CoviChron - Chronic illness interaction in transmission dynamics

Otilia Boldea, Ilse Westerhof, Eric Vos, Jantien Backer and Ganna Rozhnova

- preliminary draft -

January 16, 2025

Abstract

In this paper, we model the transmission dynamics of SARS-CoV-2 among different age groups and three risk categories: low, medium and high risk of severe disease, according to the Dutch national classification, for the purpose of assessing different vaccination strategies for the future. The model allows for waning and reinfection, and vaccination and waning of protection against infection and hospitalization due to vaccination, and is fitted to national Dutch data for the first two years of the pandemic on hospitalizations by age and by risk to pin down hospitalization rates, and to seroprevalence survey data by age and risk to pin down susceptibility rates. We compute contact matrices by age and risk, and with the ensuing parameter estimates, evaluate different vaccination strategies: uniform by age, uniform by risk, and differential by age-risk. We demonstrate that vaccinating only high and medium risk groups or older age groups (above 60) is insufficient to eliminate the public health threat unless the vaccination rates are very high (above 75%). We also find that yearly vaccinating older individuals above 60 at a rate above 50% in combination with vaccinating around 25% of medium and high risk groups in younger ages leads to peak hospitalizations that do not overwhelm the hospitals, and is the most cost-effective strategy in terms of number of vaccines used and number of vaccines needed to avoid one hospitalization. This validates the Dutch COVID-19 vaccination strategy which offers vaccines to older individuals above 60, and high and medium risk groups in lower ages, and shows that it may be beneficial for other countries with similar age-risk structure as the Netherlands to adopt this strategy.

1 Introduction

SARS-CoV-2 has become an endemic virus but it is still a public health problem. A SARS-CoV-2 infection shows pronounced differences in individual symptoms and risk, with age and chronic conditions being the most relevant factors for hospitalization and death - [1]. For the foreseeable future, it is critical to estimate hospitalization burden in chronic patients of different ages and to determine how best to protect them from severe COVID-19.

Mathematical modeling is the foremost tool for assessing the impact of interventions. Most COVID-19 modeling studies to date used age as the main risk factor - [2]-[3]. The lack of a framework integrating data on COVID-19 risk due to chronic conditions in a transmission model limits evidence-based policymaking. Modeling studies also have to address drastic changes in SARS-CoV-2 surveillance since large-scale testing was terminated in March 2022. The main SARS-CoV-2 surveillance systems currently in use in the Netherlands are the RIVM-coordinated wastewater surveillance, which cannot be done by age or risk, together with testing selected samples,

and all of these are insufficient for inferring the true infection burden for the future without modeling because transmission occurs between all groups in the population.

Our paper has three key contributions. First, we reconstruct the true infection and hospitalization burden in the Dutch population stratified by age and risk due to chronic conditions in the first two years of the pandemic. For this, we extend existing age-stratified mathematical models that were previously validated against real COVID-19 data - [4],[5], [6]. We fitted our model to a unique combination of existing national databases that were linked in [7] to produce weekly hospitalizations by 11 age categories and 3 risk groups according to the Dutch national classification by RIVM, the Royal Institute for Public Health and the Environment, allowing us to identify hospitalization rates in our model by age and risk. From a survey by RIVM, we computed seroprevalence estimates by age and risk which we then used to identify the susceptibility parameters in our model.

Second, to input a contact network into our model, we used several rounds of a large survey on close contacts before and during the pandemic collected by RIVM, and these two-strata contacts may be a useful contribution in itself, as the risk classifications are similar for other diseases like the flu, and thus can be used to model the spread of other infectious diseases.

Third, we created multiple intervention scenarios in the form of vaccination by age and risk due to chronic conditions, and investigated their effectiveness in simulations, validating them against we validated them against the actual vaccination coverage and hospitalizations in September 2023-February 2024. Based on these analyses, we formulated recommendations on the protection of chronic patients of different ages and the required hospital capacities for new patients.

To our knowledge, there are no dedicated modeling studies of this type for the Netherlands or other countries, besides [8], who consider a stochastic agent-based model with age and co-morbidities, where most disease-level parameters are either calibrated from previous studies or manually adjusted, or fitted to data for the US state of Florida using a non-standard behavioral decision model. Most of the other studies we are aware of use calibrated parameters, for which small changes in the calibration can lead to large changes in outcomes, and unfortunately do not fit existing data over longer periods, and were not rigorously validated against all available evidence - Kissler; [9], [10]. Other studies that do fit models to data did not reconstruct age-specific epidemiology [11] or were limited to specific periods of the pandemic (e.g., waves or variants) - [12], [4], [13] [14], [15]. To our knowledge, this is the first national SIR-type model by age and risk fitted to real data over longer periods, allowing us not only to quantify infection and hospitalization burden by age and risk, but also to investigate interventions targeted at chronic patients - [4],[12], Miura; [16], [17], Schoot Uiterkamp; [18]. In the future, the modeling framework can be adapted understand co-transmission dynamics of SARS-CoV-2 and seasonal influenza and to evaluate the impact of vaccines that are currently under development (e.g. combined influenza-SARS-CoV-2 and RSV vaccines). It can also be adapted to study outbreaks of other respiratory pathogens with pandemic potential (pandemic influenza).

We estimated using weekly data in February 2020-December 2021 the parameters than quantify susceptibility by age and hospitalization rates by age and risk. Our findings demonstrate not only that hospitalizations dramatically increase with age, but that the medium to high-risk groups individuals are two to three times more likely to be hospitalized upon infection, and regardless of whether they are younger or older. We find that the highest hospitalizations/infection is largest not the oldest age group, but in 70-79 years old, followed by 80-89,

and 90+, in all risk groups, due to higher estimated hospitalization rates in 70-79. Because the hospitalization rates in younger individuals at low or medium/high risk of severe disease are much lower than those in older individuals at low risk of severe disease, we find that the disease burden is much larger in older age groups.

This prompted us to contrast different vaccination strategies where the coverage is uniform across the population, and where is differs only by age, only by risk and by age and risk. We validated first our projections using the actual vaccination coverage in Netherlands in Fall 2023- Spring 2024, and could match the true hospitalizations with our projections, then ran several scenarios. We demonstrate that vaccinating only high and medium risk groups or older age groups (above 60) is insufficient to eliminate the public health threat unless the vaccination rates are very high (above 75%). In contrast to [8], we find that, maintaining a high enough vaccination level of older individuals age 60+ and above, there are large gains from vaccinating younger individuals and high and medium risk for severe disease after COVID-19. We find if the vaccination coverage in older individuals above 60 at slightly above 50% (as was the case in Netherlands in 2023-2024), then vaccination 25% of medium and high risk groups in younger ages strongly mitigates the hospitalization burden, leading to peak hospitalizations that do not overwhelm the hospitals. We also find that this is a more cost-effective strategy than vaccinating younger age groups uniformly regarding risk.

The rest of the paper is organized as follows. Section 2 explains the model. Section 3 explains the data we used to fit the model, including an extensive explanations of how to compute two-strata contact matrices from survey data and correct this survey for population non-representativeness. Section 4 explains the estimation method based on the ensemble adjustment Kalman filter. Section 5 discusses the results. Section 6 provides projections related to different vaccination strategies, and Section 7 concludes.

2 Model

Section 2.1 explains the model, and Section 2.2 explains how we modeled time variation in the parameters of the model.

2.1 Model

Let S stands for susceptible, E for exposed, I for infected, H for hospitalized, R for recovered, and N is the population size. A variable $X \in \{S, E, I, R, H, N\}$ with subscript ik denotes the number of individuals in compartment X of risk group i and age category k. We have three risk categories $i \in \{low, medium, high\}$ and 11 age groups: $k \in \{[0,5), [5,10), [10,20), [20,30), \ldots, [80,90), 90+\}$.

A superscript $a = 1, 2$ of variable $X \in \{S, E, I, R, H, N\}$ refers to the set of compartments. The first set of compartments ($a = 1$) refers to individuals who have never been infected or who have had a primary infection but the immunity conferred by this infection has not waned yet. The first set of compartments consists of individuals in the following states: fully susceptible/naive (S^1), latently infected after primary infection (E^1), primary infection (I^1), hospitalized after primary infection (H^1), and fully immune after primary infection (R^1). The second set of compartments ($a = 2$) refers to individuals who have lost immunity after primary infection or who are vaccinated. This set consists of individuals in the following states: partially susceptible after waning

of immunity after primary infection or after vaccination (S^2), latently re-infected (E^2), re-infection or infection after vaccination (I^2), hospitalized after re-infection or infection and vaccination (H^2) and fully immune after re-infection or vaccination (R^2).

If time variation in a parameter is modeled explicitly, it appears as a function of t. Otherwise, its time variation is not modeled but approximated in the estimation step by filtering methods.

[**Step 1: Vaccination**] The day t starts with vaccination (indicated as t_-). Let $V_{ik}(t)$ is the number of vaccinations on day t in age category k and risk group i. The fraction of daily vaccinated people is the number of daily vaccines over the eligible population, which excludes for simplicity infected and hospitalized, but includes waned individuals that have not yet been vaccinated:

$$v_{ik}(t) = \frac{V_{ik}(t)}{S^1_{ik}(t_-) + E^1_{ik}(t_-) + R^1_{ik}(t_-) + W^1_{ik}(t_-)}.$$

The vaccination is assumed to be administered deterministically, proportionally to the fraction of individuals in each disease state that are eligible for vaccination:

$$\begin{aligned}
W_{ik}(t_-) &\leftarrow [1 - v_{ik}(t)]\, W_{ik}(t_-) \\
X^1_{ik}(t_-) &\leftarrow [1 - v_{ik}(t)]\, X^1_{ik}(t_-), \qquad X \in \{S, E, R\} \\
X^2_{ik}(t_-) &\leftarrow X^2_{ik}(t_-) + v_{ik}(t)\, X^1_{ik}(t_-), \qquad X \in \{S, E, R\},
\end{aligned}$$

where W now denotes the pool of waned people who are yet to be vaccinated.[1] Note that we have combined waned and vaccinated individuals into the second set of compartments, denoted by $a = 2$. Therefore, the implicit model assumption is that vaccinated or previously infected individuals whose immunity has waned before being vaccinated benefit from the same reduction in susceptibility, infectivity, and hospitalization rate.

[Step 2: Disease dynamics] Susceptibles become exposed at the rate $\beta^a_{ik}(t)$, and the exposed become infectious at the rate $1/Z$ (after Z days). If not hospitalized, the infected are assumed to recover at the rate $1/D$ (after D days), after which they are no longer infectious. A proportion $\gamma^a_{ik}(t)$ of infected are hospitalized, and they are discharged from the hospital at the rate δ_k which is calibrated; for simplicity, we assume that the hospitalized participate in the contact process but that they are not infectious; for example, they continue to have contacts with the personnel and visitors but cannot infect them due to use of protective measures. The recovered in the first set of compartments wane at the rate $\eta_1(t)$ and enter the susceptibles in the second set of compartments. We assume that waning in the second compartment set appears only after Omicron, so towards the end of 2021. Let W_{ik} be the number of individuals in risk group i and age category k whose immunity from primary infection waned. We keep track of these individuals because they may also be vaccinated later. Let $\mathbf{1}[\mathbf{a} = \mathbf{1}]$ be the indicator function for the compartment set 1, and $\mathbf{1}[\mathbf{a} = \mathbf{2}]$ be the indicator function for

[1] For simplicity, we use $\leftarrow$ and the same t_- to the left and the right to show the move across states induced by vaccination

compartment set 2. The model is given by

$$\begin{aligned}
\frac{dS^a_{ik}(t)}{dt} &= -\beta^a_{ik}(t)\, S^a_{ik}(t) + \mathbf{1}[\mathbf{a}=\mathbf{2}]\,[\eta_1(t)\, R^1_{ik}(t) + \eta_2(t)\, R^2_{ik}(t)] && (1)\\
\frac{dE^a_{ik}(t)}{dt} &= \beta^a_{ik}(t) S^a_{ik}(t) - \frac{E^a_{ik}(t)}{Z} && (2)\\
\frac{dI^a_{ik}(t)}{dt} &= \frac{E^a_{ik}(t)}{Z} - \gamma^a_{ik}(t)\, I^a_{ik}(t) - \frac{1}{D} I^a_{ik}(t) && (3)\\
\frac{dH^a_{ik}(t)}{dt} &= \gamma^a_{ik}(t)\, I^a_{ik}(t) - \delta_k(t)\, H^a_{ik}(t) && (4)\\
\frac{dR^a_{ik}(t)}{dt} &= \frac{1}{D}[I^a_{ik}(t) + \delta_k(t)]\, H^a_{ik}(t) - \mathbf{1}[\mathbf{a}=\mathbf{1}]\,\eta\, R^1_{ik}(t) && (5)\\
\frac{dW_{ik}(t)}{dt} &= \eta\, R^1_{ik}(t), && (6)
\end{aligned}$$

where $\eta_1(t)\, R^1_{ik}(t)$, along with the other additive terms in the model will be subject to numerical integration in the estimation procedure.

Initialization. Let ζ be an initial range of exposed and currently infected for the entire population; we assume half of this initial range are infected, and half are exposed, and they are distributed among risk and age groups based on their relative population size. We initialize the rest of the variables to be zero, besides new infections, which are assumed to be 1/8 of the exposed in the same age-risk category (so, mostly zero due to a small initial range), and susceptibles, which are calculated as remainders.

2.2 Modelling time-varying parameters

Force of infection $\beta^a_{ik}(t)$. The time-dependent force of infection is age- and risk-group-specific and depends on whether individuals are fully susceptible (before primary infection) or partially susceptible (after vaccination or waning of immunity after primary infection). We model the force of infection as a multiplicative function of the constant probability of transmission per contact for the wild-type variant ϵ, and the susceptibility of age groups k relative to age group (60+) $f_{\epsilon,k}$, the increase in the probability of transmission per contact due to variants of concern $voc(t)$, and the time-dependent average number of contacts one individual in age group k and risk group i has per day with all individuals in other age and risk groups multiplied by the proportion of infectious individuals in those age and risk groups $\lambda^a_{ik}(t)$ as follows:

$$\beta^a_{ik}(t) = \epsilon \times f_{\epsilon,k} \times voc(t) \times \lambda^a_{ik}(t). \tag{7}$$

Due to infrequent seroprevalence survey data which does not allow us to pin down the relative susceptibility of all age groups, we group the $f_{\epsilon,k}$ into three age groups, as in [4]: $f_{\epsilon,k} = f_{\epsilon,[0,20)}$ for ages $[0, 20)$, $f_{\epsilon,k} = f_{\epsilon,[20,60)}$ for ages $[20, 60)$, and $f_{\epsilon,60+} = 1$ for the rest.

The $voc(t)$ function is specified as in [6]. As in [6], let $0 \leq p_{I,k}(t) < 1$ be the reduction in susceptibility due to vaccination or previous infection for individuals in the second set of compartments relative to individuals in the first set of compartments. For simplicity, we refer to this parameter as "protection level" against infection due to vaccination or previous infection. For simplicity, we do not further introduce a reduction in transmission

due to previous infection, therefore:

$$\lambda^1_{ik}(t) = \sum_{i'} \sum_{k'} c_{i,k,i',k'}(t) \left[I^1_{i'k'}(t) + I^2_{i'k'}(t) \right] \tag{8}$$

$$\lambda^2_{ik}(t) = \lambda^1_{ik}(t), \tag{9}$$

where $c_{i,k,i',k'}$ are time-varying contact matrices per receiver per initiator, described in the next section.

