Estimating transmission parameters for respiratory syncytial virus and predicting the impact of maternal and pediatric vaccination

Supplementary Information

Michiel van Boven, Anne C. Teirlinck, Adam Meijer, Mariëtte Hooiveld, Christiaan H. van Dorp, Rachel M. Reeves, Harry Campbell, Wim van der Hoek

May 26, 2020

Contents

1 Introduction

In this supplement we provide a detailed description of the data, further develop the basic transmission model, provide details of the estimation procedures, present a number of tables and figures mentioned in the main text, and give extended detailed results for the default model scenario as well as a variant scenario in which the expected duration of the infectious period is fixed at 1 week.

2 Data

Transmission models are fitted to three data sets. The first data set contains weekly age-stratified hospitalisations with confirmed RSV for the years 2013-2017. Data have been requested from the Dutch Hospitalisation Data organisation (DHD) using RSV specific ICD10 codes B97.4 (respiratory syncytial virus), J12.1 (respiratory syncytial virus pneumonia), J20.5 (acute bronchitis due to respiratory syncytial virus), and J21.0 (acute bronchiolitis due to respiratory syncytial virus). Coverage of the hospitalisation data is near complete in the Netherlands for the years 2013-2017 (97.3% in 2013, 98.8% in 2014, and $> 99\%$ from 2015 onwards). We exclude hospitalisation data prior to 2013 because coverage had been uncertain and much lower in 2012 and earlier years. Starting from individual-level data, we have extracted for each unique person per RSV season the first episode of RSV in the hospital using the above codes, ensuring that time of reporting is closest to the moment of infection while each infection is counted only once. To cover the RSV epidemic we take a broad range for the RSV seasons, spanning 35 weeks, starting from ISO week 40 in a given year, and extending into week 22 or 21 of the next year [1]. Including both primary and secondary diagnoses, the total number of hospitalisations is 12,038, of which the majority is in infants ($[0, 1)$ year, 8,162 cases) and young children ($[1, 5)$ years, 1,229 cases). The total number of hospitalisations with a specific RSV code ('confirmed RSV') is 1,960 in the 2012-2013 season (excluding 12 weeks in 2012), 2,538 in the 2013-2014 season, 2,432 in the 2014-2015 season, 2,466 in the 2015-2016

season, and 2,642 in the 2016-2017 season.

Second, general practitioner (GP) consultations for acute respiratory infection (ARIs) have been obtained from the Nivel Primary Care Database. Here the data are based on the weekly number of reported ARI as defined by Nivel. For each of the age classes and weeks we have at our disposal the number of cases and size of the catchment populations (covering approximately 7% of the Dutch population). The total number of incident cases included is 877,752 from a catchment population containing 300,875,954 person-weeks. A sizeable fraction of those cases (171,228) is in infants and young children ($[0, 5)$ years), and may have high likelihood being caused by RSV infection.

Third, patient age-specific virological data are available from RIVM/Nivel sentinel surveillance of influenza-like illness (ILI) and ARI [1]. These data are obtained from a small subset of approximately 40 GP practices from the Nivel Primary Care Database representing 0.8% of the population in the Netherlands. Each participating GP has been asked to send in at least two combined nose and throat swab specimens per week, of ILI (preferred) or ARI if no ILI are encountered halfway through the week. Henceforth, we will refer to these data as ARI specimens, even though the data are enriched with ILI. Details are given in [1]. Briefly, all specimens have been tested for influenza virus, RSV, rhinovirus and enterovirus using real-time reverse transcription polymerase chain reaction (RT-PCR). The total number of specimens in the five RSV seasons is 4,514, of which 378 are RSV positive. Almost half of the positive samples (159) are from infants and young children ($[0 - 5)$ years).

In addition, age-specific contact rates have been obtained from analysis of a contact survey carried out in the Netherlands in 2006-2007 [2], while demographic composition of the Netherlands in 2014 has been used for calculation of incidence of hospitalisations. We consider the population aged up to 100 years. All data are available in our github repository.

3 Transmission model

At the core of the analyses are age-structured SEIR type transmission models describing transmission of RSV in the population. In the models, individuals are classified as susceptible (S), latently infected (i.e. infected but not yet immune)(E), infected and infectious (I), or removed (i.e. immune)(R). Throughout, we consider seven age classes, viz. $[0, 1)$ years (abbreviated as 0 yr), $[1, 5)$ years (1-4 yr), $[5, 10)$ years (5-9 yr), $[10, 20)$ years (10-19 yr), $[20, 45)$ years (20-44 yr), $[45, 65)$ years (45-64 yr), and $[65, 100)$ years (65+ yr). These age groups correspond to a natural grouping of individuals by similarity of contact patterns [2] while taking into account that only limited case data are available for non-elderly adults.

For simplicity, we focus on a model without latently infected compartment as inclusion of a latent period of 2-3 days does not noticeably affect the results (not shown). Let $\mathbf{x}(t)$, $\mathbf{y}(t)$, and $\mathbf{z}(t)$ contain the age-specific relative frequencies of S, I, and R in different age groups, so that $\mathbf{z}(t) = \mathbf{1} - \mathbf{x}(t) - \mathbf{y}(t)$. Then the model is specified by ordinary differential equations given by (using dot notation)

$$\begin{aligned} \dot{\mathbf{x}} &= -\beta \; \mathit{diag}(\mathbf{x}) \; \mathbf{C} \; \mathbf{y} \\ \dot{\mathbf{y}} &= \beta \; \mathit{diag}(\mathbf{x}) \; \mathbf{C} \; \mathbf{y} - \tfrac{1}{D} \; \mathbf{y} \; , \end{aligned} \tag{1}$$

where β, D, and $\mathbf{C}$ are the transmission rate parameter, infectious period, and contact matrix. As is common practice, the contact matrix will be hard-wired into the model using Dutch person-to-person contact rates and the demographic composition of the Netherlands in 2014 [2].

If time is measured in units of the infectious period and the contact matrix is re-scaled such that its dominant eigenvalue equals 1 then the model equations can be written as

$$\begin{aligned} \dot{\mathbf{x}} &= -\mathcal{R}_0 \; \mathit{diag}(\mathbf{x}) \; \mathbf{C} \; \mathbf{y} \\ \dot{\mathbf{y}} &= \mathcal{R}_0 \; \mathit{diag}(\mathbf{x}) \; \mathbf{C} \; \mathbf{y} - \mathbf{y} \; , \end{aligned}$$

where $\mathcal{R}_0 = \beta D$ is the basic reproduction number representing the expected number of secondary infections per infection in the early stage of an epidemic in an entirely susceptible population. Notice that the above implies that D sets the time-scale, and that the total number of infections over the course

of an epidemic (the final size) depends on $\mathcal{R}_0$ only.

Now we show that the above model has a broader use, and can incorporate mild and asymptomatic infections. Consider the situation in in which a proportion m of infections is mild or even asymptomatic (I_m) and have a reduction in infectiousness r. In this case, the model equations are given by

$$\begin{aligned} \dot{\mathbf{x}} &= -\mathcal{R}_0 \; diag(\mathbf{x}) \; \mathbf{C} \; (\mathbf{y} + r \; \mathbf{y}_m) \\ \dot{\mathbf{y}} &= \mathcal{R}_0 \; (1-m) \; diag(\mathbf{x}) \; \mathbf{C} \; (\mathbf{y} + r \; \mathbf{y}_m) - \mathbf{y} \\ \dot{\mathbf{y}}_\mathbf{m} &= \mathcal{R}_0 \; m \; diag(\mathbf{x}) \; \mathbf{C} \; (\mathbf{y} + r \; \mathbf{y}_m) - \mathbf{y}_m \; , \end{aligned} \tag{2}$$

where $\mathbf{y}_m$ contains the relative frequencies of mildly infected individuals. If $r = 0$, mild infections are not infectious. We can write the basic model in terms of $\mathbf{x}^* = (1-m)\mathbf{x}$ and $\mathbf{y}$, and interpret $\mathbf{x}^*$ as the fraction of susceptible individuals that are available for symptomatic infection. Alternatively, if $0 < r \leq 1$, we re-parameterize by $\mathbf{x}^* = (1 - m + rm)\mathbf{x}$ and $\mathbf{y}^* = \mathbf{y} + r\mathbf{y}_m$, and again use the basic model without asymptomatic infection.

