

A simple branching process model for Delta-variant and non-Delta-variant incidence of COVID-19 in Germany

Johannes Bracher

July 5, 2021

This is just a sketch shared via a GitHub repository to facilitate discussion. While I think that the results are reasonable, please keep in mind that this is a rather simple model taking into account only the fact that there are two co-circulating variants. Vaccination, population behaviour, NPIs etc are ignored. Also, I set this up quite quickly and cannot guarantee that there are no technical mistakes.

1 Data and notation

The model involves the following quantities:

- To describe the epidemic process:
 - X_t : The total number of notified cases in week t , either from RKI or JHU (both by reporting date to the national authorities). Weeks run from Sunday through Saturday to ensure comparability to the Forecast Hub.
 - X_t^Δ, X_t^o : The number of Delta and other cases among the notified cases, respectively (unobserved)
 - r_t^Δ, r_t^o : One plus the weekly growth rate of the Delta variant and other variants combined, respectively.
 - ψ^Δ, ψ^o : Overdispersion parameters for the two variants.
 - $\omega_t^\Delta, \omega_t^o$: auxiliary variables needed to express the negative binomial distributions as a Poisson-gamma mixture.
 - $\lambda_t^\Delta, \lambda_t^o$: Conditional expectations of X_t^Δ, X_t^o given $X_{t-1}^\Delta, X_{t-1}^o$ and $\omega_t^\Delta, \omega_t^o$.
- To describe the sequencing:
 - N_t : The number of sequenced cases in the random sample from RKI in week t . Note: These weeks are defined as Monday through Sunday.
 - Y_t^Δ : The number among these which are from the Delta variant in week t .

Caveat: The definition of the weeks for the different data streams is not identical, and I am not sure whether one needs to shift one of them slightly to have them actually refer to the same time periods. The sequencing data are by date of sampling, while the incidence data are by date of reporting to the national authorities.

2 Model formulation

I apply a fully Bayesian model defined as follows. The total number of observed cases is decomposed into

$$X_t = X_t^\Delta + X_t^o$$

and assume negative binomial branching processes for X_t^Δ and X_t^o :

$$\begin{aligned} X_t^{(\cdot)} \mid \lambda_t^{(\cdot)} &\sim \text{Pois}(\lambda_t) \\ \lambda_t^{(\cdot)} &= \omega_t \times r_t \times X_{t-1}^{(\cdot)} \end{aligned}$$

Here, $\omega_t^{(\cdot)}$ is an auxiliary gamma-distributed random variable,

$$\omega_t^{(\cdot)} \sim \text{Gamma}(\psi^{(\cdot)}, \psi^{(\cdot)})$$

meaning that without conditioning on $\omega_t^{(\cdot)}$ the $X_t^{(\cdot)}$ are negative binomial with mean $r_t^{(\cdot)} X_{t-1}^{(\cdot)}$ and dispersion parameter $\psi^{(\cdot)}$. For the time-varying weekly multiplication factors $r_t^{(\cdot)}$ (one plus growth rate) we assume a random walk structure,

$$\begin{aligned} r_t^{(\cdot)} &= r_{t-1}^{(\cdot)} \times \exp(\epsilon_t^{(\cdot)}) \\ \epsilon_t^{(\cdot)} &\sim \text{N}(0, \sigma^{2(\cdot)}). \end{aligned}$$

We thus assume that the $r_t^{(\cdot)}$ show some smoothness and cannot change too quickly. **Caveat:** This formulation of the random walk is advantageous as it ensures that $r_t^{(\cdot)}$ is always positive. However, it may introduce a slight upward trend on the scale of $r_t^{(\cdot)}$ (as $\mathbb{E}(r_{t+1}^{(\cdot)} \mid r_t^{(\cdot)}) > r_t^{(\cdot)}$). However, I am unsure how this translates to the case scale.