Hospitalization rates $\gamma^a_{ik}(t)$. Let $0 \leq p_{H,k}(t) < 1$ be the reduction in hospitalization rate due to vaccination or previous infection, conditional on being infected, for individuals in the second set of compartments relative to individuals in the first set of compartments. We refer to this parameter as the "protection level" against hospitalization due to vaccination or previous infection. Therefore:

$$\gamma^2_{ik}(t) = [1 - p_{H,k}(t)]\, \gamma^1_{ik}(t), \qquad \gamma^1_{ik}(t) = \gamma_{ik}. \tag{10}$$

The hospitalization rates in the first compartment set are modelled as $\gamma_{ik} = f_{\gamma,ik} \times \gamma_{low,k}$, with $f_{\gamma,low,k} = 1$, and $f_{\gamma,medium,k} = f_{\gamma,high,k} > 1$. That is, the reference group is the low risk group, and the high and medium risk groups have a proportionally higher hospitalization rate, where the proportion is age specific, but, due to data availability to identify these relative increases in hospitalization rates, we group the estimates into 4 age groups $[0, 40), [40, 60), [60, 80), 80+$. We do not model the time-evolution of γ^1_{ik} but infer it during the estimation for different variants of concern. Parameters $p_{\ell,k}(t), \ell \in \{S, IH\}$ are calibrated based on external data as in [6], and the time-variation in them is due to a waning function, as in [6], but which is now age and risk-specific with mid-point of waning when 50% of vaccines have been administered in each age-risk group 20+, and for children and adolescents, a mid-point equal to those 30-39 years old.

The initial range ζ is fixed at 835, the median posterior estimate from [4].The parameters that will be estimated are $\epsilon, f_{\epsilon,k}, \gamma_{low,k}, f_{\gamma,high,k}$ (3 susceptibility parameters for ages [0,20), [20,60) and 60+, 11 low risk hospitalization rates for all age groups $[0, 5), [5, 10), [10, 20), ..., [80, 90), 90+$, and for high-risk, we group into ages [0-40), [40-60), [60-80) and 80+, to match the available data hospitalization data described in the next section). The waning rate $\eta_1(t)$ is set to 1 year (1/365.5) until arrival of Omicron (date calculated as in [6]), and to 3 months (4/365.5) afterwards. The waning rate in the second compartment set is set to zero until arrival of Omicron and 3 months (4/365.5) afterwards, to match empirical evidence that Omicron was a more immune-escape variant than the previous variants.

3 Data

3.1 Computing contact matrices by age and risk groups

In this section, we describe how we computed the contact matrices based on 11 survey PICO rounds, first pre-pandemic, second in the first lockdown, and last in 2023; the data is described in [19]. We also use data on population levels of risk by age groups from CBS. We performed the following data transformations:

- removed low vaccination region survey participants as their reported contacts may bias the results

- resampled at random at the level of participants and contact age group the number of contacts for those whose community level contacts are above 50, based on the 50%, 90%, 95% quantiles of their contacts in that contact age group. This resulted in maximum 52 contacts for each participant with a given age group.
- in the PICO contact surveys, participants self-report whether they have a particular disease and their BMI. From this, participants are classified into risk and non-risk, and an unambiguous risk is inputted as in [?]
- however, for some survey rounds, participants were not asked questions on their disease status or BMI, so there were remaining participants without information on risk. To not further distort representativeness of the sample in other dimensions, we inputted the risk for these individuals by sampling at random from the risk classifications that were available for all other participants in the same survey round, age group and gender.
- a few of these participants for which risk was inputted by resampling appear in more than one survey round. To achieve an unambiguous risk classification, for these individuals, we again randomly sampled from their inputted risk across rounds once, and considered this as a risk classification for all rounds in which they participated.

For the contact surveys, we have several participants in:

- 11 rounds $r = 0, \ldots, R$
- 10 age groups $k = 1, \ldots, K$
- the exact age of each participant is known, so within each age group, we have $s_k = 1, \ldots, S_k$ years; for the last group, we do not weight by age and therefore $s_K = S_k = 1$
- gender $G = M, F$
- whether the contacts are during the week or weekend; let the weekend indicator be: $W = N, Y$
- whether participants are at risk or not $i = N, Y$

Additionally, we use CBS population data for the first survey round to correct the contacts for representativeness. This data is available:

- by 10 age groups $k = 1, \ldots, K$
- by the exact age within each age group $s_k^* = 1, \ldots, 5$ years
- by gender $G = M, F$ for each exact age.

We obtained $f(k, i)$, the fraction of individuals of age group k and risk i in the total population, assumed constant for now.

The contacts are calculated in the following steps:

1. Count number of participants of a given gender, by week or weekend, of a given age and risk

$$n_{part}(r, k, s_k, G, W, i)$$

2. Calculate fraction of these participants in total participants of a given round:

$$frac_{part}(r,k,s_k,G,W,i) = \frac{n_{part}(r,k,s_k,G,W,i)}{\sum_{k,s_k^*,G,W,i} n_{part}(r,k,s_k^*,G,W,i)}$$

3. Calculate a corresponding fraction but for the entire population of the Netherlands within each round. Population is available by exact age in years and gender $n(r,k,s_k,G)$. Weekdays are weighted 5/7 and week-ends 2/7. From external data obtained in [7], we calculated the population fraction in each age group i of individuals with risk $i = N, Y$ (call it $f(a,i)$), but we do not have it by exact age s_k or by gender W. Then, mimicking how week-days and weekends are weighted in [20] , we have[2]

$$frac_{pop}(k,s_k,G,W,i) = f(k,i) \times g(W) \times \frac{n_{pop}(k,s_k,G)}{\sum_k \sum_{s_k^*} \sum_G n_{pop}(k,s_k^*,G)}.$$

This holds for all age groups but the last, comprising individuals aged 80+, which in the population has too many ages within that age group, but in the survey the participants are only of a few of these ages. If we would also weight by age in this age group, the results may be inaccurate. Therefore, for age group $K = 10$, we only weight by gender and weekend:

$$frac_{pop}(K,s_K,G,W,i) = f(k,i) \times g(W) \times \frac{\sum_{s_k^*} n_{pop}(K,s_k^*,G)}{\sum_{k,s_k^*,G} n_{pop}(k,s_k^*,G)}$$

4. calculate weights of participant population compared to general population:

$$weight(r,k,s_k,G,W,i) = \frac{frac_{pop}(k,s_k,G,W,i)}{frac_{part}(r,k,s_k,G,W,i)}$$

5. join participant dataset with contacts for each participants by round, and add up number of contacts of each participant in age s_k by round, gender, week-day or week-end, and with age group k' (if for a certain k' contacts are not available, we set them to zero). Note that if the contacts' risk profile was known, we could calculate the number of contacts by round, age, gender, week-day or weekend, risk group, with age group k' and risk group i':

$$n_{cont}(r,k,s_k,G,W,i,k',i')$$

We first show how to compute the contact matrices, then explain how we compute $n_{cont}(r,k,s_k,G,W,i,k',i')$.

6. calculate by round the weighted number of contacts each individual of age group k and risk i has with another of age group k' and risk i' divided by the weighted number of participants:

$$m_{obs}(r,k,i,k',i') = \frac{\sum_{s_k,G,W} weight(r,k,s_k,G,W,i) \times n_{cont}(r,k,s_k,G,W,i,k',i')}{\sum_{s_k,G,W} weight(r,k,s_k,G,W,i) \times n_{part}(r,k,s_k,G,W,i)}$$

[2]Note that CBS data is available to calculate this fraction for all survey rounds; however, we calculated it only pre-pandemic and kept it constant across surveys to ensure that contact matrices are valid in the sense that they are mutual, symmetric contacts.

7. calculate the fraction of the population of a given age-group and risk

$$frac_{pop}(k,i) = \sum_{s_k^*,G,W} frac_{pop}(k,s_k^*,G,W,i)$$

8. divide the weighted average number of contacts of age group a and risk K with age group a' and risk K' with the respective fraction of population in each contact age group a' and risk K', to smooth distribution of contacts across age groups and risk(?)

$$c_{obs}(r,k,i,k',i') = \frac{m_{obs}(r,k,i,k',i')}{frac_{pop}(i',k')}$$

9. symmetrize contacts per initiator per receiver, assuming reciprocity of contacts

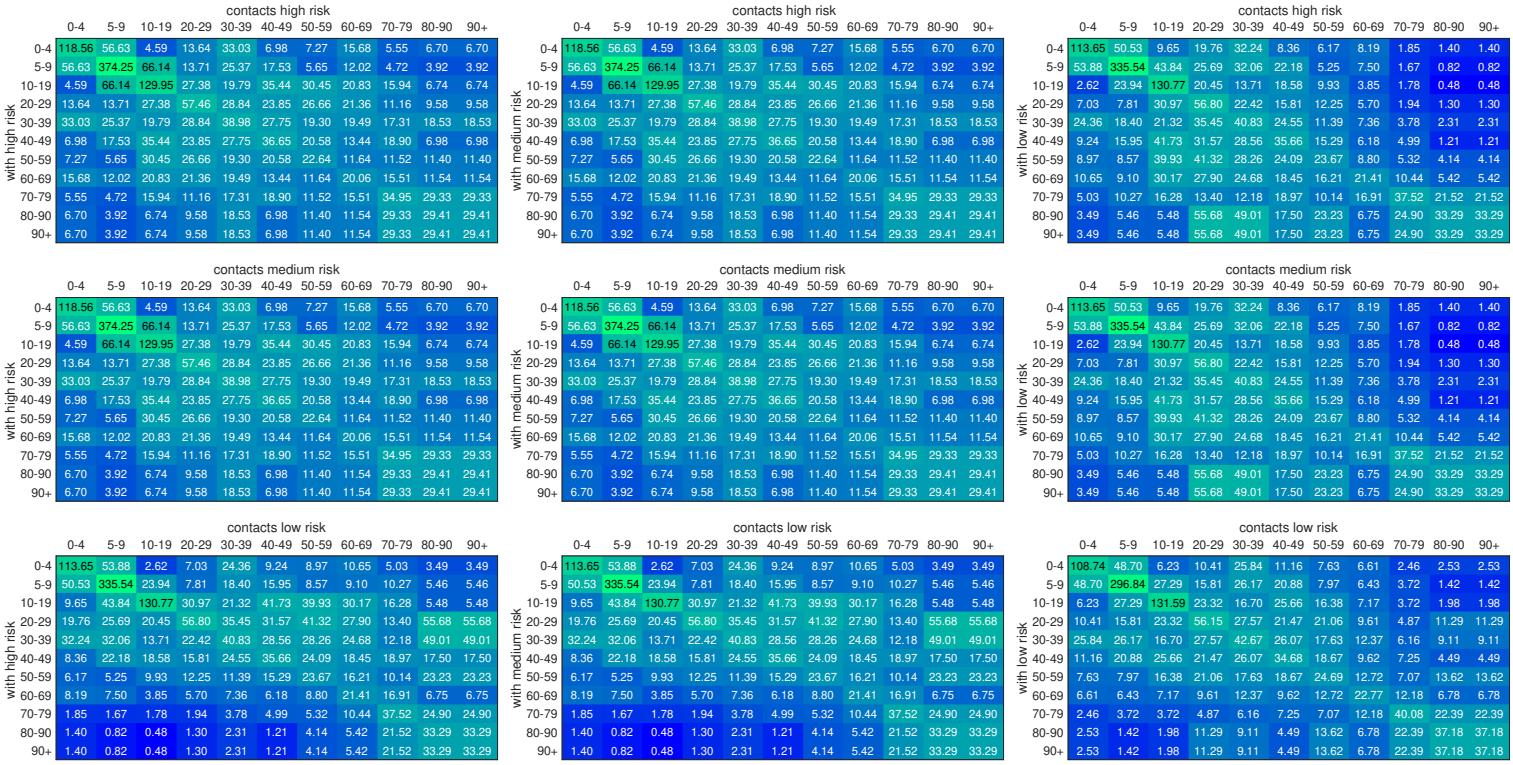
$$c_{sym}(r,k,i,k',i') = \frac{c_{obs}(r,k,i,k',i') + c_{obs}(r,k',i',k,i)}{2\sum_{k,s_k,G} n_{pop}(k,s_k,G)}$$

Computation of $n_{cont}(r,k,s_k,G,W,i,k',i')$ depends on how the populations are assumed to mix. We assume proportionate mixing, proportional to population factions of a given age group and risk, thus:

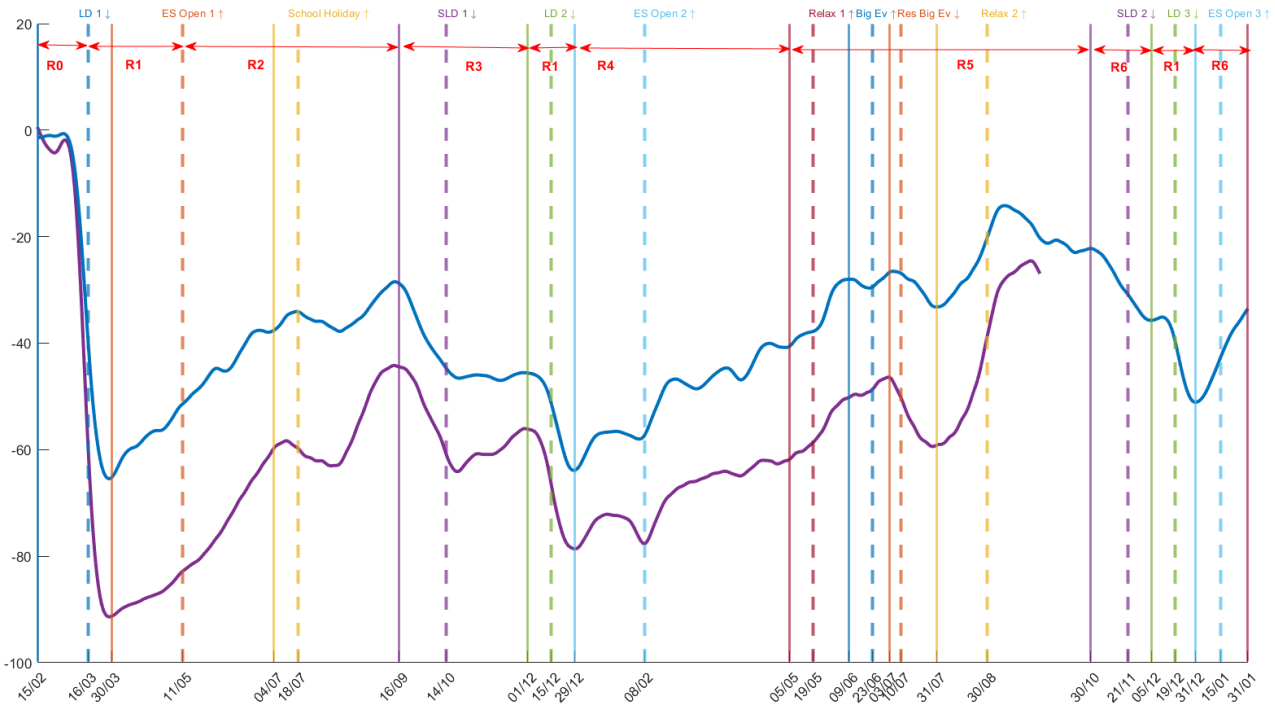
$$n_{cont}(r,k,s_k,G,W,i,k',i') = f(k',a') \times n_{cont}(r,k,s_k,G,W,i,k'),$$

where $n_{cont}(r,k,s_k,G,W,i,k')$ counts the contacts of all individuals in round r, age group k, age s_k, gender W, risk i, with individuals of age group k' and risk i'.

Now $c_{sym}(r,k,i,k',i')$ is only by no risk and risk, and does not feature the age categories $[80,90)$ and $90+$ separately, while our data on population and vaccination is by low, medium and high risk, and our hospitalization data is separate for age categories $[80,90)$ and $90+$. We therefore assume that the contacts per initiator per receiver are the same for medium and high risk as those at risk, the same for low risk as for no risk, and the same for age categories $[80,90)$ and $90+$, obtaining the final contacts used in the model $c(r,k,i,k',i')$. Below is a plot of these contacts multiplied by the total population, for the first survey round, pre-pandemic.



