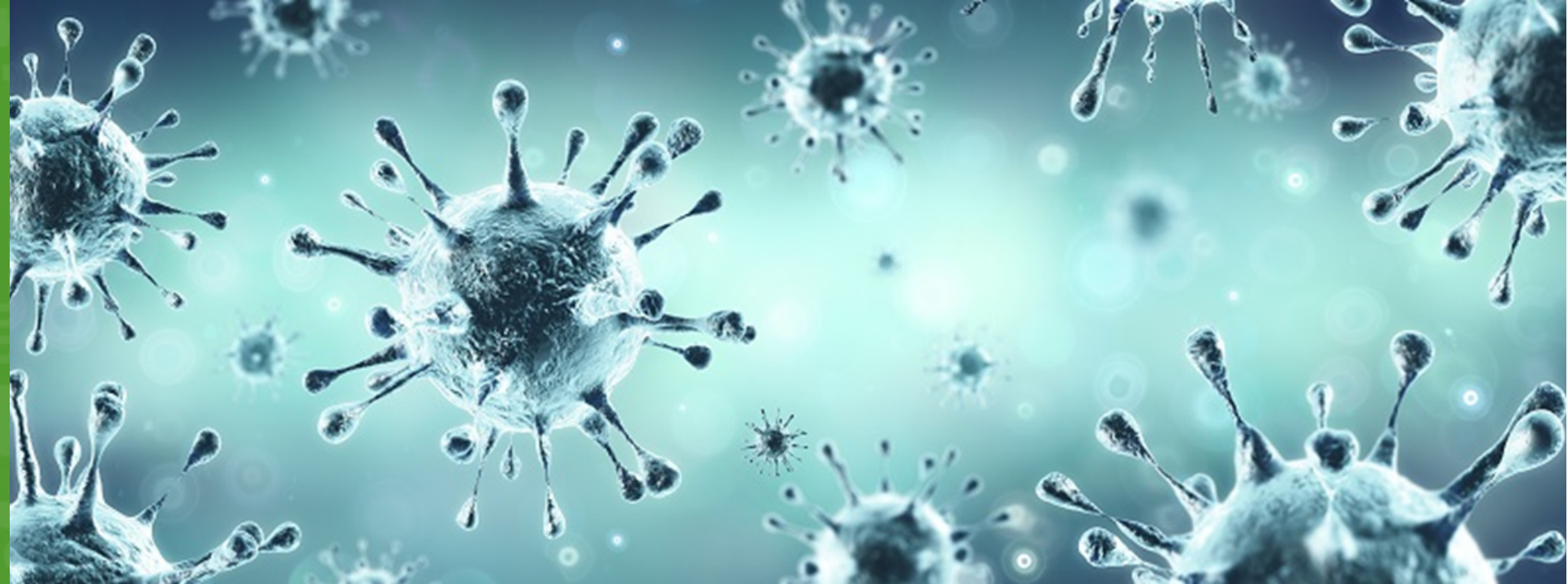
Skenaariomalleista oppiviin malleihin

Terveyden ja hyvinvoinnin laitos

Sisältö

- Koronavirusepidemian alkuvaiheen malli
 - perusteet, oletukset ja parametrit
 - skenaarioista ja niiden tulkinnasta
- Oppivat mallit
 - Periaate
 - Malleihin perustuva epidemian seuranta
 - Mallien toteutus





Epidemian alkuvaiheen malli

Alkuvaiheen mallin lähtökohdat

- Kun ilmiöstä ei vielä tiedetä paljon ja oman alueen tartunta- tai tautihavaintoja ei vielä juuri ole, on turvauduttava malleihin
- Epidemian alkuvaiheen malli perustuu tällöin
 - matemaattisen epidemiologian tarjoamiin työkaluihin epidemioiden mallintamiseksi (ks. 1. webinaari)
 - oletuksiin ja tietoihin mallin parametrien arvoista
 - tiedot koronaviruksen epidemiologiasta alueilla, joissa epidemia alkoi varhemmin: tutkimuskirjallisuus ja alkuperäiset aineistot
 - oman väestön kontaktirakenteeseen

Kaksi osamallia

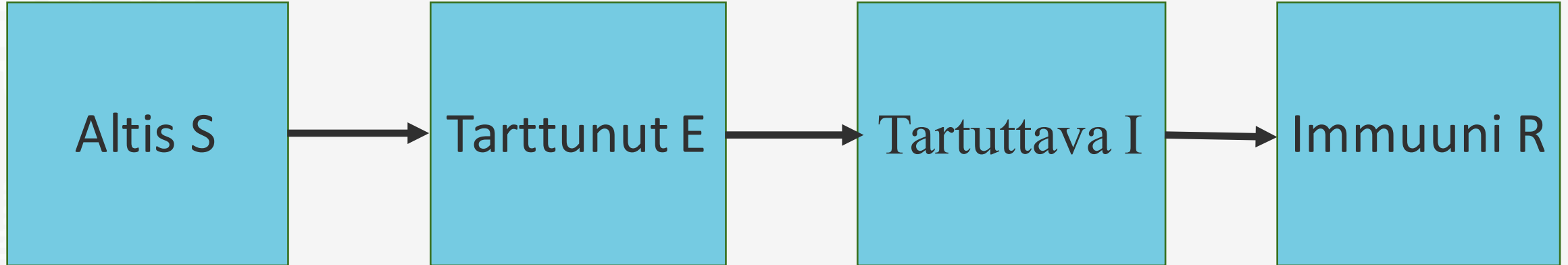
- Tarkastellaan erikseen tartuntoja ja sairastavuutta
- Koronavirustartuntojen SEIR-malli

Susceptible – Exposed – Infectious – Removed

Altis – Tartunnan saanut – Tartuttava – Immuuni

- Koronasairastavuuden malli
 - sairaalahoitojakson todennäköisyys per tartunta
 - tehohoitojakson todennäköisyys per sairaalahoitajakso
 - kuoleman todennäköisyys per tartunta
 - lisäksi arvioita viivejakaumista (esim. aika tartunnasta sairaalaan)

Tartuntojen leviämisen SEIR-malli



SEIR-mallin määrittely

- Väestön yksilöt siirtyvät ajanhetkellä t tilasta toiseen seuraavien differenssiyhtälöiden mukaisesti

alttiit: $S_{i,t+1} = S_{i,t} - (S_{i,t}/N_i) \sum_{j=1}^k (M_{i,j} I_{j,t}) / D_I$

tarttuneet: $E_{i,t+1} = E_{i,t} + (S_{i,t}/N_i) \sum_{j=1}^k (M_{i,j} I_{j,t}) / D_I - E_{i,t} / D_L$

tartuttavat: $I_{i,t+1} = I_{i,t} + E_{i,t} / D_L - I_{i,t+1} / D_I$

immuunit: $R_{i,t+1} = R_{i,t} + I_{i,t+1} / D_I$

i = ikäluokka $i = 1, \dots, 17$
 t = aika $t = 0, 1, 2, \dots, N_t$

$M_{i,j}$ on 17×17 -
jälkeläismatriisin M
ikäluokkia (i, j)
vastaava alkio
(seuraavat diat)

- Aika-askeleeksi otetaan yksi päivä
- Väestön koko N_{pop} ja ikäluokkien koot N_i oletetaan vakioiksi (2018 tieto):

$$N_{pop} = \sum_{i=1}^k N_i = \sum_{i=1}^k (S_{i,t} + E_{i,t} + I_{i,t} + R_{i,t}), t = 0, 1, 2, \dots, N_t$$

- Ikäluokkakohtaiset alkuarvot: $(S_{i,0}, E_{i,0}, I_{i,0}, R_{i,0})$, $i = 1, \dots, 17$

$1/D_I$ mallin
vauhtiparametrit
 $1/D_L$ (seuraava dia)

SEIR-mallin parametrit

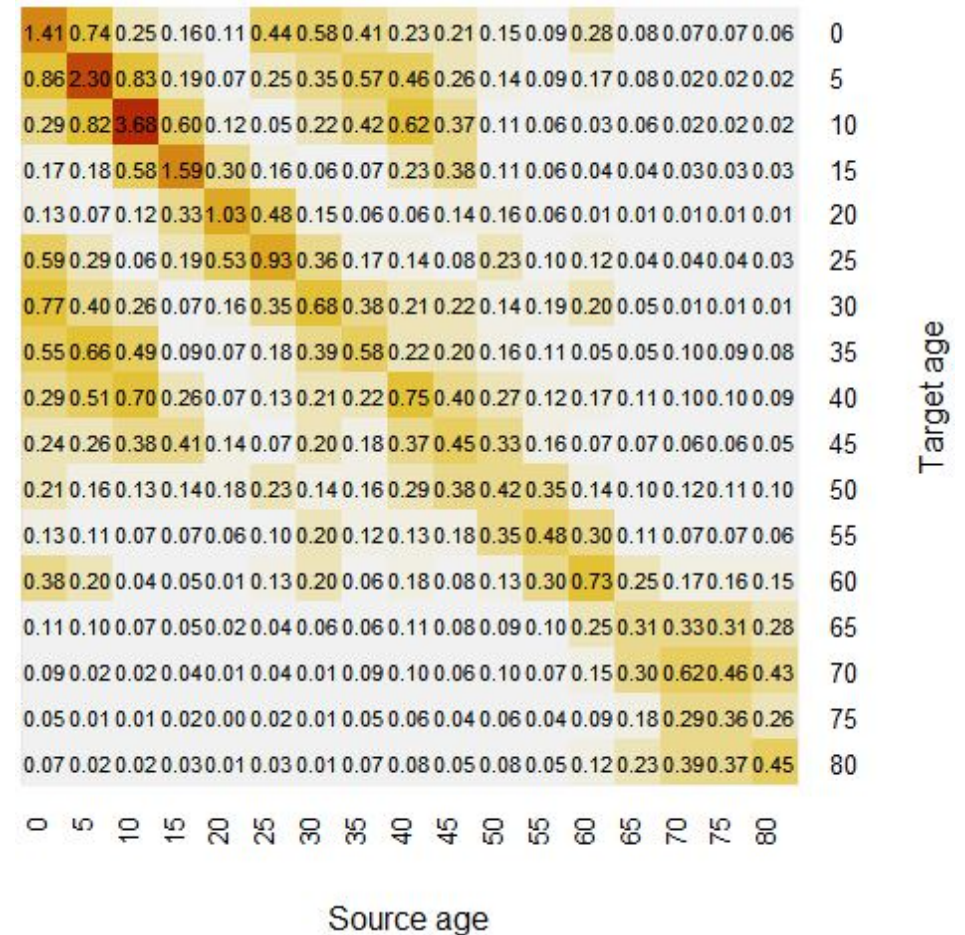
- SEIR-mallissa on kolme parametria sekä kontaktirakenne
- Parametrit määräävät epidemian koon ja epidemian vauhdin

Parametri	Merkitys	Arvo	Lähde
R_0	perustartuttavuusluku	2,4*	Kucharski et al., 2020
D_L	latentin vaiheen keskimääräinen kesto	3,0	He et al., 2020; Lauer et al. 2020
D_I	tartuttavuuden keskimääräinen kesto	7,0	He et al., 2020; Wölfel et al., 2020

*Myös vaihtoehtoisia oletuksia poikkeusolojen aikaisesta tartuttavuusluvusta ja vaihtoehtoisia arvoja epidemiaskenaarioissa

länmukainen kontaktirakenne

- 17 ikäluokkaa
- Kontaktimatriisi C
- Matriisin (i,j) alkio $C_{ij} =$
 j -ikäisen (sarake) lähikontaktit
 i -ikäisiin (rivi) per päivä



Kuva: Suomen POLYMOD –kontaktimatriisi, selvitetty tutkimuksella, jossa osallistujat pitivät kirjaa kontakteistaan (Mossong ym. 2008)

Jälkeläismatriisi saadaan kontaktimatriisista ja R_0 :sta

- Ikärakenteisessa SEIR-mallissa tartuntoja säätelee ns. jälkeläismatriisi M (*next generation matrix*)
 - Matriisin M alkio (i,j) on j -ikäisen tartuttavuusaikanaan tartuttamien i -ikäisten uusien tapausten keskimääräinen lukumäärä
- Jälkeläismatriisi saadaan skaalamalla kontaktimatriisi: $M = \alpha C$ (Wallinga ym. 2006)
- Perustartuttavuusluku on jälkeläismatriisin suurin ominaisarvo:

$$R_0 = \rho(M) = \alpha \rho(C) \quad \Rightarrow \quad \alpha = \frac{R_0}{\rho(C)}$$

- Kun R_0 :n arvo on annettu/oletettu, seuraa, että $M = \frac{R_0}{\rho(C)} C$

Koronasairastavuuden parametrit: periaate

- Arvioitiin sairaalahoidon ja kriittisen sairauden iänmukainen riski koronavirustartuntaa kohden
 - Lisäksi arvioitiin kuoleman riski koronavirustartuntaa kohden
- Aineistona Wuhanin/Hubein alueen sairaalajaksojen ja kriittisesti sairaiden kokonaismäärät 12.3. mennessä (WHO Situation Report 13.3.)
 - Vastaavasti kuolemantapausten kokonaismäärä 12.3. mennessä
- Sairaalahoitoa vaatineiden tapausten määrät jaettiin tartuntamäärillä, jotka arvioitiin Wuhanin väestössä mallinnetusta SEIR-epidemiasta
 - Kriittisesti sairastuneiden osuus arvioitiin samoin Wuhanin tietojen pohjalta
 - Vastaavasti arvioitiin kuolleisuus per tartunta (IFR = infection fatality ratio)

Tartuntamäärien arviointi

- Havaittuihin tapauksiin (Pan ym. 2020) sovitettiin SEIR-malli olettamalla, että $R_0 = 2,4$ tammikuun 23. päivään saakka ja sen jälkeen $R_0=1,2$ (Kucharski ym. 2020); ks. s. 13-14
 - Lisäksi oletettiin, että tapausten raportointi parani tammikuussa sen jälkeen, kun ilmoitettiin koronaviruksen tarttuvan ihmisten välillä (vrt. Zhao et al. 2020)
- Sovitetun epidemian mukaan saatiin tartuntatapausten arvioitu lukumäärä, joka oli n. 80-kertainen havaittujen tapausten lukumäärään verrattuna
- Tämän mukaisesti arvioitiin sairaanhoidon ja tehohoidon tarpeen riski per tartunta; vastavasti kuoleman riski
- Parametrit ovat sivulla 15

Tartuntamäärien arviointi (2)

- SEIR-mallin Wuhanin väestö- ja kontaktirakenteella tuotettu epidemiakäyrä sovitettiin lähdeviitteen (Wang; Pan, 2020) raportoituja tapauksia vastaavaan epidemiakäyrään.
- Sovitus tehtiin rajoitusten alkamiseen asti olettaen että $R_0=2,4$ (Kucharski). Näin saatiin karkea arvio Wuhanin epidemian tapausten kokonaismäärästä.
- Ennen 21.1 raportoitujen tapausten määrän on esitetty olevan vertailukelvottomia 21.1 jälkeen raportoitujen määrien kanssa. Tätä korjausta vastaavat alussa kaksinkertaistetut tpausmäärät on merkitty sinisellä (Zhao, 2020).

Tartuntamäärien arvionti (3)

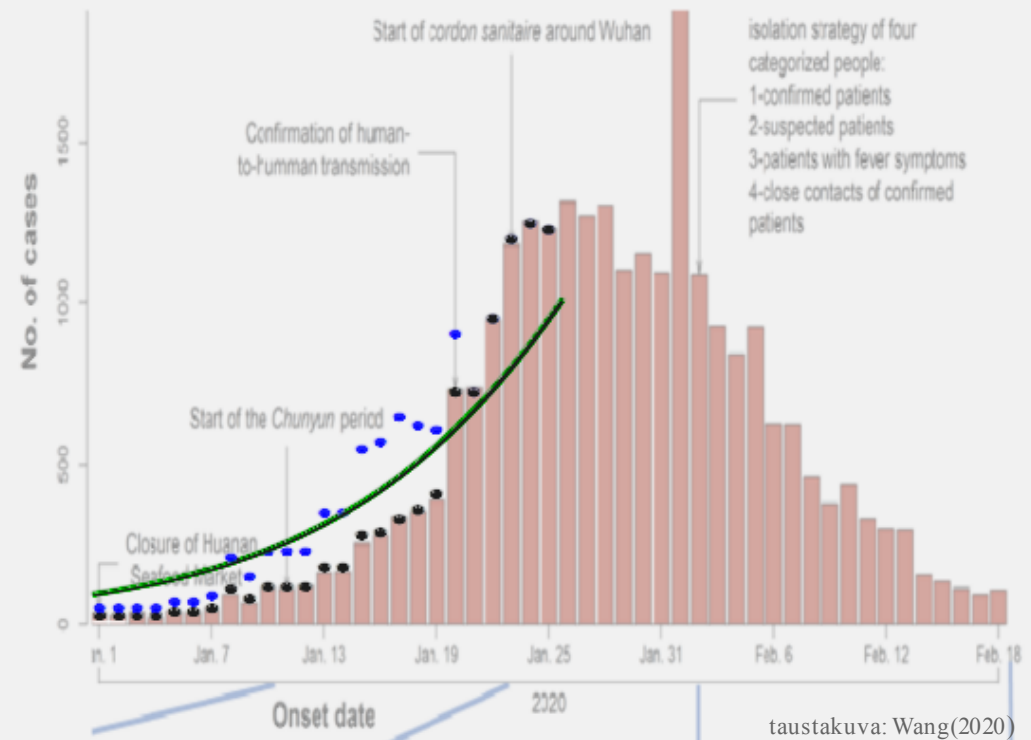
Wang (2020) Evolving Epidemiology and Impact of Non-pharmaceutical Interventions on the Outbreak of Coronavirus Disease 2019 in Wuhan, China. medRxiv. (preprint , maaliskuu)

Pan (2020) Association of Public Health Interventions With the Epidemiology of the COVID-19 Outbreak in Wuhan, China. JAMA

Kucharski (2020) Early dynamics of transmission and control of COVID-19: a mathematical modelling study: Lancet Inf. Dis.