If mild infections have a reduced probability of reporting $p_m < p$, then the rate of reporting is

$$-\left(p(1-m) + p_m m\right) \dot{\mathbf{x}} \; .$$

Now if furthermore the reporting probability is proportional to infectiousness ($p_m = r \; p$, $0 \leq r \leq 1$), then the reporting rate is

$$-\left(p(1-m) + p \; r \; m\right) \dot{\mathbf{x}} = -p \left(1 - m + r \; m\right) \dot{\mathbf{x}} = -p \; \dot{\mathbf{x}}^* \; .$$

Hence, the basic model describes a scenario with variable severity, infectiousness, and reporting if $\mathbf{x}^*$ is interpreted as susceptibility weighted by the probability of mild infection and relative infectiousness/reporting probability of mild infection. Notice that this argument extends to arbitrary number of groups, and that q is the health-care seeking probability in a reference class.

Often, $\mathbf{1} - \mathbf{x}(0) = \mathbf{1} - \mathbf{x}_0$ is interpreted as the fraction that is immune at the onset of the epidemic. Next to the basic reproduction number, the infectious period, and reporting probabilities this will be a central quantity that we aim to estimate. Now, if we let $\mathbf{x}^* = diag(\mathbf{x}_0)^{-1}\mathbf{x}$ and $\mathbf{y}^* = diag(\mathbf{x}_0)^{-1}\mathbf{y}$, then the model Eq (1) can be written as

$$\begin{aligned} \dot{\mathbf{x}}^* &= -\mathcal{R}_0 \; diag(\mathbf{x}^*) \; \mathbf{C} \; diag(\mathbf{x}_0) \; \mathbf{y}^* \\ \dot{\mathbf{y}}^* &= \mathcal{R}_0 \; diag(\mathbf{x}^*) \; \mathbf{C} \; diag(\mathbf{x}_0) \; \mathbf{y}^* - \mathbf{y}^* \; , \end{aligned}$$

and $\mathbf{x}_0$ can be interpreted as the fraction that is not fully immune, or as a reduction in infectiousness ($diag(\mathbf{x}_0)\ \mathbf{y}^*$).

Altogether, the above shows that under certain assumptions and with proper interpretation of the variables, the seemingly simple model Eq (1) has broad applicability, including situations with mild infection, variable infectiousness, and variable probability of reporting for infections with different severity [3]. The above reasoning also implies that the additional parameters of the more complex model (p and r) are not identifiable in the absence of information on the fraction of infections that are mild or asymptomatic. In the application to RSV we will further add age-specific differences in reporting.

Immunity propagation and demographic transitions

Next, we include demographic transitions and immunity propagation into the model [4, 5, 6, 7]. To this end, we separate the epidemic occurring in winter from demographic turnover and immunity losses occurring throughout the year. Starting from initial conditions in a given year describing the fraction of the population that is immune in different age classes, the epidemic is modelled by the ODEs given in Eq (1). At the end of the epidemic, susceptibility in the population will have decreased and immunity will have increased. Subsequently, losses of immunity and demographic transitions between seasons are modelled with a discrete mapping, yielding the susceptibility (initial conditions) for the next epidemic. To allow for year-to-year variations in susceptibility at the start of the epidemic (e.g., owing to variation in the rate of viral evolution), the actual age-specific susceptibility $\mathbf{x}_0$ is modelled as $\mathbf{x}_0 \sim \text{Beta}\left(\tilde{\mathbf{x}}_0\boldsymbol{\rho}, \left(\mathbf{1}-\tilde{\mathbf{x}}_0\right)\boldsymbol{\rho}\right)$, where $\tilde{\mathbf{x}}_0$ is the calculated susceptibility, and $\boldsymbol{\rho}$ the age-specific rate parameters. To prevent over-complicating the analyses, we further assume a stable uniform population distribution, so that in an age class of width w years a fraction $1/w$ is moved from that class to the next between epidemic seasons. With regard to immunity propagation, we assume that in each age class a a fraction f_a of persons with immunity retain

their immunity, and the remainder become susceptible at the start of the next RSV season. Hence, the mean duration of immunity in age class a (in the absence of demographic transitions) is given by $\frac{1}{1-f_a}$.

Preliminary analyses showed that main parameters, viz. reproduction number, infectious period and reporting rates are strongly correlated (Supplementary Information). Therefore, we include as a sensitivity analysis an additional scenario in which the infectious period is not estimated but where the mean duration is fixed at 1 week. We compare the results of the two scenarios using the WBIC information criterion [8]), and compare the parameter estimates and outcomes in terms of biological plausibility.

4 Observation model

Next we specify how observations come about. In the transmission model, each incident infection in age group a has an age-specific probability p_a^{GP} to be reported as an ARI case at the GP, and an (independent) age-specific probability p_a^{hosp} to be reported as a confirmed RSV case in the hospital. Not all ARIs at the GP are caused by RSV, however, and we add a function $b_a^{ARI}(t)$ for all other causes of ARI at the GP. These other sources include influenza, rhinovirus, enterovirus, and others, and can be highly variable between years, within a a year, and between age groups. To accommodate this variability we fit a flexible cubic spline to each epidemic season and age group. To keep the computational burden within reasonable bounds we assume that each spline is characterized by three knots (and hence 5 B-splines), yielding 5 (seasons) * 7 (age groups) * 5 (B-splines) = 175 spline weights to be estimated. We further assume that spline weights have gamma prior distributions with means given by the mean incidence of ARI in a given RSV epidemic and age group (giving a distinct empirical Bayesian flavor to the analyses), and with a single rate parameter θ that is estimated. In this manner, the analysis conservatively attributes ARI to RSV. Hence, the prior distribution for i-th weight in epidemic season y and age group a, $b_{y,a,i}$ is given by $b_{y,a,i} \sim \text{Gamma}\left(\rho\overline{ARI_{y,a}},\ \rho\right)$, where $\overline{ARI_{y,a}}$ is the mean incidence in year y and age group a, and ρ is the rate parameter.

Further, we assume that the number of RSV hospitalisations in a given week and age group is binomially distributed with totals given by the size of the age group and parameter given by the (modelled) incidence of infection multiplied by probability of hospitalisation. Likewise, the number of ARI case is binomially distributed with totals given by the size of the relevant catchment population and parameter given by sum of the ARI incidence caused by RSV and other sources. Figure S1 gives an overview of the model structure.

5 Parameter estimation

Parameters are estimated in a Bayesian framework using Hamlitonian Monte Carlo as implemented in Stan [9]. Main parameters are the basic reproduction number ($\mathcal{R}_0$), the infectious period (D), the reporting probabilities for hospitalisation and GP consultation in different age groups (p_a^{hosp} and p_a^{GP}), the age-specific probabilities that immunity is retained from one epidemic to the next (f_a), and the spline weights ($b_{y,a,i}$). Since data is scarce in non-aged adults, we have lumped reporting probabilities in the age groups 5-9 yr, 10-19 yr, 20-44 yr, and 45-64 yr. In addition, we estimate the rate parameter of the gamma prior distribution for the B-splines generating the background ARI.

To prevent unrealistic parameter excursions with GP reporting rates higher in age groups other than infants, we take for those other age groups $p_a^{\text{GP}} = g_a^{\text{GP}} \cdot p_{[0,1)}^{\text{GP}}$, where $0 \leq g_a^{\text{GP}} \leq 1$ ($a \in \{[1,5), [5,65), [65,100)\}$) is the reduction in the probability of reporting in age group a as compared to infants. For consistency we have modelled both hospital and GP reporting in this way. Prior distributions for the reduction in GP consultations are $g_a^{\text{GP}} \sim \text{Beta}\,(1,9)$, to incorporate our prior belief that the probability of reporting should be substantially lower in children and adults than in infants.

Estimates of the infectious period of RSV are scarce and variable. The Centers for Disease Control and Prevention reports that the infectious period varies from 3 to 8 days, while the European Centre for Disease Prevention and Control reports a wide range from a couple of days up to 27 days [10].

A recent study with dense sampling of Kenyan persons suggest an infectious period of approximately 1.5 weeks [11]. Here, we take a Gamma prior distribution with most mass between 0.5 and 1 week: $D \sim \text{Gamma}\,(30, 40)$ (shape and rate, mean 0.75 wk). Further, $\mathcal{R}_0 \sim \text{U}\,(1, 50)$, $D \sim \text{U}\,(0.28, 4)$ weeks (Supplementary Information for alternative scenario), and inocula seeding the epidemics are given $\text{U}\,(e^{-5}, e^{-3})$ prior distributions. Rate parameters of beta distributions are equipped with $\text{Pareto}\,(2.0, 1.5)$ prior distributions [12], and all other parameters have uniform prior distributions their domains. Details are given in the Supplementary Information.