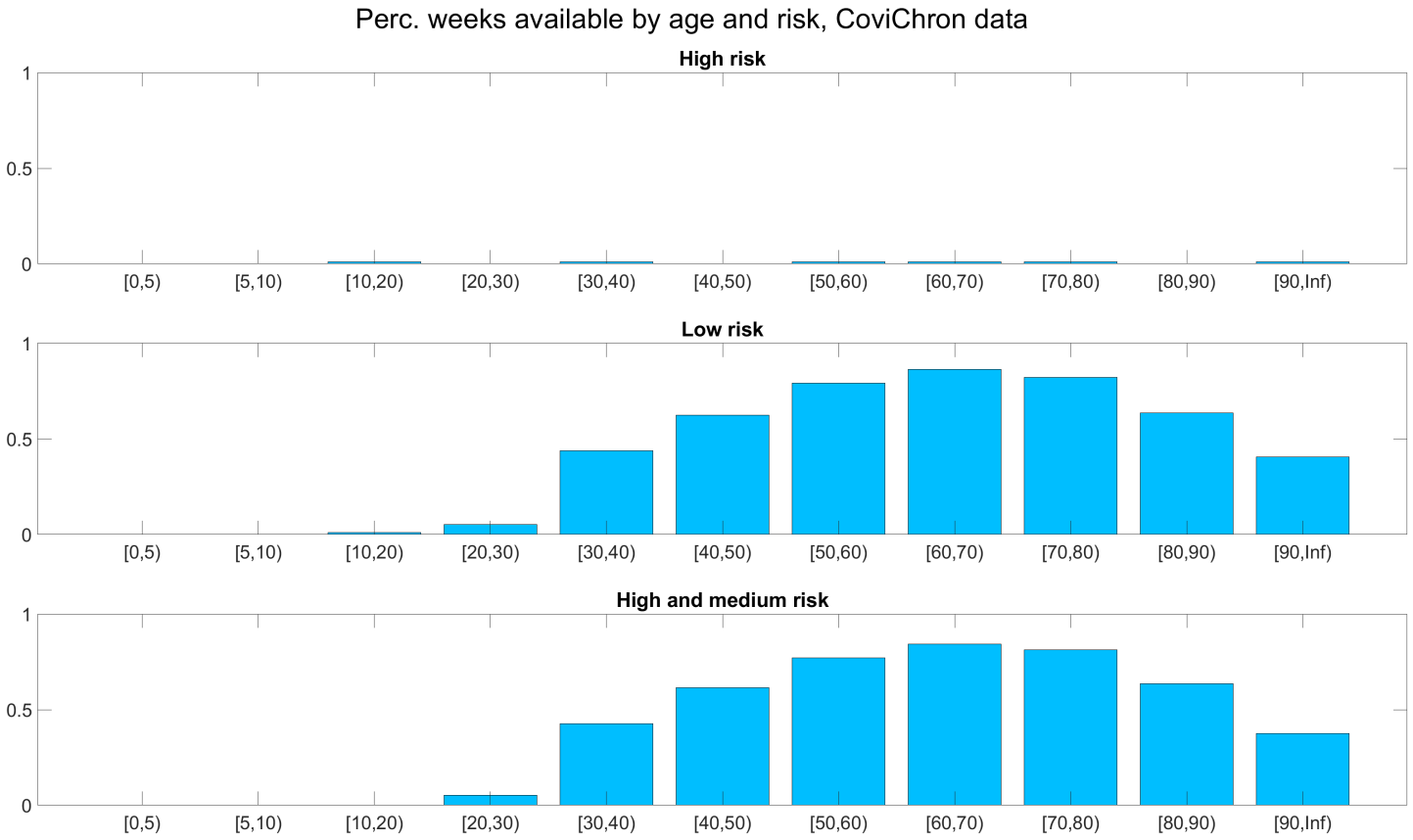
The surveys come with information on the median time when they were conducted, and we just switch between these contacts suddenly as indicated in Figure 2 below, where R stands for survey round, and R0 stands for pre-pandemic survey data, and the rest of the figure plots two mobility indexes (Google and train mobility for the Netherlands), as well as timing of vacations and non-pharmaceutical interventions for the first two years with dotted lines, and solid lines indicate choices of when a regime may start or end; data and plot if adapted from the Supplement in [6]. Because for the second lockdown there is no survey data, we simply input the contacts from the first lockdown, and multiply them by a factor of 1.3 to adjust for reduced compliance in the second lockdown compared to the first (1.3 is obtained by a grid search to ensure a reasonable fit to hospitalization data).



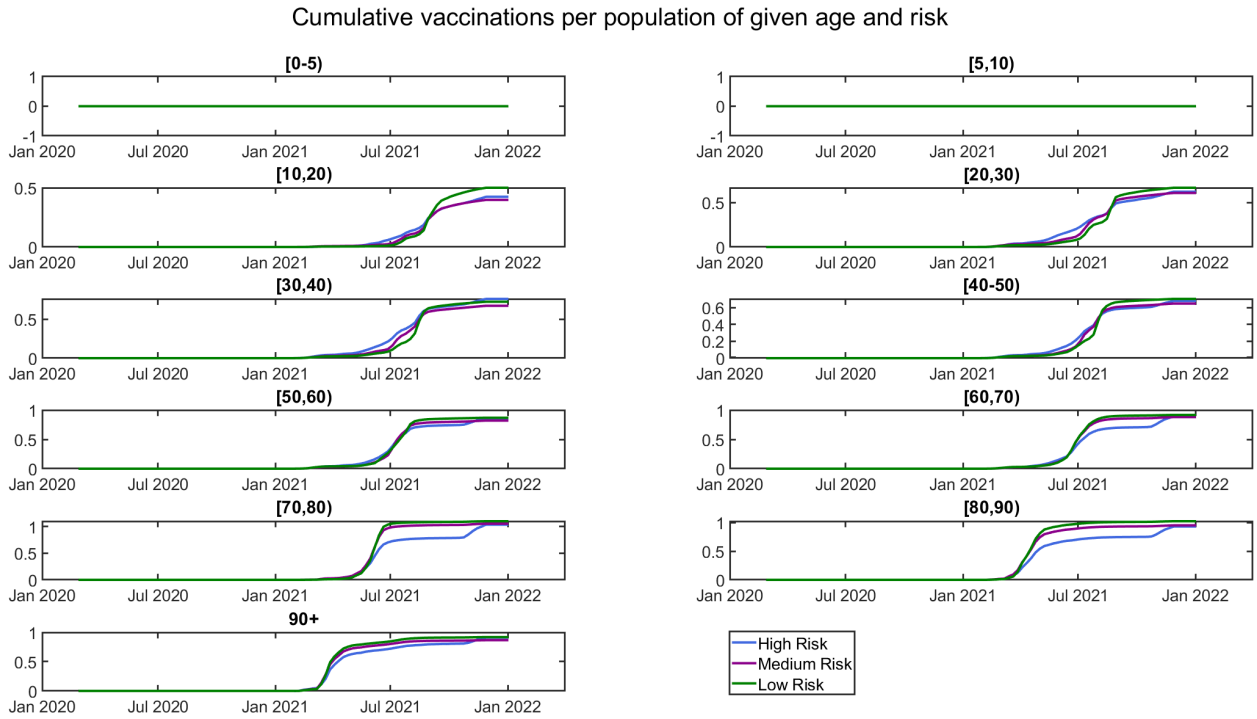
3.2 Other data used for model fitting

Hospitalizations. The data we computed in [7] linking drug prescription data with CBS hospitalization data allows us to access weekly hospitalization data for high, medium and low risk groups, by age groups $[0,5), [5,10), [10,20), ..., [80,90), 90+$ from the beginning of the pandemic (week ending in March 2, 2020) until the end of December 2021, for a total of 96 weeks. However, due to confidentiality reasons, the data is reported only if there are more than 10 hospitalizations in each age-risk group, else data is combined into multiple risk groups and sometimes also across ages. This yields a lot of missing data, as shown in the figure below. Therefore, we combine the hospitalization data by risk into high and medium risk, and into age categories [0,40), [40,60), [60,80) and 80+, and we supplement this with weekly hospitalizations only by age in $[0,10), [10,20), \ldots, [80,90), 90+$, data that is available from the Dashboard.[3] This data ($n = 1344$) will help us identify hospitalization rates for low risk in all age groups (except perhaps $[0,5)$ and $[5,10)$ separately), and for high/medium risk we estimate a relative increase that is grouped exactly as the data, into 4 age categories.

[3] See `https://data.rivm.nl/covid-19/COVID-19_ziekenhuis_ic_opnames_per_leeftijdsgroep.csv` and `https://data.rivm.nl/covid-19/COVID-19_ziekenhuis_ic_opnames_per_leeftijdsgroep_tm_03102021.csv`.



Vaccination data. The vaccination data is available weekly for the same period as the hospitalization data from [7] but this time by high, medium and low risk groups, in all ages. Because our model is run daily - but fitted to weekly data - we distribute the vaccines within a week equally across the 7 days. A plot of the cumulative vaccinations per population in different age-risk groups is below.



Seroprevalence survey data. This data comes in 6 PICO survey rounds, and after correcting for sample representativeness, we computed seroprevalence estimates in all age categories, and above 40, also by no risk or risk, with the same classification as for the contact matrix data, in survey rounds 2 and 4-6. In survey round 1 and 3 the data is available in all age categories but not by risk. During survey 5, a large share of the sample participants are vaccinated with one dose, and therefore are reported to have antibodies already for the first dose of vaccine. However, our model can only consider vaccination as complete after the first two doses in the case of Pfizer, Moderna and Astra-Zeneca vaccines, and after one dose in the case of the Janssen vaccine, because we do not have vaccination data broken down by two doses. Therefore, we do not force the model to equal the seroprevalence estimates, and we do not expect a good fit to seroprevalence data in survey 5. In total, we use $n = 90$ seroprevalence data points to pin down three susceptibility parameters.

4 Estimation Method

The model represents the transition equations for a state-space model. Let $\mathbf{Y}, \mathbf{X}$ be the vector of observed, respectively unobserved state variables, with typical elements $Y(t)$ and $X(t)$ stacked in order, and $\mathbf{\Theta}$ the vector of estimated parameters.

$$\begin{bmatrix} \mathbf{Y}(t) \\ \mathbf{X}(t) \end{bmatrix} = \mathbf{F}_t \left(\begin{bmatrix} \mathbf{Y}(t-1) \\ \mathbf{X}(t-1) \end{bmatrix}, \Theta \right), \tag{11}$$

with $\mathbf{F}_t$ a non-linear vector-valued function representing the SEIRHS model, which unlike [21] and[6], we assume to be deterministic given the parameters, and a Runge-Kutta RK4 method is employed for numerical integration for each ensemble member $e = 1, \ldots, \mathcal{E}$, where ensemble members are determined by the priors on the parameters, assumed stochastic. As a result, we obtain ensemble members with typical elements $X^e(t), Y^e(t)$. The measurement equations are:

$$Y^{obs}(t) = Y(t) + \epsilon_Y(t), \tag{12}$$

where the observed variables $Y(t)$ are:

- weekly new hospitalizations by age, and by high risk grouped into 4 age categories
- survey data, available only in 6 surveys over the two years, used at median survey data minus 14 days to allow for seroconversion.

The term $\epsilon_Y(t)$ is a mean-zero error ([22],[21]), assumed independent across series and time. We chose the variance of hospitalizations to be proportional to the mean at a given time and for a particular series. For seroprevalence estimates, we also had 95% confidence intervals, and from those we computed a variance assuming a distribution of two mixtures of normals, one for each end-point of the confidence interval, to cover this interval.

We updated the system via the ensemble adjustment Kalman filter (EAKF). This filtering method was introduced in [23] as an extension to the ensemble Kalman filter to update high dimensional state variables that are partially unobserved in atmospheric science and was subsequently used in [6], [11] and [21]. Recently, [22] showed that despite nonlinearity, the method converges in mean square error to the true process that drives the observations, for fixed ensemble size, uniformly in finite time. In the updates, $\mathcal{E}$ ensembles draws are individually propagated forward by means of closed-form updates, based on the normal distribution approximation. The observable ensemble members are updated based on their own observations, sequentially for each time period. Let $\mathcal{Z}^{e,post}(t-1)$ be an ensemble member posterior from the last period, where $e \in \{1, \ldots, \mathcal{E}\}$, and where $\mathcal{Z} \in \{Y, X\}$ can be unobserved or observed. Each member is variance inflated by a factor $\lambda > 1$ to prevent the filter from collapsing to a posterior with a single value:

$$\mathcal{Z}^e(t) = \overline{\mathcal{Z}}^{post}(t-1) + (\lambda - 1)\left(\mathcal{Z}^{e,post}(t-1) - \overline{\mathcal{Z}}^{post}(t-1)\right), \tag{13}$$

where $\overline{\mathcal{Z}}^{post}(t-1)$ is the mean over ensemble members $\mathcal{Z}^{e,post}(t-1)$. Then $\mathcal{Z}^e(t)$ completes the transition equation (SEIRHS model) and yields the prior $\mathcal{Z}^{e,prior}(t)$, $\mathcal{Z} \in \{Y, X\}$. Let $r_Y(t) = \sigma^2_Y(t)/(\sigma^2_Y(t) + \hat{\sigma}^2_{Y,prior}(t))$, where $\hat{\sigma}^2_{Y,prior}(t) = \widehat{Var}(Y^{e,prior}(t))$ is the sample variance of $Y^{e,prior}(t)$, and $\overline{Y}^{prior}(t)$ is the sample mean over $Y^{e,prior}(t)$. The observable state updates are:

$$Y^{e,post}(t) = r_Y(t)\,\overline{Y}^{prior}(t) + [1 - r_Y(t)]\,Y^{obs}(t) + \sqrt{r_Y(t)}\,[Y^{e,prior}(t) - \overline{Y}^{prior}(t)]. \tag{14}$$

Those variables X that are defined as "neighbors" of the variable Y are then sequentially based on their ensemble covariance with Y, where the neighbors are defined below. Let n_Y be the index of a neighbor of state variable Y, and X_{n_Y} a corresponding unobserved neighboring state variable. Let $\sigma^2_{X_{n_Y}}(t) = \widehat{Var}(X^{e,prior}_{n_Y}(t))$ and $\overline{X}^{prior}_{n_Y}(t)$

be the variance and the mean over ensemble members $X_{n_Y}^{e,prior}(t)$. Then, for all neighboring state variables,

$$X_{n_Y}^{e,post}(t) = X_{n_Y}^{e,prior}(t) + \frac{\widehat{Cov}(X_{N_Y}^{e,prior}(t), Y^{e,prior}(t))}{\sigma^2_{X_{n_X}}(t)}[Y^{e,post}(t) - Y^{e,prior}(t)]. \tag{15}$$

As in [6], we have several unobserved state variables per age/province ($S^a_{ik}, E^a_{ik}, I^a_{ik}, H^a_{ik}, R^a_{ik}, N^a_{ik}$, $a = 1, 2$). In principle, these could be updated at each point with the new parameter posteriors, but then an iteration needs to be employed to ensure accuracy of each time-update. Instead, we updated both parameters and state with the EAKF formula. From these unobserved states, three of these are calculated as remainders due to three equalities. One is explicit, the sum of the population in the two sets of compartments is constant. The other two are implicit: the newly recovered, $Rnew^a_{ik}(t) = \frac{1}{D}(I^a_{ik}(t) + \delta_k(t)H^a_{ik}(t)$, plus the newly waned individuals from the first set of compartments, $Wnew_{ik}(t) = \eta_1(t)R^1_{ik}(t)$, fully define the evolution of recovered on both compartment sets, as well as the susceptibles in both compartment sets given the other states. Therefore, the unobserved state variables we update based on observed state variables are $Y \in \{E^a, I^a, H^a, Rnew^a, a \in \{1, 2\}, Wnew\}$.