Zhao (2020) Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. Int J Inf Dis.

Figure 1. The daily Covid-19 onset and the control measures across different periods



Sairastavuusmallin parametrit (1)

Ikä (vuosia)	Sairaalahoido per tartunta (%)	Kriittinen sairaus sairaus per sairaalahoido (%)	Tehohoito* per kriittinen sairaus (%)	Kuolema per tartunta (%)
0-9	0.005	8	100	0.00
10-19	0.01	9	100	0.00
20-29	0.11	13	100	0.00
30-39	0.18	18	100	0.00
40-49	0.24	25	100	0.01
50-59	0.5	32	100	0.05
60-69	2.0	38	100	0.5
70-74	6.0	45	50	3.1
75-79	6.0	45	25	3.1
80+	6.5	51	10	5.6

*Suomalainen asiantuntija-arvio

THL koronamallinnusryhmä, käsikirjoitus

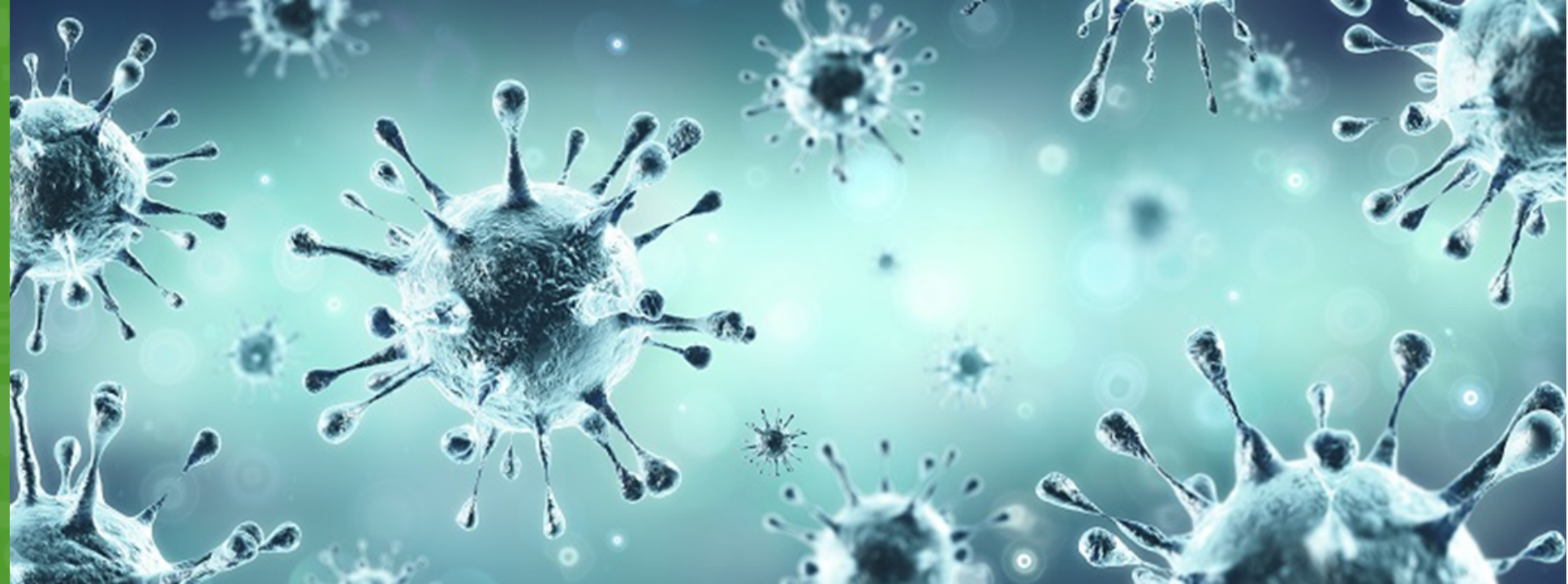
Sairastavuusmallin parametrit (2)

- Jos tartunta johtaa sairaalajaksoon ja mahdollisesti tehohoitoon, sairaanhoidon kuormituksen arvioimiseksi tarvitaan myös viivejakaumia
- Taulukossa ajat ovat päiviä tartunnasta (suluissa keskihajonta)

Potilas	Jakso	Alkupäivä vrk	Kesto vrk
vain vuodehoito tehohoitopotilas	vuodehoito	10 (4)	8 (4) *
	vuodehoito 1	10 (4)	4 (1)
	tehohoito	14 (5)	12 (4)
	vuodehoito 2	22 (6)	4 (1)
Kuolleet		Ajankohta vrk	
	kuolema	18 (6)	

Zhou ym. Lancet. 2020 March 28;1054-62
Linton ym. J Clin Med. 2020 Feb 17;9(2)
[https://www.icuregswe.org/data--resultat/
covid-19-i-svensk-intensivvard/](https://www.icuregswe.org/data--resultat/covid-19-i-svensk-intensivvard/)
Kotimaiset rekisteriaineistot

* Kaikki jakaumat diskretoituja gammajakaumia

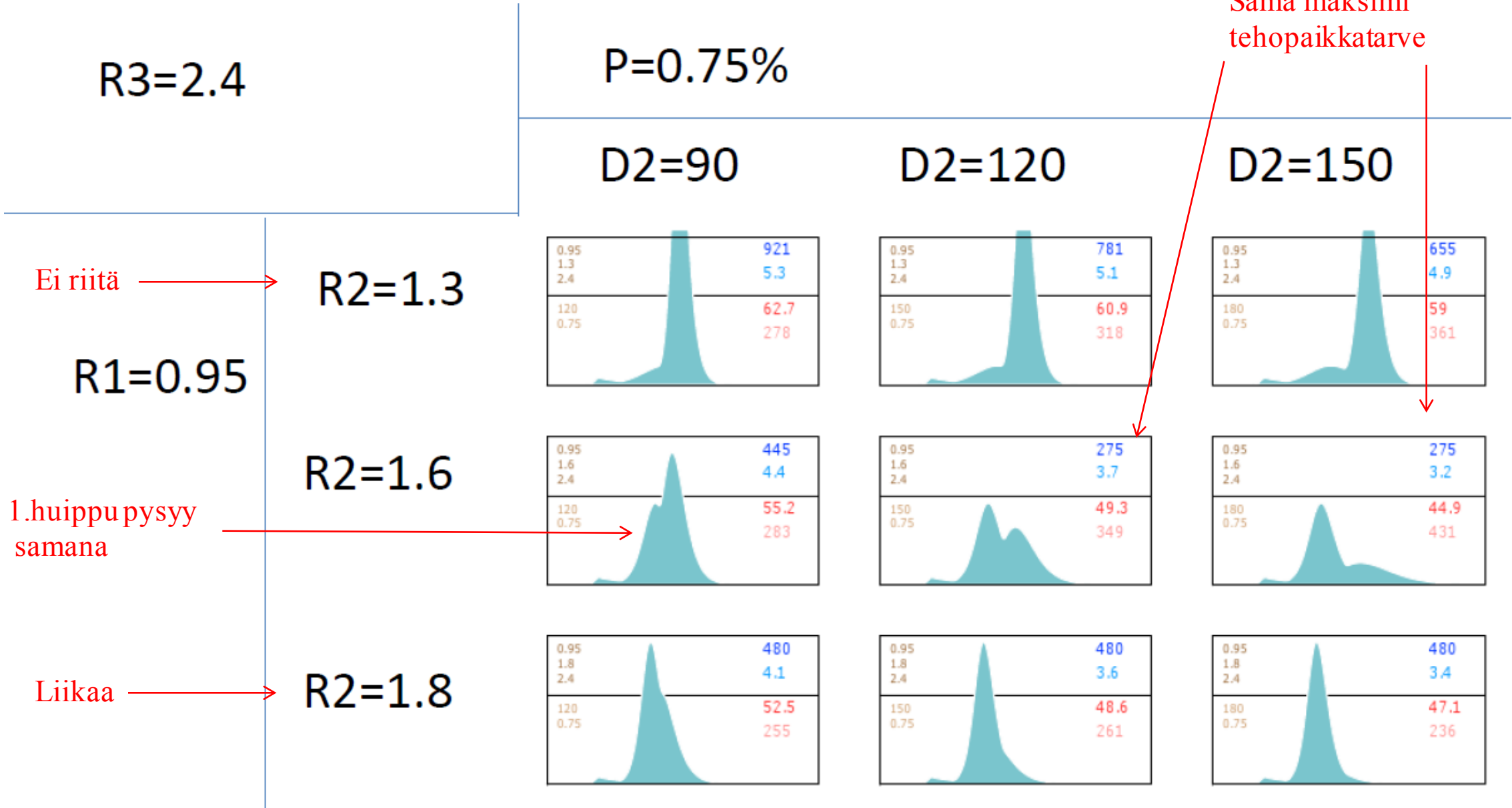


Alkuvaiheen skenaariot

Skenaarioista ja niiden parametriarvoista

- Malleilla voidaan vastata kysymyksiin: Mikä on ilmiön luonne? Mille oletuksille tulokset ovat herkkiä? Miten jonkin parametrin arvon muuttaminen vaikuttaa?
- Ennusteita voidaan antaa, mutta niiden epävarmuus heijastaa aina epävarmuuttamme parametriarvoista ja oletuksista
- Mallit eivät suoraan kerro, mitä todella tapahtuu tai mikä on totta
- Seuraavassa on erilaisia skenaarioita, joissa vaihdellaan
 - Tartuttavuusluvun arvoa poikkeusolojen aikana (R1)
 - Tartuttavuusluvun arvoa epidemian jatkovaiheessa (R2) ja tämän vaiheen kestoa (D2)
 - Tartunnan jo saaneiden eli imuunien osuutta poikkeusolojen alussa (P)

Pieni R0 poikkeusolojen aikana & hyvin pieni immuunien osuus P



Pieni R0:n arvo poikkeusolojen aikana & kohtalainen immuunien osuus P

R3=2.4

P=2.5%

pieni tehohoitoaikkujen
maksimitarve

D2=90

D2=120

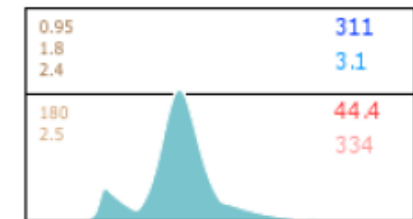
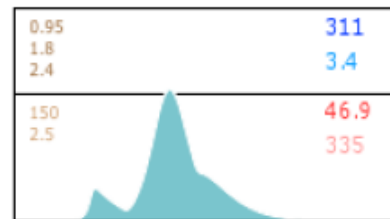
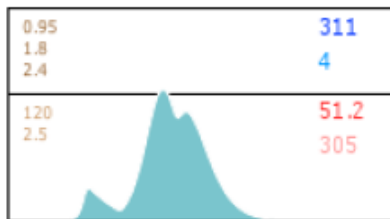
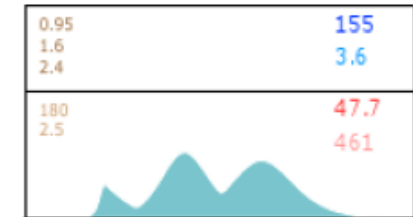
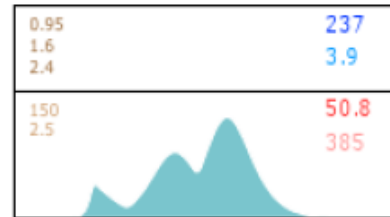
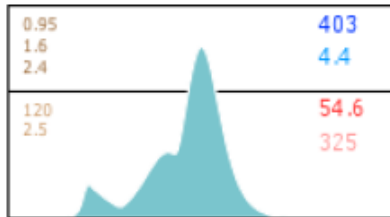
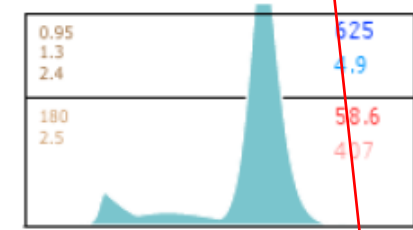
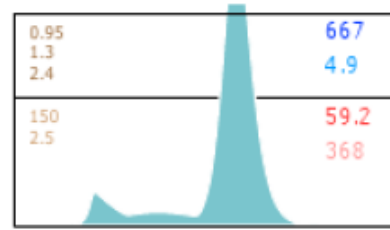
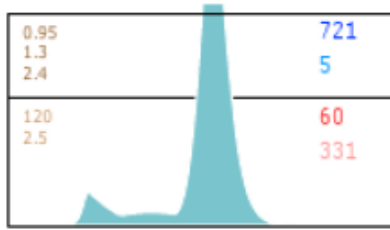
D2=150

R1=0.95

R2=1.3

R2=1.6

R2=1.8



Hetemäen raportin skenaarioissa: R0=1 ja P=2.5%

R3=2.4

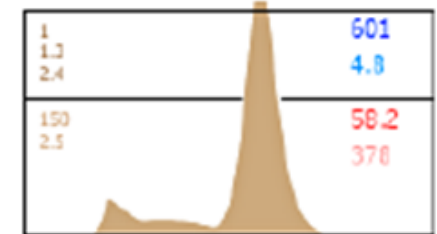
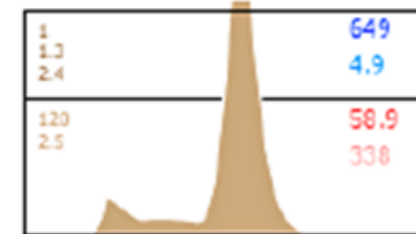
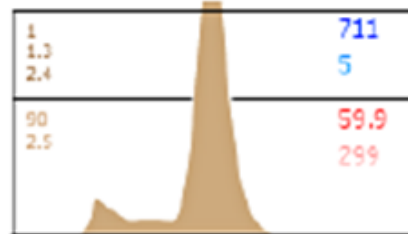
P=2.5%

D2=90

D2=120

D2=150

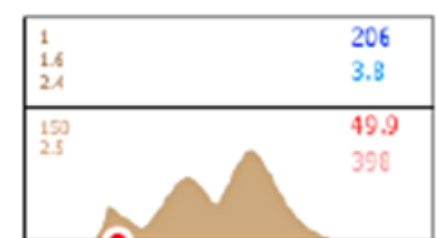
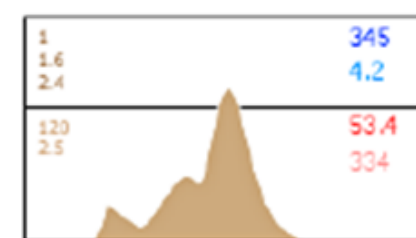
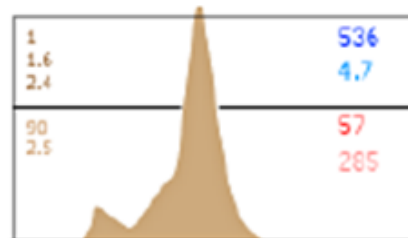
R2=1.3



