A time-varying compartmental model of COVID-19 in the Netherlands, 2020–2024

Recovering the Effective Reproduction Number from Hospital Admissions

Leto Cavé

Written for Simuleren en Modelleren · BSc Mathematics · University of Amsterdam · January 2026

Revised and published September 2026

This work began as a group assignment for Simuleren en Modelleren, carried out by four students. The model, code and text presented here were rewritten from the ground up for this version and are my own.

Abstract

We ask how much of an epidemic's transmission history can be recovered from hospital admissions alone. A SIR/SIRS model with piecewise-constant transmission is fitted to daily Dutch COVID-19 admissions for 2020-02-27 to 2024-04-15, using no case, test or contact data, and its effective reproduction number is compared against the series RIVM published independently. The reconstruction succeeds — correlation 0.804, the two agreeing on whether the epidemic was growing on 81.5 % of intervals against 50.0 % for the best constant — but only at one model complexity. Choosing the number of control values by AIC gives 219 parameters, the closest fit to admissions of any model tried, and a reproduction number correlating with the published series at 0.173. Choosing by BIC gives 75 parameters, a measurably worse fit, and nearly four times better recovery. Over this range, goodness of fit on the observed series is an actively misleading guide to fidelity of the process beneath it.

Keywords: Compartmental model, SIRS, time-varying transmission, effective reproduction number, inverse problem, COVID-19, hospital admissions

Methodological Note on the Use of Generative AI:

Every line of the model code and of this paper, as presented here, is my own. Generative AI was used as a sparring partner while debugging and while discussing implementation choices. I take full responsibility for the mathematical reasoning, the numerical results and every assertion made here.

1 Introduction

The number a government needs during an epidemic is not the number of cases. It is the effective reproduction number R_t: the average number of people each currently infectious person goes on to infect. Above one the epidemic grows, below one it recedes, and the whole purpose of an intervention is to move that number across the line. But R_t cannot be measured. It has to be inferred from something that can be counted, and the choice of what to count determines how much the answer can be trusted.

Through most of the COVID-19 pandemic the quantity counted was positive tests. That series is badly behaved as evidence about transmission: in early 2020 Dutch testing capacity was scarce enough that the case count tracked laboratory throughput as much as infection, and thereafter it moved with public willingness to be tested, with the introduction and withdrawal of free testing, and with the rise of self-tests that were never centrally recorded. Hospital admissions are recorded for a different reason and are far less sensitive to any of this. A person admitted with COVID-19 is admitted because they are ill.

This paper asks a narrow question with a checkable answer: *how much of the transmission history of the Dutch epidemic can be recovered from hospital admissions alone?* Concretely — fitting only to the daily admission counts published by RIVM, never to any case, test or reproduction-number series — can one reconstruct an effective reproduction number that agrees with the one the national institute published from independent data?

A compartmental model is the right instrument for that question, and not merely the conventional one. Admissions are a lagged, thinned image of new infections, so smoothing or differentiating the admissions curve yields a statement about admissions and nothing more. Converting a transmission rate into a reproduction number requires the susceptible fraction $S(t)/N$, by Eq. (17), and $S(t)$ is not in any dataset: it is the accumulated consequence of every infection and every lapse of immunity since the epidemic began. Only a model that carries the population's infection history forward in time can supply it. That is what the compartments are for.

What is done here goes beyond simulating a textbook SIR system in three respects. First, the transmission rate is not assumed — it is estimated. The Dutch epidemic was reshaped repeatedly by policy, so β is treated as an unknown function of time, represented by 72 control values and recovered by solving an inverse problem rather than by forward simulation. Second, the map from latent infections to observable admissions is itself part of the model and is estimated alongside the dynamics, rather than being assumed away. Third, and most importantly, the result is checked against something the fit never saw.

That check is the paper's main result, and it succeeds. Fitted to admissions alone, the model reproduces RIVM's published R_t at a correlation of 0.804 over the period from 2020-06-01, agreeing on the direction of the epidemic — whether R_t stood above or below one — on 81.5 % of intervals against 50.0 % for the best constant predictor, and landing inside RIVM's own 95 % interval half the time. Compared day by day rather than at the model's own resolution the correlation is 0.639 and the directional agreement 75.5 %. The hospital-admission record, on its own, carries most of the transmission history.

That result comes with a condition attached, and the condition is the more useful finding. It holds only at the right model complexity, and the complexity that best explains the admissions data is not it. Choosing the number of control values by AIC gives 7-day buckets and 219 parameters; that model fits the admissions series better than any other tried, and recovers R_t at a correlation of 0.173. Choosing by BIC gives 21-day buckets and 75 parameters, fits the admissions series measurably worse, and recovers R_t four times better. Goodness of fit on the observed series and fidelity to the latent process it was supposed to reveal come apart, and over the range that matters here they point in opposite directions. Nothing internal to the fit distinguishes the two; only the independent series does.

Getting there required correcting the model as much as writing it up. Verification exposed three defects that changed the results: an integrator stepping blindly across the discontinuities in $\beta(t)$ (Sec. 4.2), a reporting delay that had been discretised so coarsely that it could not be estimated at all (Sec. 4.5), and a hundred days of missing data being fitted as if no one had been admitted (Sec. 4.4). Each is documented where it was found.

The rest of the paper is arranged as follows. Sec. 2 states the model and the assumptions it rests on. Sec. 3 derives what the system does before any data is involved: its reproduction number, its equilibria, what it conserves, and which classical results do not survive. Sec. 4 covers integration and estimation, Sec. 5 checks that the implementation solves the equations it claims to, and Sec. 6 reports the fit, the choice of how many control values to use, the comparison with RIVM, and two counterfactual sweeps. Sec. 7 asks how much of the conclusion survives uncertainty in the parameters that were never fitted, and Sec. 8 says plainly what the model cannot support.

2 The model

2.1 Compartments and flows

We treat the population of the Netherlands as a closed, well-mixed collection of N individuals. At every instant t, measured in days from the start of the observation window, each individual

sits in exactly one of three compartments:

- $S(t)$ — *susceptible*: never infected, or infected long enough ago that immunity has lapsed;
- $I(t)$ — *infectious*: currently able to transmit;
- $R(t)$ — *resistant*: recovered and, for the time being, immune.

Fig. 1 shows the transitions. Two features distinguish this system from the textbook SIR model and both are forced on us by the data rather than chosen for elegance. First, the transmission rate β is a function of time: the Dutch epidemic was repeatedly reshaped by policy, and a constant β cannot produce the observed sequence of waves (Sec. 2.3). Second, the compartments themselves are never observed. What the data record is hospital admissions, which are a delayed and heavily thinned image of the infection flux; the map between the two is part of the model and is set out in Sec. 2.4.

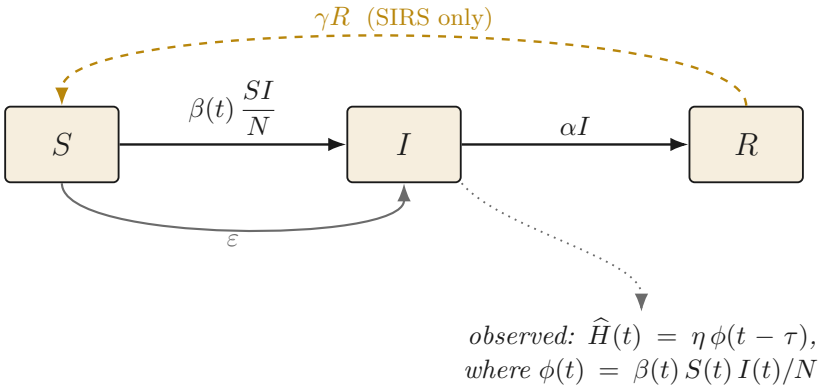

Figure 1
Compartmental structure. Solid arrows are the transitions of the classical SIR model. The dashed arrow returns resistant individuals to the susceptible pool at rate γ and is active only in the SIRS variant; setting $\gamma = 0$ recovers SIR. The grey arc is the importation flux ε of Sec. 2.2, which leaves S and enters I at a constant rate. Only the dotted quantity at the bottom is observable: the compartments themselves are latent.

2.2 The governing equations

The state (S, I, R) evolves according to

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\beta(t)\,\frac{SI}{N} + \gamma R - \varepsilon, \tag{1}$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \beta(t)\,\frac{SI}{N} - \alpha I + \varepsilon, \tag{2}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \alpha I - \gamma R, \tag{3}$$

subject to $S(0) = N - I_0$, $I(0) = I_0$ and $R(0) = 0$.

Each symbol is fixed here and used with this meaning throughout.

- N is the total population, held constant.
- $\beta(t)$, in d^{-1}, is the *transmission rate*: the expected number of contacts one infectious individual makes per day that would infect a susceptible partner. The product $\beta(t)\,SI/N$ is the mass-action incidence, i.e. the number of new infections per day.
- α, in d^{-1}, is the *recovery rate*. Under the exponential-waiting-time assumption A3 the mean infectious period is $1/\alpha$.
- γ, in d^{-1}, is the *waning rate* at which immunity lapses. Setting $\gamma = 0$ gives the SIR model with permanent immunity; any $\gamma > 0$ gives SIRS. The mean duration of immunity is $1/\gamma$.
- ε, in individuals per day, is a constant influx into the infectious compartment, balanced by an equal outflow from S.

Remark 1 (Notation)

Much of the epidemiological literature writes γ for the recovery rate and reserves β for transmission. Here α is the recovery rate and γ the waning rate. This follows the notation of the model code and is stated explicitly so that no reader has to guess; the meaning of every symbol is fixed by the list above and by Tab. 1.

The influx ε deserves a word, because it is the one term in Eqs. (1) to (3) with no epidemiological content. It exists so that the infectious compartment cannot be driven to zero and trap the trajectory at a disease-free state during the long troughs between waves — a purely numerical device standing in for importation from abroad. It is not free: it is subtracted from S, so over the whole window it removes εT individuals from the susceptible pool, and it perturbs the equilibrium structure analysed in Sec. 3.2. Both effects are quantified there and in Sec. 7.

Note that ε enters Eq. (1) and Eq. (2) with opposite signs, so the population is conserved exactly. This is proved in Sec. 3.3 and used as a numerical check in Sec. 5.

2.3 Why the transmission rate must vary

In the standard SIR model the epidemic peaks for one reason only: susceptibles are depleted until new infections no longer outpace recoveries. The Dutch hospital data of spring 2020 peaked for a different reason. Schools closed, workplaces emptied and contact rates collapsed while the overwhelming majority of the population was still susceptible. A model in which β is constant attributes every turning point to susceptible depletion, and so cannot reproduce a peak that occurred at $S/N \approx 1$, nor the repeated waves that followed.

Fig. 2 shows what has to be explained: five distinct waves over four years, of which only the first has the shape a constant-coefficient SIR model produces.

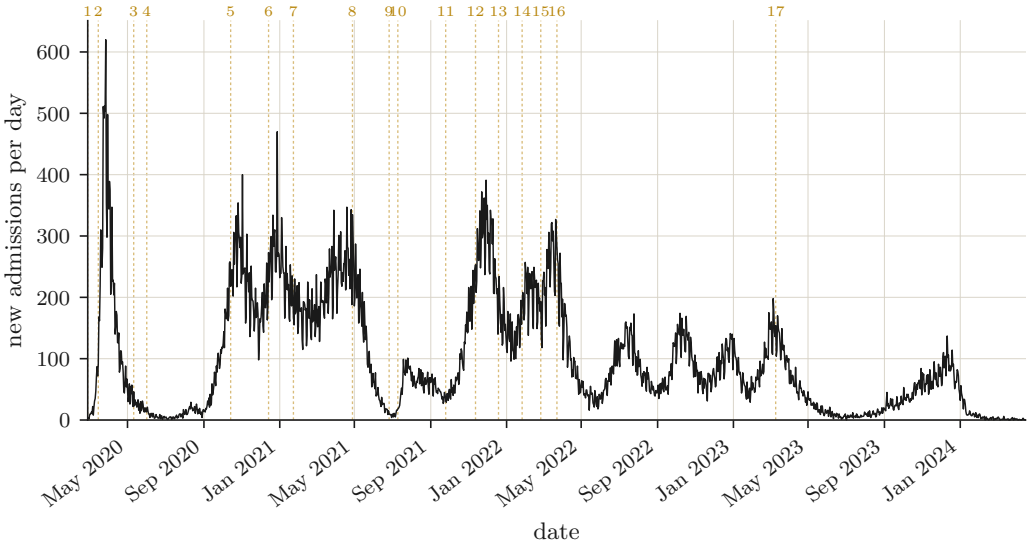

Figure 2

Daily hospital admissions with COVID-19 in the Netherlands as reported by RIVM, over the window used in this paper. Numbered rules mark the policy events of Tab. 4. The spring 2020 peak occurs while essentially the whole population is still susceptible, which is what a constant transmission rate cannot reproduce.