Parameters are estimating by augmenting the state space with the set of parameters, treated exactly as unobserved state variables with a constant transition. The second set of compartments is initialized at zero for all state variables.

For each age-risk category i, k and observable variable per age-risk category, its neighbors are defined as unobservables of the same age and the next age category, for all risks. For $t = 0$, the priors are obtained by drawing $\mathcal{E} = 300$ ensemble uniformly on a Latin Hypercube. For $t \geq 1$, the posteriors at $t - 1$ become the priors at t, after being variance inflated and completing the transition equation. The algorithm is as follows

1. retrieve posteriors at week $t - 1$
2. variance inflate each ensemble member according to equation (13), but only referring to states and parameters that are updated (recall that some parameters are only updated in particular regimes)
3. run the SEIRHS model for 7 days, integrating daily, and obtain the weekly priors for each variable
4. update observable state variables according to equation (14)
5. for each observed variable Y, update neighboring unobservable states based on their covariance with observables (equation (15)) (neighbors are the the next adjacent age category and all risk categories), and all parameters
6. do so for the two hospitalization data series and the seroprevalence estimates when available
7. update other states as remainders
8. if parameter posteriors are out of their prior bounds, redistribute them with a small variance (0.001) inside the bounds
9. $t = t + 1$; stop if $t = T$.

All model analyses were performed in MATLAB 2023A.

5 Estimation Results

5.1 Parameter estimates

The parameter posteriors at the end of the sample over their prior range are in the figures below. In the figure on hospitalization rates in low risk groups, except for the age category 90+, the rest of the parameters are reported relative to this age category. We can see a steady decrease in hospitalization rates relative to 90+ for younger age categories. It is interesting to note that despite very young children age 0-1 being typically hospitalized at a higher rate due to precaution or high fever, the hospitalization rates in low risk groups age $[0, 5)$ is similar to that of age $[10, 20)$. We also note that the hospitalization rates in $[80, 90)$ are higher than those in 90+ in the low risk groups, and this is possibly due to a policy introduced after the first wave to not hospitalize individuals in nursing homes and treat them on site, and not because of lower severity of disease for 90+. We also see that the higher risk groups have a relative hospitalization rate that is double or triple that of the smaller age group, even in younger individuals age $[0, 40)$, suggesting that it is necessary to shield the age groups from severe disease by vaccinating them regularly.

Figure 1: Posterior density of susceptibility parameters over prior range

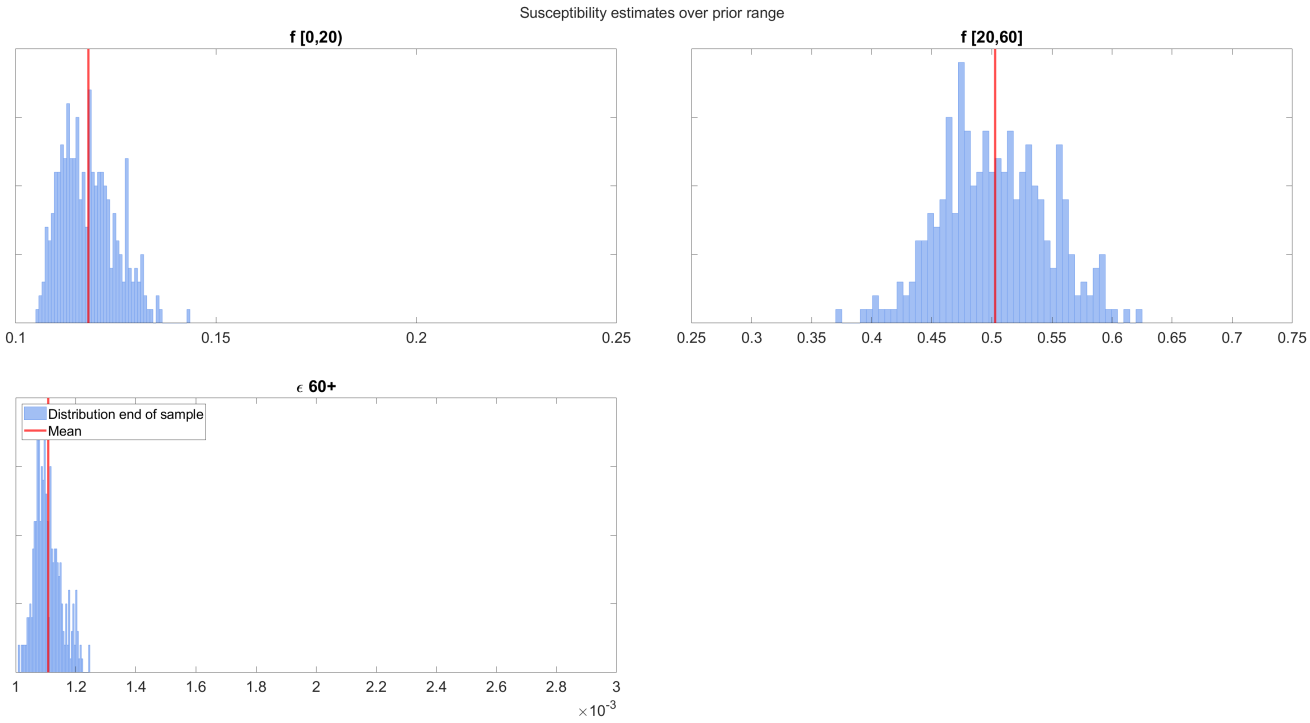


Figure 2: Posterior density of low risk hospitalization rates over prior range

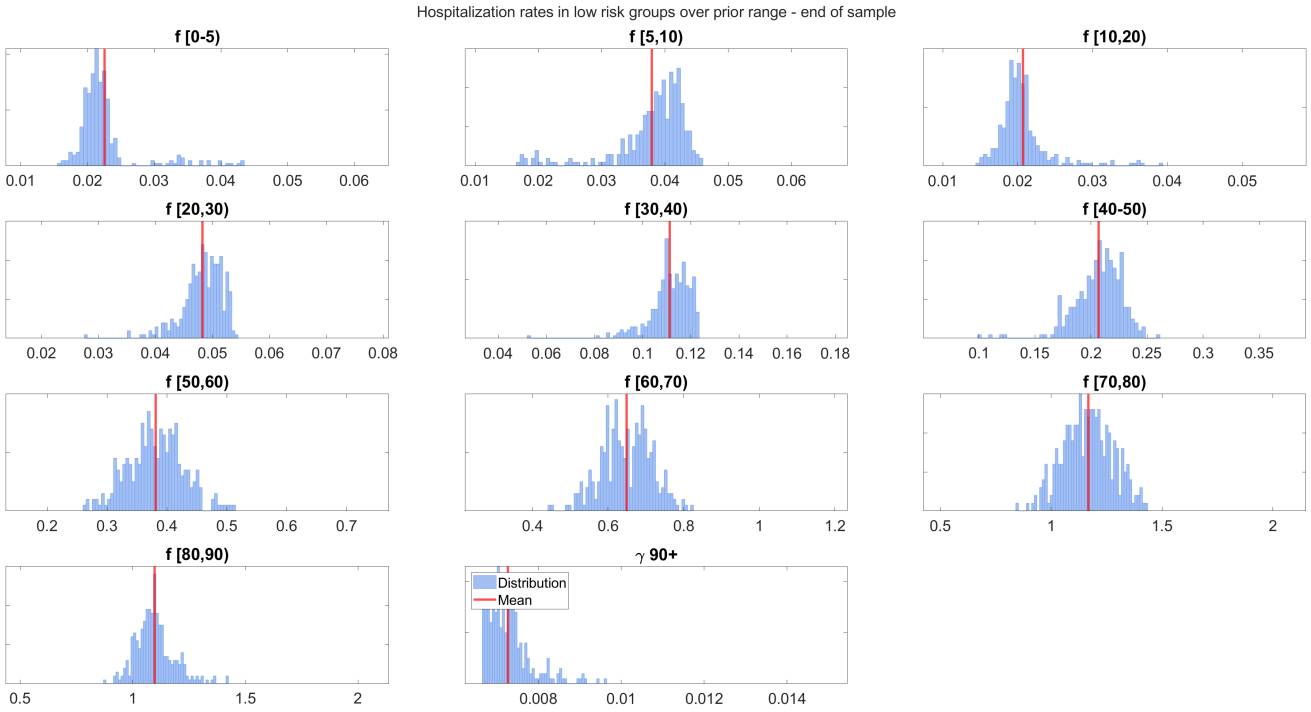
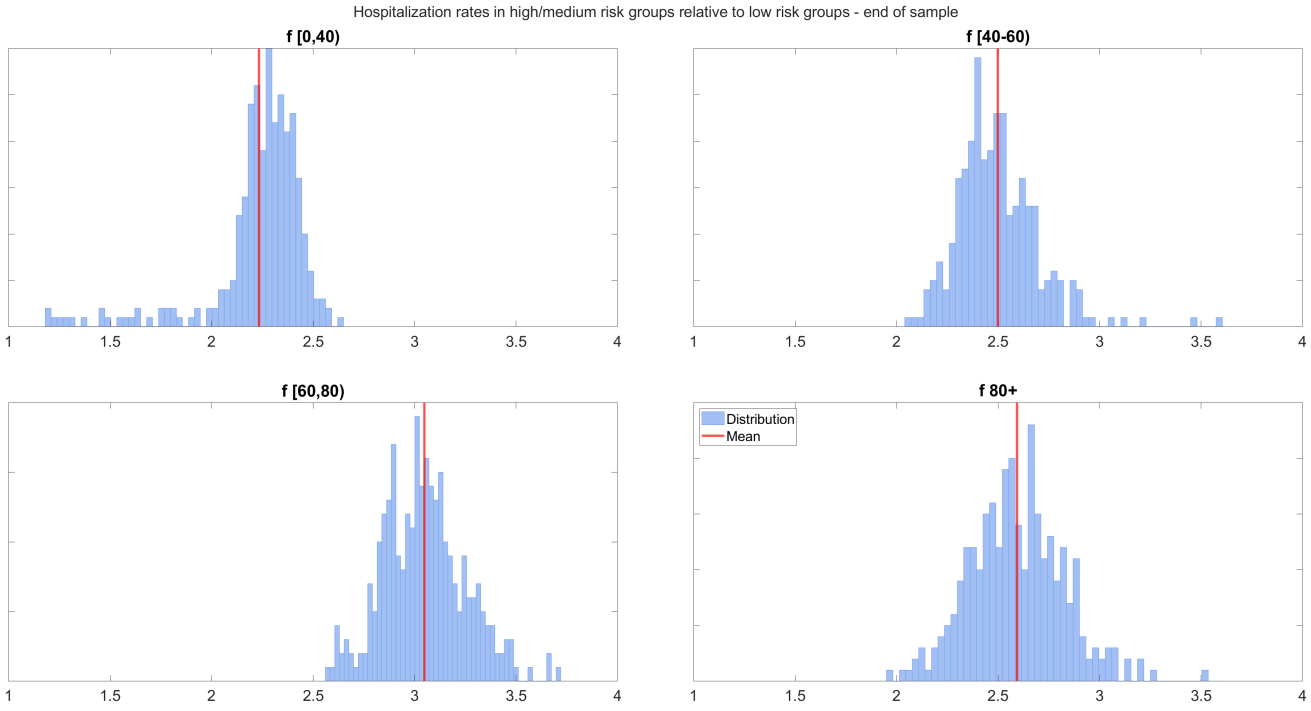


Figure 3: Posterior density of high/medium risk hospitalization rates relative to low hospitalization rates over prior range



5.2 Data fit

The fit to hospitalization data for all risk groups combined is very good except perhaps in age group [20,30). The next figure shows also a good fit for the high/medium risk groups combined. In that figure, note that the data

for low risk was not used, as their sum is already used once; nevertheless, despite this, as a validation exercise, Figure 5 shows that the low risk hospitalizations are also close to the data points.

Figure 4: Weekly hospitalizations for all risk groups combined

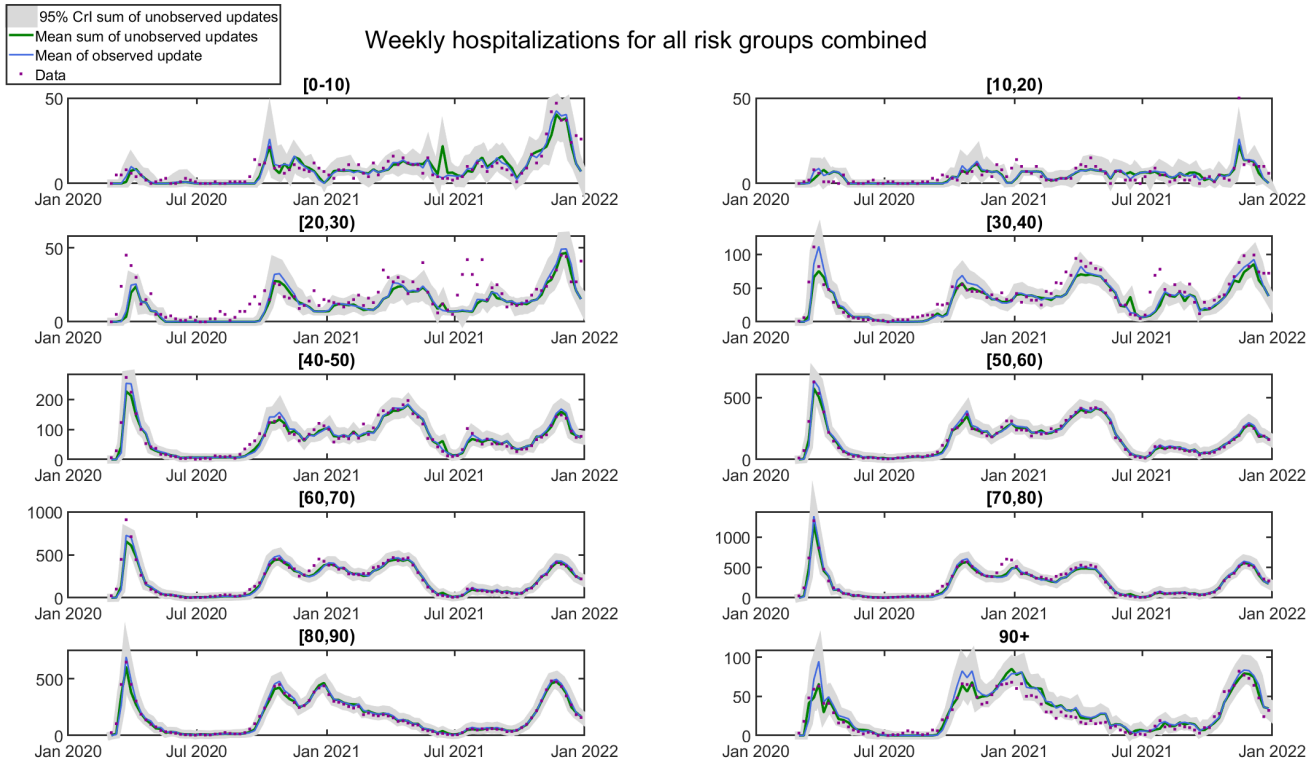
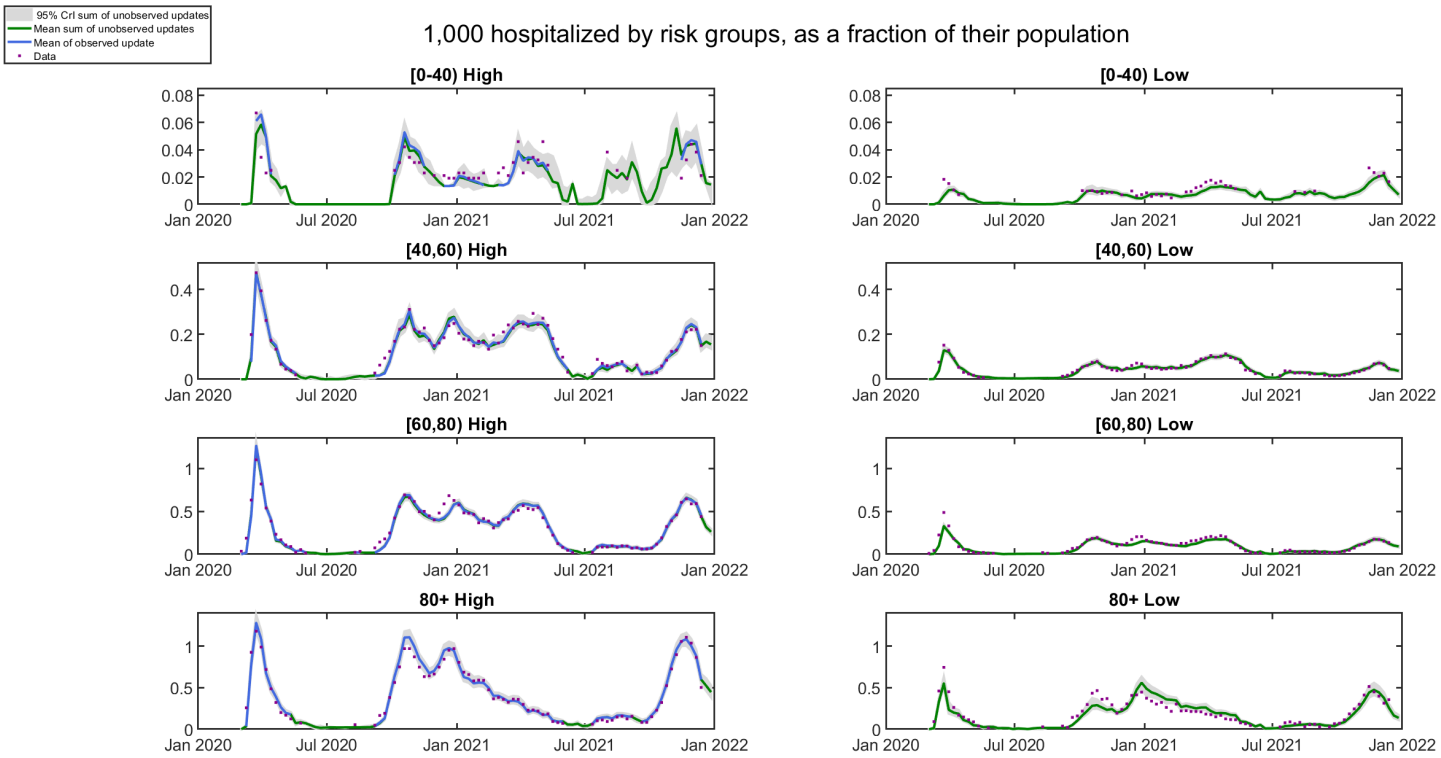