~~R1=0.95~~

R1=1

R2=1.6

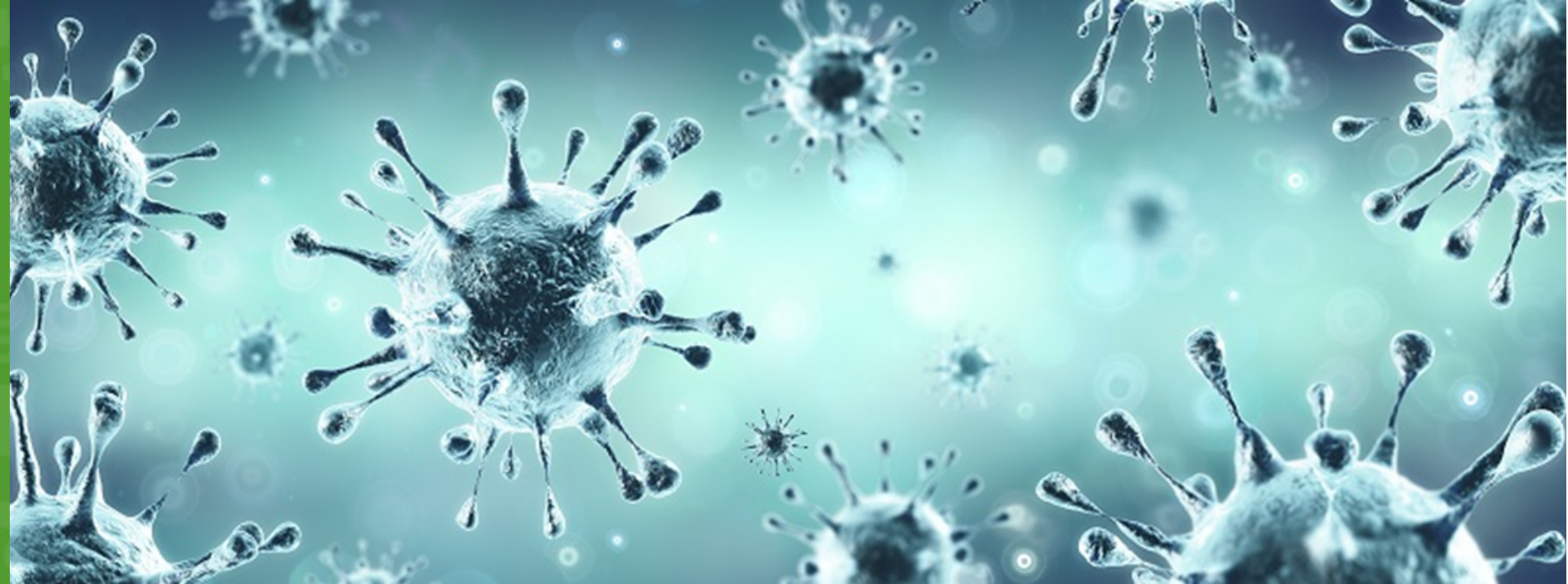


R2=1.8



Skenaarioista ja niiden parametriarvoista

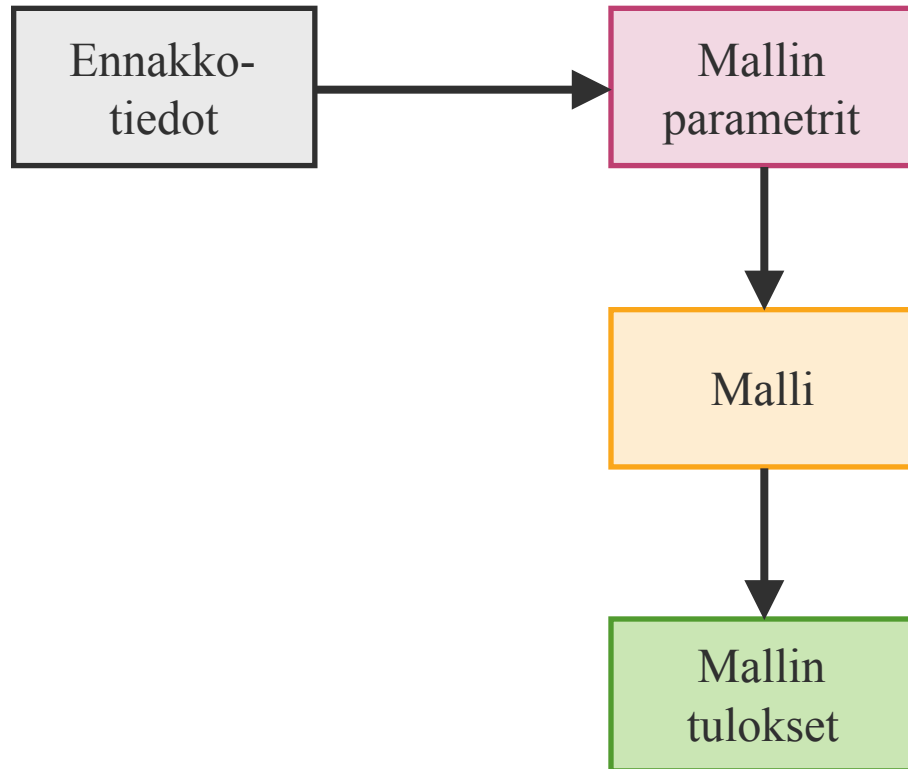
- Kun immuunien osuus on joka tapauksessa on pieni ja poikkeusolojen aikana R_0 on ollut hyvinkin matala → tehopaikat riittävät, jos epidemiaa pystytään hidastamaan niin, että tulee 3 aaltoa
- Epidemian kestolla on merkitystä ikääntyneiden suojaamisen edellytyksille
- Olisi helpointa optimoida vain jotain osaa koko prosessista, mutta se ei ole mahdollista



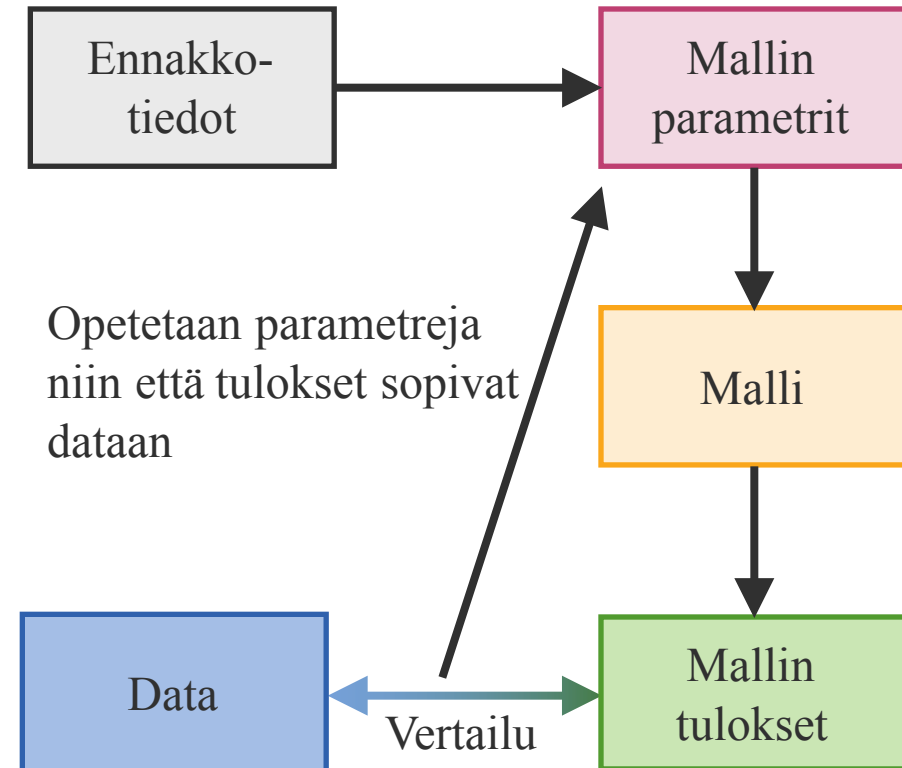
Oppiva COVID-19 malli

Oppiva COVID-19 malli – mitä tarkoittaa?

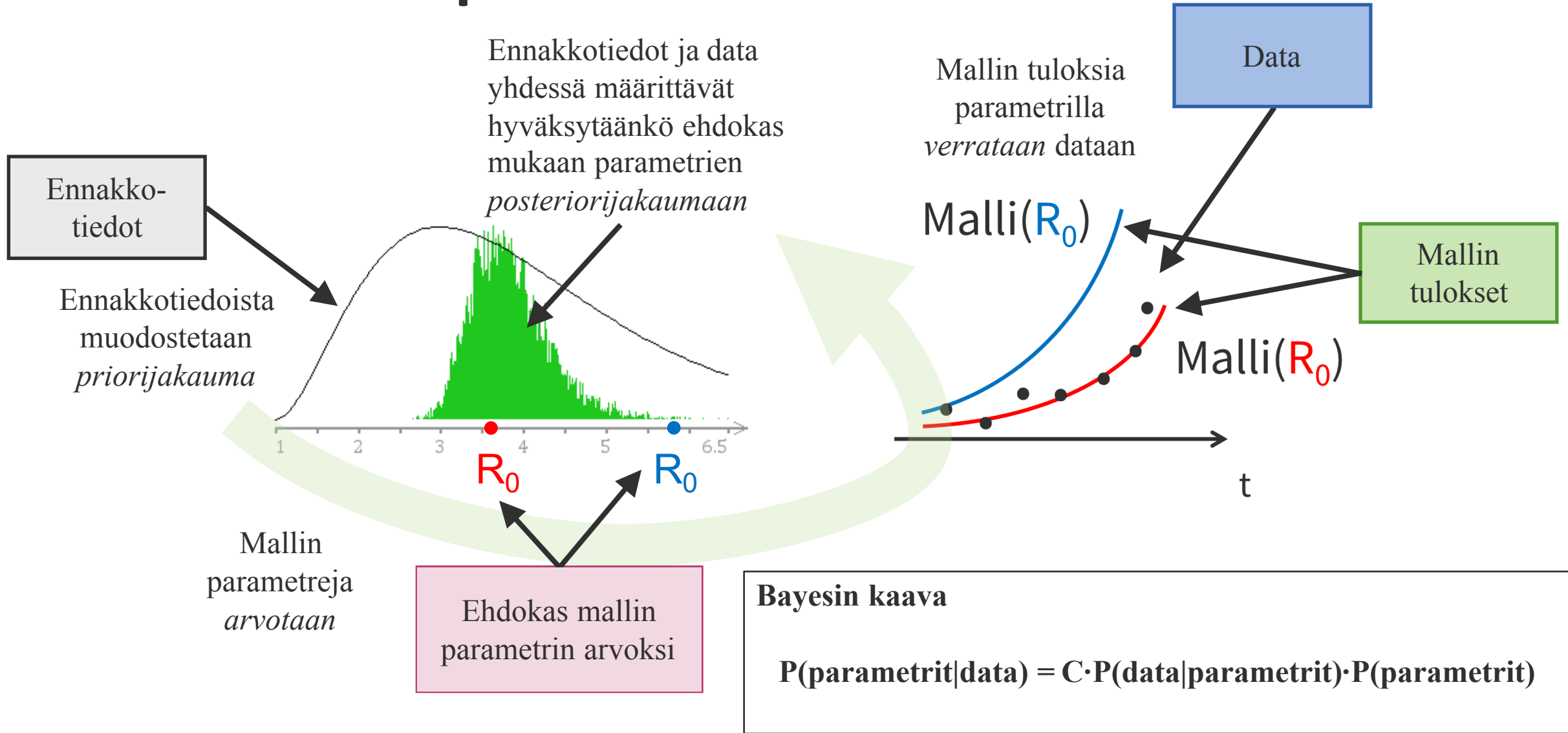
Kiinteäparametrinen malli

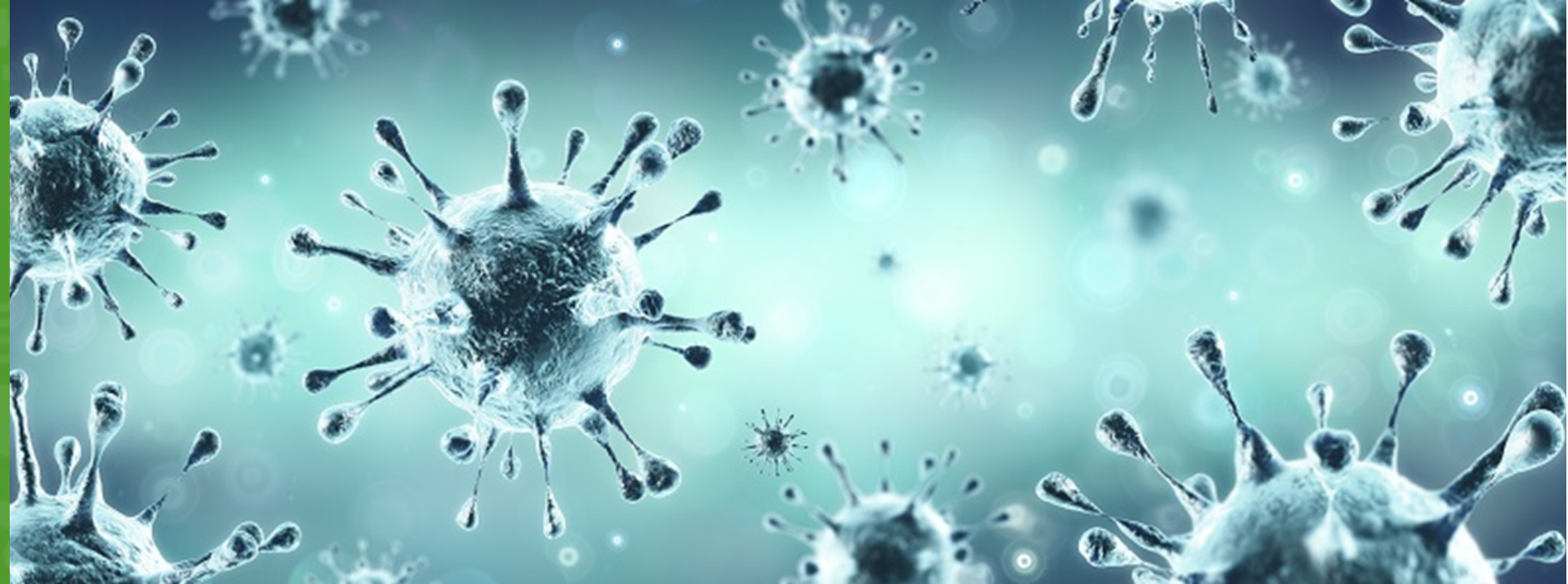


Oppiva malli



Parametrien "opettaminen"





Learning Model

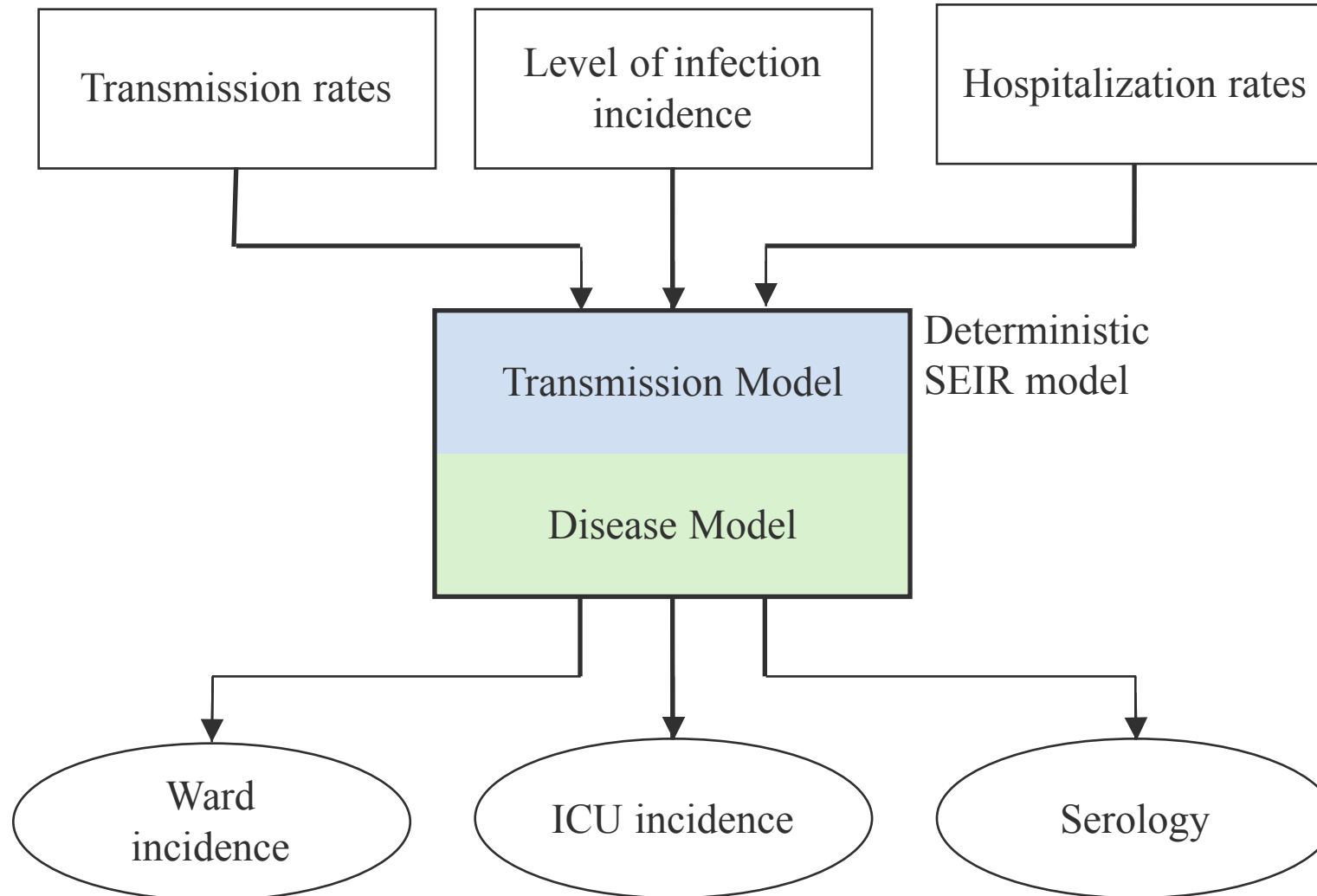
Model

- We use a deterministic model, similar to the scenario model
- We calibrate the model, using actual data on hospitalisation and intensive care
- Uncertainty in the estimates derives from uncertainty in the parameters

Deterministic Model

- SEIR model
- Discrete-time dynamics, time step equal to one day
 - Modelling starts from 15 February
- Population: either HUS or whole Finland
 - Structured into 9 age groups: 0-9, 10-19, ... , 70-79, 80+
 - Focus on HUS as it has the most data
 - Estimates could be transferred to other regions
- Two parts: Transmission model and Disease model

Model overview



Estimated parameters

- Transmission rates (R_0)
 - Is epidemic growing? At which rate?
 - How effective were the interventions?
- Infection incidence
 - How many people have been infected already?
 - How many are infectious right now?
- Hospitalization/infection, ICU/hospitalization rates
 - How severe is the infection?
 - What would be the burden on healthcare system?