6 Tables and figures

Tables S1-S2 and Figures S1-S3 below give the results presented and discussed in the main text.

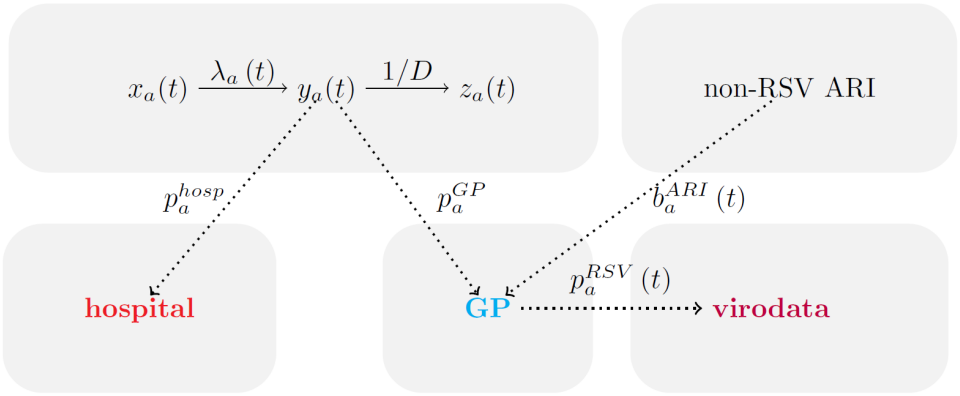

Figure S1. Overview of model structure and data. The boxes on top show the transmission model (top left-hand) and other sources of ARI reported at the GP (top right-hand). The boxes at the bottom show the data sources. Dotted arrows depict relations between model and data.

Parameter	2.5%	50%	97.5%
$\mathcal{R}_0$	20.1	22.9	25.7
D (wk)	2.2	2.5	2.8
p_0^{hosp}	0.013	0.014	0.015
p_{1-4}^{hosp}	0.0013	0.0014	0.0017
p_{5-64}^{hosp}	0.00058	0.00070	0.00084
p_{65+}^{hosp}	0.0012	0.0014	0.0017
p_0^{GP}	0.21	0.23	0.25
p_{1-4}^{GP}	0.17	0.19	0.21
p_{5-64}^{GP}	0.14	0.17	0.20
p_{65+}^{GP}	0.12	0.15	0.18
f_0	0.024	0.47	0.97
f_{1-4}	0.10	0.86	0.99
f_{5-9}	0.67	0.94	1.0
f_{10-19}	0.76	0.94	1.0
f_{20-44}	0.78	0.95	1.0
f_{45-64}	0.70	0.93	0.98
f_{65+}	0.64	0.85	0.96

Table S1. Parameter estimates of the default scenario. Estimates are represented by posterior medians and 2.5%-97.5% posterior quantiles. $\mathcal{R}_0$ and D are the basic reproduction number and infectious period, p_a^{hosp} and p_a^{GP} are the probabilities of hospitalisation and GP consultation in age group a, and f_a is the fraction that retains its immunity from one epidemic to the next.

Age group	Infection attack rate	Vaccination impact
No intervention		
0 yr	0.84 (0.73, 0.94)	NA
1-4 yr	0.24 (0.16, 0.42)	NA
5-9 yr	0.085 (0.024, 0.23)	NA
65+ yr	0.13 (0.040, 0.34)	NA
Maternal vaccination		
0 yr	0.62 (0.52, 0.70)	0.27 (0.25, 0.30)
1-4 yr	0.27 (0.19, 0.44)	-0.10 (-0.31, -0.0029)
5-9 yr	0.088 (0.022, 0.24)	-0.0022 (-0.038, -0.032)
65+ yr	0.13 (0.033, 0.34)	0.00026 (-0.013, 0.010)
Pediatric vaccination		
0 yr	0.59 (0.47, 0.69)	0.30 (0.27, 0.36)
1-4 yr	0.19 (0.12, 0.31)	0.24 (0.20, 0.27)
5-9 yr	0.078 (0.019, 0.23)	0.078 (-0.0020, 0.21)
65+yr	0.14 (0.031, 0.36)	0.0025 (-0.011, 0.013)

Table S2. Impact of vaccination. Shown are the infection attack rates (AR) in different age groups (95%CrI between brackets) in a scenario without intervention and with maternal or pediatric vaccination. Results are based on 1,000 samples from the posterior distribution 20 years after initiation of the vaccination program. Infection attack rates are shown for the main age groups with high RSV circulation. Within each age group vaccination impact is calculated as 1 minus the ratio of the infection attack rates in the scenario with and without intervention. NA: not applicable.

Figure S2. Weekly age-stratified number of hospitalisations with confirmed RSV infection in the Netherlands (dots) with model fit (lines) in infants (under 1 year), young children (1-5 years), and older adults (over 65 years). Number of hospitalisations are very low in all other age groups (see below). Results are shown for the 2012/2013 up to and including the 2016/2017 epidemics (left to right). Bold lines correspond to posterior medians, and shaded areas represent the 95% credible ranges of the posterior distribution.

Figure S3. Weekly age-stratified fraction of virological samples that are positive for RSV (dots) with model fit of the probability that an ARI presenting at the GP is caused by RSV (lines). Bold lines correspond to posterior medians, and shaded areas represent 95% credible ranges of the posterior distribution. Size of the dots indicate the sample size.

7 Extended results

Here we provide additional results for the default scenario presented in the main text and the variant scenario with fixed infectious period, including trace plots, plots of pairwise correlations for key parameters, and figures of data and model fits for hospitalisations and GP consultations for all age groups. We also provide tables with more complete results of parameters estimates. Finally, for both scenarios we also report (an estimate of) the widely applicable Bayesian information criterion (WBIC) to be able to select the more likely data generating process [8].

The model has been coded in Stan [9], and is available on reasonable request.

In both scenarios, we run 10 chains of 2,000 iterations in parallel. The first 1,000 iterations (warmup) are discarded. We then sample every tenth iteration, yielding a total of 1,000 samples for both scenarios. Convergence of chains is assessed visually by inspection of trace plots and pair plots, and by assessment of the empirical variance within and between chains (Rhat) [13]. To prevent the occurrence of divergent transitions we set `adapt_delta = 0.97` and take `max_treedepth = 15` for more efficient sampling.

7.1 Default scenario

Table S3 shows estimates of transmission parameters and reporting rates with convergence diagnostics. For all parameters Rhat is close to 1, and the effective sample size is usually close to the actual sample size, suggesting adequate convergence. This is corroborated by Figures S4 and S5, which show trace plots of main parameters, and pair plots of the basic reproduction number, infectious period, and reporting probabilities for hospitalisations and GP consultations in infants. Interestingly, the pair plots show a positive correlation between the basic reproduction number and infectious period, and strong negative correlations between the basic reproduction number/infectious period and reporting probabilities, especially with regard to hospitalisations.

Figures S6 and S7 give results of model fits to hospitalisations and GP consultations for all age classes (cf. Figure 1 and Figures S2 and S3). Figure S6 shows that RSV coded hospitalisations occur very infrequently (usually less than 10 per week) in persons 5-44 years, even at the peak of RSV epidemics. In persons 45-65 years the number of cases is somewhat higher, although it should be noted that this age group is relatively large in the Netherlands. Figure S7 shows the model fit to GP consultations for ARI in the Netherlands. In all epidemics and age groups, GP consultations for ARI peak in winter, but in most age groups only a small fraction is ascribed to RSV, due to the very low fraction of ARI samples testing positive for RSV.

Finally, Table S4 shows, for each of the epidemics, estimates of the infection attack rates in different age groups. Infection attack rates are low in most age groups, but invariably very high in infants and, to a lesser extent, also

in 1-4 year-old children. In the 2015-2016 and 2016-2017 epidemics infection attack rates also seem increased in older adults (65 years and older) and to a lesser extent also in persons 45-65 years.