We therefore let β be piecewise constant on buckets of fixed width h:

$$\beta(t) = \beta_k \qquad \text{for } t \in [(k-1)h,\ kh), \qquad k = 1, \dots, K, \qquad K = \lceil T/h \rceil\,, \tag{4}$$

where T is the length of the observation window. The β_k are free parameters, estimated from data. In the language of the assignment these are *control values*: the coarsest possible interpolation of an unknown function from finitely many degrees of freedom. Choosing h trades bias against variance — too few controls and the model cannot follow the policy history, too many and it fits noise. We fix $h = 21$ d for the baseline, a choice made by an information criterion over a sweep of candidate widths rather than by hand, and one that turns out to matter a great deal more than it looks like it should; Sec. 6.2 makes the argument.

2.4 From infections to admissions

The compartments are latent. What RIVM publishes is a daily count of new hospital admissions, and the model has to say how that count arises from the dynamics. Write the *incidence*, the number of new infections per day, as

$$\phi(t) \;=\; \beta(t)\,\frac{S(t)\,I(t)}{N}. \tag{5}$$

An individual infected at time t is hospitalised, if at all, some days later. We model this as a deterministic delay τ combined with a constant thinning factor η:

$$\widehat{H}(t) \;=\; \eta\,\phi(t-\tau), \qquad \widehat{H}(t) = 0 \;\; \text{for } t < \tau. \tag{6}$$

Here η is the infection–hospitalisation ratio and τ the mean delay from infection to admission, both in the units given in Tab. 1. Together, Eq. (6) is an affine map from the latent flux onto the observable series, parameterised by the pair (η, τ).

Two simplifications are buried in Eq. (6) and both cost something. The delay is a single number rather than a distribution: in reality the interval from infection to admission is spread over days, so the true relation is a convolution of ϕ against a delay kernel, and using a point mass sharpens the predicted peaks relative to the data. And η is constant across four years during which variants, vaccination and clinical practice all changed the probability that an infection ends in hospital. Both are recorded as limitations in Sec. 8 rather than repaired here.

2.5 Assumptions

The list below is the model. Everything in Sec. 3 onwards is a consequence of it, and every one of these statements is false in some respect about the real epidemic; what matters is whether the departures are small relative to the signal being measured.

A1 **Closed population.** There are no births, deaths or migration, and N is fixed for the whole window. Vital dynamics and COVID-19 mortality are neglected rather than modelled. Over four years each of these moves N by a fraction of a per cent, which is small beside the transmission signal being measured, but it is a neglect rather than an approximation with an error bound attached.

A2 **Homogeneous mixing.** Every infectious individual is equally likely to contact every susceptible one, so incidence takes the mass-action form $\beta SI/N$. Age structure, household structure and geography are all absent.

A3 **Exponential waiting times.** Time spent in I is exponentially distributed with mean $1/\alpha$, and time spent in R with mean $1/\gamma$. This is what makes the system a finite set of ordinary differential equations rather than an integro-differential one.

A4 **No latent period.** There is no exposed compartment: individuals become infectious the moment they are infected. The incubation period is partially absorbed into the observation delay τ of Eq. (6).

A5 **Single, homogeneous immunity.** Immunity from infection and immunity from vaccination are the same state R and wane at the same constant rate γ. Vaccination is not modelled as a flow; its effect on transmission is absorbed into $\beta(t)$.

A6 **Piecewise-constant transmission.** β is constant on each of K buckets of width h, as in Eq. (4). Behaviour is therefore assumed to change in steps at bucket boundaries rather than smoothly.

A7 **Stationary observation model.** The ratio η and the delay τ in Eq. (6) are constant over the full window, and the delay is deterministic.

A8 **Constant importation.** Infectious individuals arrive at the constant rate ε, independent of the state of the epidemic anywhere.

A9 **Unbiased reporting.** Reported admissions are the true admissions. Reporting delays and revisions in the RIVM series are ignored, and dates on which no figure was published are excluded from the fit rather than treated as zeros (Sec. 4.4).

2.6 Parameters

Tab. 1 lists every quantity in Eqs. (1) to (6), separating those fixed *a priori* from those estimated from the data. The distinction matters: a parameter that is fixed by assumption carries no uncertainty in the fit but every bit of its uncertainty into the conclusions, which is why α and γ are the primary targets of Sec. 7.

3 Analysis

Before fitting anything we establish what the system of Eqs. (1) to (3) does on its own: where its equilibria are, when they are stable, what it conserves, and which classical results about SIR models do and do not survive the modifications of Sec. 2. Two results in this section carry the argument — the decomposition behind Prop. 2 and the growth identity Eq. (18) — and the rest is algebra recorded so the reader need not redo it.

Throughout this section, and only in this section, we freeze the transmission rate at a constant β and set the importation term $\varepsilon = 0$. Both restrictions are lifted in Sec. 3.5, which extends the reproduction numbers to time-varying β and closes by stating what $\varepsilon > 0$ does to the threshold established here.

3.1 The basic reproduction number

The threshold quantity of a compartmental model is not simply read off the equations; it depends on how one splits the flows into *new infections* and *everything else*. The next-generation

Table 1

Model parameters. The upper block is fixed before fitting; the lower block is estimated. K is the number of transmission buckets implied by Eq. (4), so the fit has $K + 3$ free parameters in total.

Symbol	Meaning	Value	Unit	Source / justification
Fixed before fitting				
N	Population size	17 400 000	individuals	Netherlands. CBS records 17 407 585 on 1 January 2020 [1]; the rounded value differs by 0.04 %.
α	Recovery rate	0.142 9	d^{-1}	Mean infectious period $1/\alpha = 7$ d, from the 7.5 d serial interval estimated for the ancestral lineage [2]. Later variants were faster; see Sec. 8.
γ	Waning rate	0.007 3	d^{-1}	Mean immunity duration $1/\gamma \approx 137$ d. Not derived from these data and deliberately uncited: no single value is defensible for a process that varies by variant, vaccine and individual. Swept instead in Sec. 7.3.
ε	Importation influx	0.1	individuals d^{-1}	Numerical device, Sec. 2.2. Not estimated.
h	Bucket width	21	d	Not chosen by hand: selected by BIC over the sweep of Sec. 6.2, which also shows the choice matters more than it appears to.
Estimated from the admissions series				
$\beta_1 \dots \beta_K$	Transmission rates	–	d^{-1}	One control value per bucket, Eq. (4).
η	Infection–hospitalisation ratio	–	dimensionless	Thinning factor in Eq. (6).
τ	Infection-to-admission delay	–	d	Delay in Eq. (6).
I_0	Initial infectious	–	individuals	State at the first observed date.

construction of Diekmann, Heesterbeek and Metz, in the form given by van den Driessche and Watmough, makes that split explicit [3, 4].

Proposition 2 (Basic reproduction number)

For the system Eqs. (1) to (3) with constant β and $\varepsilon = 0$, the basic reproduction number at the disease-free equilibrium $E_0 = (N, 0, 0)$ is

$$R_0 \;=\; \frac{\beta}{\alpha}. \tag{7}$$

Proof. There is a single infected compartment, I. Write its equation as $\dot{I} = \mathcal{F} - \mathcal{V}$, where $\mathcal{F}$ collects flows that represent *new* infections and $\mathcal{V}$ collects every other flow into or out of the compartment:

$$\mathcal{F}(S, I) \;=\; \beta \frac{SI}{N}, \qquad \mathcal{V}(S, I) \;=\; \alpha I. \tag{8}$$

This split is the whole content of the argument. Recovery αI removes individuals from I without creating an infection, so it belongs in $\mathcal{V}$; had it been placed in $\mathcal{F}$ the construction would return a different and meaningless quantity.

Linearising at E_0, where $S = N$, gives the 1×1 matrices

$$F \;=\; \left.\frac{\partial \mathcal{F}}{\partial I}\right|_{E_0} \;=\; \beta \left.\frac{S}{N}\right|_{S=N} \;=\; \beta, \qquad V \;=\; \left.\frac{\partial \mathcal{V}}{\partial I}\right|_{E_0} \;=\; \alpha. \tag{9}$$

The next-generation matrix is $K = FV^{-1} = \beta/\alpha$, and R_0 is its spectral radius, which for a 1×1 matrix is the entry itself. ∎

The factorisation $K = FV^{-1}$ is worth reading rather than merely deriving. The entry $V^{-1} = 1/\alpha$ is the mean time an individual spends infectious, by A3; $F = \beta$ is the rate at which one infectious individual generates new infections while the population is wholly susceptible. Their product is the expected number of secondary infections caused by one infectious individual introduced into a fully susceptible population, which is what R_0 is supposed to mean.

3.2 Equilibria and their stability

Because the population is conserved (Prop. 4) we may eliminate $R = N - S - I$ and study the planar system

$$\dot{S} = -\beta \frac{SI}{N} \;+\; \gamma\,(N - S - I), \tag{10}$$

$$\dot{I} = \;\; \beta \frac{SI}{N} \;-\; \alpha I. \tag{11}$$

The SIRS case ($\gamma > 0$). Setting $\dot{I} = 0$ gives $I = 0$ or $S = \alpha N/\beta = N/R_0$. The first branch forces $S = N$, giving the disease-free equilibrium $E_0 = (N, 0)$. The second gives an endemic equilibrium: substituting $\beta S^*/N = \alpha$ into $\dot{S} = 0$ yields $\gamma(N - S^*) = (\alpha + \gamma)I^*$, hence

$$S^* = \frac{N}{R_0}, \qquad I^* = \frac{\gamma N}{\alpha + \gamma}\left(1 - \frac{1}{R_0}\right), \qquad R^* = \frac{\alpha N}{\alpha + \gamma}\left(1 - \frac{1}{R_0}\right). \tag{12}$$

These sum to N, as they must, and lie in the positive orthant precisely when $R_0 > 1$.

Proposition 3 (Threshold)

For $\gamma > 0$ and $\varepsilon = 0$, the disease-free equilibrium E_0 is locally asymptotically stable if $R_0 < 1$ and unstable if $R_0 > 1$. The endemic equilibrium (12) exists in the positive orthant if and only if $R_0 > 1$, and is locally asymptotically stable whenever it exists.

Proof. The Jacobian of Eqs. (10) to (11) is

$$J(S, I) = \begin{pmatrix} -\beta I/N - \gamma & -\beta S/N - \gamma \\ \beta I/N & \beta S/N - \alpha \end{pmatrix}. \tag{13}$$

At $E_0 = (N, 0)$ this is upper triangular, so its eigenvalues are read off the diagonal:

$$\lambda_1 = -\gamma < 0, \qquad \lambda_2 = \beta - \alpha = \alpha\,(R_0 - 1). \tag{14}$$

Hence E_0 is stable exactly when $R_0 < 1$, and the instability that appears as R_0 crosses 1 is in the I direction, as one would expect.

At the endemic equilibrium the relation $\beta S^*/N = \alpha$ makes the lower-right entry vanish, so

$$J(S^*, I^*) = \begin{pmatrix} -\beta I^*/N - \gamma & -(\alpha + \gamma) \\ \beta I^*/N & 0 \end{pmatrix}, \qquad \begin{aligned} \operatorname{tr} J &= -\left(\beta I^*/N + \gamma\right) < 0, \\ \det J &= (\alpha + \gamma)\,\beta I^*/N > 0, \end{aligned} \tag{15}$$

the inequalities holding for any $I^* > 0$. A planar equilibrium with negative trace and positive determinant has both eigenvalues in the open left half-plane, so E^* is locally asymptotically stable. ■

The two equilibria exchange stability at $R_0 = 1$, where they also collide: this is a transcritical bifurcation, and it is the mathematical content of the familiar statement that an epidemic takes hold only once R_0 exceeds one.

The SIR case ($\gamma = 0$). Here the picture degenerates in a way worth stating, because it is the model actually used for one of the two fits reported in Sec. 6. With $\gamma = 0$, $\dot{S} = 0$ requires $I = 0$, so *every* point $(S^*, 0)$ is an equilibrium and there is no endemic state at all. The Jacobian there has eigenvalues 0 and $\beta S^*/N - \alpha$, so the equilibrium is non-hyperbolic; the epidemic decays if and only if $S^* < N/R_0$. An SIR epidemic therefore burns out and leaves a permanent residue of susceptibles, whereas the SIRS epidemic settles onto (12). Over a four-year window with repeated waves this difference is not cosmetic, which is why both variants are fitted and compared.

3.3 Conservation of the population

Proposition 4 (Conserved total population)

For any bounded $\beta(\cdot)$ and any $\alpha, \gamma, \varepsilon \geq 0$, the solution of Eqs. (1) to (3) satisfies $S(t) + I(t) + R(t) = N$ for all t.