Figure 5: 1,000 Hospitalized by risk groups, as a fraction of their population by age



The fit to seroprevalence for all risk combined is also good except for age category $[0, 5)$, which can be explained by the fact that the children often have mild infections and antibodies are not fully detectable with sero-surveys. The fit to seroprevalence in older age groups by risk is also good, except for survey 5, where we discussed that due to vaccinations, there is divergence between our definition of seroprevalence as after the vaccines are fully effective versus the survey definition which detects antibodies due to vaccination even after one vaccine dose.

Figure 6: Weekly hospitalizations for all risk groups combined

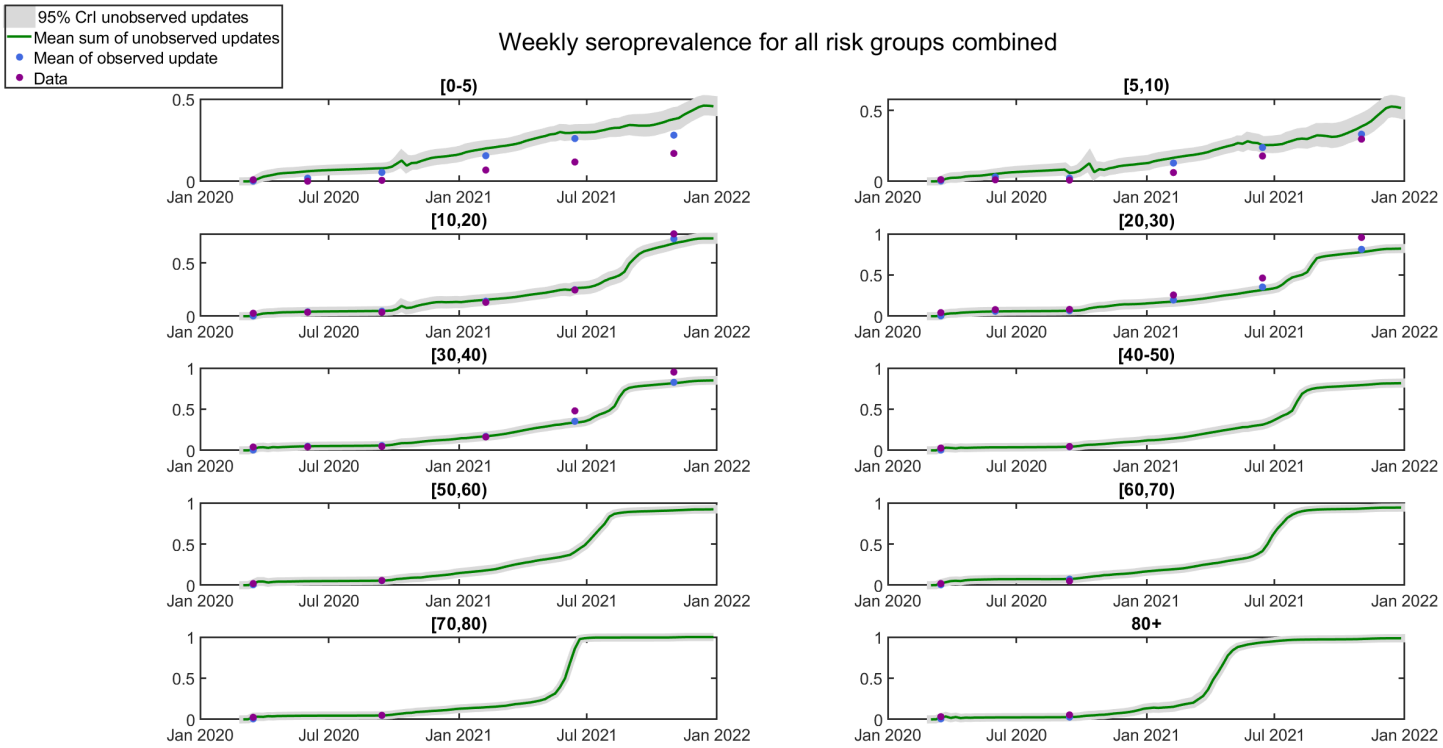
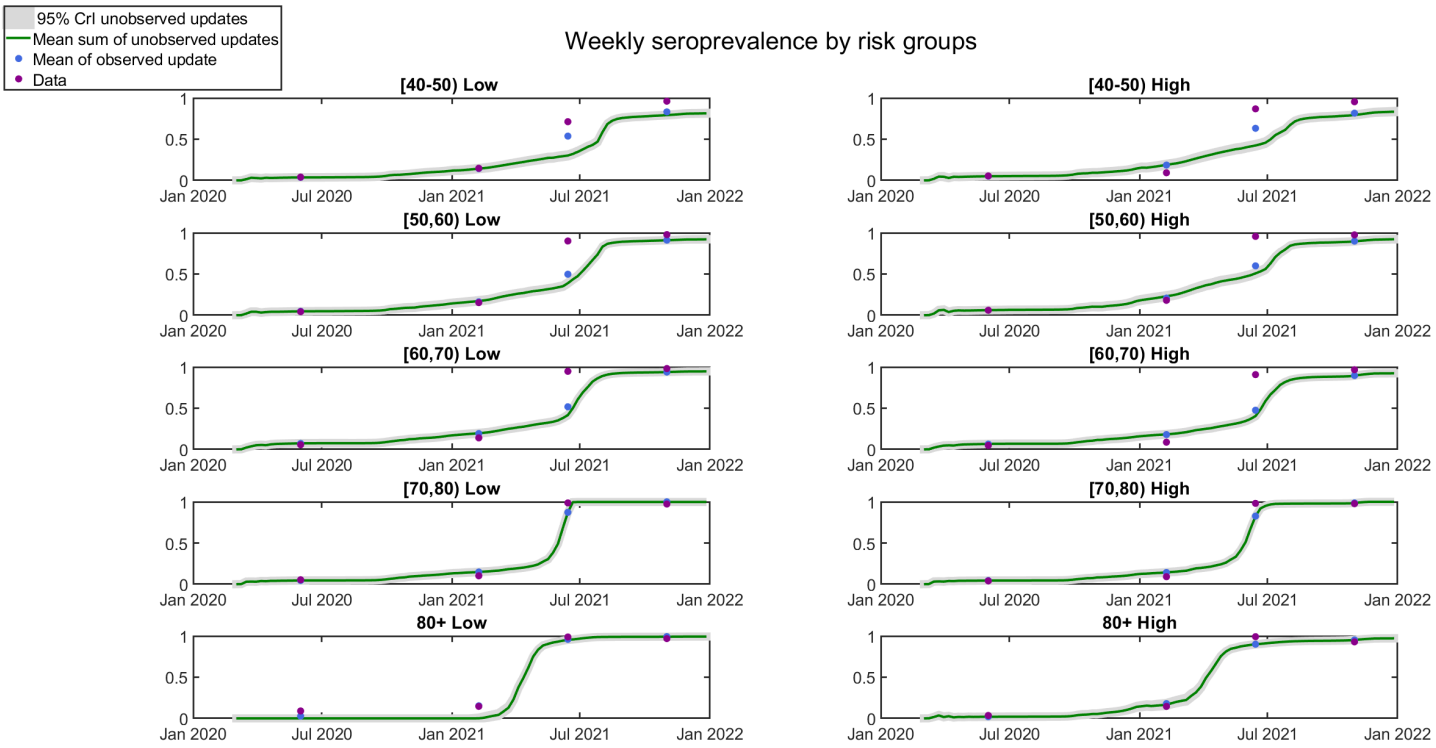


Figure 7: Weekly hospitalizations by risk



5.3 Quantifying infections and hospitalizations by age and risk

Infections. Figure 8 shows model estimated infections in a given age-risk group divided by the population of that age-risk group. First we observe that the infections per population are similar across medium and high risk groups. This is not surprising, as in a fully deterministic model, they would be equal without any vaccination differences, as we assumed that all other parameters and contacts are equal across these two groups. The small differences are purely due to vaccinating high-risk groups earlier or later (see also the vaccination figure). Second, the estimated infections are higher in the high/medium risk groups compared to low risk groups, and this can be explained by the fact that since vaccines are administered earlier in high/medium risk groups for many of the age groups, they also wane faster, and we have introduced a waning rate that is age and risk specific, with complete waning over a year, and mid-point equal to when 50% of the vaccines were administered in that particular age-risk group. Moreover, the survey data indicates that people at risk have in fact in general slightly more contacts in some age categories (see Figure 9), making them more vulnerable to infections (on average across the population, the contacts are the same, as demonstrated in [19], but when broken down by age, younger age individuals at higher risk have more contacts). Finally, we note that the higher age groups, 80+, have the highest infection rate, which can also be explained by this group not being able or willing to lower their number of contacts during lockdowns as much as the older risk groups; in higher risk groups, these contacts even increased during the first lockdown, as indicated by Figure 9.

Figure 8: Cumulative infections by age and risk groups 2020-2021

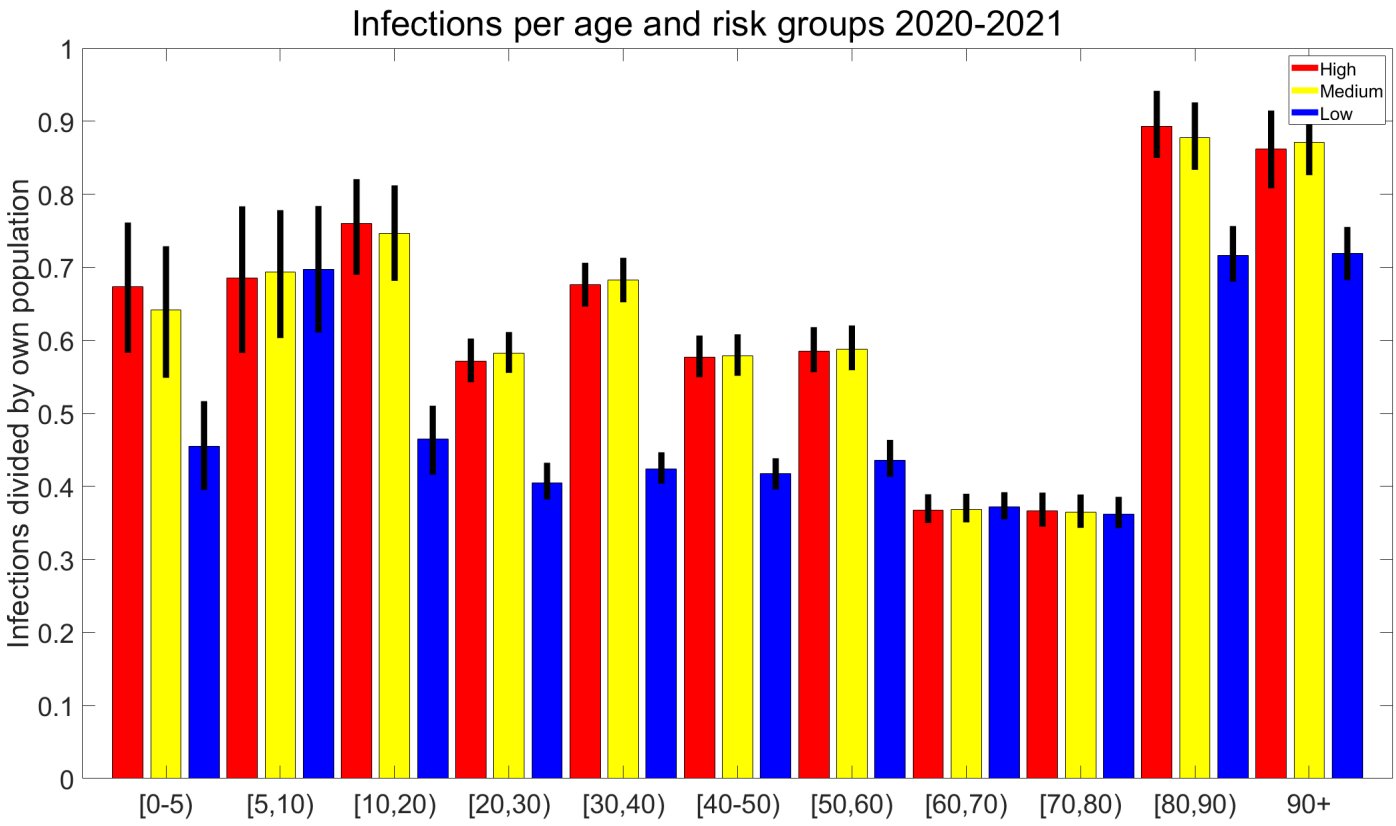
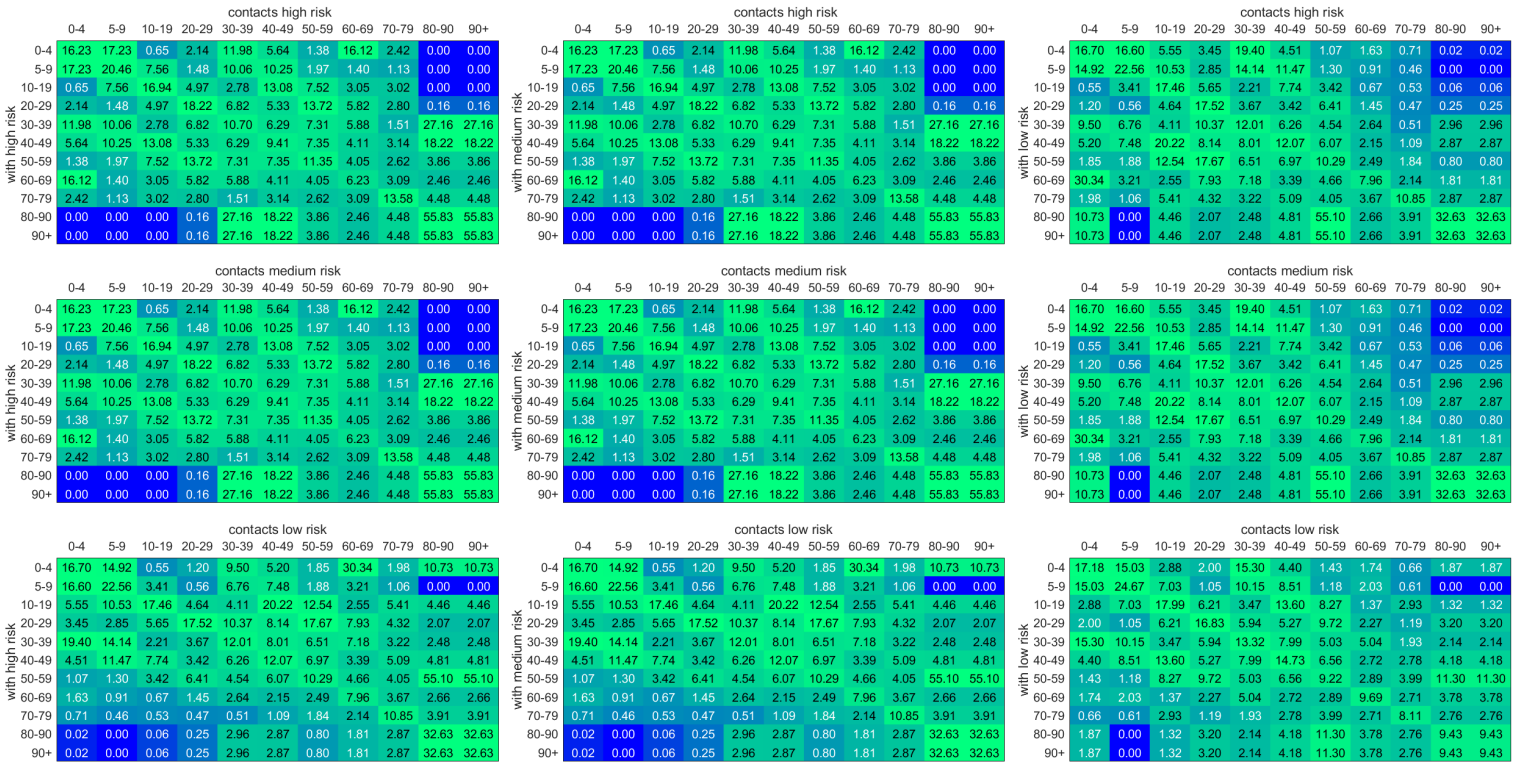