Important sources of information

□

Parameter	Growth dynamics of hospitalization and ICU cases	Serological data: quick test results	Numbers of hospitalizations and ICU cases
R_0	X		
Infection incidence		X	X
hosp/infection		X	X
ICU/hosp			X

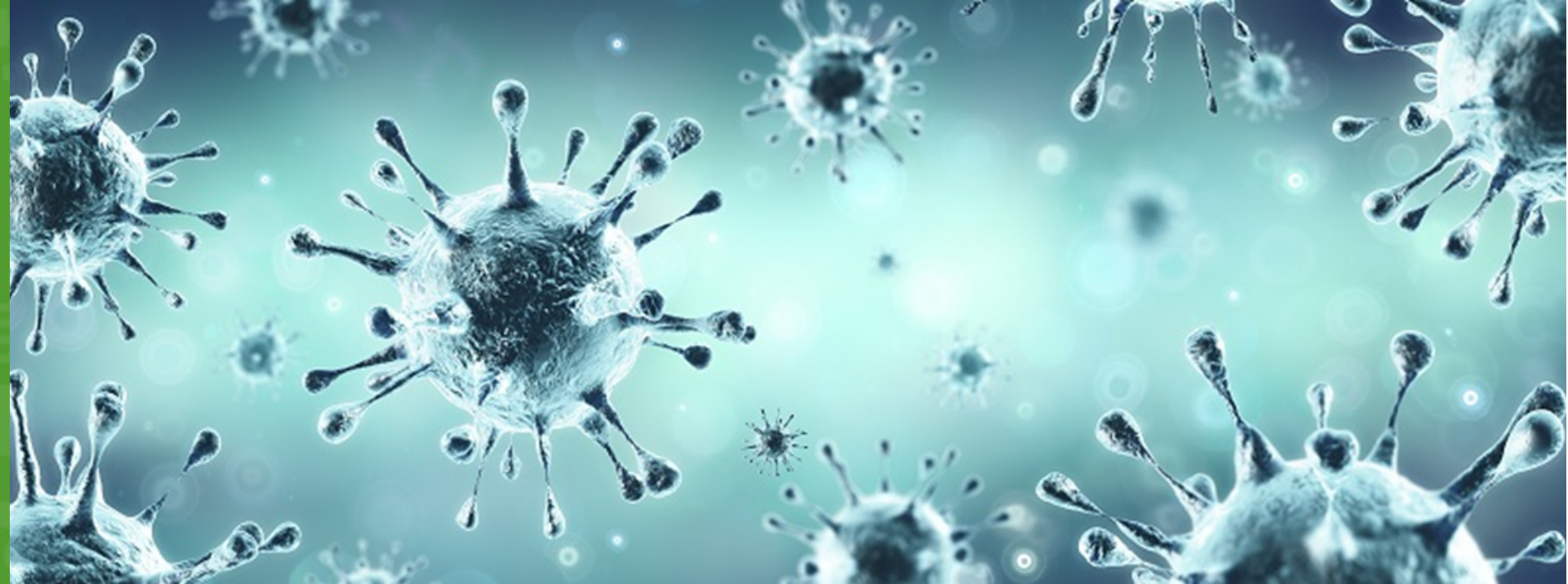
Parameters learned from elsewhere

- Exposed and Infectious period
 - Taken from literature
- Period of hospital and ICU stay
 - Estimated separately from local individual-based data
 - Used to predict hospital burden
- Contact rates in the population
 - Estimated separately from our own survey (new study)

Data not utilized

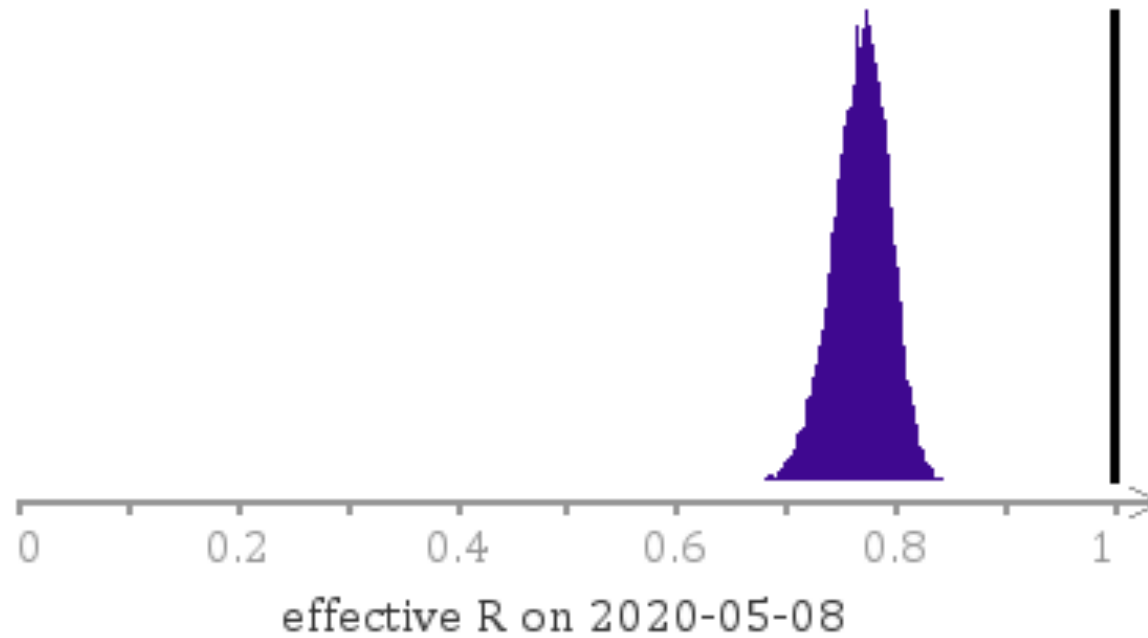
- Confirmed Cases
 - Too stochastic, depends on the number of tests taken, unknown underreporting, selection rules and test rules change all the time
- Mortality Data
 - Too sporadic. Majority of deaths comes from elderly homes, these events do not reflect the process of spreading of epidemic through the population
- Hospital ward and ICU occupations
 - Derivatives from hospital and ICU incidence, which we are using

Although we don't fit this data explicitly, we keep an eye on these numbers

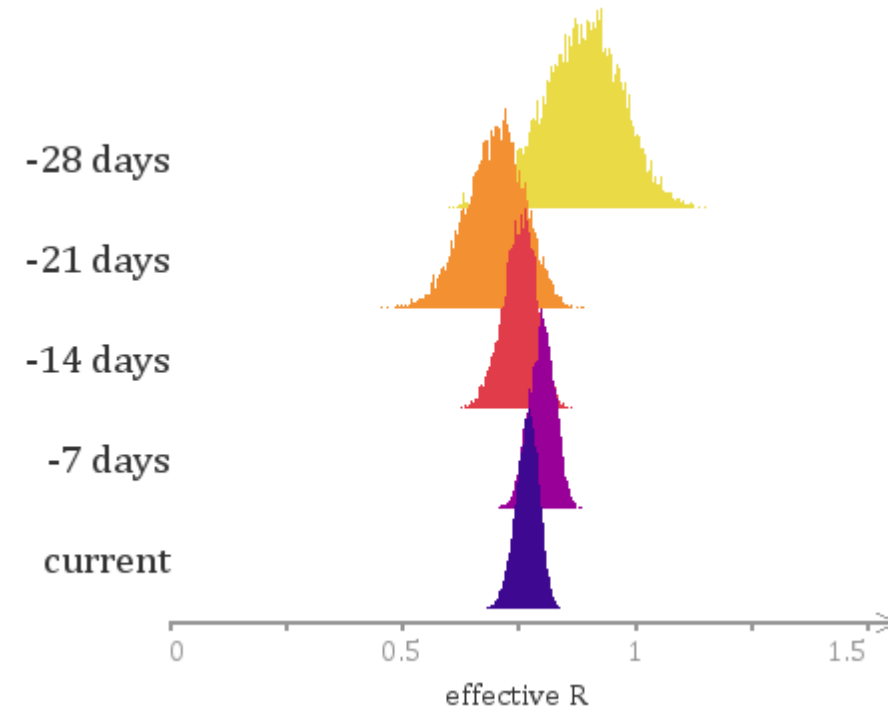
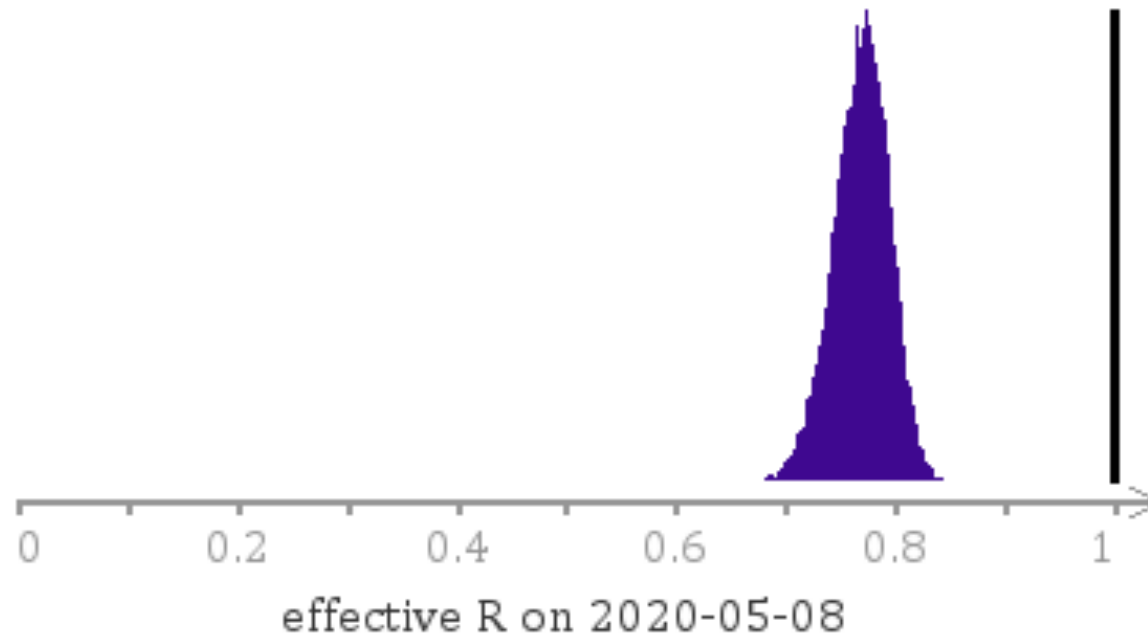


Results (for HUS area)

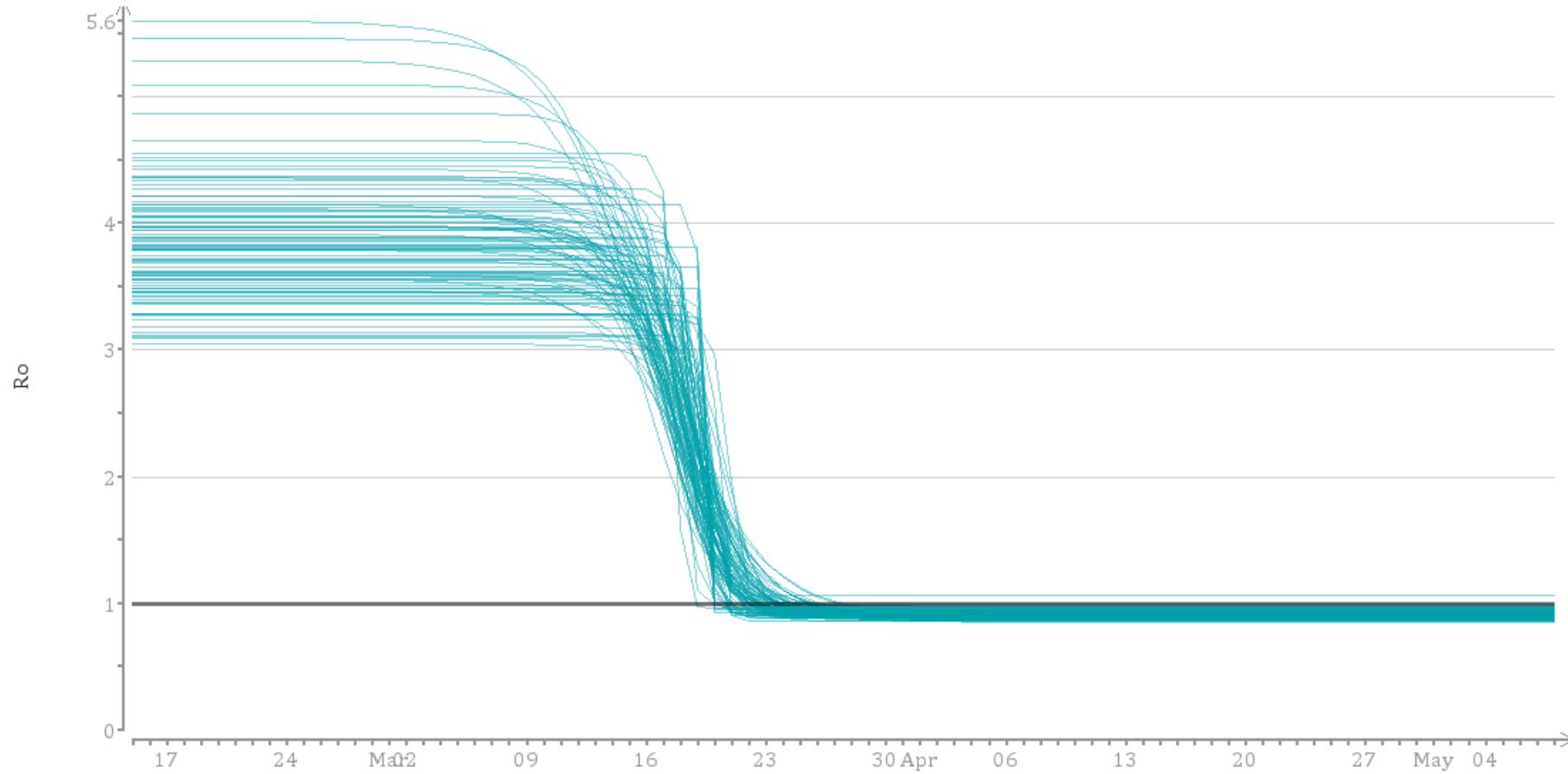
Effective reproduction number



Effective reproduction number



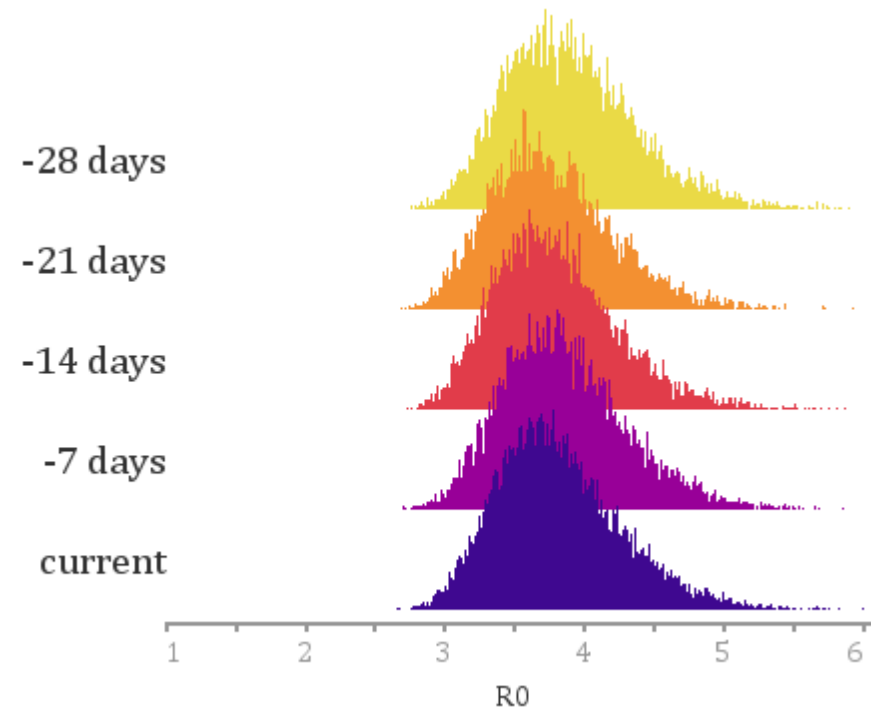
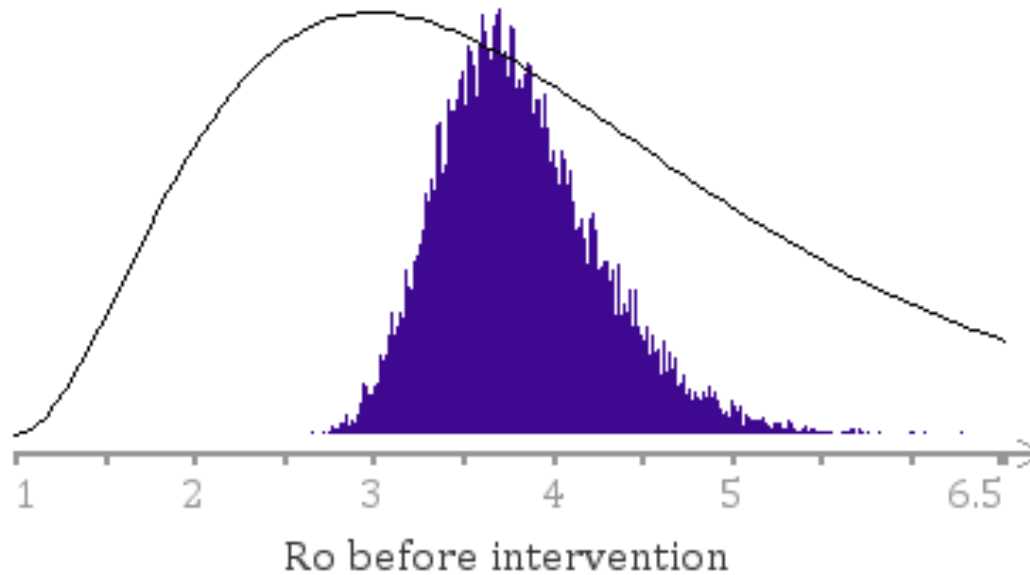
R dynamics