Proof. Add the three equations. Every term cancels in pairs: the incidence $\beta(t)SI/N$ leaves S and enters I; the recovery αI leaves I and enters R; the waning flow γR leaves R and enters S; and the importation ε is subtracted from S and added to I. Hence $\frac{\mathrm{d}}{\mathrm{d}t}(S + I + R) \equiv 0$, and the initial condition of Sec. 2.2 fixes the constant at $S(0) + I(0) + R(0) = N$. ∎

Two things follow. First, Prop. 4 is an exact invariant of the continuous system, so any drift in $S + I + R$ computed by the integrator is pure numerical error; this is the check run in Sec. 5.1. Second, it is the reason the planar reduction of Sec. 3.2 is legitimate.

The invariant says nothing about positivity, which has to be argued separately. When $\varepsilon > 0$ the infectious compartment cannot be extinguished, since $\dot{I} = \varepsilon > 0$ whenever $I = 0$; this is precisely the property ε was introduced for. The susceptible compartment enjoys no such protection — $\dot{S} \leq -\varepsilon$ when $I = 0$ — so S could in principle be driven negative. Over the window used here it is not: $\varepsilon T \approx 151$ individuals against $N = 17\,400\,000$, and the fitted trajectories keep S at the order of 10^6 throughout.

3.4 There is no final-size relation here

The closed SIR model admits a celebrated invariant relating the total number ever infected to R_0 [5]. It is obtained by dividing $\dot{S}$ by $\dot{R}$, which for constant β, $\gamma = 0$ and $\varepsilon = 0$ removes both I and t from the problem:

$$\frac{\mathrm{d}S}{\mathrm{d}R} = \frac{-\beta SI/N}{\alpha I} = -R_0 \frac{S}{N} \quad \Longrightarrow \quad \ln\frac{S_0}{S_\infty} = R_0\left(1 - \frac{S_\infty}{N}\right). \tag{16}$$

None of the three cancellations survives in the present model, and it is worth being explicit about why rather than quietly omitting the result:

1. β **varies.** With $\beta = \beta(t)$ the quotient becomes $-(\beta(t)/\alpha)S/N$, which still contains t. The equation is no longer autonomous in (S, R) and does not separate.

2. **Immunity wanes.** For $\gamma > 0$ the numerator acquires $+\gamma R$ and the denominator becomes $\alpha I - \gamma R$. The factor I no longer cancels, so no relation between S and R alone exists.

3. **There is an influx.** With $\varepsilon > 0$ the numerator acquires $-\varepsilon$, and, more fundamentally, $I \not\to 0$, so the limit S_∞ on which Eq. (16) is built does not exist.

Beyond the algebra there is a modelling point: with waning immunity the epidemic has no final size. It approaches the endemic equilibrium (12) rather than exhausting itself, so the question Eq. (16) answers is not one this model can be asked. The invariant available for verification is therefore Prop. 4, not a final-size identity, and Sec. 5 is built accordingly.

3.5 Reproduction numbers when transmission varies

Prop. 2 concerns a system with constant coefficients, but the model of Sec. 2 has none. Two distinct quantities are used in the rest of the paper and conflating them is the most common error in this kind of write-up, so we define both.

Definition 5 **(Instantaneous and effective reproduction numbers)**

For the time-varying system, define

$$R_0(t) \;:=\; \frac{\beta(t)}{\alpha}, \qquad\qquad R_t \;:=\; R_0(t)\,\frac{S(t)}{N} \;=\; \frac{\beta(t)}{\alpha}\,\frac{S(t)}{N}. \tag{17}$$

$R_0(t)$ is *not* a reproduction number of the non-autonomous system. It is the value that Prop. 2 would return for the frozen-coefficient system obtained by holding β at $\beta(t)$ forever, in a wholly susceptible population. It therefore measures contact behaviour and intrinsic transmissibility — the things policy and variants change — and deliberately ignores accumulated immunity. Because β is piecewise constant by A6, $R_0(t)$ is a step function with jumps at the bucket boundaries.

R_t discounts $R_0(t)$ by the susceptible fraction, and is the quantity that national health institutes publish. Since $S(t)$ is continuous, R_t inherits its jumps only from β.

The reason R_t is the right object, rather than merely the conventional one, is the following identity. Substituting Eq. (17) into Eq. (2),

$$\frac{\mathrm{d}I}{\mathrm{d}t} \;=\; \beta(t)\frac{SI}{N} - \alpha I + \varepsilon \;=\; \alpha I\left(\frac{\beta(t)}{\alpha}\frac{S}{N} - 1\right) + \varepsilon \;=\; \alpha I(R_t - 1) + \varepsilon. \tag{18}$$

Remark 6 (The threshold is instantaneous, not asymptotic)
With $\varepsilon = 0$, Eq. (18) gives $\operatorname{sign}(\dot{I}) = \operatorname{sign}(R_t - 1)$ at *every* instant. The condition $R_t = 1$ is thus not an asymptotic statement about equilibria but a pointwise criterion separating growth from decay, which is what justifies drawing the line $R_t = 1$ across every trajectory plot in Sec. 6.

The importation term blunts this threshold, and Eq. (18) says by exactly how much. Setting $\dot{I} = 0$ for $R_t < 1$ gives a nonzero floor below which the infectious compartment cannot fall:

$$I_{\text{floor}} = \frac{\varepsilon}{\alpha\,(1 - R_t)}. \tag{19}$$

At the values of Tab. 1, $\varepsilon = 0.1$ and $\alpha = 0.1429$, a sustained $R_t = 0.9$ holds roughly 7 individuals infectious and $R_t = 0.5$ roughly 1.4. Against $N = 17\,400\,000$ and against fitted infectious populations of order 10^5 this is negligible, so the convenience of A8 is bought cheaply — but it does mean the model has no exact disease-free state, and that the sharp transcritical bifurcation of Prop. 3 is smoothed into a rapid but continuous crossing. Sec. 6.8 examines what that looks like numerically.

4 Numerical method

Two numerical decisions shape every result that follows: how the differential equations are integrated, and how the parameters are searched for. Neither is routine here. The first is complicated by the fact that Eq. (4) makes the right-hand side discontinuous; the second by a 75-dimensional non-convex objective.

4.1 Integrating the system

We integrate Eqs. (1) to (3) with the explicit Runge–Kutta pair of Dormand and Prince [6], as implemented by `solve_ivp` in SciPy [7].

An explicit method is the right choice because the system is not stiff, and it is worth saying why rather than assuming it. The time scales present are the infectious period $1/\alpha = 7\,\text{d}$, the immunity duration $1/\gamma \approx 137\,\text{d}$, and the force of infection $\beta(t)\,S/N$, which over the fitted window stays below α in order of magnitude. The ratio of slowest to fastest is therefore around twenty. Stiffness becomes a problem at ratios many orders of magnitude larger; at twenty an explicit solver takes steps limited by accuracy rather than by stability, which is exactly the regime it is built for.

4.2 The discontinuity that does cause trouble

The real difficulty is elsewhere. Adaptive step-size control works by comparing two estimates of the same step and treating the difference as the local error, an argument that rests on the solution being smooth over the step. Piecewise constant transmission violates that at every bucket boundary: with $K = 72$ buckets there are $K - 1$ interior points at which β jumps and the right-hand side is discontinuous. A solver given the whole window in one call knows nothing about them, steps across, and produces an error estimate that means nothing on those steps.

The consequence is measurable, and it is not subtle. Integrating the fitted trajectory in a single call and tightening the tolerance does not reliably reduce the error against a high-accuracy reference. Going from rtol $= 10^{-7}$ to 10^{-8} moves the maximum error from 30.49 to 30.50 admissions per day — the wrong way. A convergence study performed this way would licence any tolerance one liked.

The fix is to stop hiding the discontinuities from the solver. We integrate segment by segment: on each interval where β is constant the right-hand side is smooth, the integrator is restarted from the state carried across the boundary, and standard error control applies again. Consecutive buckets holding the same value are merged into one segment, so a counterfactual scenario built from day-wide buckets but only three distinct transmission levels (Sec. 6.7) costs three segments rather than 365.

This is both more accurate and faster. It restores monotone refinement (Fig. 3) and, because the solver no longer wastes steps being rejected at jumps, runs 2.7 times quicker than the single-call version at the same tolerance.

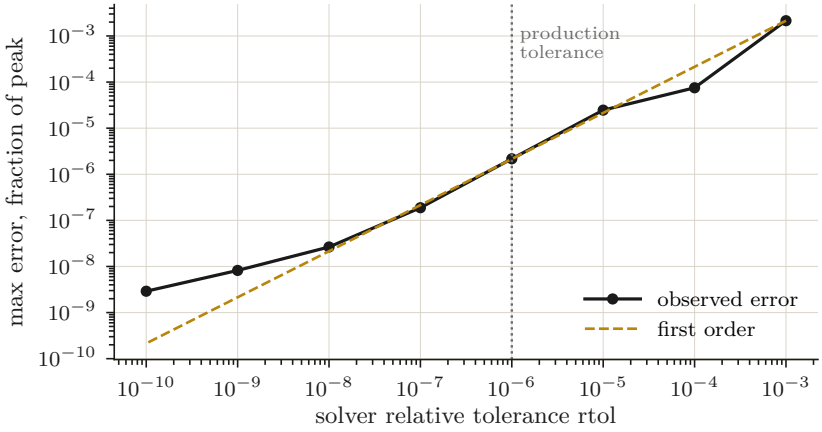

Figure 3
Maximum error in the predicted admissions against a reference integration at rtol $= 10^{-12}$*, as a fraction of the peak. Segment-aware integration converges cleanly until it meets the floor set by the absolute tolerance. The dotted line marks the production setting used for every result in this paper.*

4.3 Choice of tolerances

We use rtol $= 10^{-6}$ with an absolute tolerance of 10^{-3} individuals. The absolute tolerance is set in people rather than left at a default because the state variables are of order 10^{7}: a tolerance of one thousandth of a person is far below any quantity the model could be asked to resolve.

At that setting the predicted admissions differ from the reference integration by at most 0.0014 admissions per day, or 2.2×10^{-06} of the peak. The fitted model's own root-mean-square error against the data is 25.1 admissions per day (Sec. 6.1). Numerical error is therefore some six orders of magnitude below the error that actually limits the result, which is the condition a tolerance choice has to satisfy.

4.4 The observation window

The consolidated data frame is assembled by joining sources with different date coverage and then filling gaps with zeros. For a series of counts that is a dangerous default: a date on which RIVM published nothing becomes a date on which nobody was admitted, and least squares cannot tell the difference.

RIVM reported admissions from 2020-02-27 to 2024-04-15. The original pipeline fitted to 2024-07-30, which handed the optimiser the 105 days from 2024-04-16 to 2024-07-29 as consecutive fabricated zeros — a stretch of apparent eradication that pulls the whole tail of the fitted transmission path downwards. We therefore clip the window to the reported span, which leaves 1510 days, and carry an explicit mask: the 15 dates inside that span for which no figure was published contribute no residual. All fitting and all goodness-of-fit statistics use the 1 495 genuinely observed days.

The same correction applies to the comparison series. RIVM stopped publishing the effective reproduction number after 2023-07-09; later values in the frame are fill values and are excluded from Sec. 6.6 rather than plotted as a collapse to zero.

4.5 Estimating the parameters

The 75 parameters — 72 transmission values plus η, τ and I_0 — are estimated by bounded nonlinear least squares, using the trust-region reflective algorithm of `least_squares`. Bounds are not a numerical convenience but a modelling statement: they confine each parameter to a range in which it retains its epidemiological meaning, and they are listed in Tab. 2. They are also a diagnostic. A first attempt used a narrower box, and the fit pinned both the opening transmission bucket and I_0 to their ceilings, trading one against the other; the estimated initial R_0 was then a property of the constraint rather than of the data. The bounds below are wide

enough that no fitted value rests on one, and any that did would be reported as a failure of the bound.

Table 2

Search bounds for the estimated parameters, taken directly from the model configuration. Each is a statement about what values would still describe a plausible epidemic. A fitted value resting on a bound is treated as a failure of the bound rather than a result; see Sec. 4.5.

Parameter	Lower	Upper	Reasoning
β_k	0.01	1.2	Transmission rate, d^{-1}. Corresponds to $R_0(t)$ between 0.07 and 8.40, spanning hard suppression through the most transmissible variants.
η	0.00001	0.05	Fraction of infections admitted, from one in 100 000 to one in 20.
τ	1	28	Infection to admission, in days.
I_0	10	200000	Infectious individuals present when reporting began.

Two details of the objective matter enough to record.