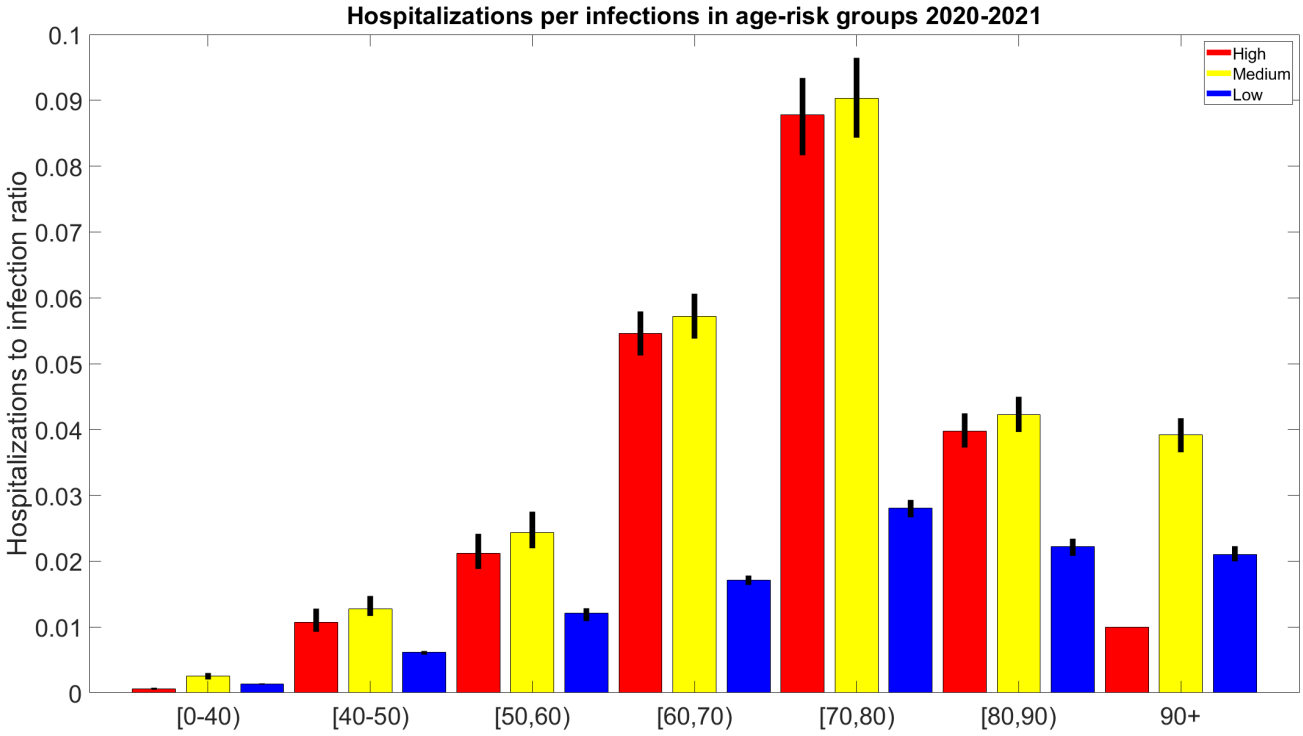
Figure 9: Contact matrices first lockdown in March 2020 - per initiator per receiver, multiplied by the population



Hospitalizations. The figure below shows estimated hospitalizations divided by estimated infections in each age-risk group. The results mirror the estimated hospitalization rates which increase with age. First, we find that the high and medium risk groups have similar hospitalizations per number of infections, and twice to three-times higher than those of low risk, in all age categories. In the 90+ age category, we see as many infections per population in the medium and high risk groups as for [80,90) years old, but because the high-risk population twice as small for 90+ than it is for [80,90) years old, the higher risk groups has less hospitalizations/infection. Note that this last finding is due to the assumption of proportionate mixing, and may be different should the higher risk groups have smaller contacts per initiator per receiver compared to the medium risk groups. Second, and more importantly, we find that the hospitalizations per infections are higher in the [60,80) age group compared to 80+ in high and medium risk groups, suggesting that large reductions in hospitalizations could be achieved by vaccinating [60,80) individuals at higher rates (the true uptake in 2023-2024 is in fact higher in this age groups, as reported in [24], Figure 2.1).

6 Evaluating vaccination strategies for the future

In this section, we evaluate different vaccination strategies. We assume that the new variants are equally transmissible to Omicron BA.1, as the sewage data indicate waves of qualitatively the same magnitude. We use the same model as for estimation, but individuals that are vaccinated start in the second compartment set (so they are assumed vaccinated before the wave begins). The initial range is assumed to be 1000 individuals and, for lack



of good estimates, we assume they are equally distributed across age-risk groups proportional to their population. We use the parameter posteriors at the end of the sample.

In the scenarios we create, the two model compartment sets change their meaning: the first one includes individuals that were previously infected or vaccinated or both, assuming all the population falls into one of these categories. In this compartment set, with multiply ϵ, the probability of infection, with $(1 - p_{1,I})$, where $p_{1,I} = 0.36$, according to [25] (review is not available differentially by age or risk), to indicate remaining protection from previous infection (estimated in the survey around 40 weeks after a previous infection), assuming everyone was previously infected (sero-surveys in [24] indicate that indeed most of the population was infected). We also introduce in this compartment set an (unconditional) protection level from a previous infection against hospitalization, $p_{1,H,UNC}$, which we set to 0.83, following [25]. From this unconditional protection, conditional ones are derived with the same formula as for the estimation section. The second compartment set now includes people who are assumed, for simplicity, to be vaccinated at once before running our scenarios. They benefit from higher protection against infection. Following [24], which gives estimates of additional protection against infection and hospitalization for 2023-2024, comparing to a population previously infected or vaccinated, the additional protection for infection of a vaccine is 37% in people age below 60, and 51% in those above 60. Therefore, the total protection against infection in the second compartment set is $1 - (1 - 0.37) * (1 - 0.36) = 0.5968$ for people below 60, and $1 - (1 - 0.51) * (1 - 0.37) = 0.6864$ for individuals aged 60 or above. The additional protection against hospitalization is taken from [24], 63% for individuals below 60; for those 60-79 years old, protection is 73% for high-risk groups, and 67% for low and medium risk groups, while for those 80 or above, it is 64% for high and medium risk groups, and 69% for those in low risk groups. Similarly to the protection from infection, we calculate from these the unconditional protection for hospitalization, which are, in the same order as above:

0.9371 for 0-60, for 60-80 they are 0.9541 for high risk and 0.9439 for medium and low risk, while for 80+, they are 0.9388 for high and medium risk, and 0.9473 for low risk individuals.

The current vaccination strategy in the Netherlands is to offer vaccines for all individuals aged 60 or above, for medium and high risk adults, for high risk children, and for medical personnel, with current (latest available: 2023-2024, [24]) vaccination coverage around 50% or higher for 60+ and higher for older age groups, and with lower coverage among individuals below 60 that are eligible for a vaccine. This strategy roughly corresponds to vaccinating a fraction of the 60+ population and of the medium and high risk individuals below 60. To check the external validity of our estimates, we approximate from [24], Figure 2.1 and 2.3, the vaccination coverage in each age category for 2023-2024, the latest numbers available to us. Table 1 gives these approximate vaccination coverages which differ by age and risk in most groups except some older ones:

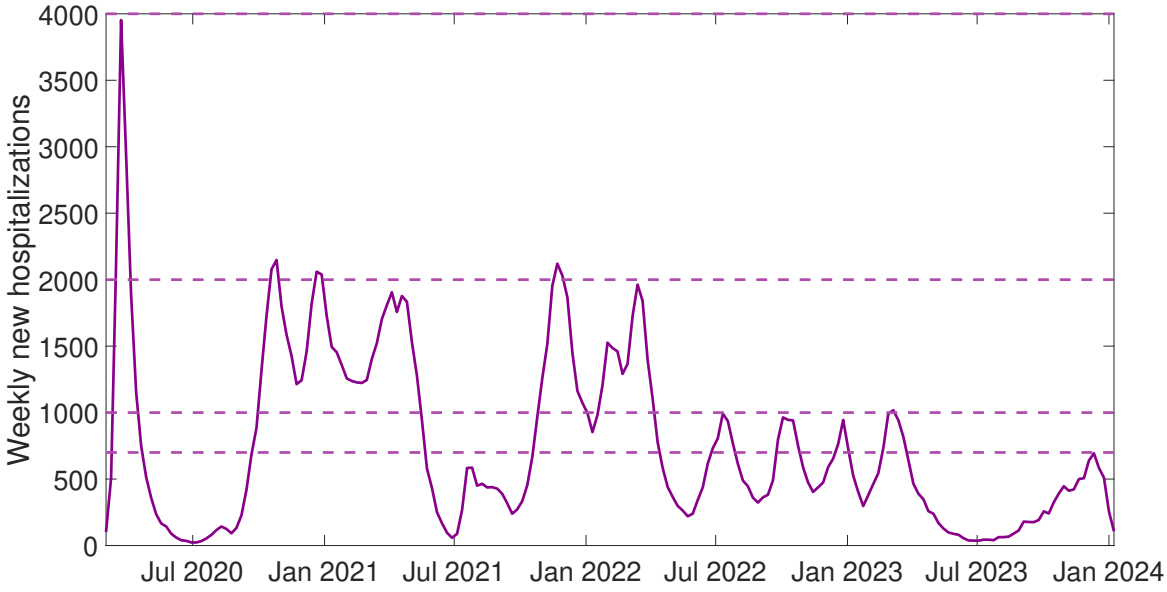
Table 1: Vaccination coverage 2023-2024

Age/Risk	High	Medium	Low
0-20	0	0	0
20-60	0.10	0.05	0.02
60-70	0.50	0.5	0.4
70-80	0.65	0.65	0.65
80-90	0.70	0.64	0.64
90+	0.56	0.56	0.56

With this vaccination coverage, we project the national weekly new hospitalizations, and check if the peak matches the peak in 2023-2024, the latest data available, when there was only one wave (we do not look at total hospitalizations because the data in 2023-2024 is incomplete). The available data for all years is plotted in Figure 10. This figure shows that the peak weekly new hospitalizations were 690, while our projection with the calibrated protection levels and vaccination coverage gives a mean estimate of peak new hospitalizations of 630 (80% CrI $[0, 2241]$), which is quite close to the real peak, validating our estimates.

Figure 10 also shows that in the largest wave, the first one, the peak weekly hospitalizations were approximately 4000 and completely overwhelmed the health care system. On the other hand, the peak weekly hospitalizations in the fall wave in 2023 were around 700, and the year before around 1000, in both years this being manageable for the health care system, with some burden still evident in fall 2022 due to multiple waves.

Figure 10: National weekly new hospitalizations



In our projections, we vary the vaccination coverage, as well as which risk groups and age groups are eligible, and report cumulative infections per population over a 6 month period, percentage change relative to no vaccination, peak weekly new hospitalizations, the cumulative weekly hospitalizations over a 6 month period over the population, assuming for simplicity no waning, the relative change compared to no vaccination, the number of vaccines needed to avoid 1 hospitalization in each scenario (relative to the case of no vaccination), and the total vaccines per population used in each strategy. We evaluate these scenarios mainly based on whether the projected peak new hospitalizations are below 1000 (deemed acceptable, while above 1000, the scenario is deemed unacceptable due to potentially overwhelming the health care system), and on their cost-effectiveness measured by the number of vaccines used to avoid one hospitalization.

6.1 Varying the vaccination coverage for all age-risk groups

In this section, we assume the entire population is vaccinated with different coverages, and identify what vaccine coverage is acceptable to avoid overwhelming the health system. Table 2 presents the results for varying the vaccination coverage and indicates that, should the coverage be uniform across the entire population, a coverage below 40% is not acceptable, and certainly not vaccinating anyone is also not acceptable for countries with a large older population like the Netherlands, as it would lead to a wave double the size of the first COVID-19 wave in 2020. A 50% coverage seems sufficient to not overwhelm the system, with peak new hospitalizations estimated at 400, but with some upside risk due to large credible intervals (note that while the individual parameter posteriors are quite tight, due to relatively many parameters, this can result in scenarios where there are many hospitalizations). We find that a coverage of 60% is sufficient for no wave to occur. New cumulative hospitalizations by age-risk are plotted in Figure 11, which shows, that, unlike for adults, the wave in children can be slightly higher when more individuals are vaccinated, and that is because their susceptibility is lower, so they get infected at a slower rate, and the longer the wave lasts, the more likely it is for them to accumulate new infections. We also found that without vaccination, the hospitalization wave takes off in 4 weeks, giving authorities time to conduct mass vaccination. If instead vaccination would stretch over more than 4 weeks, it it will result in a higher wave, and we will investigate this later.