R dynamics

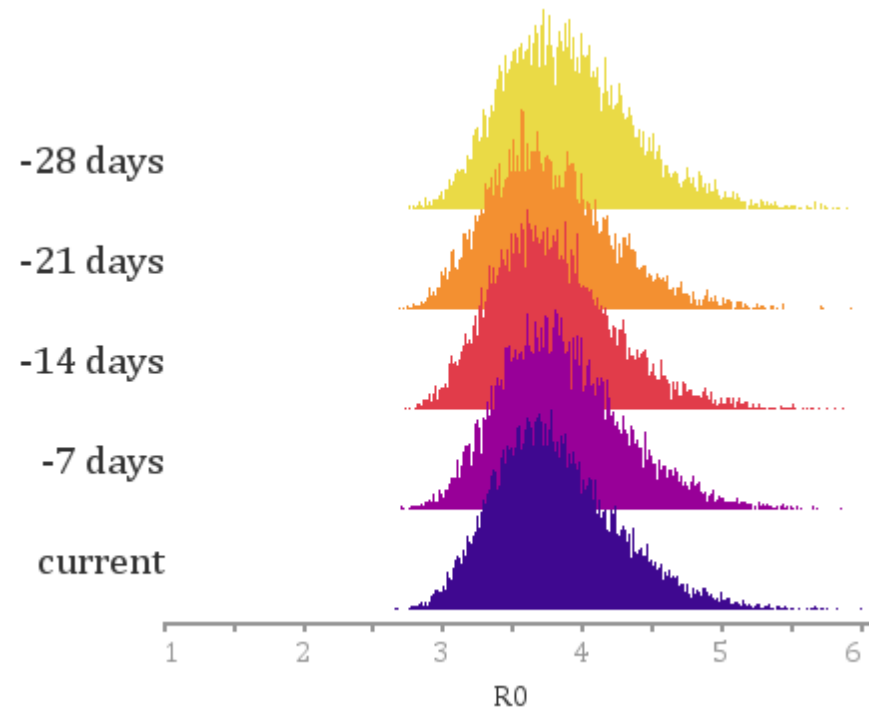
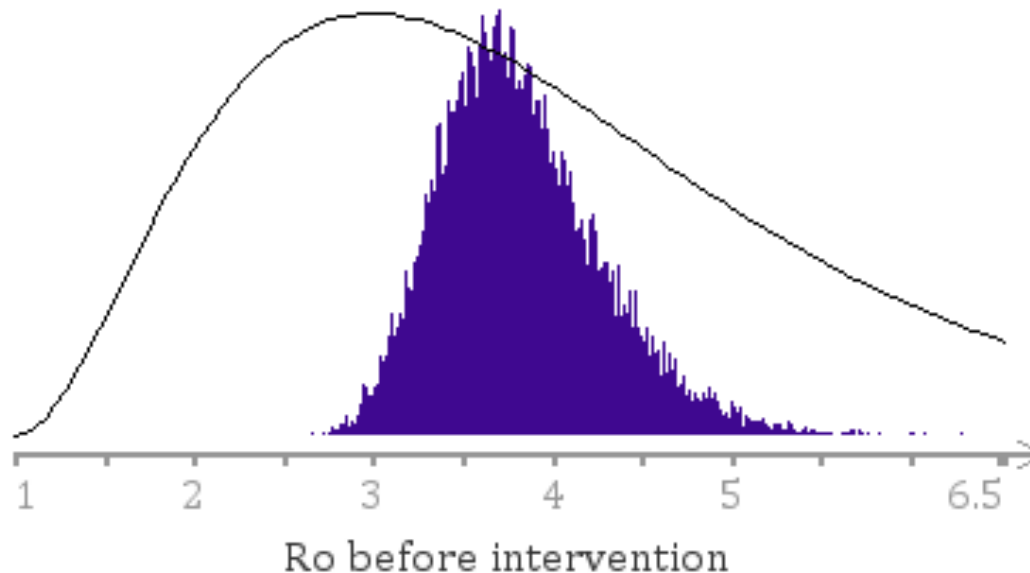
- We can estimate current effective R quite confidently
 - data have now accumulated from a long enough period
- Less confidence in estimating R_0 at the beginning of the epidemic
 - Restrictions were implemented so early that there are no direct Finnish data from a pure restriction-free situation, in practice
 - Infection in February was affected by processes we are not modelling explicitly (imported cases)

Initial Reproduction Number

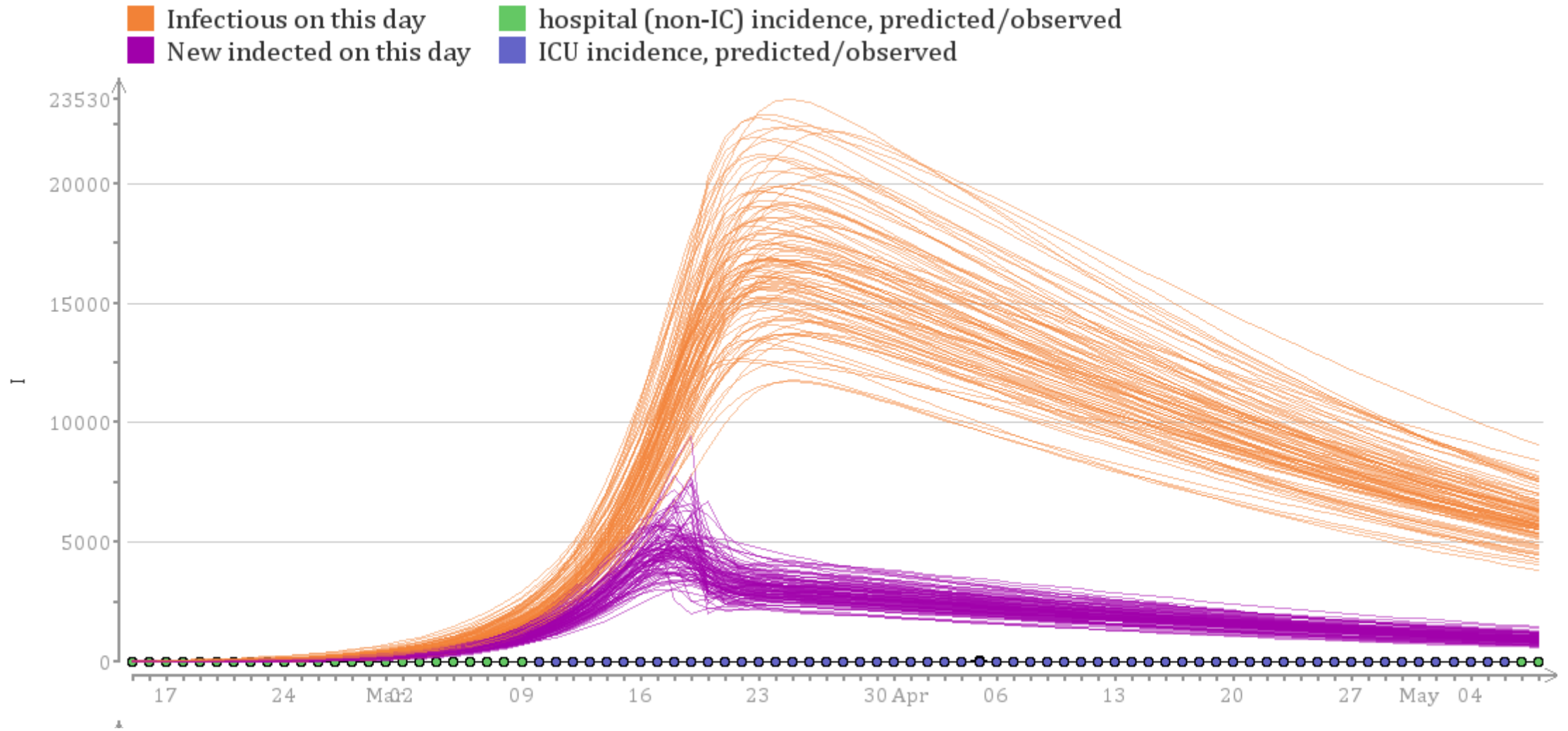


Initial Reproduction Number

- Initial (basic) R_0 would be important for long-term planning
- Other evidence suggest it should be bigger than 4

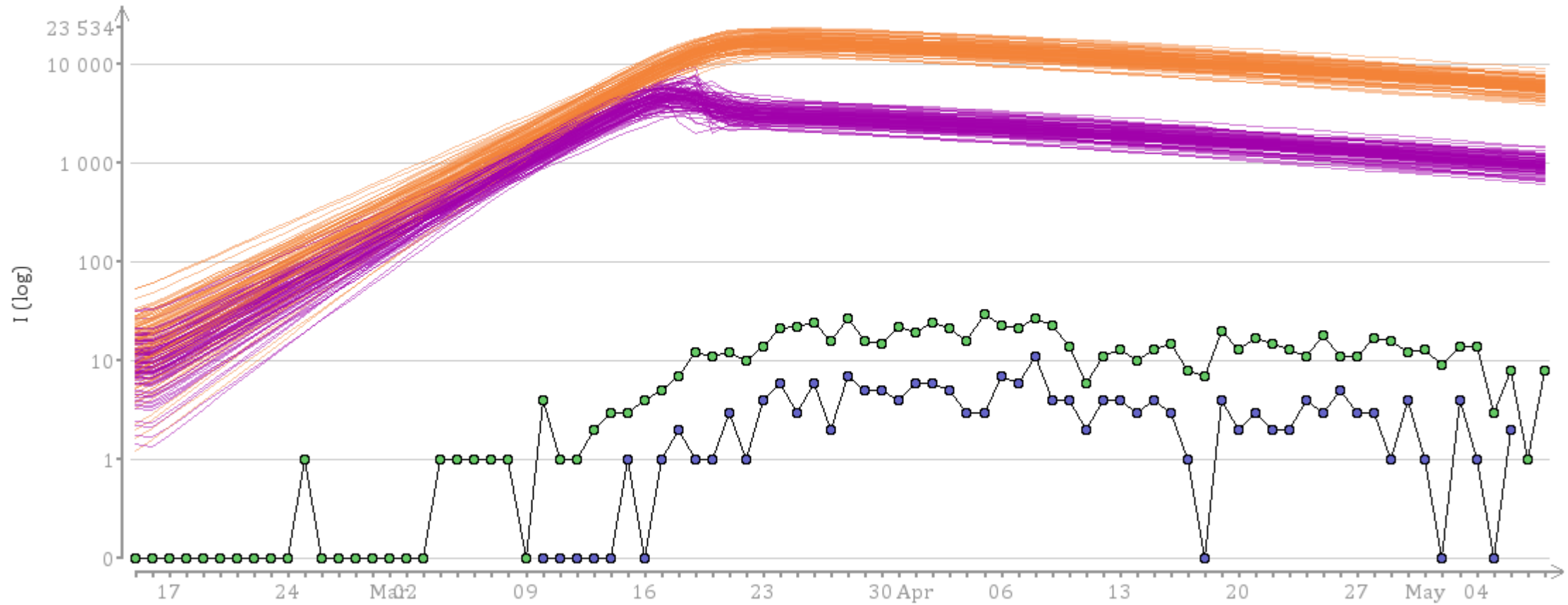


Incidence

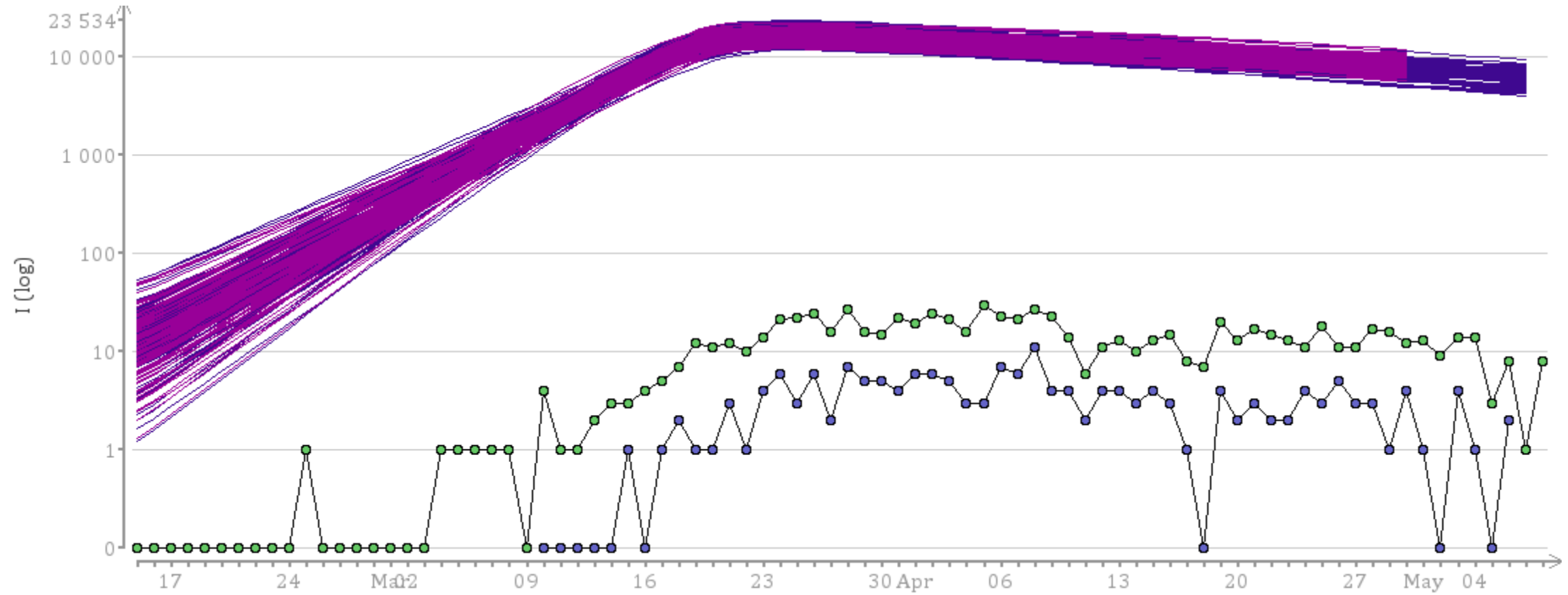


Incidence

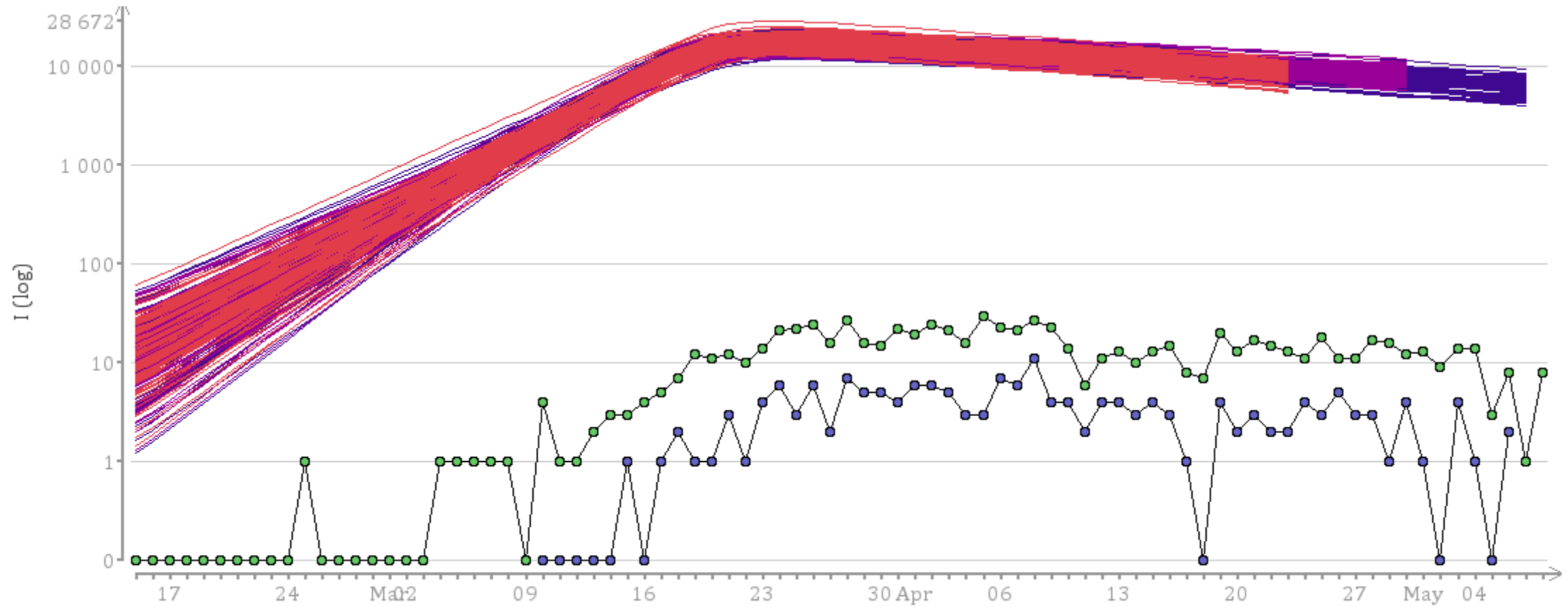
- Infectious on this day
- New indected on this day
- hospital (non-IC) incidence, predicted/observed
- ICU incidence, predicted/observed