First, the delay τ enters through a linear interpolation of the infection flux, as written in Eq. (6). An earlier version of this model rounded τ to whole days before applying the shift. That makes the residual piecewise constant in τ: the predicted series is bit-identical for every τ in $[13.5, 14.5)$, the finite-difference derivative with respect to τ is exactly zero, and the optimiser has no gradient to follow. Any lag reported under that scheme would be an artefact of where the search happened to start. Interpolating instead makes τ a genuine estimated parameter, and Sec. 5.3 tests whether it is identifiable in practice.

Second, the objective is not convex, and with 75 parameters there is no prospect of proving that any particular optimum is global. We therefore restart the search from 48 independent random starting points drawn uniformly inside the bounds, with a fixed seed, and keep the lowest-error result. Reporting the mean of several fitted trajectories, as an earlier version of this pipeline did, is not a defensible alternative: no member minimises that mean, and averaging over a hundred piecewise-constant transmission paths flattens exactly the peaks the model is being asked to reproduce. The spread across restarts is reported in Sec. 6.1 for what it is — a measure of how reliably the optimiser finds the same solution, not a confidence interval.

Each restart is also capped at 600 function evaluations. Without a cap a single restart that wanders into a flat region can consume more time than all the others together, and SciPy's default allowance here would be around 11 000; restarts that converge normally use well under half the cap, and one that reaches it simply records a worse error and loses to its neighbours.

5 Verification

Fitting a model to data tells you how closely the code you wrote reproduces the series you gave it. It does not tell you whether that code solves the equations you think it does. This section

asks the second question, and asks it without reference to the data: each check below compares the implementation against something known independently — an exact invariant, a closed-form solution, or a parameter set the fit was given in advance and should be able to give back.

Convergence under refinement is the fourth such check and has already been reported, in Sec. 4.2, because it is what motivated the segment-aware integrator described there.

5.1 Conservation of the population

By Prop. 4 the quantity $S + I + R$ is exactly constant for the continuous system, whatever $\beta(\cdot)$, γ and ε may be. Any departure computed numerically is therefore attributable entirely to the integrator.

Integrating the fitted trajectory over the full window, the largest departure from N at the production tolerance is 1.5×10^{-15} in relative terms — 26.1 billionths of one person. That is machine precision: the integrator neither creates nor destroys population.

The check earns its place for a second reason. The importation term ε appears twice in Eqs. (1) to (2), once negative and once positive, and the cancellation on which Prop. 4 rests holds only if both signs are right. A transposition there would be invisible in the fit — the model would simply describe a population that slowly leaks — but shows up here immediately as a drift growing linearly in time.

5.2 Agreement with the analytic early-growth solution

Conservation tests bookkeeping. It would be satisfied by a system with the wrong dynamics, provided the wrong terms still cancelled. The second check therefore compares against a case where the solution is known in closed form.

Take β constant and $\varepsilon = 0$, and start from a small seed $I_0 \ll N$. While the susceptible pool is essentially untouched, $S/N \approx 1$, Eq. (2) linearises to

$$\frac{\mathrm{d}I}{\mathrm{d}t} \;\approx\; (\beta - \alpha)\, I, \qquad \text{so} \qquad I(t) \;=\; I_0\, e^{(\beta-\alpha)t}, \tag{20}$$

an exact exponential with growth rate $r = \beta - \alpha$. We integrate the full nonlinear system for three transmission rates spanning growth and decay, and compare against Eq. (20) over the interval in which the susceptible fraction has fallen by less than 10^{-4}, so that the linearisation is legitimate rather than merely convenient.

Across all three, over as much as 31 days of simulated time, the largest relative departure is 2.5×10^{-04}. The implemented right-hand side reproduces the analytic growth rate to the accuracy of the integrator, which is what this check is for.

5.3 Recovery of known parameters

The two checks above concern the solver. This one concerns the estimator, and it is the one that matters most for a paper whose result is a parameter path.

The baseline fit carries 75 free parameters against 1 495 observations. A good fit does not establish that those parameters are identifiable: a model with enough freedom can reproduce a curve through many different combinations of values, and the goodness-of-fit statistics of Sec. 6.1 cannot distinguish that case from a well-determined one. The way to find out is to give the estimator a problem whose answer is already known.

The protocol is as follows. Treat the fitted parameter vector as ground truth and generate the admissions series it implies. Corrupt that series with Poisson counting noise — admissions are counts, so counting noise is the right noise model, and its magnitude scales with the signal in the way real reporting noise does. Then run the ordinary fitting procedure of Sec. 4.5 on the synthetic series, from fresh random starting points, and compare what comes back with what went in. Three independent noise realisations are used, so that a single lucky or unlucky draw cannot carry the conclusion.

The parameters come back. Across 3 noise realisations the recovered transmission path correlates with the true one at no worse than 0.995, with a mean absolute error of 5.3 per cent per bucket. The three observation parameters are recovered to within 4.7 per cent for η, 0.5 per cent for τ and 12.4 per cent for I_0. At the complexity used for the baseline, the estimator is doing what it claims to.

That last qualification is not decoration. Run at the finer bucket widths that Sec. 6.2 considers, this same test degrades badly and erratically: replicates that reach a comparable error still return transmission paths correlating with the truth at anywhere from 0.05 to 0.99, depending on whether the optimiser happens to find the global basin. The identifiability established here is a property of the model *at this resolution*, not of the model in general, and it is one of the two independent grounds on which Sec. 6.2 settles the resolution.

A second observation falls out of the same runs, and it needs stating carefully because it is easy to over-read. I_0 is recovered to within a few per cent, which settles one question the baseline fit raises: the value of 63348 infectious individuals in late February 2020 is not an artefact of a flat objective or of a bound pressed up against, but a quantity the admissions data determines sharply *given this model*.

That is identifiability, not credibility, and the distinction is the whole of it. The synthetic series was generated by the model itself, so this experiment asks whether the estimator can invert its own forward map — not whether the forward map describes the Netherlands. A model

with no exposed compartment must hold every already-infected person in I, and one that cannot represent a falling hospitalisation ratio must express early severity through the size of the seed; both push I_0 upward for reasons internal to the model's structure. The comparison between the SIR and SIRS columns in Tab. 3 makes the same point from the other direction: two fits of near-identical quality disagree about I_0 by a factor of 3.4. A reader should take 63348 as what this model requires, not as an estimate of how many Dutch residents were infectious on that date.

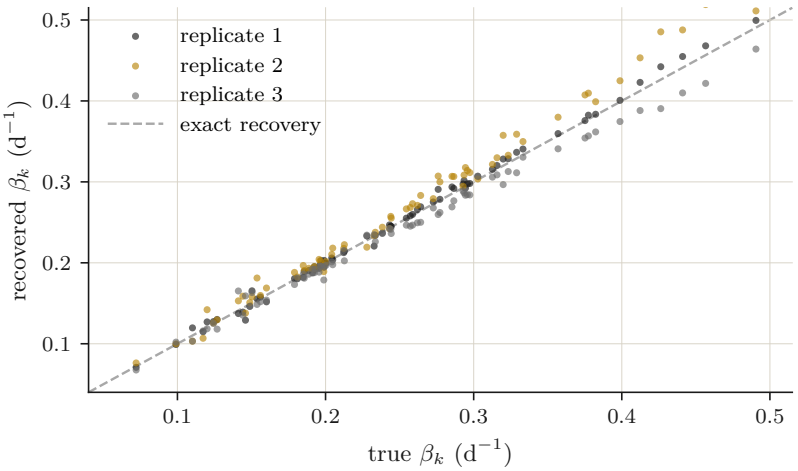

Figure 4
Transmission values recovered from synthetic data against the values used to generate it, for three independent noise realisations. Points on the dashed line are exact recovery.

6 Experiments

Six investigations. The first two establish what the model is and how much freedom it should be given; the third and fourth are what the paper is for; the last two ask what the fitted model implies about situations that did not occur, and are read as statements about the model rather than about the Netherlands.

6.1 The baseline fit

Both variants — SIR ($\gamma = 0$) and SIRS ($\gamma > 0$) — were fitted to the daily admissions series over 2020-02-27 to 2024-04-15, using 48 random restarts each and the procedure of Sec. 4.5. Tab. 3 reports both.

Fig. 5 shows the fitted SIRS trajectory against the data.

The two variants are far closer than the raw structural difference suggests. SIRS wins on both criteria, but by 50 in AIC — decisive, yet a fraction of the gap that separates either from a badly chosen bucket width (Sec. 6.2). Waning immunity earns its parameter, and not much more.

Table 3

The two baseline fits. Both use the same window, the same bucket width and the same number of restarts; they differ only in whether immunity wanes. Lower AIC and BIC are better.

	SIRS	**SIR**
Free parameters	75	75
SSE	9.390e+05	9.707e+05
RMSE (admissions per day)	25.1	25.5
AIC	9 782	9 831
BIC	10 180	10 230
η (hospitalisation ratio)	0.209%	0.920%
τ (delay, days)	9.1	9.1
I_0 (initial infectious)	63 348	18 671
mean $R_0(t)$	1.72	3.31
range of $R_0(t)$	0.50–3.43	0.55–8.40

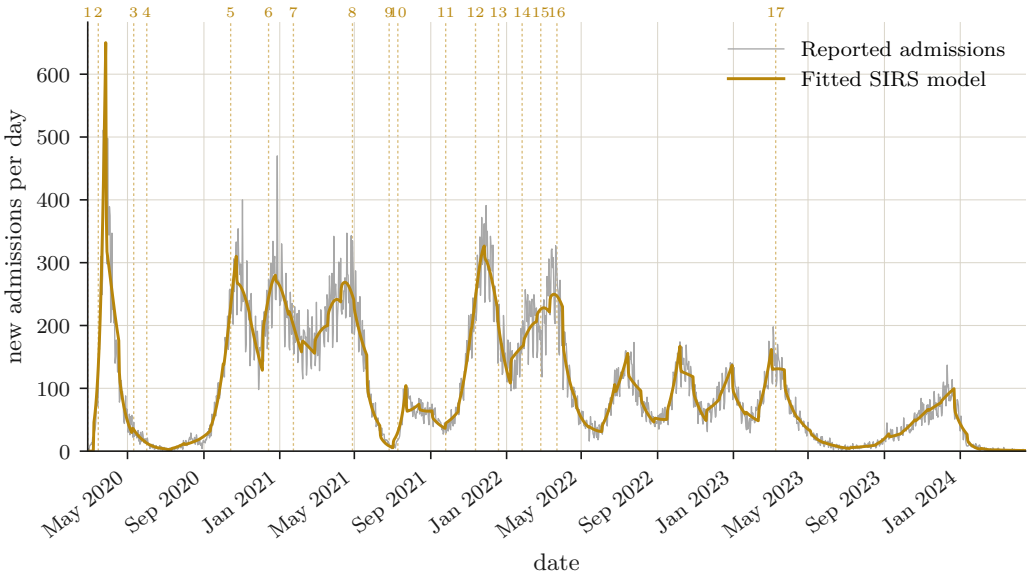

Figure 5

The baseline SIRS fit against reported daily admissions. Dashed vertical rules mark the policy events of Tab. 4, numbered as there.

Two features of the fit deserve comment before anything is read into it. The fitted delay from infection to admission, 9.1 days, is the one estimated quantity with an external sanity check available, and it passes: it agrees with the gap of 8 days measured in Sec. 6.5 between the modelled infection peak and the observed admissions peak, which is a different calculation on the same fit. The initial infectious population, 63348, is large — far larger than the case counts of late February 2020. Sec. 5.3 shows that it is well determined *within the model*, which is a narrower claim than it may appear and is taken up below.

The sharpest evidence in this table is easy to miss, because it is a comparison between the columns rather than a number in either. The two models fit almost equally well — root-mean-square errors of 25.1 and 25.5 admissions per day — yet they disagree about the hospitalisation ratio by a factor of 4.4 (0.209 % against 0.920 %) and about the initial infectious population by a factor of 3.4 (63348 against 18671). Two descriptions of the same data, indistinguishable by fit, disagree by a factor of four about how many infections there were. The admissions series pins down the *product* of η with the size of the latent epidemic far better than either factor separately, which is the degeneracy Sec. 8 describes, made concrete. Quantities that depend on the shape of the transmission path — R_t, the timing of turning points — are on much firmer ground than quantities that depend on its absolute level.

The restarts are also worth reporting honestly. Of 48 independent starting points, 27 reached the global basin — 56 per cent. The objective is not merely non-convex; it has one narrow deep minimum among many much worse ones, and a search with a handful of restarts will usually miss it. For the SIR variant the rate is 6 per cent. This is a property of the problem, not of this implementation, and it is the reason Sec. 4.5 uses as many restarts as it does.