Parameter	Rhat	n_eff	mean	sd	2.5%	50%	97.5%
$\mathcal{R}_0$	1.00	856	22.9	1.4	20.2	22.9	25.7
D (wk)	1.00	928	2.5	0.15	2.2	2.5	2.8
p_0^{hosp}	1.00	895	0.014	0.00049	0.013	0.014	0.015
p_{1-4}^{hosp}	1.00	938	0.0015	0.000099	0.0013	0.0014	0.0017
p_{5-64}^{hosp}	1.00	1046	0.00070	0.000067	0.00058	0.00070	0.00084
p_{65+}^{hosp}	1.00	935	0.0014	0.00014	0.0012	0.0014	0.0017
p_0^{GP}	1.00	943	0.23	0.010	0.21	0.23	0.25
p_{1-4}^{GP}	1.00	920	0.19	0.011	0.17	0.19	0.21
p_{5-64}^{GP}	0.99	990	0.17	0.015	0.14	0.17	0.20
p_{65+}^{GP}	1.00	1047	0.15	0.016	0.12	0.15	0.18
f_0	1.00	1118	0.49	0.28	0.024	0.47	0.97
f_{1-4}	0.99	1088	0.84	0.10	0.60	0.86	0.99
f_{5-9}	1.00	956	0.92	0.067	0.76	0.94	1.0
f_{10-19}	1.01	941	0.92	0.065	0.76	0.94	1.0
f_{20-44}	1.00	796	0.93	0.062	0.78	0.95	1.0
f_{45-64}	1.00	1079	0.90	0.075	0.70	0.93	0.98
f_{65+}	1.00	988	0.84	0.083	0.64	0.85	0.96

Table S3. Parameter estimates for the default scenario (cf. main text). For all parameters the R-hat convergence diagnostic is close to 1*, and effective sample sizes are approximately* $1,000$*. Estimates are represented by posterior medians and 2.5%-97.5% posterior quantiles.* $\mathcal{R}_0$ *and* D *are the basic reproduction number and infectious period,* p_a^{hosp} *and* p_a^{GP} *are the probabilities of hospitalisation and GP consultation in age group* a*, and* f_a *is the fraction that retains its immunity from one epidemic to the next.*

7.2 Fixed infectious period

In the default scenario estimates of the infectious period are in excess of two weeks, which is on the high end of what is considered reasonable. High values for estimates of the infectious period tend to come with high estimates for the reproduction number to get the timing of the epidemic right (see the pair plots of the default scenario), and consequently with relatively low reporting rates for GP consultations and hospitalisations. To investigate how the results would be affected if the infectious period would be lower, we analyzed a scenario in which the infectious period is fixed at 1 week.

Table S5 shows estimates of transmission parameters and reporting rates with convergence diagnostics. Again, for all parameters Rhat is close to 1, and the effective sample size is close to the actual sample size, suggesting adequate convergence. Figures S8 and S9 show trace plots of the main parameters and pair plots for selected parameters, while Figures S10 and S11 show the model fits to hospitalisations and GP consultations for all age classes. Superficially, model fits are close to the default model scenario.

Interestingly, even though model fits of hospitalisations and GP consultations are similar in the default scenario and variant scenario, parameter estimates show substantial differences. Specifically, in the scenario with fixed infectious period, estimates of the basic reproduction number are much lower and estimates of the reporting probabilities are much higher (Table S5). In addition, infection attack rates in the scenario with fixed infectious period are much lower in infants and older adults (65 years and older) than in the default scenario, while infection attack rates rates are much higher in children 1-4 years. Similar results are obtained when the infectious period is fixed at 1 week, and in addition informative prior distributions are used for the (age-specific) fractions of the population that retain their immunity (not shown).

Comparison of both scenarios with WBIC indicates that the default model is favoured over the variant models on statistical grounds (WBIC $= 32,177$ versus WBIC $= 32,286$ in the scenario without informative priors). The variant models also yield unrealistically low infection attack rates in infants (well under 40%), thus suggesting that the default model should also be

favoured on biological grounds.

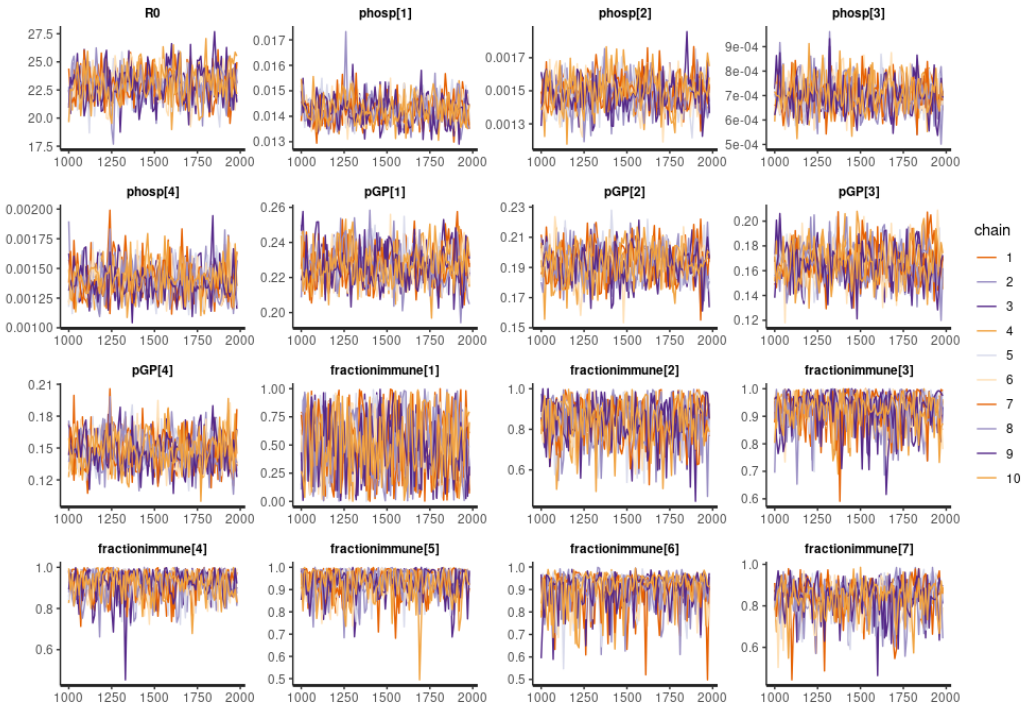

Figure S4. Trace plots of the main parameters for the default scenario. Shown are trace plots of the basic reproduction number, the age-specific probabilities of hospitalisation and GP consultation, and age-specific fractions of the population that retain their immunity from one epidemic to the next.

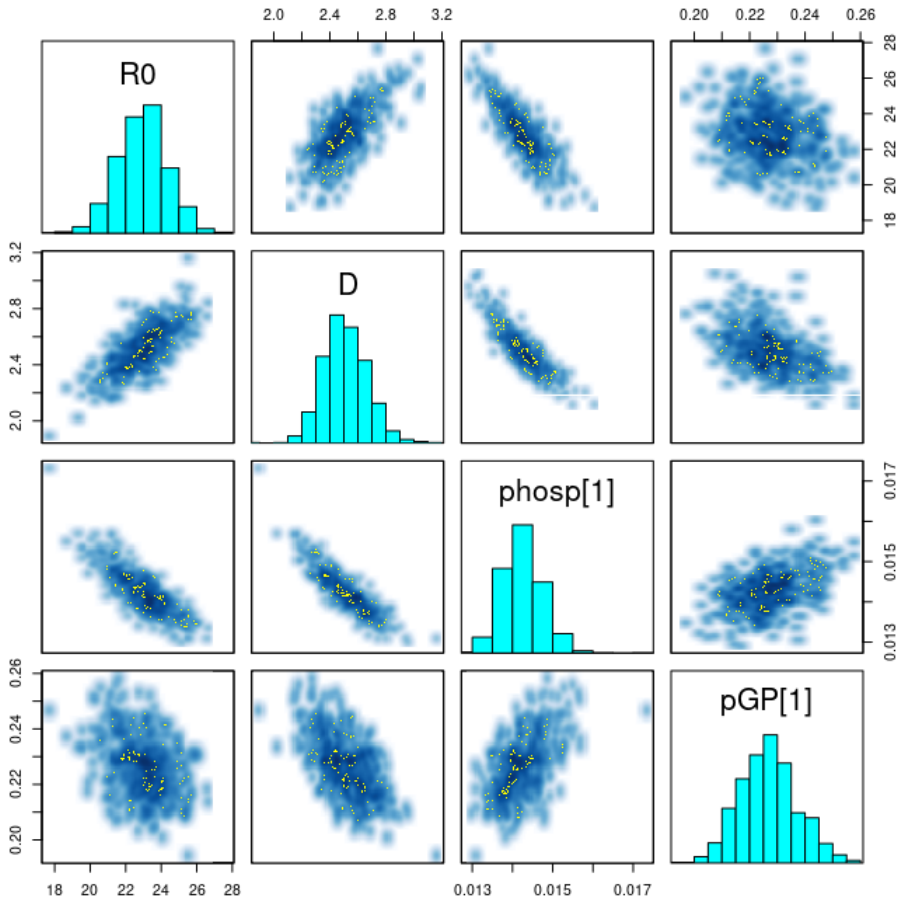

Figure S5. Pair plots of selected parameters for the default scenario.