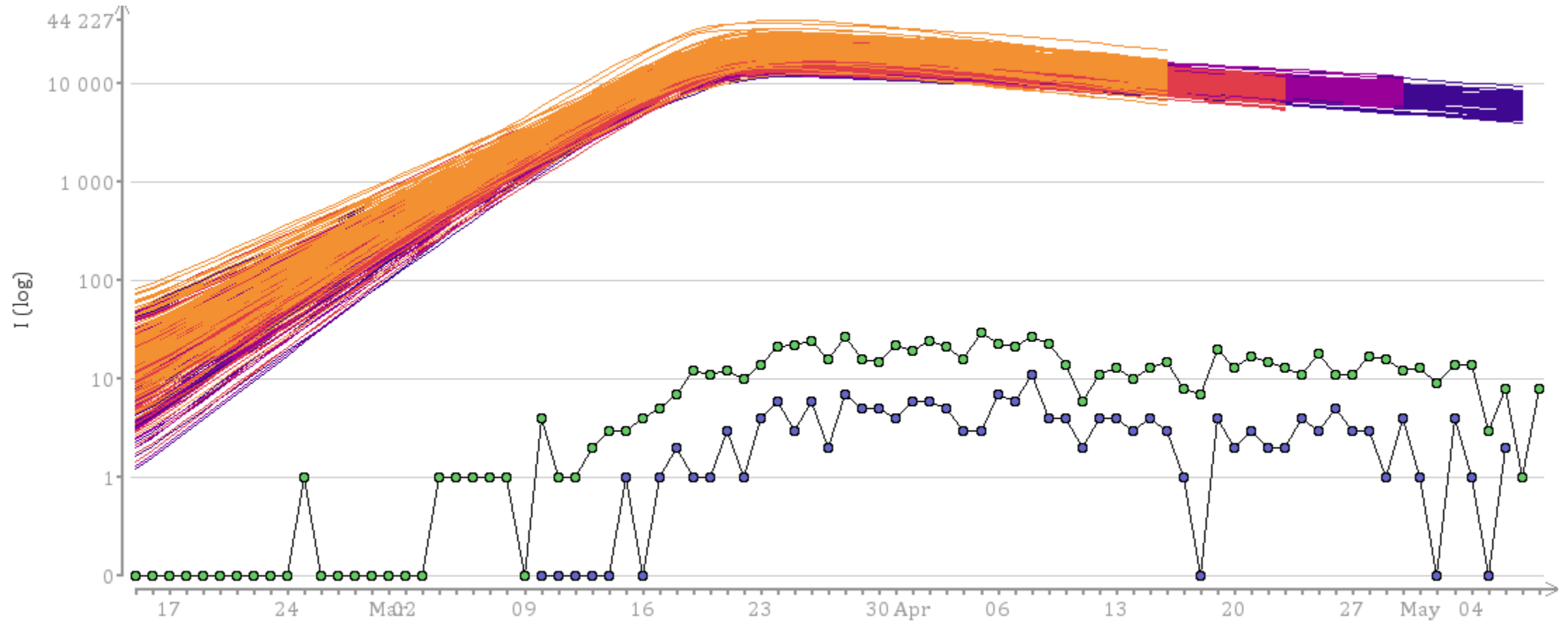
Infectious



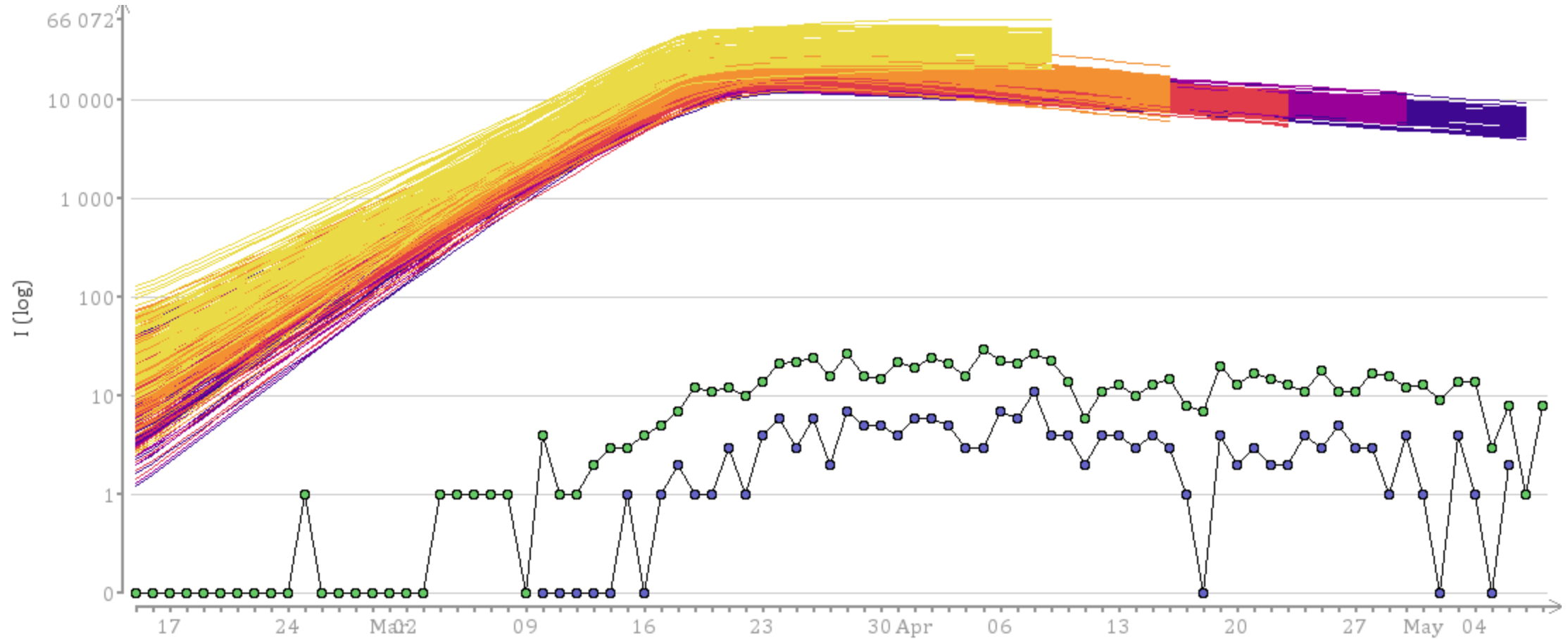
Infectious



Infectious



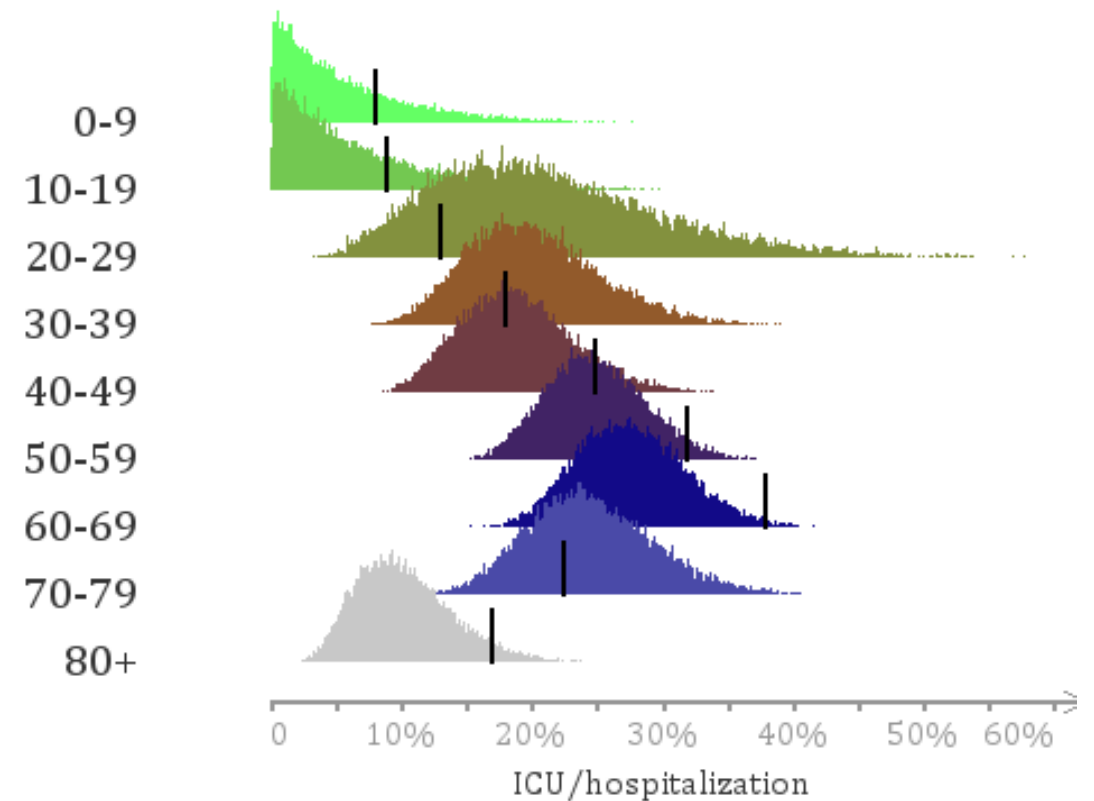
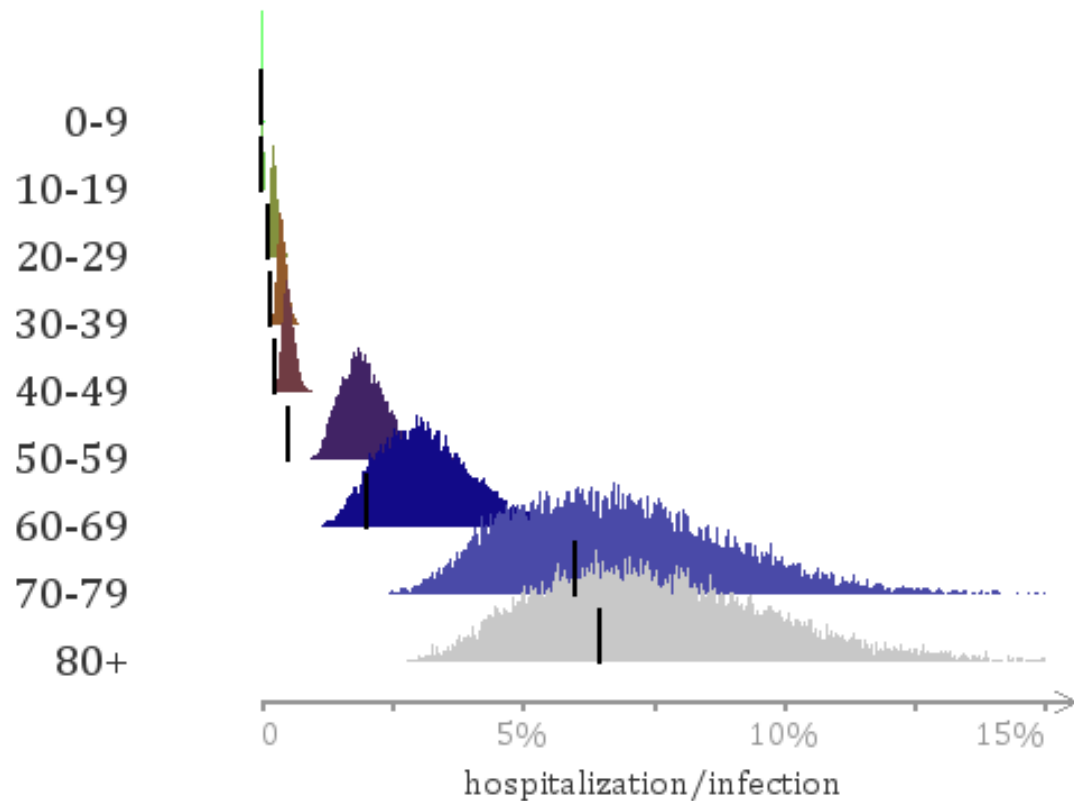
Infectious



Incidence and seroprevalence

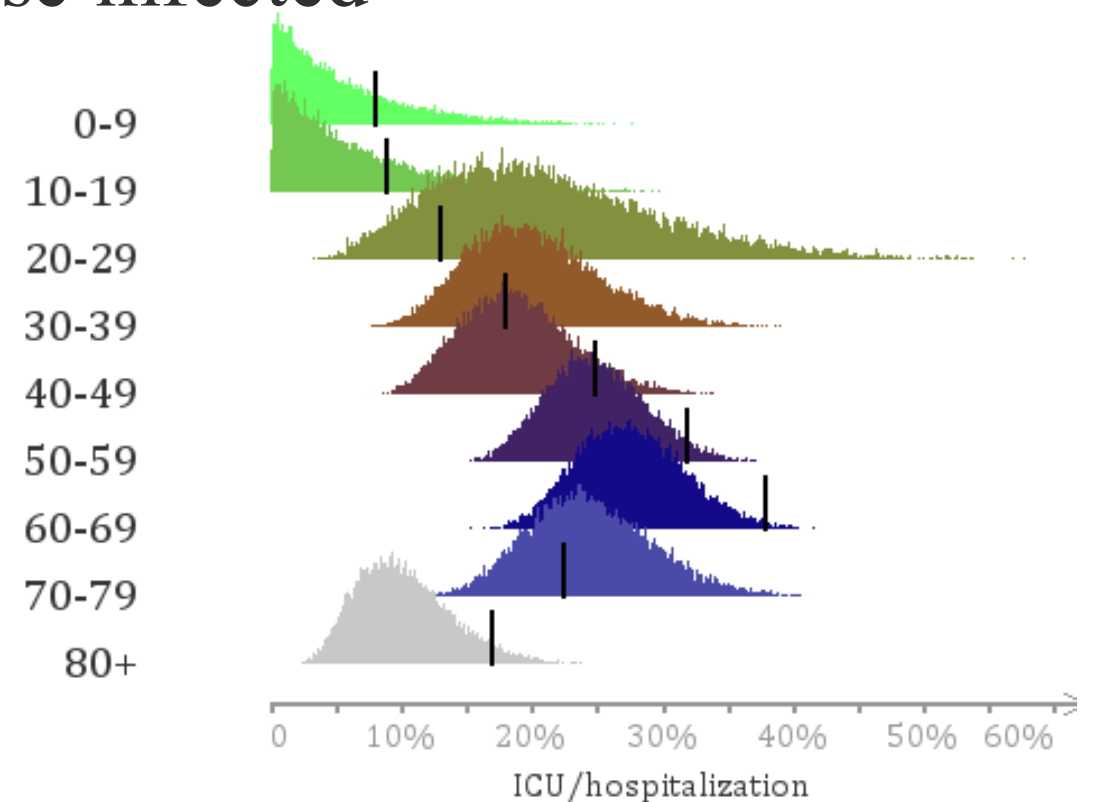
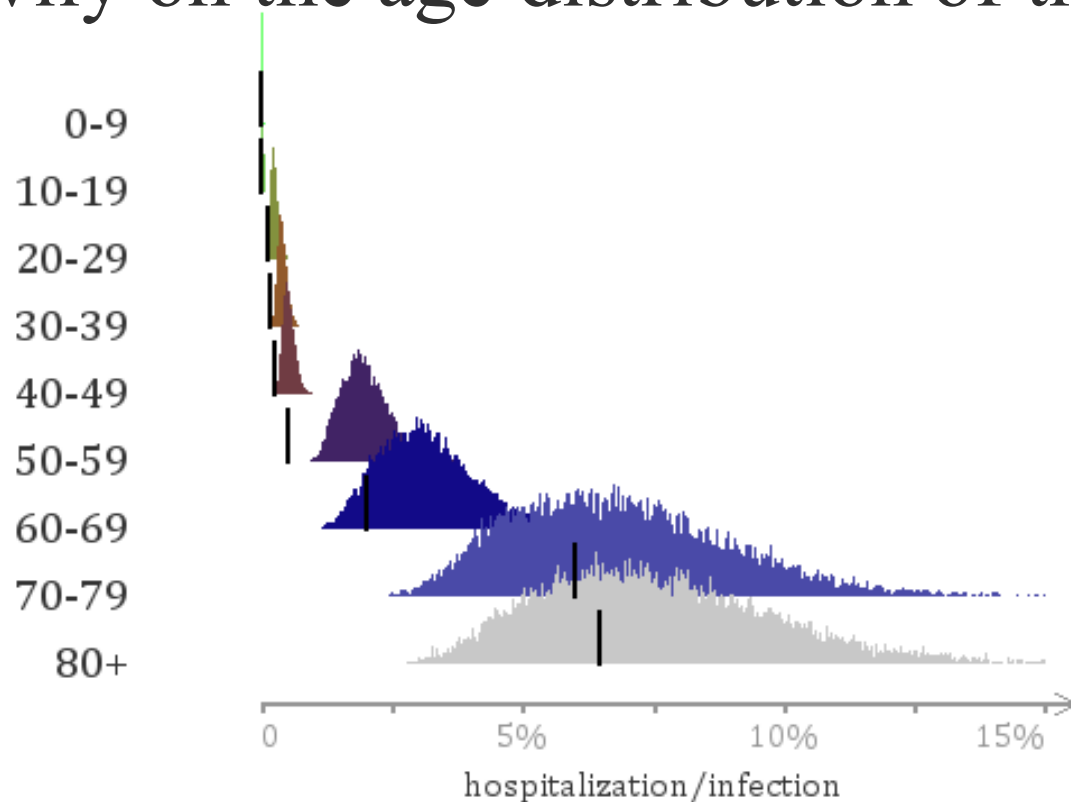
- There is lots of uncertainty in the estimates, mostly due to uncertainty in serology
 - Unknown false-positive and false-negative rates
 - Still need to understand how antibodies develop
- We use hospitalization rates estimated from Wuhan and our own serology. They are consistent, so we have some grounds for being confident.