6.2 How many control values?

Sec. 2.3 adopted a bucket width h without yet saying where it came from. The argument is owed here, because h is the single choice that determines how much freedom the transmission path has. Too few control values and the model cannot follow a policy history that changed direction seventeen times; too many and it fits reporting noise, and the estimated $R_0(t)$ acquires structure that reflects nothing real.

We refit the model at bucket widths from 7 d to 84 d, holding everything else fixed, with 48 restarts at each width. Fig. 6 puts every resulting quantity on one axis, and it is the central figure of this paper.

The criteria disagree, and the disagreement is not marginal. AIC selects 7 d buckets with 219 parameters — the finest width tried. BIC selects 21 d with 75. Fit quality alone cannot

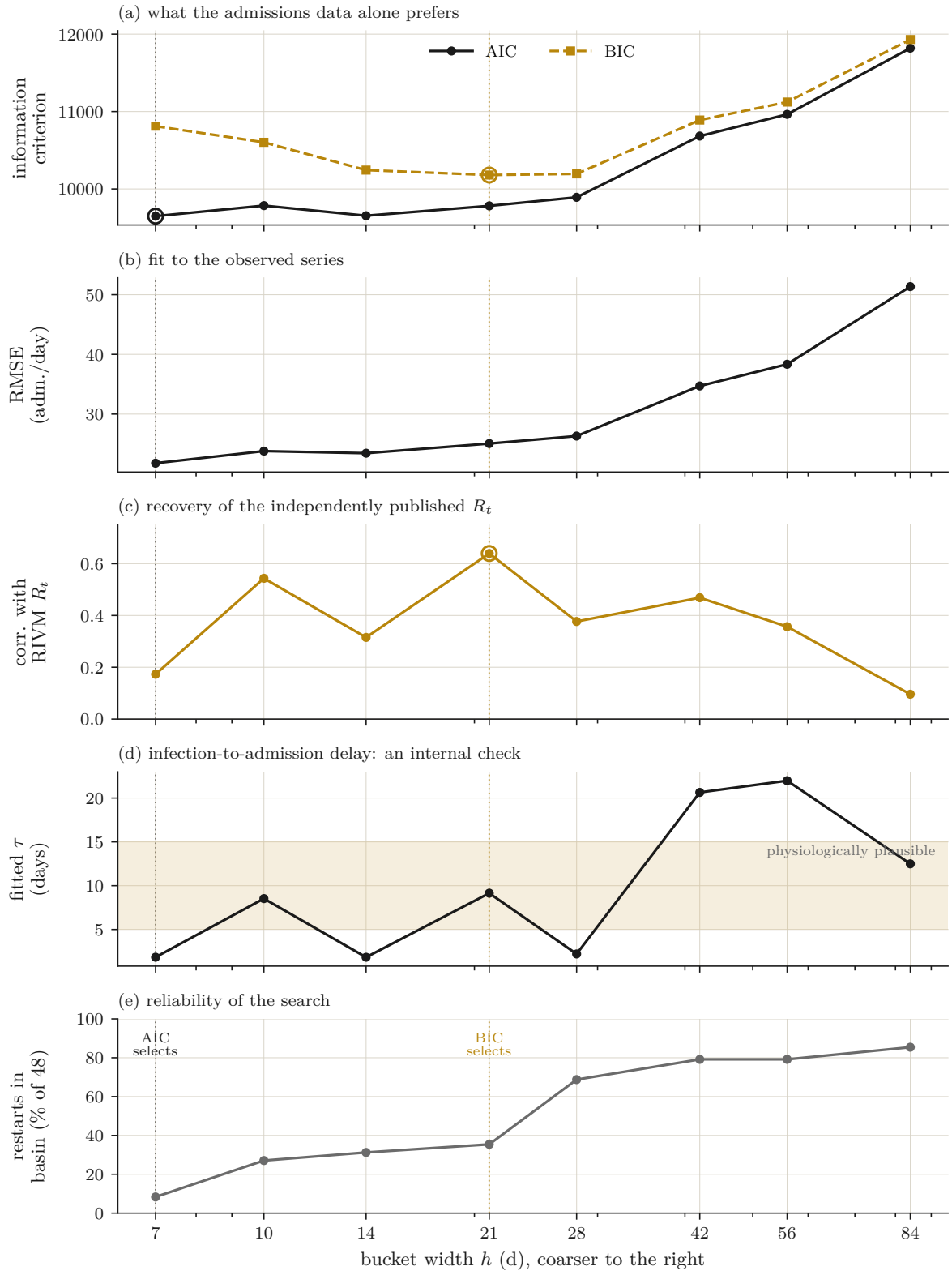

Figure 6

Everything that depends on model complexity, against one axis. (a) The two information criteria, computed from the admissions data alone; circles mark their minima and the dotted guides carry those two widths through every panel. (b) Fit to the observed series. (c) Correlation between the recovered R_t and the series RIVM published independently. (d) The fitted infection-to-admission delay, with a physiologically plausible band shaded. (e) The fraction of restarts reaching the global basin. Panels (a) and (b) are computable by anyone with the admissions data; panel (c) requires an external series; panel (d) requires only knowing roughly how long it takes an infected person to reach hospital.

settle it: root-mean-square error falls steadily as buckets narrow, from 51.4 admissions per day at 84 d to 21.8 at 7 d, so whichever criterion one trusts, the choice comes down to how heavily a parameter is priced.

The external check does not follow either criterion down. Panel (c) is the point. Correlation with the published R_t is 0.173 at AIC's chosen width — the *worst* of every width tried, worse than the 84 d model with a tenth as many parameters. It rises to 0.639 at 21 d and falls away again beyond that, so the curve does have an interior optimum rather than continuing to improve with coarseness: recovery is worst at the finest widths, best in the 10 d to 21 d range, and poor again once buckets grow too coarse to follow the epidemic at all. The width BIC chose and the width that recovers transmission best are the same, but that coincidence is worth stating carefully, and Sec. 6.4 does so.

Why the extra parameters hurt. Beyond a certain resolution the additional control values are not resolving finer structure in transmission; they are absorbing reporting noise in the admissions series, and every unit of admissions error they remove is bought by deforming the latent transmission path that produced it. Because the deformation makes the residuals *smaller*, no statistic computed from the fit can detect it. AIC, which prices a parameter at two units of deviance, does not price this at all. BIC's heavier penalty lands in a better place here, but it does so by being conservative, not by knowing anything about transmission.

Two limits on how hard the individual points can be pressed. First, the width-to-width variation is noisy: each point is a single optimisation of a non-convex problem, and panel (e) shows that the reliability of the search is itself a function of complexity, falling from 41 of 48 restarts reaching the global basin at 84 d to 4 of 48 at 7 d. At the finest widths, then, we cannot rule out that the reported optimum is not the global one, and the reported AIC there should be read as an upper bound on the fit achievable at that width. That cuts against the finding rather than for it: if anything, the true AIC at 7 d is lower still, which would make AIC prefer that width more strongly while its recovery of R_t stayed the worst on the chart. The 10 d point is a demonstrated case: refitting it from a different seed reaches a root-mean-square error below the 14 d value, reversing the order of those two widths.

Second, no single point should carry the argument. What survives the noise is the shape of panel (c) against panels (a) and (b): the criteria computable from the data alone point one way, and the check against reality points another.

6.3 A diagnostic that needs no external series

The comparison in panel (c) required a published R_t to compare against. Most people fitting a model like this will not have one — that is usually the reason for fitting it. If the only way to detect the failure in Sec. 6.2 were an external validation series, the finding would be a curiosity rather than something anyone could act on.

Panel (d) is the reason it is not. The infection-to-admission delay τ is estimated from the same data as everything else, and unlike a transmission rate it has a value a reader already knows something about: the interval between catching COVID-19 and being admitted to hospital is on the order of one to two weeks. It cannot be two days.

At several widths, it is. The fitted delay collapses to 1.8 d to 2.2 d at 7 d, 14 d and 28 d buckets, and sits in a plausible 8.5 d to 9.1 d at 10 d and 21 d. The widths at which τ is physiologically impossible are — with one exception discussed below — the widths at which recovery of R_t is poor, and the two widths where τ is credible are the two where correlation with the published series is highest. The diagnostic and the external check agree, and only one of them needs anybody else's data.

The mechanism is the same one described above. A transmission path with enough freedom can reproduce a wave's timing by shifting β rather than by delaying the observation, so the fitted τ is free to drift to whatever value lets the extra buckets absorb noise most efficiently. The delay stops being an estimate of a physiological interval and becomes a nuisance parameter. Its departure from plausibility is a visible symptom of that.

Two honest qualifications. The relationship is not monotone in complexity — τ does not degrade steadily as buckets narrow, it becomes *unstable*, taking implausible values at some widths and credible ones at others. And it is a one-sided test: at 84 d the fitted delay is credible while recovery is the worst on the chart, because that model fails for an unrelated reason, being too coarse to represent the epidemic at all. An implausible τ is therefore good evidence that a parameterisation should be rejected; a plausible one is not on its own evidence that it should be accepted. Used that way — as a filter rather than a criterion — it is available to anyone, at no cost, and it would have excluded AIC's choice here.

6.4 What was selected on what

One point of intellectual hygiene, because the alternative is for a reader to work it out and wonder what else went unsaid.

The baseline width of 21 d is BIC's choice, computed from the admissions data alone, and BIC would have selected it whether or not any comparison series existed. But this analysis

was not pre-registered, and the correlations in panel (c) were computed before that width was written down as the baseline. The honest claim is therefore that two criteria agree — a penalised-likelihood criterion internal to the fit, and an external comparison against an independent series — not that the external comparison was withheld from the selection and then passed as a blind test. The agreement is corroboration, and worth something; it is not the same thing as prediction, and is not offered as such.

6.5 Transmission against the policy timeline

Fig. 7 shows the estimated $R_0(t) = \beta(t)/\alpha$ across the whole window, with the policy events marked. Read with the caveats of Sec. 8 — in particular that $R_0(t)$ after early 2021 conflates behaviour, variant and vaccination — it is a reconstruction of the transmission history from admissions alone.

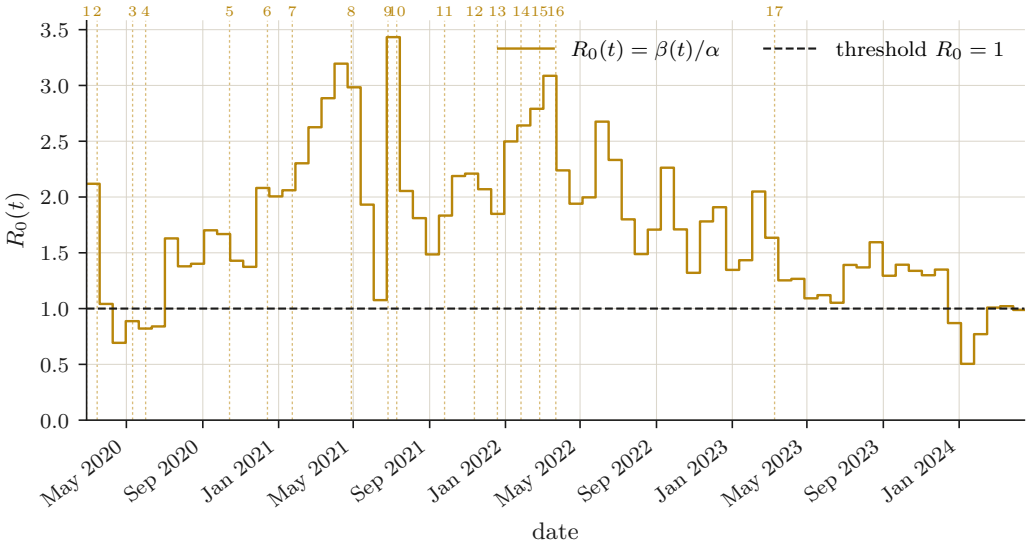

Figure 7
Estimated instantaneous basic reproduction number. The step structure is the piecewise-constant transmission of Eq. (4)*: each tread is one 21 d bucket. Numbered rules are the events of Tab. 4.*

The first wave answers a question the raw series cannot. Measures took effect on 2020-03-15; admissions peaked 11 days later, on 2020-03-26 by a centred seven-day mean. That lag has been read in public discussion as the time the measures took to work, which conflates two very different delays. The model separates them. The fitted infection flux peaks on 2020-03-18, only 3 days after the measures, and the first estimated fall in β is placed in the bucket beginning 2020-03-19. The remaining 8 days are not epidemiology at all — they are the delay between infection and admission, independently estimated as 9.1 days.

So the measures turned transmission within days. What took a fortnight was the hospital system registering that they had.