Parameter	Rhat	n_eff	mean	sd	2.5%	50%	97.5%
AR_2012[1]	1.00	892	0.73	0.02	0.69	0.73	0.78
AR_2012[2]	1.01	931	0.30	0.02	0.27	0.30	0.34
AR_2012[3]	1.00	1060	0.01	0.00	0.00	0.01	0.01
AR_2012[4]	1.00	906	0.01	0.00	0.00	0.01	0.01
AR_2012[5]	1.00	1080	0.01	0.00	0.01	0.01	0.01
AR_2012[6]	1.00	1034	0.03	0.00	0.02	0.03	0.04
AR_2012[7]	1.01	908	0.03	0.00	0.02	0.03	0.04
AR_2013[1]	1.00	883	0.69	0.02	0.64	0.69	0.73
AR_2013[2]	1.00	993	0.25	0.02	0.22	0.25	0.28
AR_2013[3]	1.00	981	0.01	0.00	0.01	0.01	0.02
AR_2013[4]	1.00	925	0.00	0.00	0.00	0.00	0.01
AR_2013[5]	1.01	964	0.01	0.00	0.00	0.01	0.01
AR_2013[6]	1.01	897	0.03	0.00	0.02	0.03	0.03
AR_2013[7]	1.00	867	0.03	0.00	0.03	0.03	0.04
AR_2014[1]	1.00	885	0.67	0.03	0.62	0.67	0.73
AR_2014[2]	1.00	984	0.19	0.01	0.16	0.19	0.22
AR_2014[3]	1.00	1173	0.03	0.01	0.02	0.03	0.04
AR_2014[4]	1.01	1072	0.00	0.00	0.00	0.00	0.01
AR_2014[5]	1.00	880	0.01	0.00	0.01	0.01	0.01
AR_2014[6]	1.01	984	0.05	0.01	0.04	0.05	0.06
AR_2014[7]	1.00	899	0.06	0.01	0.05	0.06	0.08
AR_2015[1]	1.00	903	0.70	0.02	0.66	0.70	0.74
AR_2015[2]	1.00	967	0.21	0.01	0.18	0.21	0.24
AR_2015[3]	1.00	1065	0.01	0.00	0.01	0.01	0.02
AR_2015[4]	1.00	944	0.01	0.00	0.00	0.01	0.01
AR_2015[5]	1.00	989	0.01	0.00	0.01	0.01	0.02
AR_2015[6]	1.00	1074	0.05	0.01	0.04	0.05	0.06
AR_2015[7]	1.00	966	0.11	0.01	0.09	0.11	0.13
AR_2016[1]	1.00	902	0.79	0.02	0.75	0.79	0.84
AR_2016[2]	1.00	1005	0.32	0.02	0.29	0.32	0.36
AR_2016[3]	0.99	1137	0.01	0.00	0.01	0.01	0.02
AR_2016[4]	1.00	1121	0.02	0.00	0.01	0.02	0.03
AR_2016[5]	1.00	1032	0.01	0.00	0.01	0.01	0.02
AR_2016[6]	1.01	1024	0.07	0.01	0.06	0.07	0.08
AR_2016[7]	1.00	985	0.18	0.02	0.15	0.18	0.21

Table S4. Estimates of the infection attack rates for the default scenario in different years and age groups. Age groups (indexed by square brackets) are as in the main text.

Parameter	Rhat	n_eff	mean	sd	2.5%	50%	97.5%
$\mathcal{R}_0$	1.00	872	5.7	0.12	5.5	5.7	6.0
p_0^{hosp}	1.00	702	0.036	0.00097	0.034	0.036	0.038
p_{1-4}^{hosp}	1.00	944	0.00076	0.00003	0.00071	0.00076	0.00081
p_{5-64}^{hosp}	1.01	506	0.00073	0.00022	0.00048	0.00066	0.00133
p_{65+}^{hosp}	1.01	659	0.0014	0.00026	0.00092	0.0014	0.0019
p_0^{GP}	1.00	929	0.43	0.031	0.37	0.43	0.50
p_{1-4}^{GP}	1.00	880	0.12	0.0069	0.11	0.12	0.14
p_{5-64}^{GP}	1.01	534	0.16	0.046	0.10	0.15	0.28
p_{65+}^{GP}	1.00	720	0.15	0.026	0.099	0.14	0.20
f_0	1.00	1017	0.48	0.29	0.023	0.48	0.97
f_{1-4}	1.00	1082	0.29	0.19	0.018	0.26	0.75
f_{5-9}	1.00	1022	0.94	0.058	0.78	0.95	0.99
f_{10-19}	1.00	925	0.91	0.078	0.72	0.93	1.0
f_{20-44}	1.00	962	0.92	0.070	0.75	0.95	1.0
f_{45-64}	1.00	1050	0.87	0.095	0.63	0.90	0.98
f_{65+}	1.00	931	0.70	0.15	0.38	0.72	0.95

Table S5. Parameter estimates for the variant scenario with infectious period of fixed expected duration. For all parameters the R-hat convergence diagnostic is close to 1*, and effective sample sizes are approximately* $1,000$*. Estimates are represented by posterior medians and 2.5%-97.5% posterior quantiles.* $\mathcal{R}_0$ *is the basic reproduction number,* p_a^{hosp} *and* p_a^{GP} *are the probabilities of hospitalisation and GP consultation in age group* a*, and* f_a *is the fraction that retains its immunity from one epidemic to the next.*

Parameter	Rhat	n_eff	mean	sd	2.5%	50%	97.5%
AR_2012[1]	1.00	635	0.29	0.01	0.27	0.29	0.32
AR_2012[2]	1.00	957	0.52	0.01	0.49	0.52	0.54
AR_2012[3]	1.01	627	0.01	0.00	0.00	0.01	0.01
AR_2012[4]	1.00	894	0.01	0.00	0.00	0.01	0.01
AR_2012[5]	1.00	610	0.02	0.01	0.01	0.02	0.03
AR_2012[6]	1.01	570	0.04	0.01	0.02	0.04	0.06
AR_2012[7]	1.01	559	0.04	0.01	0.02	0.04	0.06
AR_2013[1]	1.00	676	0.27	0.01	0.26	0.27	0.29
AR_2013[2]	1.00	769	0.49	0.01	0.46	0.49	0.51
AR_2013[3]	1.00	794	0.02	0.01	0.01	0.02	0.03
AR_2013[4]	1.00	878	0.00	0.00	0.00	0.00	0.01
AR_2013[5]	1.00	704	0.01	0.00	0.00	0.01	0.01
AR_2013[6]	1.01	595	0.03	0.01	0.01	0.03	0.05
AR_2013[7]	1.00	662	0.04	0.01	0.02	0.04	0.06
AR_2014[1]	1.01	704	0.27	0.01	0.25	0.27	0.28
AR_2014[2]	1.00	904	0.44	0.01	0.42	0.44	0.46
AR_2014[3]	1.00	899	0.03	0.01	0.01	0.03	0.05
AR_2014[4]	1.00	950	0.00	0.00	0.00	0.00	0.01
AR_2014[5]	1.00	628	0.01	0.00	0.01	0.01	0.02
AR_2014[6]	1.01	597	0.05	0.01	0.02	0.05	0.07
AR_2014[7]	1.01	642	0.07	0.01	0.04	0.06	0.09
AR_2015[1]	1.00	677	0.27	0.01	0.25	0.27	0.28
AR_2015[2]	0.99	982	0.42	0.01	0.40	0.42	0.44
AR_2015[3]	1.00	790	0.01	0.00	0.01	0.01	0.02
AR_2015[4]	1.00	781	0.01	0.00	0.00	0.01	0.01
AR_2015[5]	1.01	645	0.01	0.00	0.01	0.01	0.02
AR_2015[6]	1.01	563	0.05	0.01	0.03	0.06	0.08
AR_2015[7]	1.01	613	0.11	0.02	0.08	0.11	0.16
AR_2016[1]	1.00	817	0.33	0.01	0.32	0.33	0.34
AR_2016[2]	1.00	939	0.57	0.01	0.54	0.57	0.58
AR_2016[3]	1.00	797	0.01	0.01	0.01	0.01	0.03
AR_2016[4]	1.00	725	0.02	0.00	0.01	0.02	0.03
AR_2016[5]	1.01	604	0.01	0.00	0.01	0.01	0.02
AR_2016[6]	1.01	586	0.07	0.01	0.04	0.07	0.09
AR_2016[7]	1.01	616	0.19	0.03	0.13	0.18	0.26

Table S6. Estimates of the infection attack rates for the variant scenario in different years and age groups. Age groups are as in the main text.