Because the higher risk groups have a higher hospitalization risk, and the older age groups are also more susceptible for infection, vaccinating with priority risk and groups while using less vaccines may be desirable. In the next sections, we investigate vaccinating differentially by risk only, by age only, and by age-risk.

Table 2: Changing the vaccination coverage for all age and risk groups

Vaccination Coverage	Statistic	Infections		Hospitalizations			Vaccines/hosp avoided
		Total/pop	Rel. ↓	Peak new*	Total/pop	Rel. ↓	
0%	Mean	0.68	-	7800	0.0029	-	-
	10%	0.64	-	6900	0.0026	-	-
	90%	0.72	-	8600	0.0031	-	-
20%	Mean	0.55	0.19	4900	0.0022	0.23	303
	10%	0.50	0.16	3800	0.0020	0.21	279
	90%	0.60	0.21	5800	0.0024	0.25	329
30%	Mean	0.45	0.34	3500	0.0018	0.39	273
	10%	0.37	0.27	2300	0.0014	0.33	230
	90%	0.53	0.42	4200	0.0020	0.47	314
40%	Mean	0.21	0.70	1900	0.0008	0.74	200
	10%	0.03	0.44	500	0.0001	0.50	151
	90%	0.40	0.95	2800	0.0015	0.96	270
50%	Mean	0.04	0.95	400	0.0001	0.96	184
	10%	0.00	0.81	0	0.0000	0.84	166
	90%	0.14	1.00	1800	0.0005	1.00	201
60%	Mean	0.00	1.00	0	0.0000	1.00	210
	10%	0.00	1.00	0	0.0000	1.00	194
	90%	0.00	1.00	0	0.0000	1.00	229

*Rounded to lowest 100. Note: the first column indicates the vaccination coverage by age-risk, with L (low), M (medium), H (high) indicating the risk group of in each age group.

Note: the legend indicates the vaccination coverage by age-risk, with L (low), M (medium), H (high) indicating the risk group of in each age group.

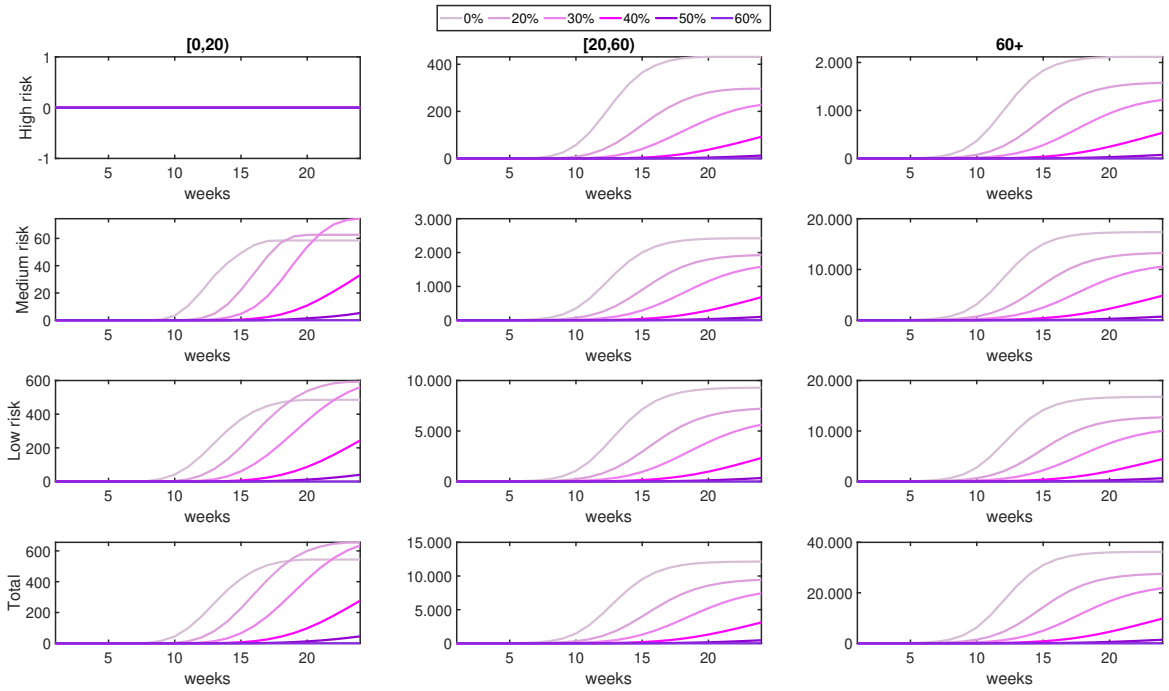


Figure 11: Cumulative Hospitalizations - by age and risk

6.2 Varying the vaccination coverage by risk groups only

In this section, we compare strategies where the vaccination coverage is the same or larger in higher risk groups.

Table 3 and Figure 12 demonstrate that it is insufficient to vaccinate the high and medium risk groups at a high rate, and, because the low risk group is large, a portion of them should be vaccinated as well to not overwhelm the public health system.

Table 3: Changing the vaccination coverage for risk groups

Vaccination Coverage	Statistic	Infections		Hospitalizations			Vaccines	
		Total/pop	Rel. ↓	Peak new*	Total/pop	Rel. ↓	per hosp avoided	per pop
0% L, 0% M, 0% H	Mean	0.68	-	7800	0.0029	-	-	-
	10%	0.64	-	6900	0.0026	-	-	
	90%	0.72	-	8600	0.0031	-	-	
50% L, 50% M, 50% H	Mean	0.04	0.95	400	0.0001	0.96	184	0.50
	10%	0.00	0.81	0	0.0000	0.84	166	
	90%	0.14	1.00	1800	0.0005	1.00	201	
0% L, 0% M, 50% H	Mean	0.62	0.08	6100	0.0026	0.08	31	0.01
	10%	0.58	0.06	5200	0.0024	0.07	26	
	90%	0.67	0.09	6800	0.0029	0.10	35	
0% L, 0% M, 75% H	Mean	0.58	0.14	5300	0.0025	0.14	27	0.01
	10%	0.53	0.10	4100	0.0022	0.11	22	
	90%	0.64	0.17	6100	0.0027	0.18	33	
0% L, 50% M, 50% H	Mean	0.50	0.26	3600	0.0018	0.37	60	0.06
	10%	0.42	0.19	2600	0.0015	0.31	51	
	90%	0.58	0.34	4300	0.0021	0.44	70	
0% L, 50% M, 75% H	Mean	0.37	0.46	2700	0.0013	0.54	44	0.07
	10%	0.23	0.29	1700	0.0008	0.40	34	
	90%	0.51	0.64	3400	0.0018	0.70	56	
0% L, 75% M, 75% H	Mean	0.11	0.84	1100	0.0003	0.88	38	0.09
	10%	0.01	0.56	0	0.0000	0.67	33	
	90%	0.31	0.99	2400	0.0010	0.99	46	
25% L, 75% M, 75% H	Mean	0.01	0.99	0	0.0000	1.00	110	0.31
	10%	0.00	0.99	0	0.0000	1.00	101	
	90%	0.00	1.00	0	0.0000	1.00	120	

*Rounded to nearest 100. Note: the percentages given in the first column refer to how many people are vaccinated in the low (L), medium (M) and high (H) risk groups.

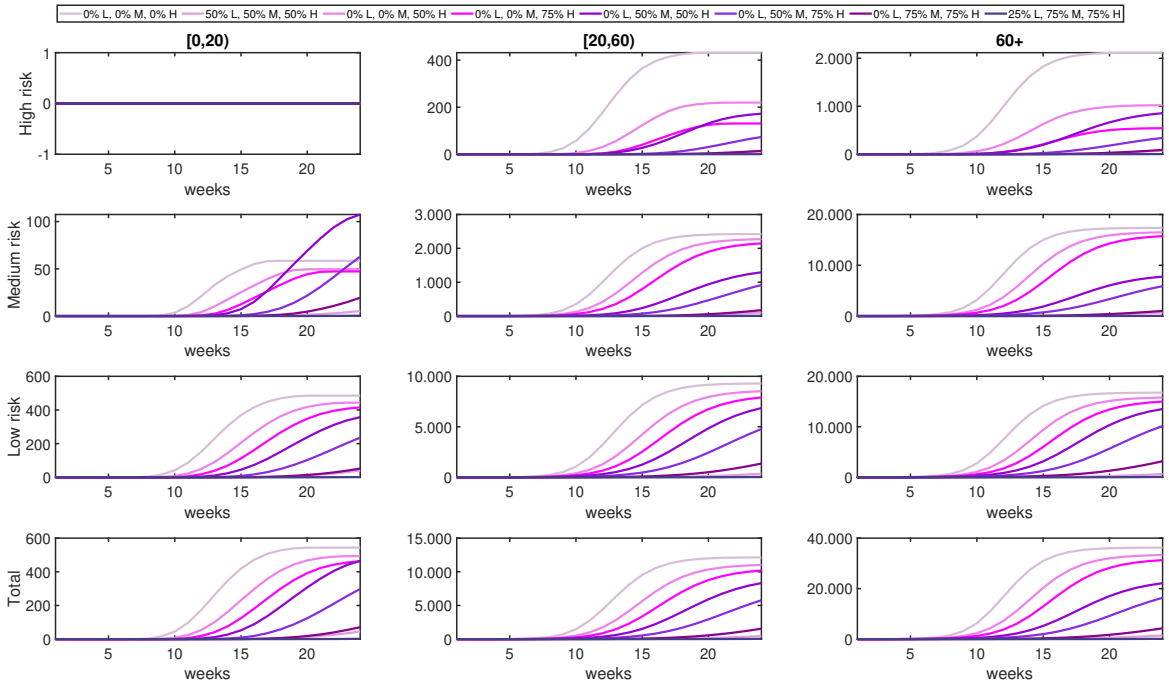


Figure 12: Cumulative Hospitalizations - by age and risk
Note: the percentages given in the legend refer to how many people are vaccinated in the low (L), medium (M) and high (H) risk groups.

6.3 Varying the vaccination coverage by age groups only

In this section, we compare strategies where the vaccination coverage is the same or larger in older age groups.

Tables 4 and Figure 13 show that a larger vaccination coverage in the older age groups is desirable, and a 75% vaccination in 60+ is sufficient to not overwhelm the whole system. However, such a high vaccination coverage may not be attainable in all these age groups (it was not in 2023-2024 in the Netherlands), suggesting that vaccinating a share of the lower age groups is also necessary.

Table 4: Changing the vaccination coverage by age groups

Vaccination Coverage	Statistic	Infections		Hospitalizations			Vaccines	
		Total/pop	Rel. ↓	Peak new*	Total/pop	Rel. ↓	per hosp avoided	per pop
0% all	Mean	0.68	-	7800	0.0029	-	-	-
	10%	0.64	-	6900	0.0026	-	-	
	90%	0.72	-	8600	0.0031	-	-	
50% all	Mean	0.04	0.95	400	0.0001	0.96	184	0.50
	10%	0.00	0.81	0	0.0000	0.84	166	
	90%	0.14	1.00	1800	0.0005	1.00	201	
75% 70+	Mean	0.18	0.74	1200	0.0004	0.85	40	0.09
	10%	0.01	0.36	0	0.0000	0.62	33	
	90%	0.46	0.99	2400	0.0012	1.00	51	
75% 60+	Mean	0.03	0.95	200	0.0001	0.98	66	0.19
	10%	0.00	0.86	0	0.0000	0.95	61	
	90%	0.10	1.00	800	0.0002	1.00	72	
75% 50+	Mean	0.01	0.99	0	0.0000	1.00	103	0.29
	10%	0.00	0.99	0	0.0000	1.00	96	
	90%	0.01	1.00	0	0.0000	1.00	112	

*Rounded to nearest 100. Note: the percentages given in the first column refer to how many people are vaccinated above a certain age.

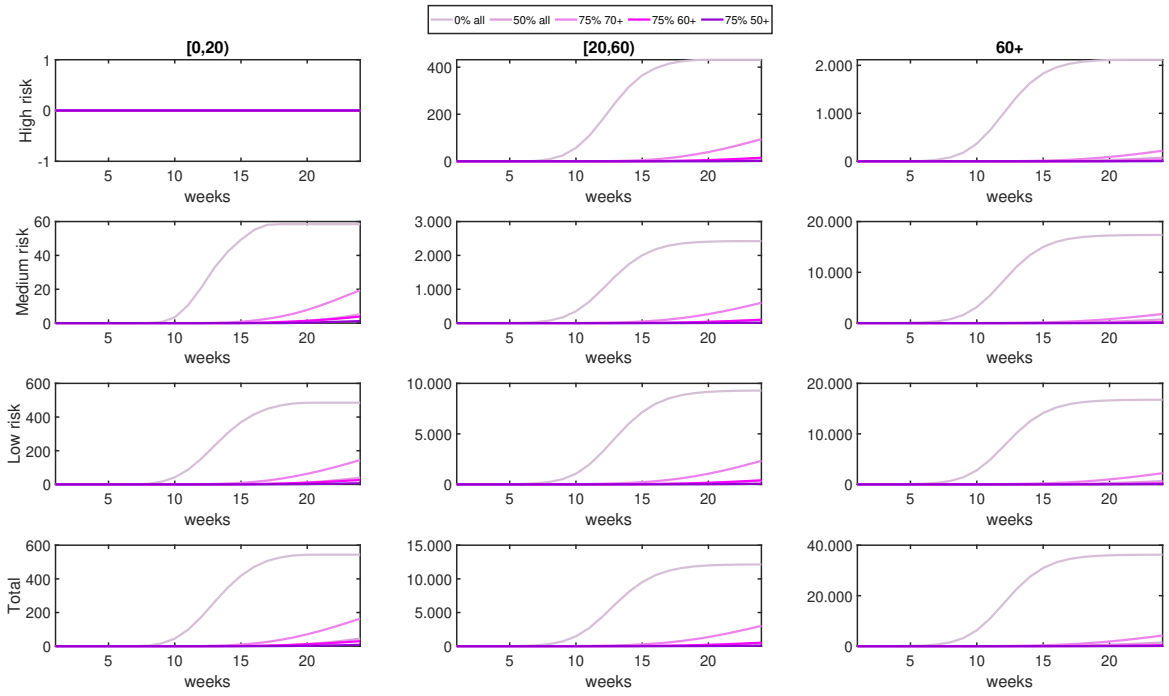


Figure 13: Cumulative Hospitalizations - by age and risk
Note: the percentages given in the legend refer to how many people are vaccinated above a certain age.

6.4 Varying the vaccination coverage by age-risk groups

In this section, we explore vaccination strategies that are closest to the strategy employed in Netherlands: offering vaccines to individuals age 60+, below 60, to high risk groups, and between [20-60), also to medium risk groups. In all scenarios, we maintain that 60+ individuals should be vaccinated with some coverage, as the previous section demonstrates large benefits in doing so. We also consider only strategies that vaccinate simultaneously high and medium risk groups, one, because the gains from vaccinating children and adolescents at high risk of severe disease are projected to be negligible in our model, and two, because Section 6.2 demonstrated that the gains of vaccinating medium risk groups in reducing hospitalizations are sizable. We also report the 2023-2024 scenario which is closest to the vaccination attained in the Netherlands.