Hospitalization ratios

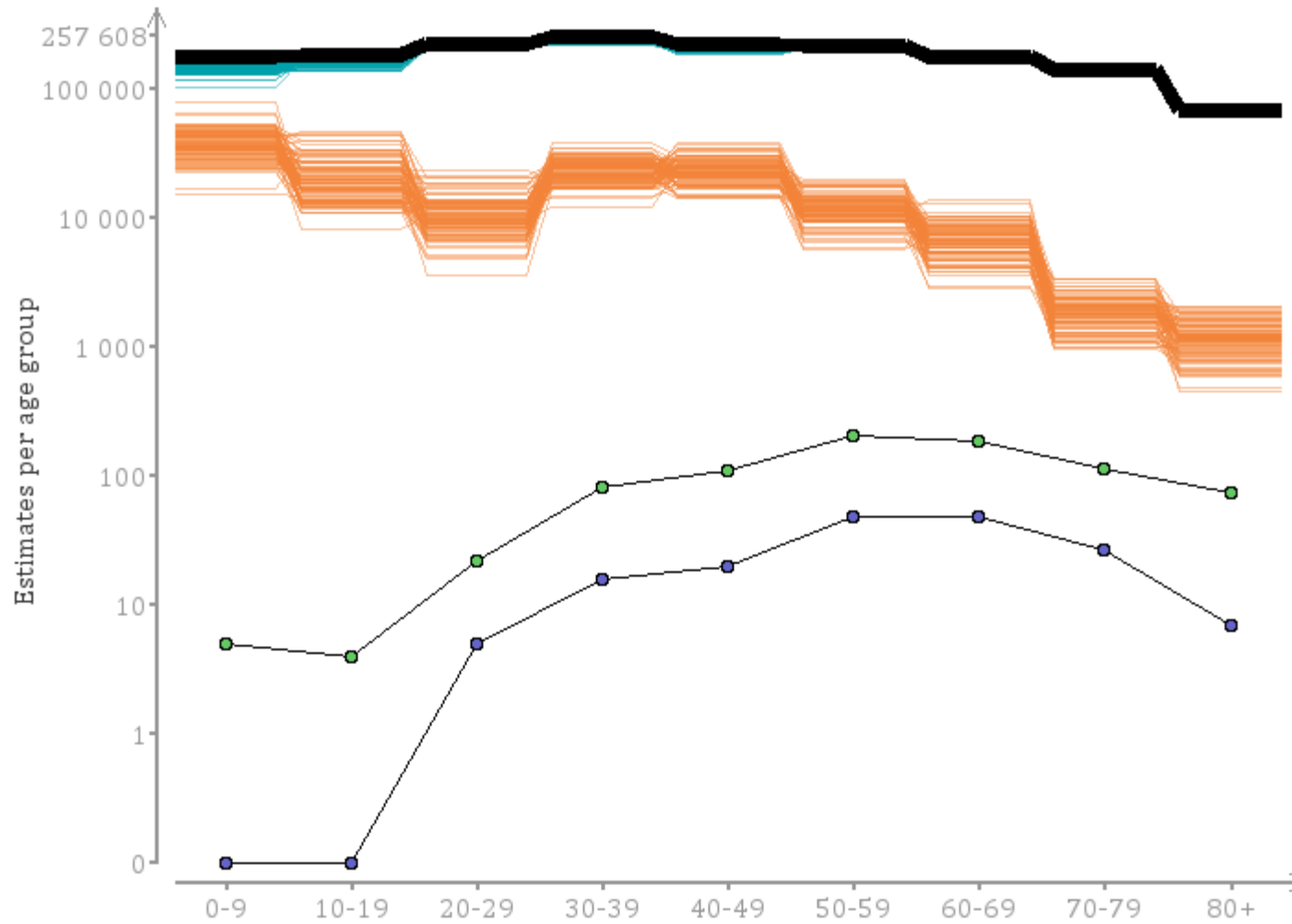


Average hospitalization ratios

With such a difference, average hospitalization rate would depend heavily on the age distribution of those infected



- Pop Size
- total Infected
- Susceptible on 2020-05-08
- hospital burden
- ICU burden

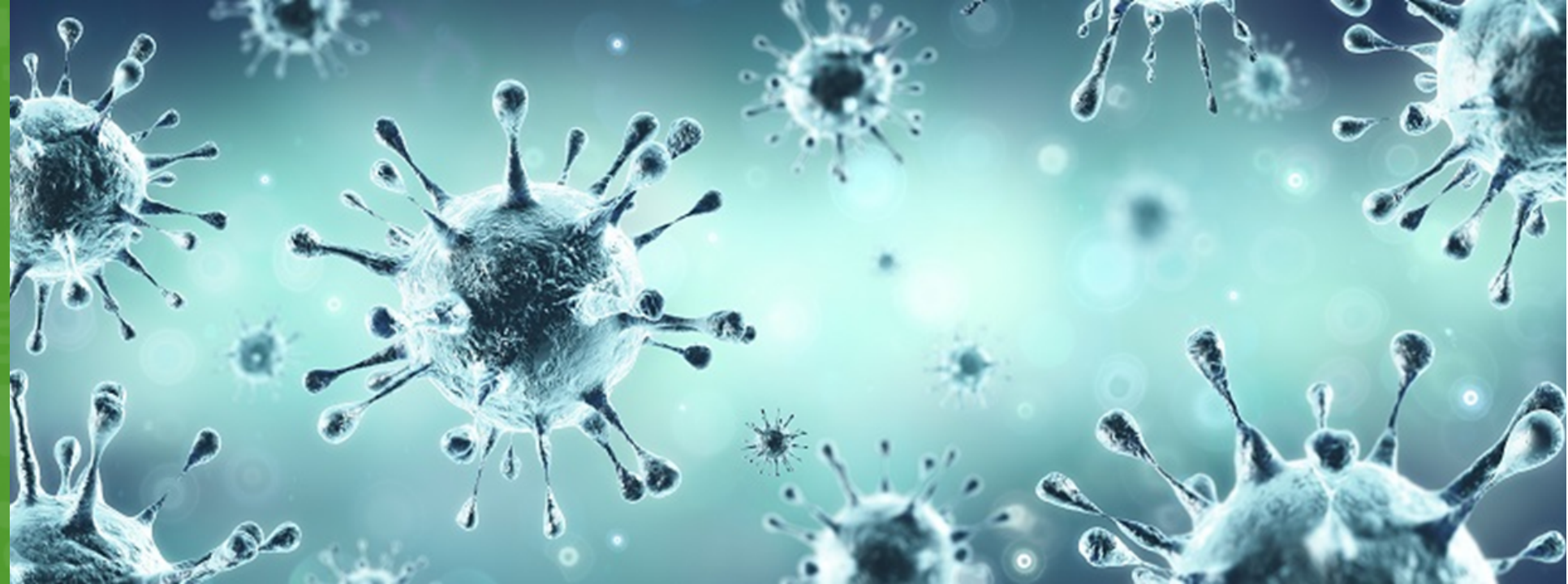


Incidence and severity in children

- Not enough our own data to estimate directly
 - Some external evidence that children are less susceptible, this can be implemented in to the model
- Assumed relationship:
 - Children get more contacts (contact survey)
 - Children are less likely to develop severe symptoms

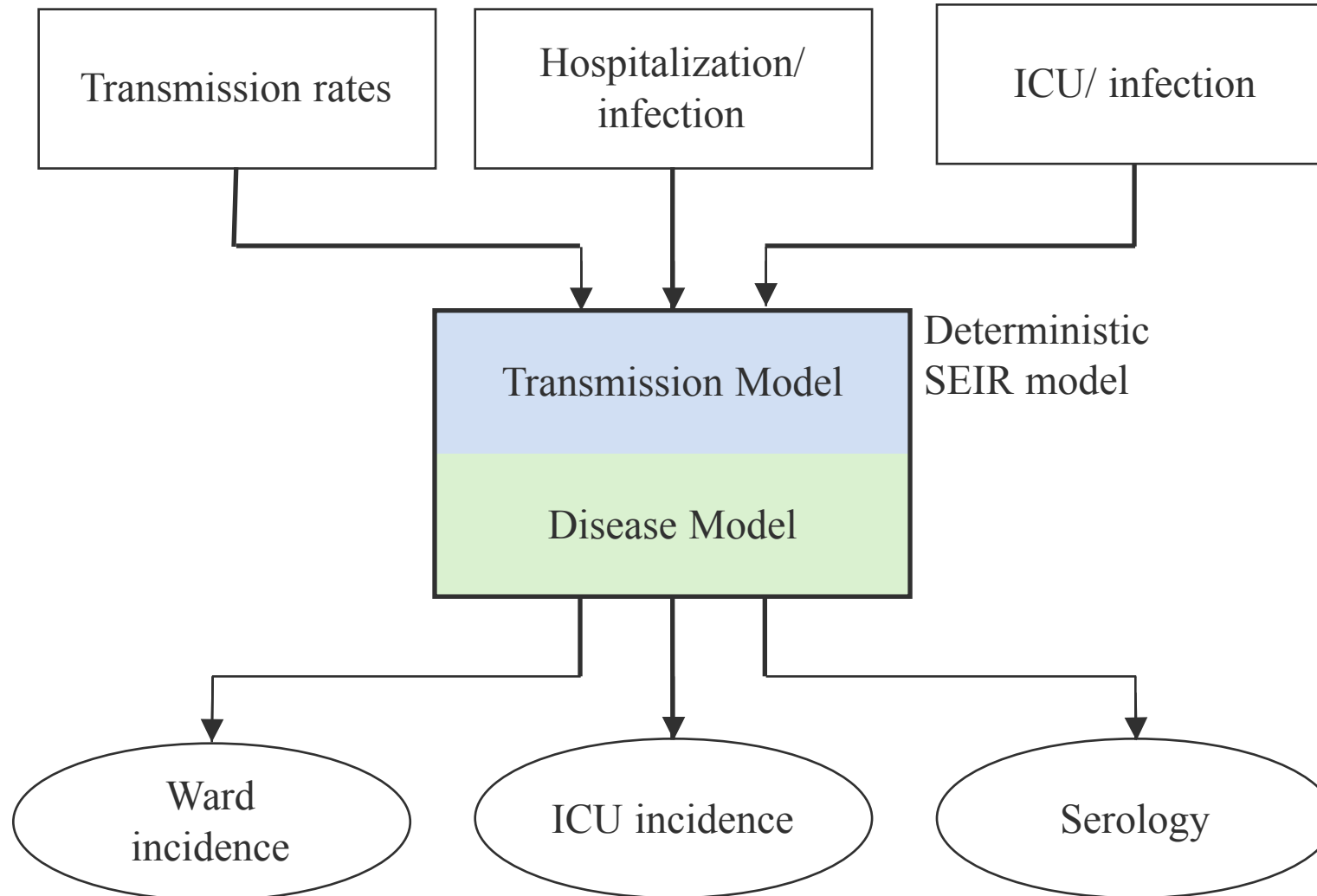
Predicting the future and scaling to other regions

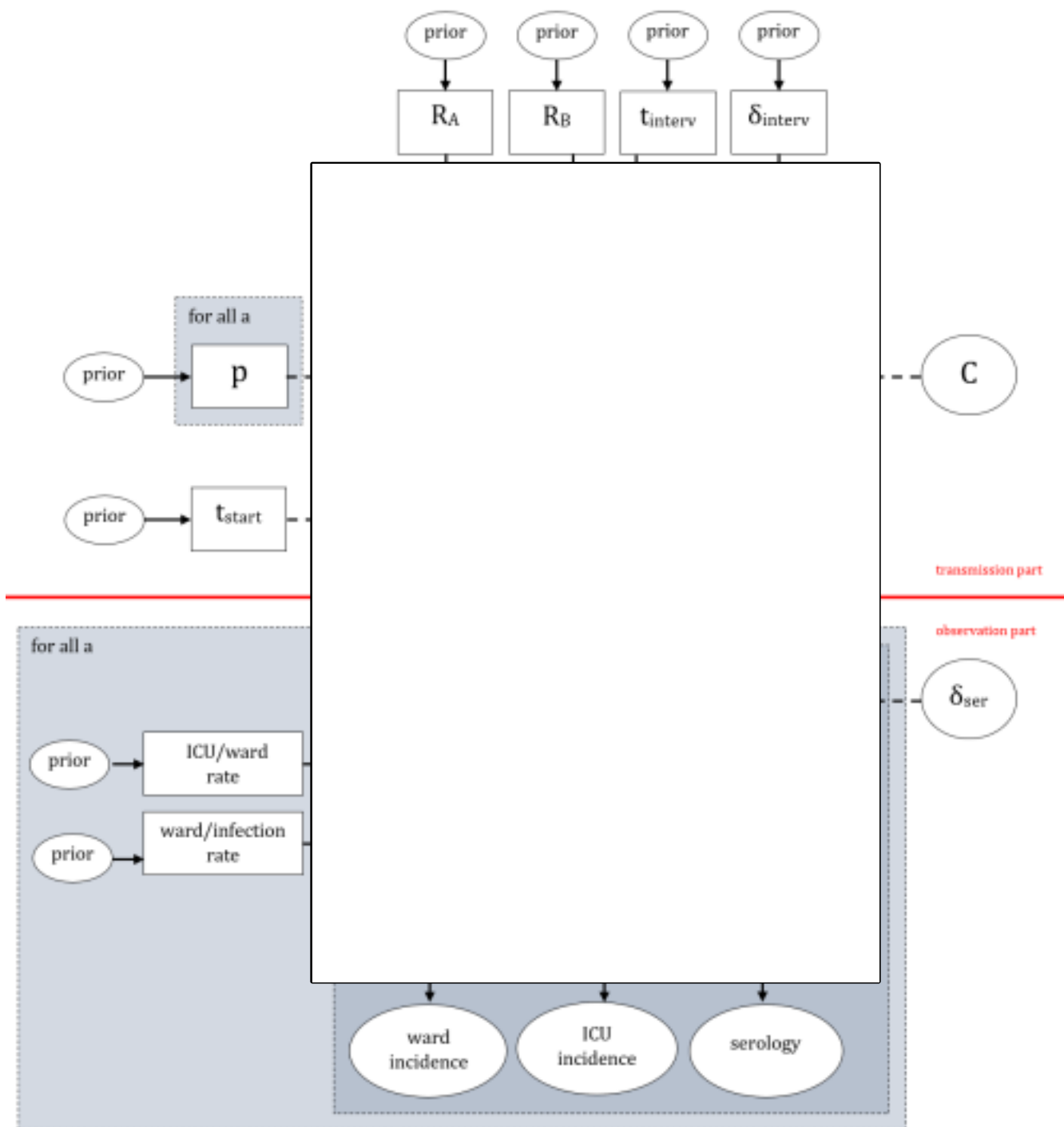
- Currently, model supports one change-point for R_0 , we would need more
- Delay from R_0 change and possibility to estimate it
- R_0 in other regions could be lower