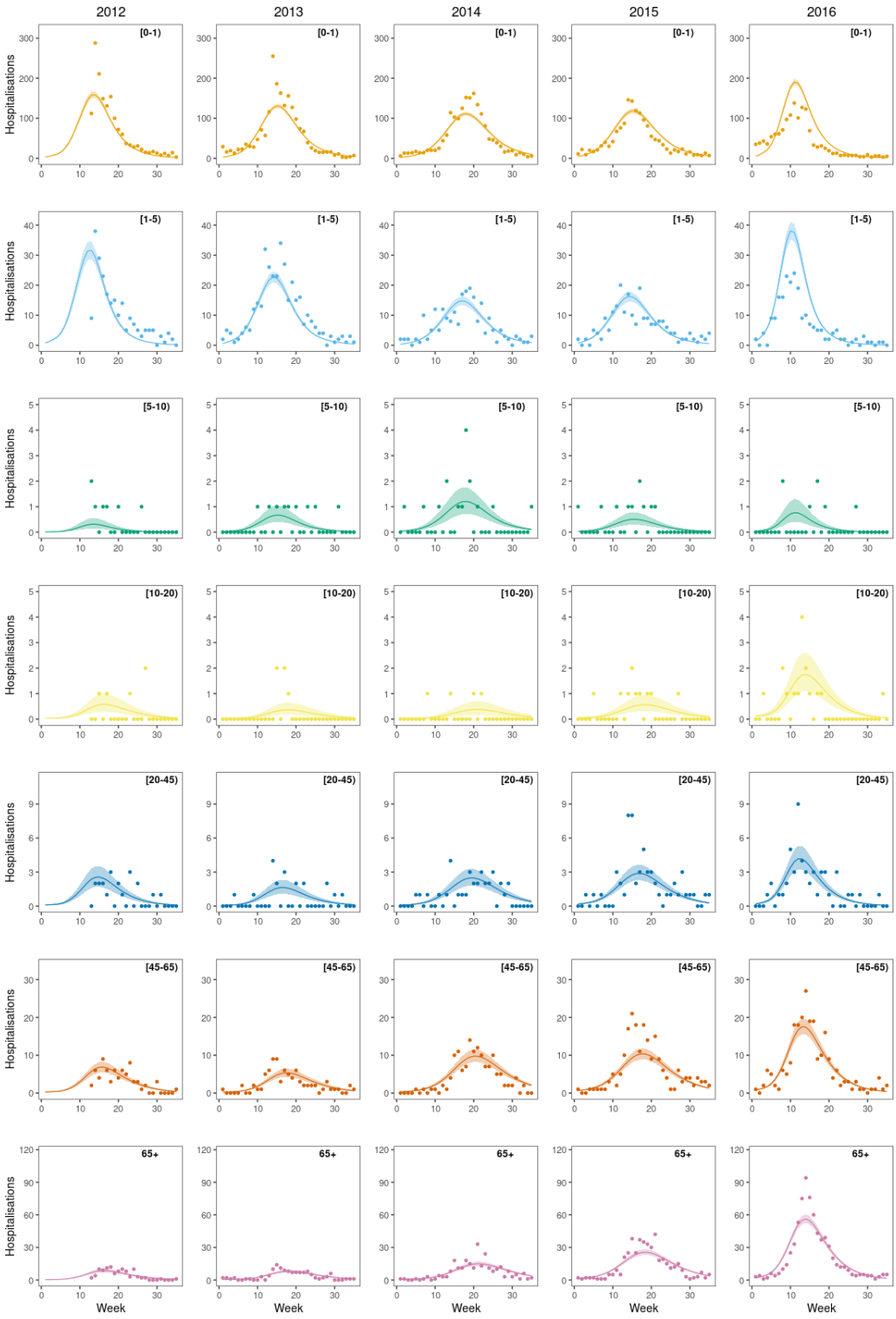

Figure S6. Weekly age-stratified number of hospitalisations with confirmed RSV infection in the Netherlands (dots) with fit of the default model (lines) for all age groups (Figure S2 for details).

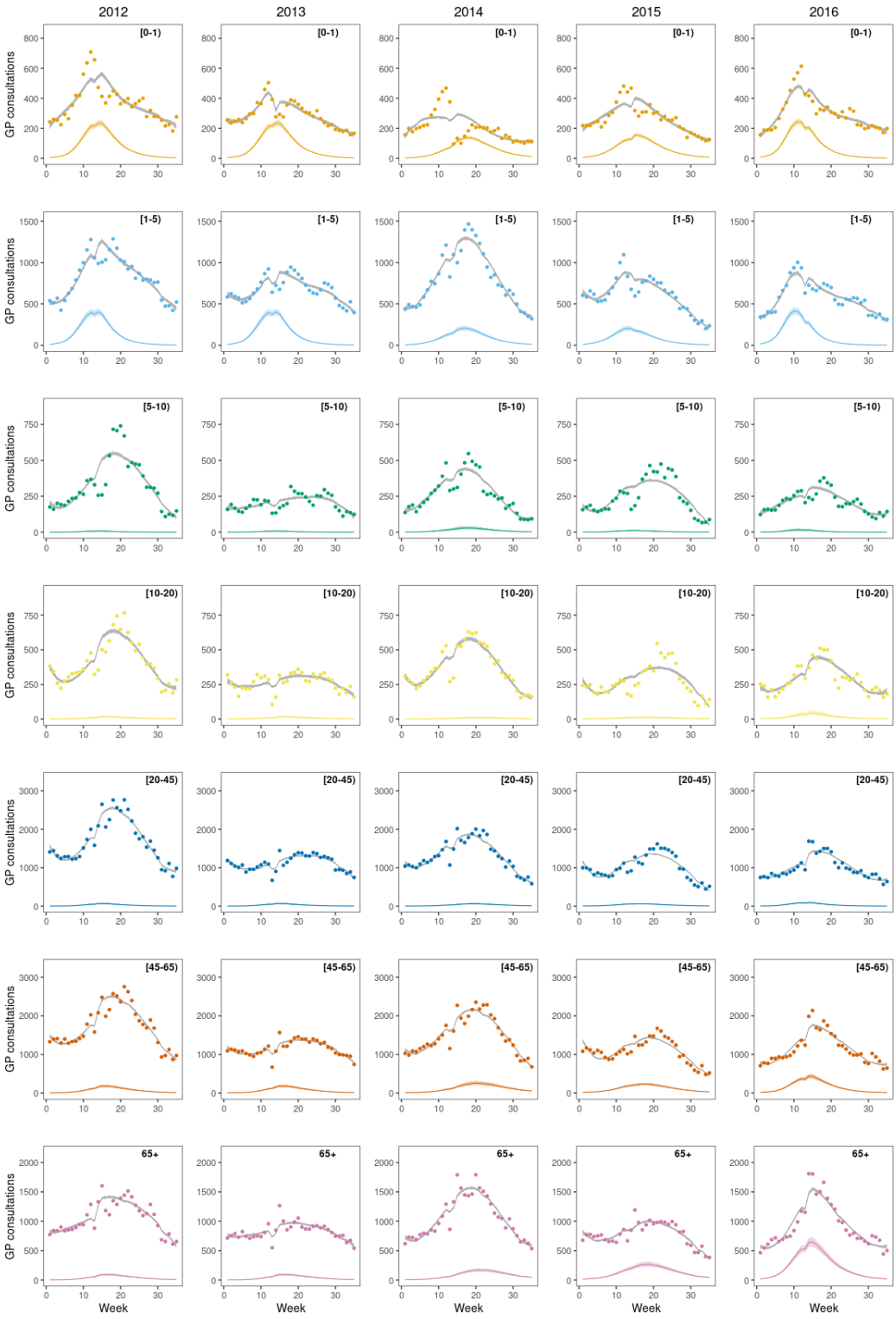

Figure S7. Weekly age-stratified number of GP consultations for ARI in the Netherlands (dots) with fit of the default model (lines) for all age groups (Figure 1 for details).