Tables 5 and Figure 14 demonstrate that with a combination of high vaccination coverage among individuals aged 60+ (above 50%) and additional vaccination of younger individuals in high medium risk groups, larger than 25%, the public threat can be substantially mitigated. This also underscores that a vaccination strategy where individuals at high/medium risk or above 60 are encouraged to vaccinate at higher rates, is more fruitful and wastes less vaccines than offering everyone a vaccine but where the uptake is lacking in older age groups or high/medium risk individuals.

Table 5: Changing the vaccination coverage by age/risk groups

Vaccination Coverage	Statistic	Infections		Hospitalizations			Vaccines	
		Total/pop	Rel. ↓	Peak new*	Total/pop	Rel. ↓	per hosp avoided	per pop
0% all	Mean	0.68	-	7800	0.0029	-	-	-
	10%	0.64	-	6900	0.0026	-	-	
	90%	0.72	-	8600	0.0031	-	-	
50% 60+	Mean	0.38	0.44	2200	0.0011	0.63	72	0.12
	10%	0.16	0.22	900	0.0004	0.46	53	
	90%	0.56	0.74	3000	0.0017	0.84	90	
25% HM <60, 50% 60+	Mean	0.15	0.78	1200	0.0004	0.86	59	0.14
	10%	0.00	0.40	0	0.0000	0.59	48	
	90%	0.43	0.99	2600	0.0012	1.00	79	
25% HM <60, 75% 60+	Mean	0.01	0.99	0	0.0000	1.00	66	0.19
	10%	0.00	0.99	0	0.0000	1.00	61	
	90%	0.01	1.00	0	0.0000	1.00	72	
50% HM <60, 50% 60+	Mean	0.07	0.91	600	0.0002	0.94	59	0.15
	10%	0.00	0.66	0	0.0000	0.76	52	
	90%	0.24	0.99	2100	0.0007	1.00	68	
50% HM <60, 75% 60+	Mean	0.00	1.00	0	0.0000	1.00	76	0.22
	10%	0.00	1.00	0	0.0000	1.00	70	
	90%	0.00	1.00	0	0.0000	1.00	82	
25% HM<60, 50% 60-80, 75% 80+	Mean	0.01	0.98	100	0.0000	0.99	53	0.15
	10%	0.00	0.99	0	0.0000	1.00	49	
	90%	0.00	1.00	0	0.0000	1.00	58	
2023-2024	Mean	0.10	0.87	600	0.0002	0.93	59	0.15
	10%	0.00	0.54	0	0.0000	0.74	52	
	90%	0.33	0.99	2200	0.0008	1.00	70	

*Rounded to nearest 100.

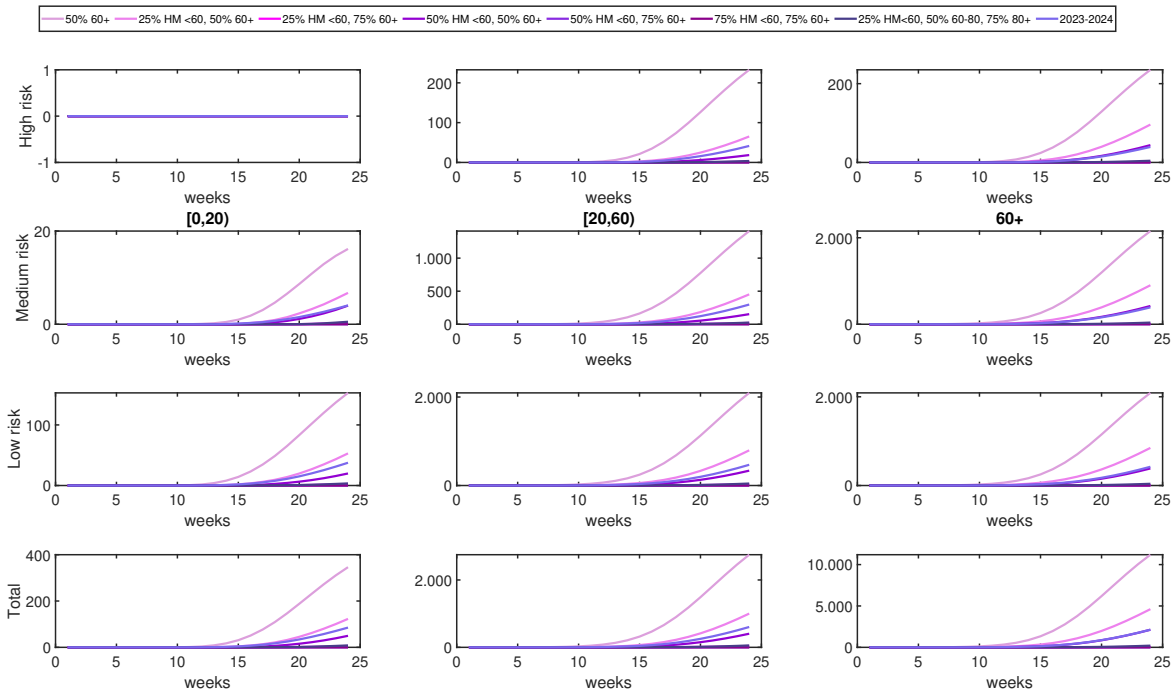


Figure 14: Cumulative Hospitalizations - by age and risk
Note: the percentages given in the legend refer to how many people are vaccinated in the low (L), medium (M) and high (H) risk groups.

7 Conclusions

We developed a national model of the spread of COVID-19 stratified by age and risk of severe disease. We found that the largest disease burden is in medium and high risk groups in most age categories, with more infections per population and more hospitalizations per population than the corresponding lower risk groups. We also found that hospitalizations per infection increase dramatically with age, regardless of the risk classification. We compared various vaccination scenarios in which individuals are vaccinated uniformly across age groups, differentially by age, by risk and by age-risk. We validated our estimates against external data: hospitalizations in low risk individuals in 2020-2021 in the estimation stage and hospitalization peak in 2023-2024 in the scenario stage. We found that the most cost-effective vaccination strategy is to achieve a high coverage in older individuals, and a reasonable coverage for medium-high risk groups in younger individuals.

There are some drawbacks to our study. First, because of privacy concerns, we did not obtain sufficiently fine-grained data by age and risk to precisely estimate susceptibility and hospitalization rate in each age-risk categories, and we had to group these estimates by age-risk groups, based on data availability and previous studies. This reflects in relatively wide confidence intervals for our vaccination scenarios. Second, for the projections, we assumed that everyone starts in the first compartment set, meaning they still have some protection from a previous infection but this protection waned over at least 9 months. Should more individuals benefit from recent infections, the scenarios we simulated would achieve the same goal with less vaccination coverage. Third, the vaccination scenario assume that everyone is vaccinated at once, before a wave starts. Should the vaccination occur more gradually or start after the start of a COVID-19 wave, we would observe more infections, and more

vaccination coverage would be required to mitigate the wave. Finally, in the scenarios we did not allow for waning from a previous infection once the wave starts, but we plan to do so in the future.

References

[1] M. Flook, C. Jackson, E. Vasileiou, C. R. Simpson, M. D. Muckian, U. Agrawal, C. McCowan, Y. Jia, J. L. K. Murray, L. D. Ritchie, C. Robertson, S. J. Stock, X. Wang, M. E. J. Woolhouse, A. Sheikh, and H. R. Stagg. Informing the public health response to COVID-19: a systematic review of risk factors for disease, severity, and mortality. *BMC infectious diseases*, 21(1):342–342, Apr 2021.

[2] Kate M. Bubar, Kyle Reinholt, Stephen M. Kissler, Marc Lipsitch, Sarah Cobey, Yonatan H. Grad, and Daniel B. Larremore. Model-informed COVID-19 vaccine prioritization strategies by age and serostatus. *Science*, 371(6532):916–921, 2021.

[3] Laura Matrajt, Julia Eaton, Tiffany Leung, and Elizabeth R. Brown. Vaccine optimization for COVID-19: Who to vaccinate first? *Science Advances*, 7(6):eabf1374, 2021.

[4] G. Rozhnova, C.H. van Dorp, P. Bruijning-Verhagen, M. Bootsma, J. van de Wijgert, M. Bonten, and M. E. Kretzschmar. Model-based evaluation of school and non-school-related measures to control the COVID-19 pandemic. *Nature Communications*, 12(1):1614, 2021.

[5] J. Viana, C. H. van Dorp, A. Nunes, M. C. Gomes, M. van Boven, M. E. Kretzschmar, M. Veldhoen, and G. Rozhnova. Controlling the pandemic during the SARS-CoV-2 vaccination rollout. *Nature Communications*, 12(1):3674, 2021.

[6] Otilia Boldea, Amir Alipoor, Sen Pei, Jeffrey Shaman, and Ganna Rozhnova. Age-specific transmission dynamics of SARS-CoV-2 during the first 2 years of the pandemic. *PNAS Nexus*, 3(2):pgae024, 01 2024.

[7] Ilse Westerhof, Annemarijn de Boer, Angela Lupattelli, Isabel Slurink, Otilia Boldea, Hedvig Marie Egeland Nordeng, Jizzo R. Bosdriesz, Frank Pijpers, Maarten Schim van der Loeff, Mirjam Knol, Janneke van de Wijgert, Patricia Bruijning, and Ganna Rozhnova. Medical risk classification for severe COVID-19 based on chronic medical conditions: A comparative analysis. *medRxiv*, 2024. `https://www.medrxiv.org/content/early/2024/09/06/2024.09.06.24313189`.

[8] Thomas J Hladish, Alexander N Pillai, Carl AB Pearson, Kok Ben Toh, Andrea C Tamayo, Arlin Stoltzfus, and Ira M Longini. Evaluating targeted COVID-19 vaccination strategies with agent-based modeling. *PLOS Computational Biology*, 20(5):e1012128, 2024.

[9] Chadi M. Saad-Roy, Caroline E. Wagner, Rachel E. Baker, Sinead E. Morris, Jeremy Farrar, Andrea L. Graham, Simon A. Levin, Michael J. Mina, C. Jessica E. Metcalf, and Bryan T. Grenfell. Immune life history, vaccination, and the dynamics of SARS-CoV-2 over the next 5 years. *Science*, 370(6518):811–818, 2020.

[10] Jennie S. Lavine, Ottar N. Bjornstad, and Rustom Antia. Immunological characteristics govern the transition of COVID-19 to endemicity. *Science*, 371(6530):741–745, 2021.

[11] Sen Pei, Teresa K Yamana, Sasikiran Kandula, Marta Galanti, and Jeffrey Shaman. Burden and characteristics of covid-19 in the united states during 2020. *Nature*, 598(7880):338–341, Oct 2021.

[12] Mark M Dekker, Luc E Coffeng, Frank P Pijpers, Debabrata Panja, and Sake J de Vlas. Reducing societal impacts of SARS-CoV-2 interventions through subnational implementation. *Elife*, 12:e80819, 2023.

[13] Nathanaël Hozé, Juliette Paireau, Nathanaël Lapidus, Cécile Tran Kiem, Henrik Salje, Gianluca Severi, Mathilde Touvier, Marie Zins, Xavier de Lamballerie, Daniel Lévy-Bruhl, Fabrice Carrat, and Simon Cauchemez. Monitoring the proportion of the population infected by SARS-CoV-2 using age-stratified hospitalisation and serological data: a modelling study. *The Lancet Public Health*, 6(6):e408–e415, Jun 2021.

[14] Paolo Bosetti, Cécile Tran Kiem, Alessio Andronico, Vittoria Colizza, Yazdan Yazdanpanah, Arnaud Fontanet, Daniel Benamouzig, and Simon Cauchemez. Epidemiology and control of SARS-CoV-2 epidemics in partially vaccinated populations: a modeling study applied to france. *BMC Medicine*, 20(1):33, Jan 2022.

[15] Raphael Sonabend, Lilith K. Whittles, Natsuko Imai, Pablo N. Perez-Guzman, Edward S. Knock, Thomas Rawson, Katy A. M. Gaythorpe, Bimandra A. Djaafara, Wes Hinsley, Richard G. FitzJohn, John A. Lees, Divya Thekke Kanapram, Erik M. Volz, Azra C. Ghani, Neil M. Ferguson, Marc Baguelin, and Anne Cori. Non-pharmaceutical interventions, vaccination, and the SARS-CoV-2 delta variant in england: a mathematical modelling study. *The Lancet*, 398(10313):1825–1835, Nov 2021.

[16] Kylie E. C. Ainslie, Jantien Backer, Pieter de Boer, Albert Jan van Hoek, Don Klinkenberg, Hester Korthals Altes, Ka Yin Leung, Hester de Melker, Fuminari Miura, and Jacco Wallinga. The impact of vaccinating adolescents and children on COVID-19 disease outcomes. *medRxiv*, 2021.

[17] Michiel van Boven, Wouter A. Hetebrij, Arno M. Swart, Erwin Nagelkerke, Rudolf F.H.J. van der Beek, Sjors Stouten, Rudolf T. Hoogeveen, Fuminari Miura, Astrid Kloosterman, Anne-Merel R. van der Drift, Anne Welling, Willemijn J. Lodder, and Ana M. de Roda Husman. Modelling patterns of SARS-CoV-2 circulation in the netherlands, august 2020-february 2022, revealed by a nationwide sewage surveillance program. *medRxiv*, 2022.

[18] Sake J. de Vlas and Luc E. Coffeng. Achieving herd immunity against COVID-19 at the country level by the exit strategy of a phased lift of control. *Scientific Reports*, 11(1):4445, Feb 2021.

[19] Jantien A Backer, Eric RA Vos, Gerco den Hartog, Cheyenne CE van Hagen, Hester E de Melker, Fiona RM van der Klis, and Jacco Wallinga. Contact behaviour before, during and after the COVID-19 pandemic in the netherlands: evidence from contact surveys in 2016-2017 and 2020-2023. *medRxiv*, pages 2024–03, 2024.

[20] Jantien A Backer, Liesbeth Mollema, Eric RA Vos, Don Klinkenberg, Fiona RM van der Klis, Hester E de Melker, Susan van den Hof, and Jacco Wallinga. Impact of physical distancing measures against COVID-19 on contacts and mixing patterns: repeated cross-sectional surveys, the Netherlands, 2016-17, April 2020 and June 2020. *Eurosurveillance*, 26(8), 2021.

[21] Ruiyun Li, Sen Pei, Bin Chen, Yimeng Song, Tao Zhang, Wan Yang, and Jeffrey Shaman. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science*, 368(6490):489–493, 2020.

[22] Theresa Lange. Derivation of ensemble Kalman Bucy filters with unbounded nonlinear coefficients. *Nonlinearity*, 35:1061–1092, 2021. https://doi.org/10.1088/1361-6544/ac4337.

[23] Jeffrey L. Anderson. An Ensemble Adjustment Kalman Filter for data assimilation. *Monthly Weather Review*, 129(12):2884 – 2903, 2001.

[24] RIVM. COVID-19 vaccination.Background information for the Health Council of the Netherlands. *RIVM-briefrapport*, 2024-0055, 2024. https://www.rivm.nl/bibliotheek/rapporten/2024-0055.pdf.

[25] F.A. Niessen, P.C.J.L. Bruijning-Verhagen, M.J.M. Bonten, V. Peters, H.S.M. Ammerlaan, C.E. Delsing, N.G. Dijkstra, P. Bresser, H. Kramer, D. Rusch, E.A. bij de Vaate, P. Veen, A. Dofferhoff, M. Bruns, L.C. Terpstra, W. Thijs, R. Hassing, D.H. Huijbens, E. Jong, and M.J. Knol. Vaccine effectiveness against COVID-19 related hospital admission in the netherlands by medical risk condition: A test-negative case-control study. *Vaccine*, 42(15):3397–3403, 2024.