Across the rest of the window the estimated $R_0(t)$ moves between 0.50 and 3.43 with a mean of 1.72. Two cautions govern how much can be read into any particular movement. First, by A5 vaccination is not modelled, so from 2021 onward a fall in $R_0(t)$ may be behaviour, a variant, or immunity; the model cannot say which. Second, Sec. 6.2 has just shown that transmission structure finer than the bucket width is not reliable even when it improves the fit, so individual buckets should not be read against individual events. What the figure supports is the broad correspondence: sustained falls following each tightening, sustained rises following each relaxation.

Table 4

Policy events marked on Figs. 2, 5 and 7. Dates follow the RIVM chronology.

	date	event
1	2020-02-27	First COVID-19 case in NL ↓
2	2020-03-15	Schools, catering closed; work from home ↓
3	2020-05-11	First easing (primary schools) ↑
4	2020-06-01	Catering and theatres reopen, restricted ↑
5	2020-10-14	Partial lockdown (catering closed) ↓
6	2020-12-14	Hard lockdown (non-essential shops, schools) ↓
7	2021-01-23	Curfew introduced ↓
8	2021-04-28	Curfew ends, terraces reopen ↑
9	2021-06-26	Substantial easing ↑
10	2021-07-10	Easing reversed (nightlife closed) ↓
11	2021-09-25	Distance rule dropped; entry pass ↑
12	2021-11-12	Evening lockdown reintroduced ↓
13	2021-12-19	Lockdown, Omicron emerges ↓
14	2022-01-26	Easing (shops, culture reopen) ↑
15	2022-02-25	Almost all national measures removed ↑
16	2022-03-23	Final measures abolished ↑
17	2023-03-10	All specific advice ends ↑

6.6 Comparison with the published reproduction number

This is the result the paper exists for.

Everything to this point used one data series: daily hospital admissions. RIVM separately published an effective reproduction number, estimated from a different data stream by a different method, with a 95 % interval, up to 2023-07-09. That series enters neither the model nor the fit at any point. Comparing the two is therefore an out-of-sample test of whether the admissions record alone carries the transmission signal.

We score the agreement three ways, in increasing order of how much each should convince a sceptic, over the period from 2020-06-01 — RIVM's earliest estimates carry intervals wide enough to be uninformative, and including them would flatter the comparison.

At the model's own 21 d resolution — the finest at which it can say anything, since β is constant within a bucket by construction — the model tracks the published series at a correlation of 0.804 across 54 intervals. It falls inside RIVM's 95 % interval on 50.0 per cent of them, against 38.9 per cent for the best constant, and agrees on the direction of the epidemic on 81.5 per

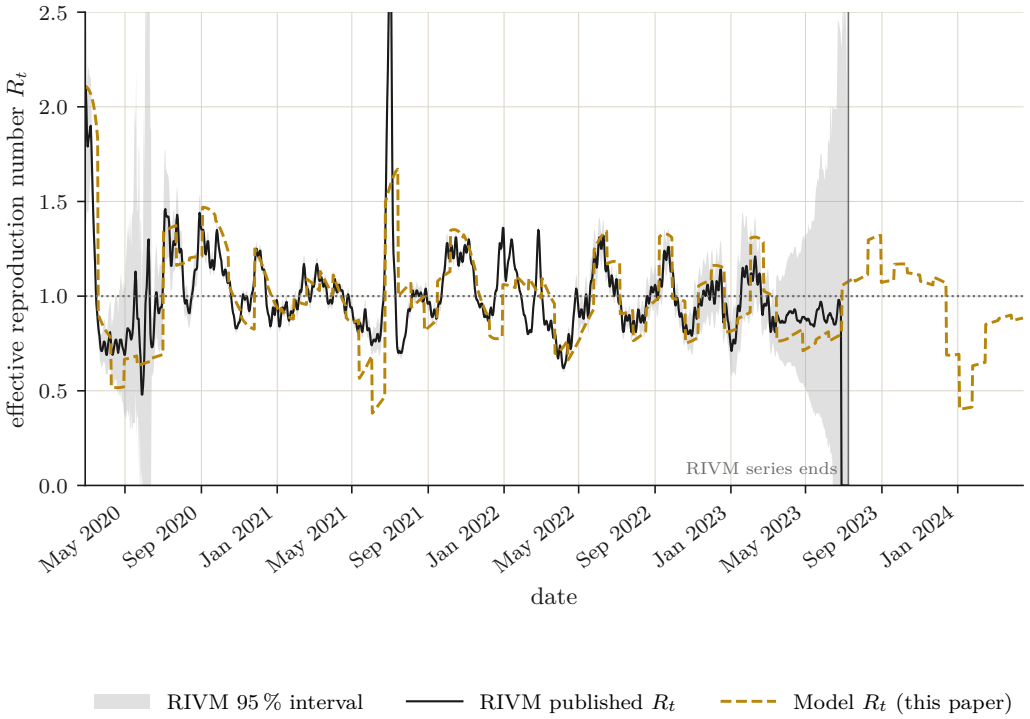

Figure 8

The model's effective reproduction number R_t, computed from Eq. (17) using the fitted transmission path and the modelled susceptible fraction, against the series RIVM published. The model is fitted to hospital admissions only and has never seen the RIVM estimate. The shaded band is RIVM's own 95 % interval; values after 2023-07-09 are omitted because RIVM stopped publishing.

cent against 50.0 per cent. Compared day by day, which charges the model for high-frequency structure it was never given the freedom to express, the correlation is 0.639 and the directional agreement 75.5 per cent against 42.1 per cent.

The constant control matters more than it might appear. RIVM's R_t spends the whole epidemic near one — that is what it means for an epidemic to persist without exploding — and its published interval has a mean width of only 0.252. A flat line at the series mean therefore scores respectably on interval coverage by doing nothing at all, which is why coverage alone would be a poor test. The directional score is the one that separates a model from a constant, and there the gap is large.

What this does and does not establish is worth stating precisely. It establishes that the daily count of hospital admissions contains enough information to reconstruct the transmission history of the epidemic to within about a tenth of a unit in R_t, using no case data, no test data and no contact data. It does not establish that this model would have been useful during the epidemic: the whole window is fitted at once, so the March 2020 estimate uses data from 2024, while RIVM's estimates were produced in real time (Sec. 8).

6.7 Intervention timing against intervention strength

The remaining two experiments are forward simulations of the fitted model. They describe what the model implies, not what would have happened; the distinction matters, because the model's assumptions (Sec. 2.5) are exactly the ones a genuine counterfactual would need to relax.

The setup holds transmission at the value estimated for the first bucket of the epidemic, $R_0 = 2.12$, until a single intervention begins on day T_i, reduces β by a fraction s for 90 days, and then lifts. We sweep T_i and s and record the peak daily admissions and the total over 365 days. With no intervention at all the model gives a peak of 1 099 admissions per day.

The surface is not what one would guess, and its most interesting feature is a failure of monotonicity.

The hardest intervention is not the one that lowers the peak. Peak admissions are minimised at a transmission reduction of 36 % beginning on day 12, giving 287 admissions per day against 1 099 with no intervention at all. Push the reduction to 90 % from day zero and the peak climbs back to 1 083 — barely distinguishable from doing nothing.

The mechanism is not a discovery. That a suppression which is lifted returns the epidemic to a population whose susceptibility it has preserved, producing a rebound comparable to the wave that was avoided, was set out at the start of the pandemic as a known property of this class of model [8]; it also follows directly from Prop. 3, since an epidemic re-entering a wholly

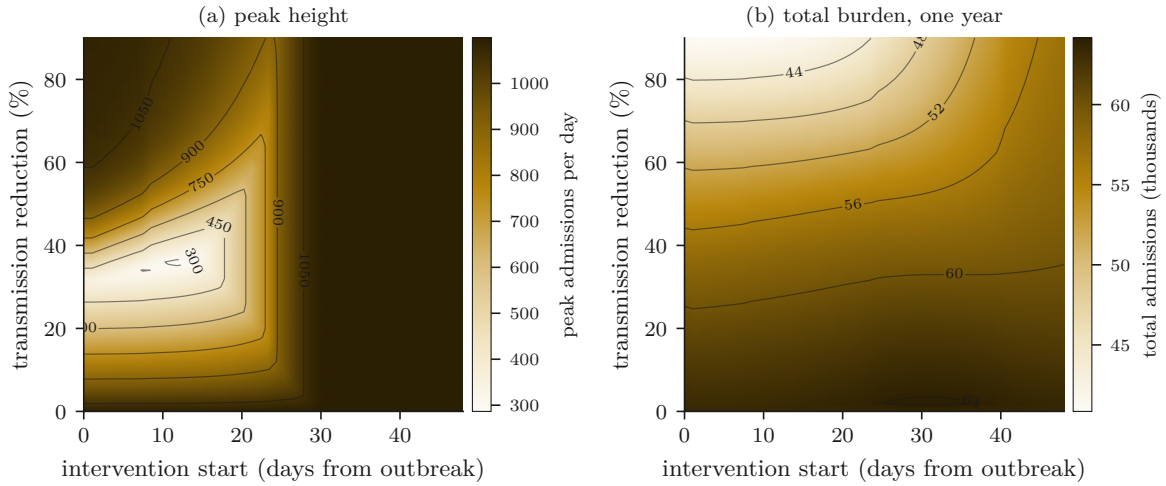

Figure 9

Peak height and one-year burden across intervention start day and transmission reduction, from the fitted model. Contours are drawn at even intervals of the plotted quantity.

susceptible population faces the same R_0 it faced the first time. What the sweep adds is the quantitative shape for this fitted model, not the phenomenon.

Three conditions the number depends on. The statement that a 90 % reduction performs no better than no intervention is true only under conditions that travel badly if the sentence is quoted without them. First, the intervention has a *fixed* 90-day duration set by construction, and is then removed completely; nothing here models a measure that is relaxed gradually or held until some condition is met. Second, no vaccine arrives — there is no immunising process in the model other than infection, so a suppression cannot be buying time for one. Third, transmission returns exactly to its pre-intervention value on release, with no residual behavioural change.

The first of those matters most, and can be measured. Sweeping the duration while holding everything else fixed, the peak-minimising reduction moves from 16 % at 30 days to 32 % at 180 days: a short intervention is best used gently, a long one somewhat harder, and the optimum settles around a third once the duration exceeds roughly three months. Against the waning rate the optimum is markedly more stable — 29 % at an assumed immunity duration of 60 days against 33 % at 730 days, a range spanning more than a factor of ten in γ. So the *existence* of an interior optimum is robust to both, and to the parameter Sec. 7 identifies as the model's weakest; its *location* depends on how long the measure lasts, which is a policy variable rather than a property of the epidemic.

Peak and total pull in opposite directions. The annual total does fall monotonically with strength, from 63 465 admissions with no intervention to 41 109 at the strongest — roughly a third less. The strongest intervention is therefore close to the best available policy if the objective

is cumulative burden and close to the worst if the objective is peak hospital load. Those are different objectives, optimised by different policies, and a model reporting only one of them would mislead about the other.

Timing matters early, and then not at all. Along the timing axis the contours are steep for roughly the first four weeks and flat thereafter. Under this parameterisation the unmitigated epidemic peaks within about a month, so an intervention begun after that finds the susceptible pool already depleted and changes little. Within that first month delay is expensive: the contours run closer together along timing than along strength, so a fortnight's hesitation costs more than a substantial weakening of the measure.

None of this is evidence about the Netherlands. It is a fitted model run forward under Sec. 2.5, and a real counterfactual would have to relax several of them.

6.8 Behaviour through the threshold

Prop. 3 predicts a transcritical bifurcation at $R_0 = 1$: a stable disease-free state below, a stable endemic state above, given by Eq. (12). That derivation assumed $\varepsilon = 0$, which the fitted model does not satisfy. Sec. 3.5 argued that the importation term removes the disease-free state entirely and smooths the bifurcation into a continuous crossing. This experiment checks both claims numerically: we hold β constant, integrate to a long horizon on each side of the threshold, and compare the settled state against Eq. (12), with and without ε.