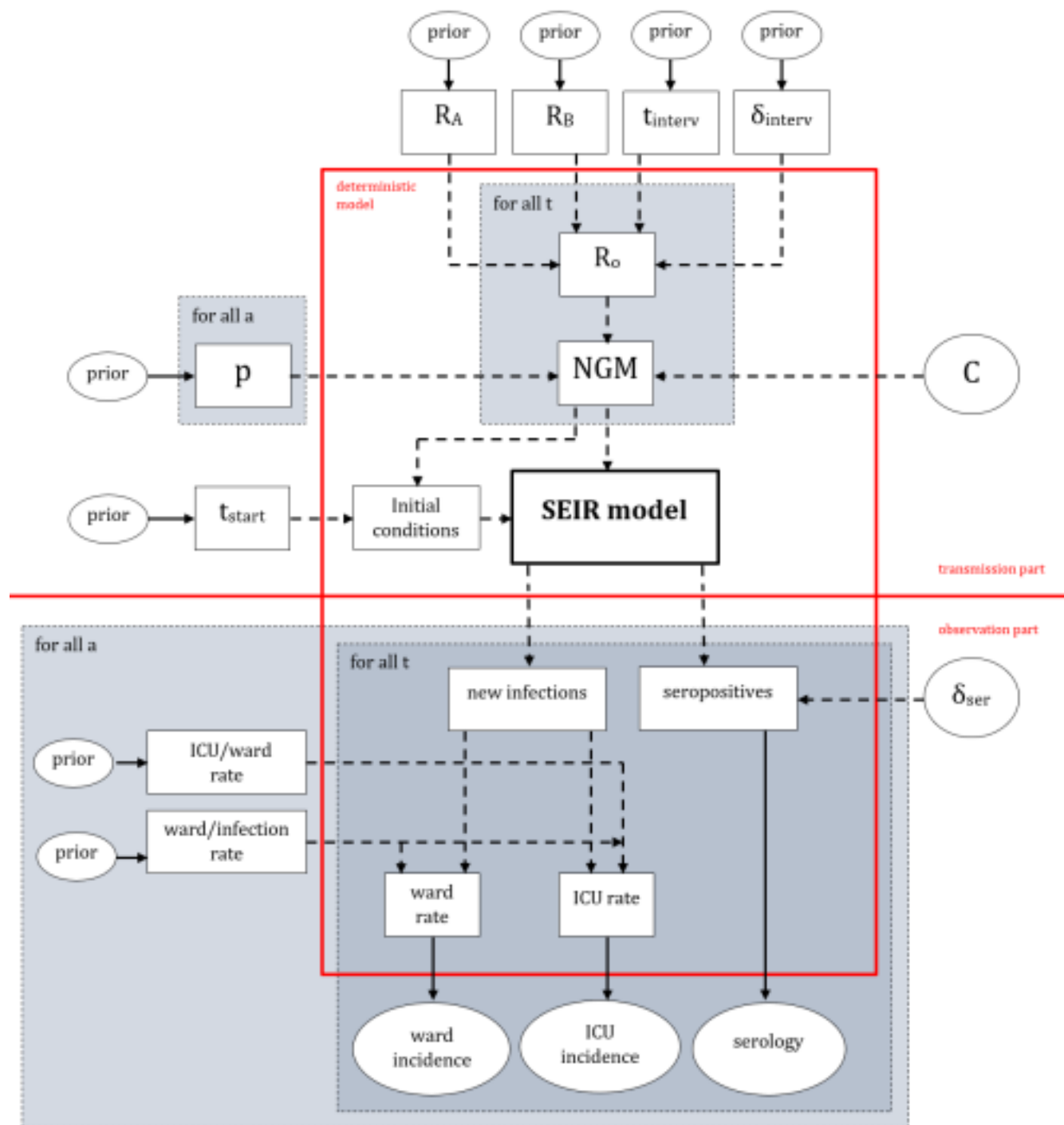


Methods

Model overview

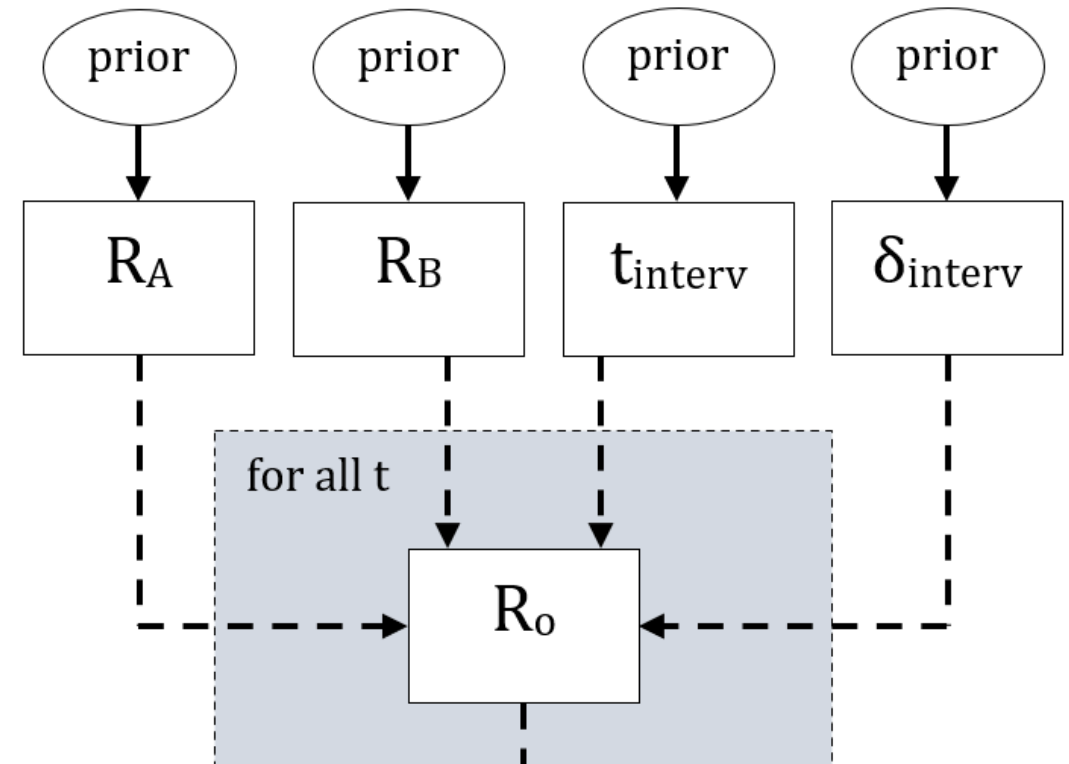






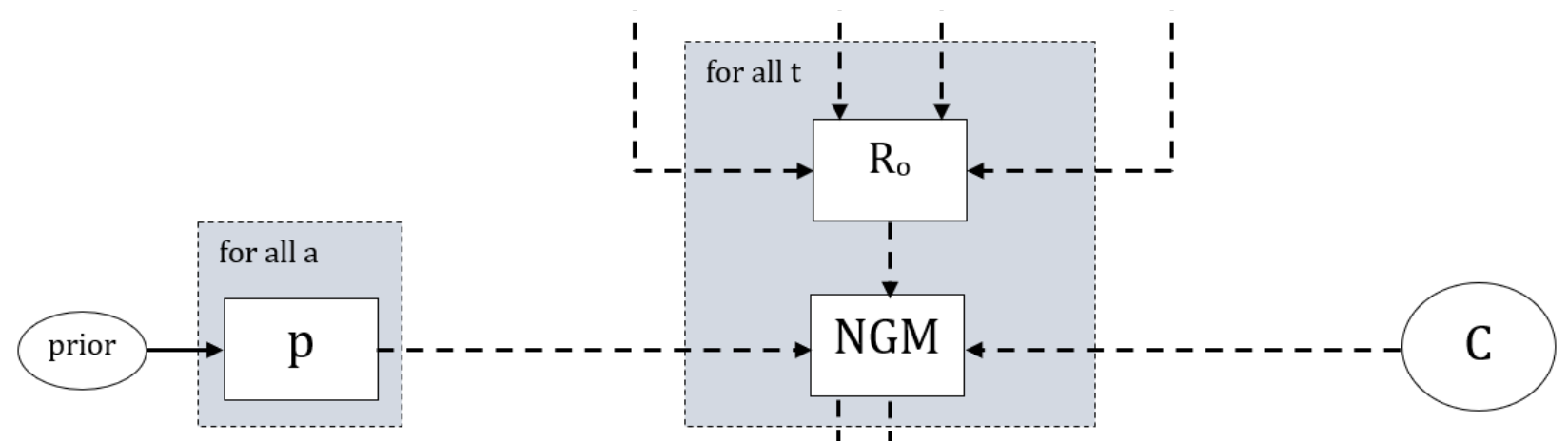
Parameters: time-dependent reproduction rate R_t

- R changes logistically between R_A and R_B
- $R_A \sim 1 + \text{Gamma}(3, 1)$
- $R_B \sim \text{Gamma}(1.4, 1)$
- We use uninformative prior for t_{interv} and δ_{interv}



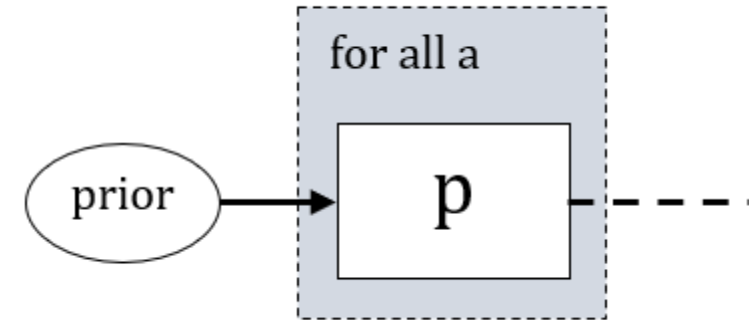
Parameters: Next Generation Matrix

- Time-dependent NGM_t
- Allows to simulate age-dependent control measures
- Depends on R_0 , contact Matrix C_t and age-dependent susceptibilities



Parameters: age-dependent susceptibilities p_a

- Represents biological differences, but also corrects for uncertainty in the contact matrix
- Given uninformative prior
 $\log(p_a) \sim \text{Norm}(0, 1/4)$
- Due to the way we construct NGM, only their relative size matters



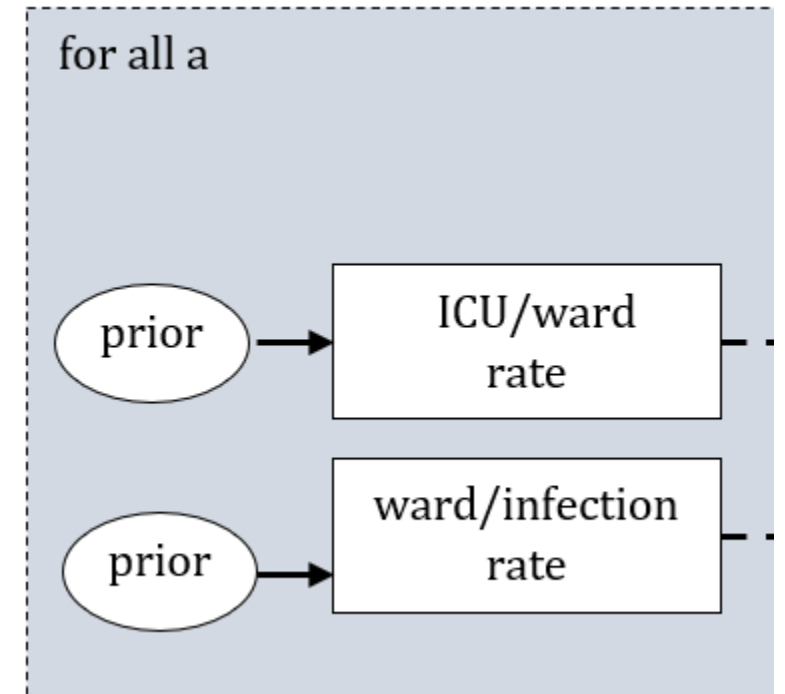
$$NGM_t^* = C_t \begin{pmatrix} p_1 & & \\ & \dots & \\ & & p_9 \end{pmatrix}$$

$$NGM_t = NGM_t^* / l.eigenv(NGM_{t=0}^*) \times R_{0t}$$

$$R_{0t} = R_A \logit \frac{t_{intervention} - t}{\delta_{intervention}} + R_B \logit \frac{t - t_{intervention}}{\delta_{intervention}}$$

Parameters: hospitalization rates

- Ward/infection rate and ICU/ward rate
- Given exponential priors, with the means equal to values from “scenario” model
- Prior is informative, but not binding



Disease model

- Ward and ICU incidence are sampled from Poisson distribution
 - Poisson's rate is a gamma convolution from a number of new cases
- Number of positive sero tests are sampled from Binomial distribution, taking sensitivity and specificity into the account

$$Ward_{at} \sim \text{Poisson} \left(r_a^{ward} \sum_{\tau=0}^{t-1} \text{gamma}(t - \tau | \delta_{mean\ ward}, \delta_{std\ ward})(S_{a,\tau-1} - S_{a\tau}) \right)$$

$$ICU_{at} \sim \text{Poisson} \left(r_a^{ward} r_a^{ICU} \sum_{\tau=0}^{t-1} \text{gamma}(t - \tau | \delta_{mean\ ICU}, \delta_{std\ ICU})(S_{a,\tau-1} - S_{a\tau}) \right)$$

$$SeroPositive_{at} \sim \text{Binomial} \left(SeroTaken_{at}; \frac{R_{a,t-\tau_{sero}}}{N_a} (\delta_{sens} + \delta_{spec} - 1) + (\delta_{spec} - 1) \right)$$

Parameter		Prior
R_A	Basic reproduction number before intervention. Weakly informative prior with mean = 4 and sd = 1	$R_A - 1 \sim \text{Gamma}(3, 1)$
R_B	Basic reproduction number after intervention. Weakly informative prior with mean = 1.4 and sd = 1	$R_B \sim \text{Gamma}(1.4, 1)$
p_a	Relative susceptibilities by age group. Used to scale contact matrix to receive <i>NGM</i> . Only relative values are relevant, p and $p \times \text{const}$ would produce the same <i>NGM</i> .	$p_a \sim \text{logNormal}(0, 1/4)$
t_{start}	Time when the number of infectious reaches 1000.	uninformative
$t_{intervention}$	Time when $R_{0t} = \frac{R_A + R_B}{2}$.	uninformative
$\delta_{intervention}$	Slope of logistic change from R_A to R_B .	uninformative, > 0
r_a^{ward}	Ward hospitalizations/infection ratio. See below for prior means.	$s \sim \text{Exponential}(1/\hat{r}_a^{ward})$
r_a^{ICU}	ICU/hospitalizations ratio.	$s \sim \text{Exponential}(1/\hat{r}_a^{ICU})$

2020, close contacts

	0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80+
0-9	0.7	0.09	0.09	0.7	0.4	0.06	0.08	0.02	0.008
10-19	0.09	0.5	0.07	0.05	0.4	0.1	0.01	0.002	0.004
20-29	0.08	0.06	0.4	0.08	0.03	0.07	0.01	0.005	0.003
30-39	0.6	0.04	0.07	0.4	0.1	0.03	0.02	0.008	0.003
40-49	0.3	0.4	0.03	0.1	0.3	0.1	0.03	0.02	0.03
50-59	0.05	0.08	0.07	0.03	0.1	0.3	0.07	0.01	0.008
60-69	0.06	0.009	0.01	0.02	0.03	0.07	0.2	0.04	0.003
70-79	0.02	0.002	0.007	0.01	0.02	0.01	0.06	0.2	0.02
80+	0.02	0.007	0.007	0.007	0.06	0.02	0.008	0.04	0.04

Parameter	Value	Description
$\delta_E =$	3	Mean time spend in E
$\delta_I =$	5	Mean time spend in I
$\tau_{sero} =$	14	Mean time from Removal to positive sero test
$\delta_{mean\ ward} =$	10	Mean time from Exposed to Ward hospitalization
$\delta_{std\ ward} =$	4	Std of time from Exposed to Ward hospitalization
$\delta_{mean\ ICU} =$	11	Mean time from Exposed to ICU hospitalization
$\delta_{std\ ICU} =$	4	Std of time from Exposed to ICU hospitalization
$\delta_{sens} =$	99.3%	Sensitivity (true positive rate) of a serology test
$\delta_{spec} =$	95%	Specificity (true negative rate) of a serology test

MCMC details

- We fit the model using a pseudo-adaptive MCMC algorithm.
- We allow chains to run for 5,000 warm-up iterations, adapting the proposal distributions to get an acceptance rate of about 1/3.
- We run chains for **100,000** iteration, recording every 10th
- Standard practice

Code details

- Code written in Python 3
- Independent implementation in Matlab
 - Some differences in e.g.
 - Imported cases in February – early March
 - Susceptibility
 - Local ERVA estimates