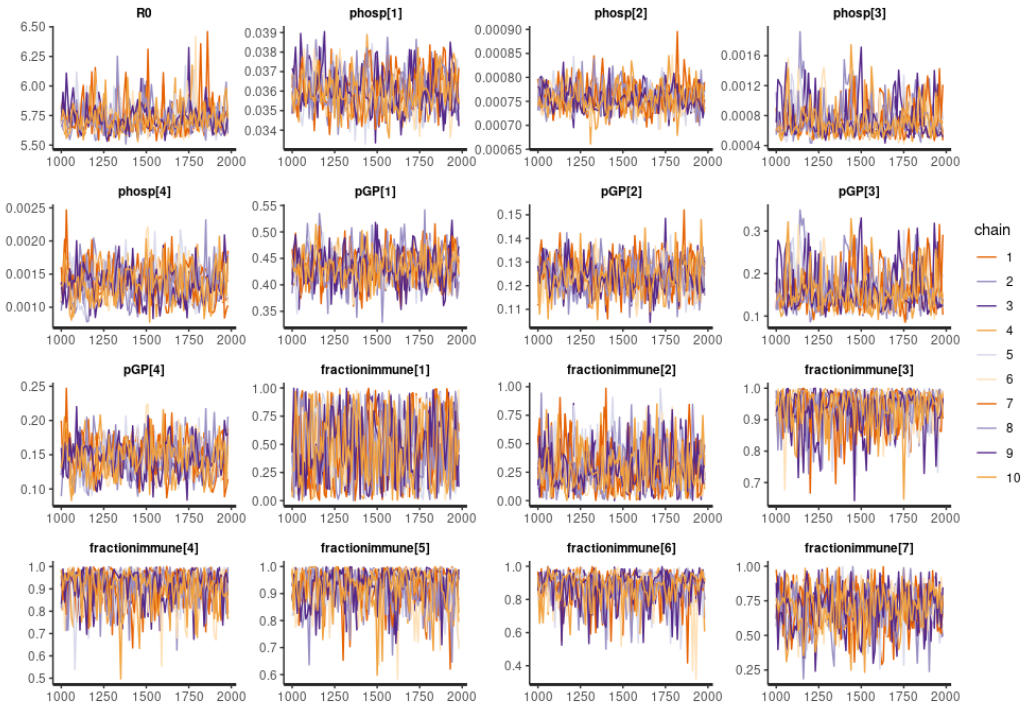

Figure S8. Trace plots of the main parameters for the variant scenario with fixed infectious period. See Figure S4 for details.

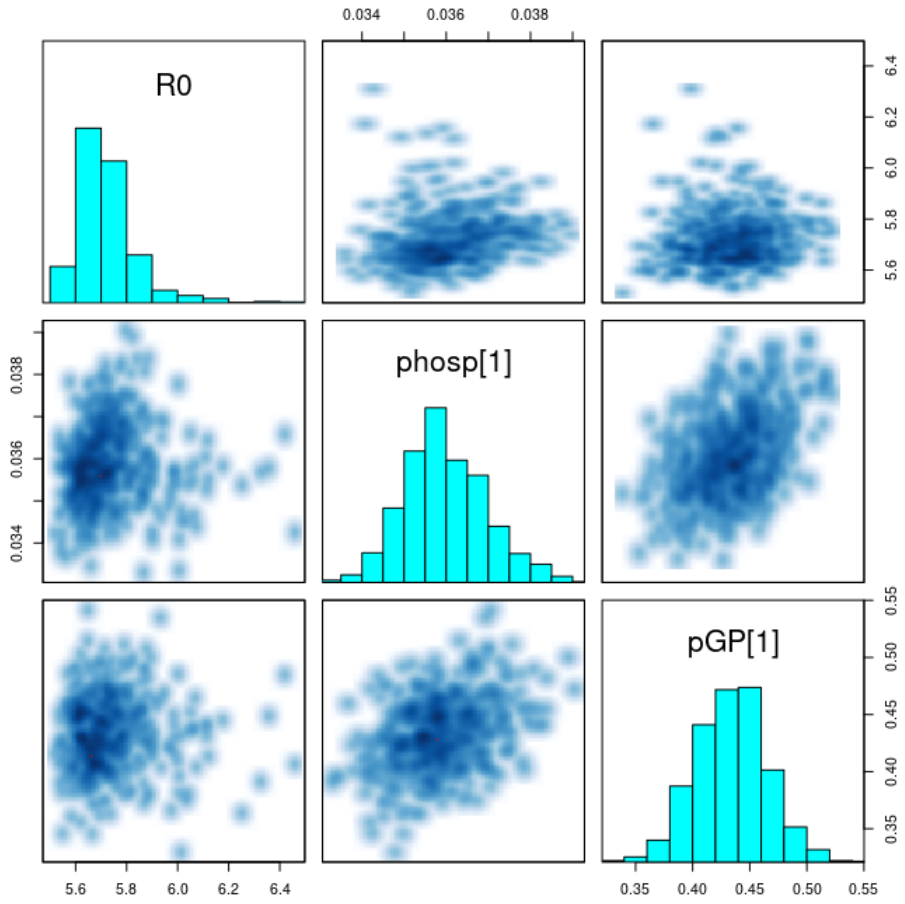

Figure S9. Pair plots of selected parameters for the variant scenario with fixed infectious period. See Figure S5 for details.