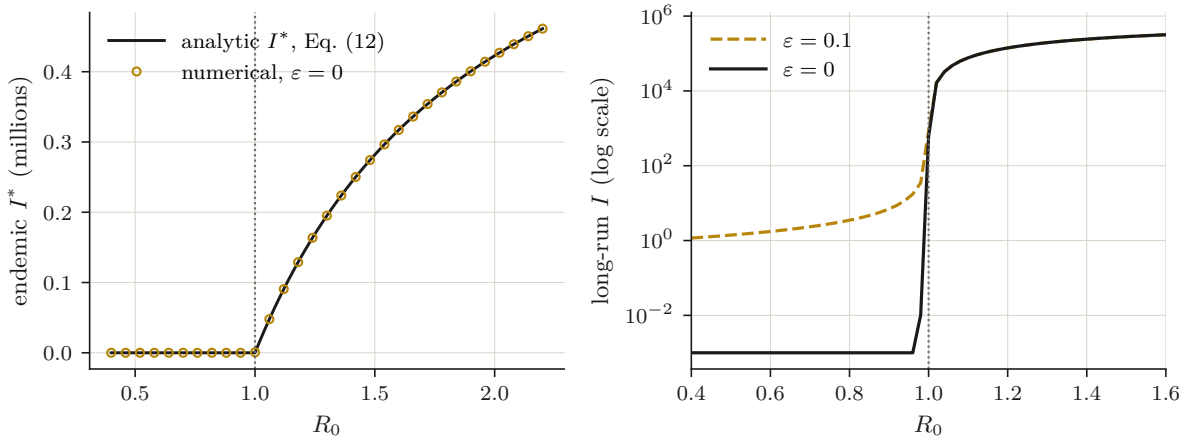

Figure 10

Left: the long-run infectious population against R_0, with the closed form of Eq. (12) and the numerical result at $\varepsilon = 0$. Right: the same on a logarithmic scale, showing what the importation term does below the threshold.

Both claims hold. With $\varepsilon = 0$ the numerical long-run state matches the closed form of Eq. (12) to a maximum relative departure of 5.8×10^{-06} across every point above the threshold, so the endemic branch derived in Sec. 3.2 is the one the integrator finds. Below the threshold

the trajectory collapses to the disease-free state, and the transcritical structure of Prop. 3 is reproduced exactly.

With the fitted $\varepsilon = 0.1$ the disease-free state is gone, as Eq. (19) predicts. The importation term sustains a floor that reaches at most 11.7 infectious individuals, at R_0 just below one where the floor is largest. On a population of 17 400 000 that is epidemiologically nothing, and the right panel of Fig. 10 shows the bifurcation surviving as a very rapid but continuous crossing rather than a genuine discontinuity. The convenience of A8 costs about a dozen notional people and no qualitative structure.

7 Sensitivity

Four quantities in Tab. 1 — N, α, γ and ε — were fixed before fitting. They therefore carry no uncertainty in the estimation and all of it into the conclusions, and they are what this section is for.

Elasticities are nonetheless computed for every parameter, fitted ones included, because a ranking is only informative if nothing large is left out of it: an assumed constant that moves an output by one per cent is unimportant if a fitted parameter moves it by five. The fitted parameters appear in Fig. 11 for that comparison, but their uncertainty is a different question, addressed by the recovery experiment of Sec. 5.3 rather than here. The waning rate is then treated separately in Sec. 7.3, because for that one parameter a local elasticity is not enough.

7.1 Method

For an output Q and a parameter θ the elasticity

$$E_\theta(Q) \;=\; \frac{\partial Q}{\partial \theta} \cdot \frac{\theta}{Q} \tag{21}$$

is the percentage change in Q per percentage change in θ. It is dimensionless, which is the point: it lets a population size and a rate constant be ranked against each other. We evaluate Eq. (21) by central differences at $\pm 1\,\%$ around the fitted values, for three outputs — the peak daily admissions, the total admissions over the window, and the mean effective reproduction number.

One caveat governs how these numbers should be read. They are elasticities of the model output holding the *fitted* parameters fixed. If α really were one per cent larger, the estimation of Sec. 4.5 would have returned different transmission values, partly absorbing the change. The elasticities therefore measure how much each assumed constant matters to the model's behaviour, not how much the published result would move if the assumption were corrected and everything refitted. For one parameter that distinction can be closed exactly, and it is the most

important one: $R_0(t) = \beta(t)/\alpha$ by definition, so an error of a given proportion in α rescales every reproduction number reported in this paper by that same proportion, whatever the fit does. A ten per cent error in the assumed infectious period is a ten per cent error in every R_0 quoted here.

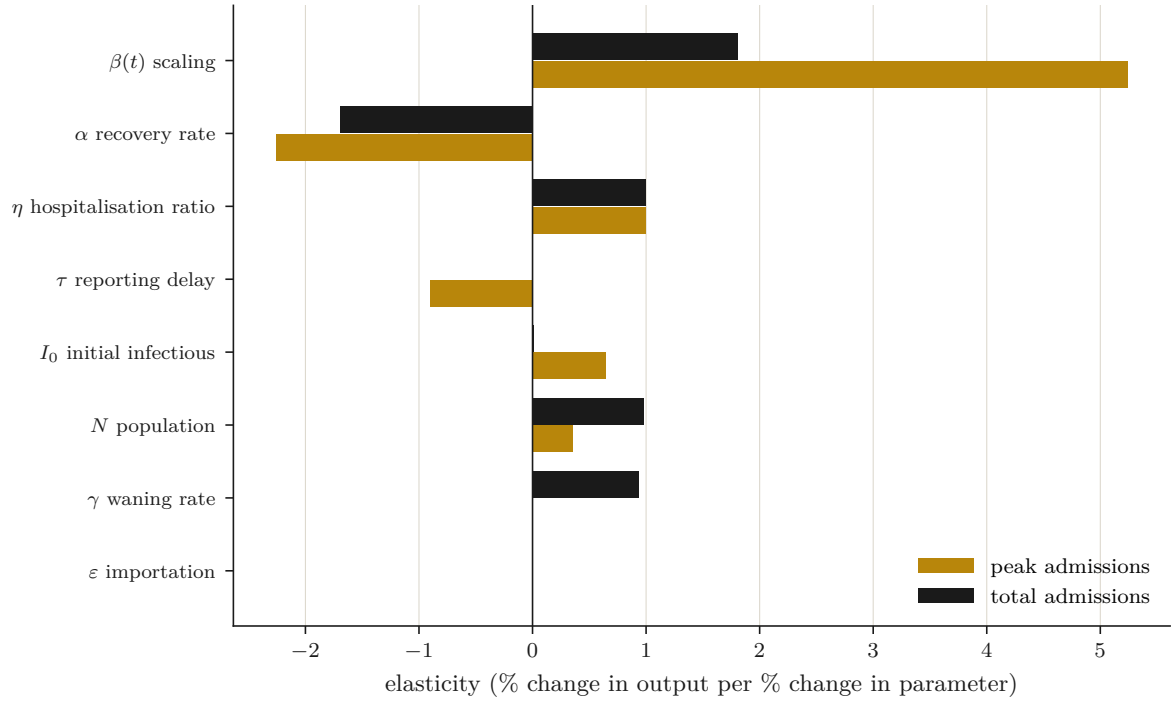

Figure 11
Elasticities of the peak and the total admissions with respect to every model parameter, ranked by the magnitude of the effect on the peak. A bar of length one means a one per cent change in the parameter moves the output by one per cent.

7.2 What the ranking shows

Three things stand out, and the ranking is not the one intuition suggests.

The peak belongs to transmission, the total belongs to waning. Scaling the whole transmission path moves the peak with an elasticity of 5.24 — by far the largest single effect — while the waning rate γ barely touches it, at 0.01. For the annual total the ordering reverses: γ carries an elasticity of 0.94, close to one for one. This is the same fact that Sec. 6.7 reads off the intervention surface, arriving from the other direction. How high the epidemic rises is a question about transmission; how many people it reaches in the end is a question about how fast immunity lapses. And γ is the parameter in Tab. 1 with the weakest justification, so the annual burden is the least secure quantity this model produces.

The recovery rate matters twice. α has an elasticity of -2.26 on the peak and -1.70 on the total, second only to transmission itself — and that is before the definitional effect noted above, by which any error in α passes straight into every reported R_0. It is assumed, not fitted.

The importation term is inert. ε registers an elasticity indistinguishable from zero on every output. Taken with Eq. (19) and Sec. 6.8, the numerical device introduced in Sec. 2.2 can be said to cost nothing that this analysis can detect. That is the outcome one wants from a term with no epidemiological content, and it is worth having checked rather than assumed.

The elasticity of η on both outputs is exactly one, which is a check rather than a finding: η multiplies Eq. (6) and nothing else, so anything other than unity would indicate an error in the implementation.

7.3 The waning rate, measured rather than cited

γ is the least defensible number in this model. It is not estimated, it was not derived from these data, and it stands in for a process that differs by variant, by vaccine, by dose and by individual. Tab. 1 therefore gives it without a citation, which is a deliberate choice: no single published value would honestly support 137 d as *the* duration of immunity to SARS-CoV-2, and citing one selected to look defensible would be worse than citing none. What can be done instead is to measure what depends on it.

We refit the model from scratch at seven assumed immunity durations from 60 to 730 days — a range spanning a factor of 12 in γ and bracketing every value seriously proposed — and at $\gamma = 0$, which is the SIR limit. Refitting rather than perturbing matters: the transmission path is free to compensate, which is what would have happened had a different value been assumed at the outset. Two claims are at stake.

Does SIRS still beat SIR? Yes, at 6 of the 7 durations tried, with the best at 60 days (AIC 9 746 against SIR's 9 831). The preference for waning immunity is not an artefact of the particular value assumed. It is also not a large effect: the AIC gap between the best SIRS fit and the SIR limit is smaller than the gap between good and bad choices of bucket width in Sec. 6.2, so waning immunity earns its parameter without being the most consequential structural decision in the paper.

Does the reproduction number survive? This is the more important question, and the answer is that it barely notices. Across the whole sweep the correlation between the recovered R_t and the published series stays between 0.801 and 0.843, and the fitted delay stays between 9.1 and 12.4 days. Even at $\gamma = 0$ — no waning at all, a structurally different model — the correlation is 0.810. Root-mean-square error moves only from 24.8 to 26.4 admissions per day.

The conclusion of Sec. 6.6 therefore does not rest on γ. That is worth more than a citation would have been: a reference would have asserted that 137 d is right, whereas this shows that the

paper's central result would be unchanged if it were wrong by a factor of ten in either direction. The quantity that *does* depend on γ is the cumulative burden, with an elasticity near unity (Sec. 7.2), and every statement about annual totals in this paper — particularly in Sec. 6.7 — should be read with that dependence attached.

8 Limitations

The assumptions of Sec. 2.5 are all false about the real epidemic. What follows is an account of which of those falsehoods actually constrain what may be concluded, roughly in decreasing order of how much damage they do.

A constant hospitalisation ratio forces changes in severity into the transmission path. This is the most consequential limitation in the paper. Assumption A7 holds η fixed across four years over which the probability that an infection ends in hospital certainly did not stay fixed: vaccination reduced it substantially from 2021 onwards, Omicron was intrinsically less severe than Delta, and admission thresholds moved with hospital pressure. The model has no way to represent a falling η, so a period in which the same number of infections produced fewer admissions can only be reproduced by lowering $\beta(t)$. The estimated $R_0(t)$ after early 2021 is therefore biased downward relative to true transmission, by an amount this study cannot quantify. That R_t nonetheless tracks the RIVM series across that period (Sec. 6.6) is evidence the bias is not large enough to destroy the signal — it is not evidence that the bias is absent.

Vaccination is invisible, so $R_0(t)$ conflates three things. By A5 vaccination is not a flow in the model; vaccinated people are not moved into R, and vaccine-derived and infection-derived immunity are the same state waning at the same rate. Whatever vaccination did to transmission is absorbed into $\beta(t)$. A fall in the estimated $R_0(t)$ during 2021 can therefore mean reduced contact, a less transmissible variant, or vaccine-derived protection, and nothing in this analysis separates them. The interpretation offered in Sec. 6.5 is accordingly restricted to the periods where a policy change and a change in $R_0(t)$ coincide closely enough in time that the alternatives are implausible.

This is a retrospective reconstruction, not a nowcast. The whole window is fitted at once, so the estimate of β in March 2020 uses admissions data from 2024. RIVM's published R_t was produced in real time, with only the data available on the day. The agreement reported in Sec. 6.6 is therefore between a retrospective estimate and a real-time one, which is a fair test of whether the admissions series carries the transmission signal, but says nothing at all about

whether this model would have been useful during the epidemic. Nothing here should be read as a claim about forecasting.

Homogeneous mixing misrepresents both spread and severity. Assumption A2 treats contacts as uniform across the whole population. Two consequences matter. Spatially, a well-mixed model spreads infection faster in the early phase than a structured population does, because real epidemics start in clusters. Demographically — and more seriously for this model — admissions are concentrated in the elderly, so the observation model of Eq. (6) is applying a population-average η to an infection flux whose age composition shifted considerably between waves. An age-stratified model would separate these; this one cannot.

The reporting delay is a point, not a distribution. The interval between infection and admission varies across individuals, so the correct relation between incidence and admissions is a convolution against a delay kernel. Eq. (6) uses a single τ, which leaves predicted peaks sharper than a distributed delay would produce and gives the fitted τ the character of a central tendency rather than a mean with a meaningful spread.

Two rate constants are assumed, and one of them is barely defensible. Neither α nor γ is estimated. The infectious period of $1/\alpha = 7\,\mathrm{d}$ is at least a conventional value. The waning rate $\gamma = 0.007\,3\,\mathrm{d}^{-1}$, an immunity duration of about 137 d, is a single number standing in for a process that differs by variant, by vaccine, by dose and by individual, and it was not derived from these data. Because $R_0(t) = \beta/\alpha$ is defined through α, any error in α rescales every reproduction number reported here by exactly the same factor. Sec. 7 quantifies both dependencies; they are the reason those two rows exist.

The transmission path has many parameters and should not be read pointwise. The baseline carries 72 transmission values against 1 495 observations. Sec. 5.3 tests directly whether that is too many by asking the fit to recover a path it was given, and Sec. 6.2 asks what the information criteria prefer. The outcome of both should govern how much weight is placed on any individual β_k; broad movements across several buckets are on much firmer ground than the value in any single fortnight.

The importation term has no epidemiological content. ε exists to keep the infectious compartment from collapsing numerically between waves (A8). Eq. (19) shows the floor it imposes is a handful of individuals, and Sec. 6.8 shows what it does to the threshold, but it

remains a device rather than a model of importation, which in reality is episodic and tied to travel.

Absolute scale is weaker than shape. The model never sees a measurement of how many people were infected, only how many were admitted. The product of η with the size of the latent epidemic is much better determined than either factor separately, so the absolute level of $I(t)$ should be treated with more caution than its shape. An independent anchor — the PIENTER seroprevalence surveys are the obvious candidate — would break that degeneracy and is the single most useful thing that could be added to this model. It is not used here.

9 Conclusion

The question was how much of an epidemic's transmission history can be recovered from hospital admissions alone. The answer is: most of it, and the reconstruction can be checked. A time-varying SIRS model fitted to nothing but the daily Dutch admission counts reproduces the effective reproduction number that RIVM published from independent data at a correlation of 0.804, agrees with it about whether the epidemic was growing on 81.5 per cent of intervals against 50.0 per cent for the best constant, and locates the March 2020 turning point in transmission 3 days after the measures — the remaining 8 days to the hospital peak being reporting delay, not epidemiology. None of that used a single case count, test result or contact survey.

The condition attached to that answer is the part worth carrying away. It holds at one particular model complexity and fails at others, and the complexity that best explains the observed series is not the one that best recovers the process underneath it. AIC selects 7 d buckets and 219 parameters, fits the admissions data better than anything else tried, and returns a reproduction number correlating with the published series at 0.173. BIC selects 21 d buckets and 75 parameters, fits measurably worse, and recovers R_t four times better. The extra control values are not resolving finer structure in transmission; they are absorbing reporting noise, and they do it by deforming the latent path that generated the data. Because that deformation makes the residuals smaller, no quantity computed from the fit can reveal it. Only the external series can, and most compartmental fitting exercises do not have one.

Three implementation faults found along the way changed the numbers as much as any modelling choice did: a hundred days of missing data being fitted as zeros, a reporting delay discretised so coarsely that its gradient was exactly zero, and an integrator stepping blind across the discontinuities in $\beta(t)$. Each was found by a check that compared the code against something known independently, and none would have shown up in the goodness of fit.

The single claim this paper makes is this: hospital admissions alone are sufficient to reconstruct

the effective reproduction number of an epidemic, but only at a model complexity that the fit to those admissions cannot itself identify — the resolution that best explains the data recovers the underlying transmission four times worse than a coarser one that explains it slightly less well.

A The fitted transmission path

Tab. 5 lists every estimated transmission value with the dates it covers and the implied instantaneous basic reproduction number $R_0(t) = \beta_k/\alpha$. It is the full content of Fig. 7, given numerically so that any claim made about a particular wave can be checked against the value the model actually used.

Table 5

Fitted transmission values for the baseline daily SIRS model. Each bucket spans 21 d from the date shown; β_k is in d^{-1} and $R_0 = \beta_k/\alpha$ is dimensionless.

k	from	β_k	R_0	k	from	β_k	R_0
1	2020-02-27	0.303	2.12	37	2022-03-24	0.320	2.24
2	2020-03-19	0.149	1.04	38	2022-04-14	0.277	1.94
3	2020-04-09	0.099	0.69	39	2022-05-05	0.285	2.00
4	2020-04-30	0.127	0.89	40	2022-05-26	0.382	2.68
5	2020-05-21	0.117	0.82	41	2022-06-16	0.333	2.33
6	2020-06-11	0.120	0.84	42	2022-07-07	0.257	1.80
7	2020-07-02	0.233	1.63	43	2022-07-28	0.213	1.49
8	2020-07-23	0.197	1.38	44	2022-08-18	0.244	1.71
9	2020-08-13	0.200	1.40	45	2022-09-08	0.323	2.26
10	2020-09-03	0.243	1.70	46	2022-09-29	0.244	1.71
11	2020-09-24	0.238	1.67	47	2022-10-20	0.189	1.32
12	2020-10-15	0.204	1.43	48	2022-11-10	0.255	1.78
13	2020-11-05	0.196	1.37	49	2022-12-01	0.273	1.91
14	2020-11-26	0.297	2.08	50	2022-12-22	0.192	1.35
15	2020-12-17	0.287	2.01	51	2023-01-12	0.205	1.43
16	2021-01-07	0.294	2.06	52	2023-02-02	0.293	2.05
17	2021-01-28	0.329	2.30	53	2023-02-23	0.234	1.63
18	2021-02-18	0.375	2.63	54	2023-03-16	0.179	1.25
19	2021-03-11	0.412	2.89	55	2023-04-06	0.181	1.27
20	2021-04-01	0.457	3.20	56	2023-04-27	0.156	1.09
21	2021-04-22	0.426	2.98	57	2023-05-18	0.160	1.12
22	2021-05-13	0.276	1.93	58	2023-06-08	0.150	1.05
23	2021-06-03	0.154	1.08	59	2023-06-29	0.199	1.39
24	2021-06-24	0.491	3.43	60	2023-07-20	0.196	1.37
25	2021-07-15	0.294	2.05	61	2023-08-10	0.228	1.59
26	2021-08-05	0.259	1.81	62	2023-08-31	0.185	1.29
27	2021-08-26	0.212	1.49	63	2023-09-21	0.199	1.39
28	2021-09-16	0.262	1.83	64	2023-10-12	0.191	1.34
29	2021-10-07	0.313	2.19	65	2023-11-02	0.186	1.30
30	2021-10-28	0.316	2.21	66	2023-11-23	0.193	1.35
31	2021-11-18	0.296	2.07	67	2023-12-14	0.124	0.87
32	2021-12-09	0.264	1.85	68	2024-01-04	0.072	0.50
33	2021-12-30	0.357	2.50	69	2024-01-25	0.110	0.77
34	2022-01-20	0.377	2.64	70	2024-02-15	0.144	1.01
35	2022-02-10	0.399	2.79	71	2024-03-07	0.146	1.02
36	2022-03-03	0.441	3.09	72	2024-03-28	0.141	0.99

B Reproducing these results

Everything below assumes a clone of the repository (Sec. D) and Python 3.11 or later.

Environment.

```
python -m venv .venv && source .venv/bin/activate
pip install numpy scipy pandas matplotlib
```

The figures are typeset through LaTeX by matplotlib's `pgf` backend, so a TeX installation providing `pdflatex`, `lmodern` and `siunitx` must be on the path. Without one, the figure scripts fail; the numerical results do not depend on it.

Data. The raw RIVM files are about 0.8 GB and are not in the repository.

```
python fetch_data.py
```

downloads them from https://data.rivm.nl/covid-19/ and applies the renaming described in Sec. C. The results in this paper use the copies retrieved on 19–20 January 2026; RIVM has since stopped updating several of the series, so a fresh download may cover a different span. The fitting window is derived from the data rather than hard-coded, so it will follow.

Running. In order, with approximate wall-clock times on eight cores:

```
python src/Pipelines/run_baseline_fit.py        # the two baseline fits
python src/Experiments/verify_integrator.py     # conservation, convergence
python src/Experiments/verify_recovery.py       # parameter recovery
python src/Experiments/exp_bucket_sweep.py      # control-value sweep
python src/Experiments/exp_rt_agreement.py      # the RIVM comparison
python src/Experiments/exp_threshold.py         # threshold behaviour
python src/Experiments/exp_intervention.py      # counterfactual sweep
python src/Experiments/exp_sensitivity.py       # elasticities
```

Each writes JSON, and in two cases NPZ, under `results/`. Nothing downstream re-runs the optimiser.

Rebuilding the paper.

```
cd paper
python make_numbers.py       # results/ -> generated/numbers.tex
python figures/make_figures.py
latexmk -pdf COVID_project_compartmental_model.tex
```

Every number quoted in the text is a macro defined in `generated/numbers.tex` and generated from the stored results; none is typed by hand. If a result file is missing the corresponding macro is undefined and the build fails rather than printing a stale value.

Randomness. The only stochastic element is the choice of starting points for the optimiser, and the synthetic noise in Sec. 5.3. Both are drawn from `numpy.random.default_rng` with the fixed seeds recorded in the `protocol` block of each result file, so repeated runs reproduce exactly.

C Data sources

All data comes from the open-data portal of the Dutch National Institute for Public Health and the Environment (RIVM) at https://data.rivm.nl/covid-19/, retrieved 19–20 January 2026. RIVM revised its reporting on 4 October 2021 and archived the earlier series under a `_tm_03102021` suffix; the two halves are concatenated on load.

Hospital admissions — `COVID-19_ziekenhuisopnames.csv` and `COVID-19_ziekenhuisopnames_tm_03102021.csv`. Daily admissions, concatenated across the 2021 reporting change. This is the *only* series the model is fitted to.

Reproduction number — `COVID-19_reproductiegetal.json` and `COVID-19_reproductiegetal_tm_03102021.json`. RIVM's published effective reproduction number with its 95 % interval, ending 2023-07-09. Used exclusively for the comparison in Sec. 6.6, and never in the fit.

Loaded but unused — `COVID-19_prevalentie.json`, `COVID-19_rioolwaterdata.csv`, `COVID-19_uitgevoerde_testen*.csv`, `ARI_Infectieradar_symptomen_per_dag.csv`, `COVID-19_Infectieradar_symptomen_per_dag.csv` and `COVID-19_ic_opnames*.csv`. Available to the data pipeline for exploration; none enters any result reported here.

D Repository

The model code, the experiment scripts, the figure scripts and the LaTeX source of this paper are at

https://github.com/LetoCave/covid-compartmental-model

The commit that produced this version of the paper is recorded in the repository's tag history. The raw data is excluded for size and is recovered with `fetch_data.py`.

References

[1] Statistics Netherlands (CBS). *Population; key figures, 1950–2022*. Table 37296eng. Population of the Netherlands on 1 January 2020: 17 407 585. CBS StatLine. 2022. URL: https://opendata.cbs.nl/statline/#/CBS/en/dataset/37296eng/table (visited on 09/06/2026).

[2] Qun Li, Xuhua Guan, Peng Wu, et al. “Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia”. In: *New England Journal of Medicine* 382.13 (2020), pp. 1199–1207. DOI: 10.1056/NEJMoa2001316.

[3] Odo Diekmann, Johan A. P. Heesterbeek, and Johan A. J. Metz. “On the Definition and the Computation of the Basic Reproduction Ratio R_0 in Models for Infectious Diseases in Heterogeneous Populations”. In: *Journal of Mathematical Biology* 28.4 (1990), pp. 365–382. DOI: 10.1007/BF00178324.

[4] Pauline van den Driessche and James Watmough. “Reproduction Numbers and Sub-threshold Endemic Equilibria for Compartmental Models of Disease Transmission”. In: *Mathematical Biosciences* 180.1–2 (2002), pp. 29–48. DOI: 10.1016/S0025-5564(02)00108-6.

[5] William Ogilvy Kermack and Anderson Gray McKendrick. “A Contribution to the Mathematical Theory of Epidemics”. In: *Proceedings of the Royal Society of London A* 115.772 (1927), pp. 700–721. DOI: 10.1098/rspa.1927.0118.

[6] John R. Dormand and Peter J. Prince. “A Family of Embedded Runge–Kutta Formulae”. In: *Journal of Computational and Applied Mathematics* 6.1 (1980), pp. 19–26. DOI: 10.1016/0771-050X(80)90013-3.

[7] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, et al. “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”. In: *Nature Methods* 17 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

[8] Roy M. Anderson, Hans Heesterbeek, Don Klinkenberg, and T. Déirdre Hollingsworth. “How Will Country-Based Mitigation Measures Influence the Course of the COVID-19 Epidemic?” In: *The Lancet* 395.10228 (2020), pp. 931–934. DOI: 10.1016/S0140-6736(20)30567-5.