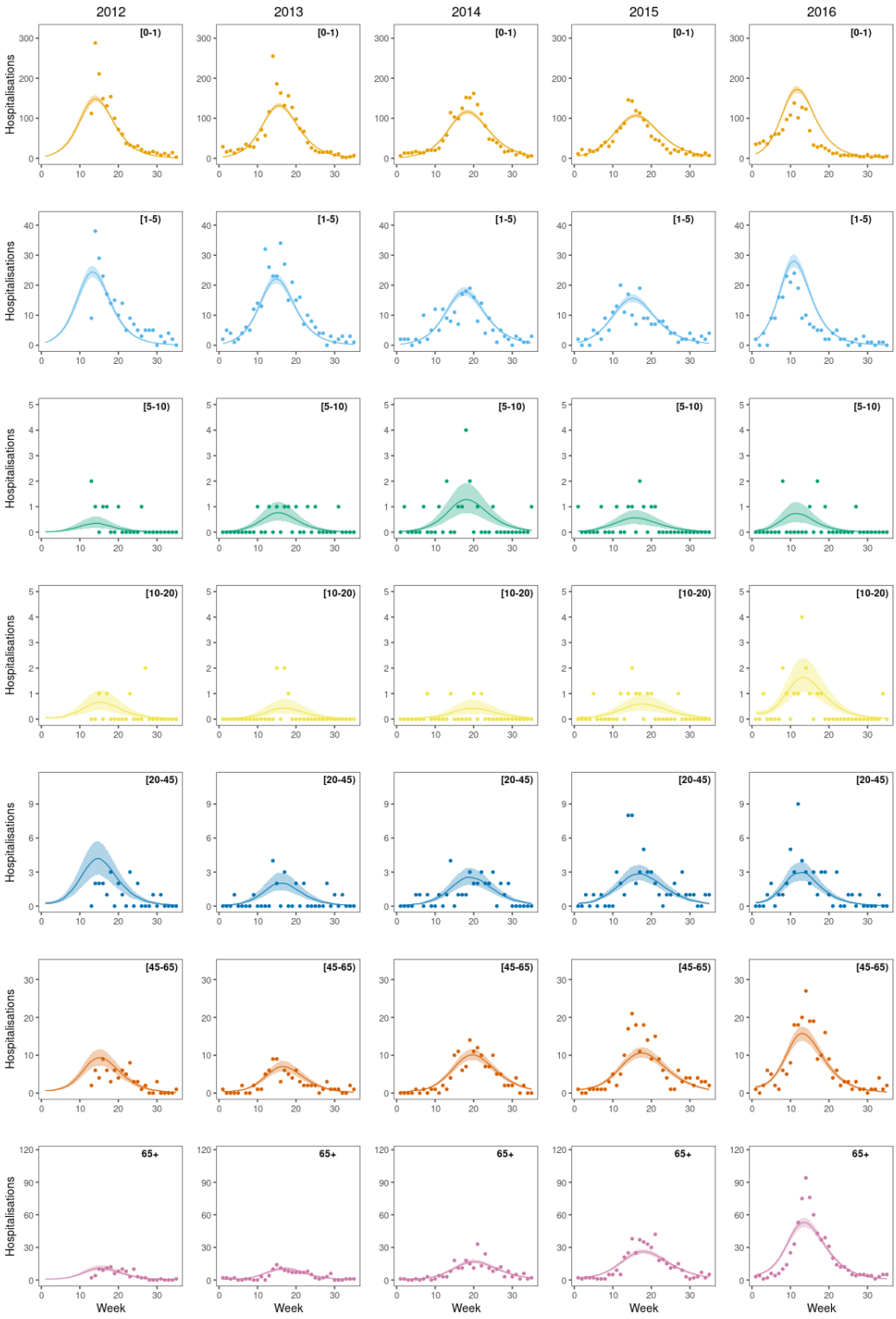

Figure S10. Weekly age-stratified number of hospitalisations with confirmed RSV infection in the Netherlands (dots) with fit of the variant model (lines) for all age groups.

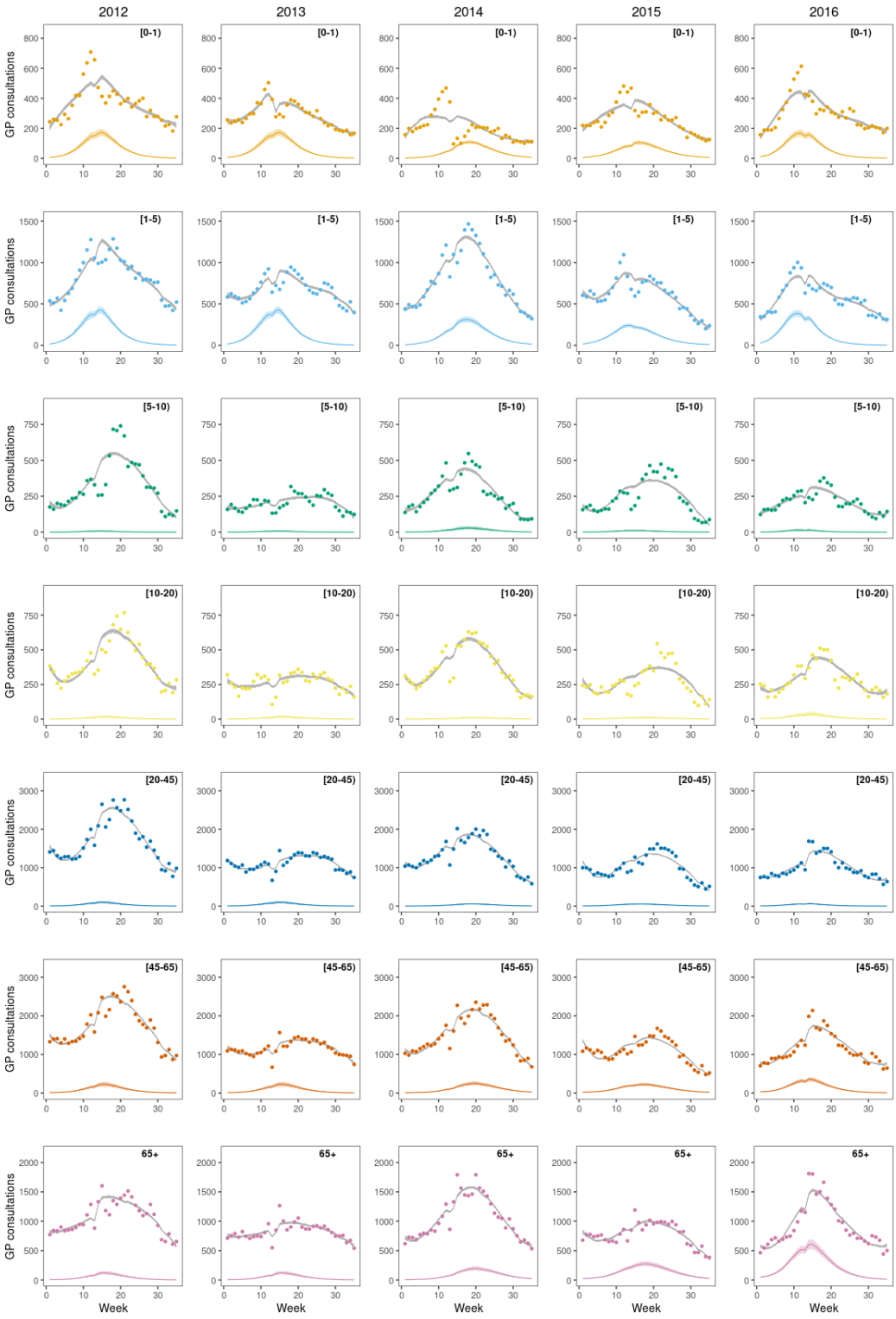

Figure S11. Weekly age-stratified number of GP consultations for ARI in the Netherlands (dots) with fit of the variant model (lines) for all age groups.

8 References

[1] L. M. Vos, A. C. Teirlinck, J. E. Lozano, T. Vega, G. A. Donker, A. I. Hoepelman, L. J. Bont, J. J. Oosterheert, and A. Meijer. Use of the moving epidemic method (MEM) to assess national surveillance data for respiratory syncytial virus (RSV) in the Netherlands, 2005 to 2017. *Euro Surveill.*, 24(20), May 2019.

[2] Jan van de Kassteele, Jan van Eijkeren, and Jacco Wallinga. Efficient estimation of age-specific social contact rates between men and women. *Ann. Appl. Stat.*, 11(1):320–339, 03 2017.

[3] Christiaan H. van Dorp, Rutger G. Woolthuis, Jeffrey H. C. Yu, Rob J. de Boer, and Michiel van Boven. Estimation of age-specific susceptibility to influenza in the netherlands and its relation to loss of cd8+ t-cell memory. *bioRxiv*, 2020.

[4] R. G. Woolthuis, J. Wallinga, and M. van Boven. Variation in loss of immunity shapes influenza epidemics and the impact of vaccination. *BMC Infect. Dis.*, 17(1):632, 09 2017.

[5] E. M. Hill, S. Petrou, S. de Lusignan, I. Yonova, and M. J. Keeling. Seasonal influenza: Modelling approaches to capture immunity propagation. *PLoS Comput. Biol.*, 15(10):e1007096, Oct 2019.

[6] J. A. Backer, J. Wallinga, A. Meijer, G. A. Donker, W. van der Hoek, and M. van Boven. The impact of influenza vaccination on infection, hospitalisation and mortality in the Netherlands between 2003 and 2015. *Epidemics*, 26:77–85, 03 2019.

[7] J. A. Backer, M. van Boven, W. van der Hoek, and J. Wallinga. Vaccinating children against influenza increases variability in epidemic size. *Epidemics*, 26:95–103, 03 2019.

[8] Sumio Watanabe. A widely applicable bayesian information criterion. *J. Mach. Learn. Res.*, 14(1):867–897, March 2013.

[9] Bob Carpenter, Andrew Gelman, Matthew Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. *Journal of Statistical Software, Articles*, 76(1):1–32, 2017.

[10] *Systematic review on the incubation and infectiousness/shedding period of communicable diseases in children.* ECDC, 2016.

[11] P. K. Munywoki, D. C. Koech, C. N. Agoti, N. Kibirige, J. Kipkoech, P. A. Cane, G. F. Medley, and D. J. Nokes. Influence of age, severity of infection, and co-infection on the duration of respiratory syncytial virus (RSV) shedding. *Epidemiol. Infect.*, 143(4):804–812, Mar 2015.

[12] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin. *Bayesian Data Analysis, Third Edition.* Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, 2013.

[13] Andrew Gelman and Donald B. Rubin. Inference from iterative simulation using multiple sequences. *Statist. Sci.*, 7(4):457–472, 11 1992.