For the sequencing results we add the following observation process:

$$Y_t^\Delta \mid N_t \sim \text{Bin}(N_t, p_t^\Delta),$$

where

$$p_t^\Delta = \frac{X_t^\Delta}{X_t}.$$

To complete the model we need to specify some prior distributions for the parameters and initialization of the epidemic process:

$$\begin{aligned} 1/\sigma^{2(\cdot)} &\sim \text{Unif}(5, 200) \\ \psi^{(\cdot)} &\sim \text{Unif}(0, 100) \\ X_{16}^\Delta &\sim \text{Unif}(0, 1000.) \end{aligned}$$

These are intended to be “uninformative”, motivating priors for these parameters is not straightforward.

3 Inference

I generate samples from the posterior distribution of the various unobserved and predicted quantities using the JAGS software via the R package `runjags` (see `JAGS_model.R` for the code). I generate 50,000 samples from 5 chains after a burn-in period of 4,000 samples.

4 Results

I ran the model on the RKI and JHU data and obtained the results shown in Figures 1 and 2. The RKI data are by reporting date to the local authorities, while JHU data seem to be by reporting date to national authorities (but the two data sources look ery similar in recent weeks). While the data imply slightly different current case levels, in both cases the extrapolation of trends implies a rather sudden flattening of the overall case counts which had previously been decaying steadily. The data which have accumulated since the last

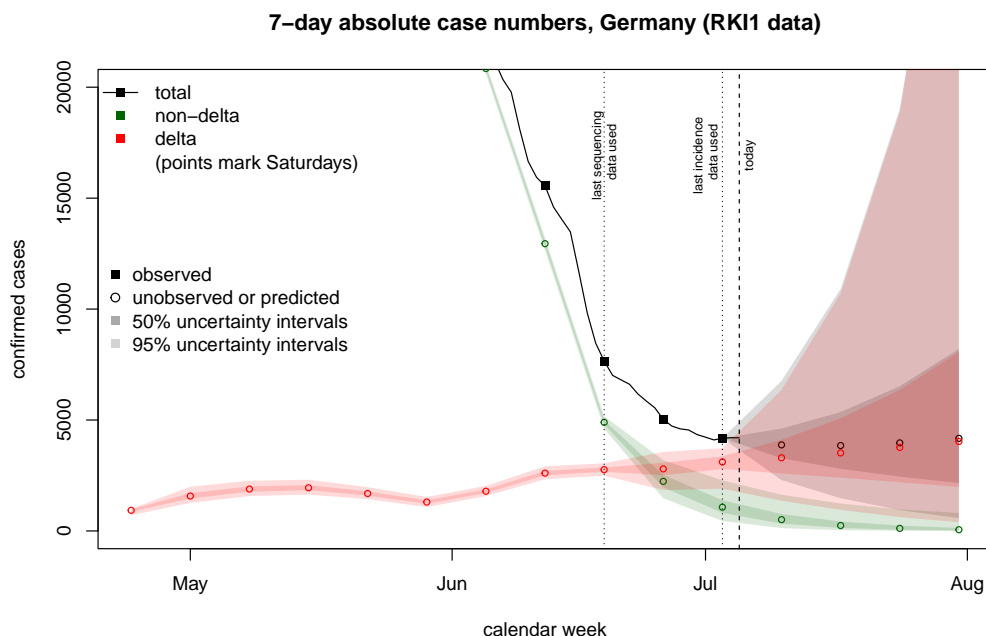


Figure 1: Results based on RKI data, 5 July 2021. Note that predictions have been generated up to 4 weeks ahead to conform with the Forecast Hub format, but I recommend focusing on one and two-week-ahead predictions only.

used data point point into this direction, too, even though the flattening may not be as abrupt as implied by the point predictions. Note that the model implies a lot of uncertainty and for more than one week ahead actually makes very uninformative predictions. In three weeks the Delta variant will have become strongly dominant.

Note that the availability of sequencing results lags behind the available incidence data and there is always at least one week for which the overall case count is already known while the sequencing results are not. This is reflected by a widening of the uncertainty intervals for the strain-specific numbers.

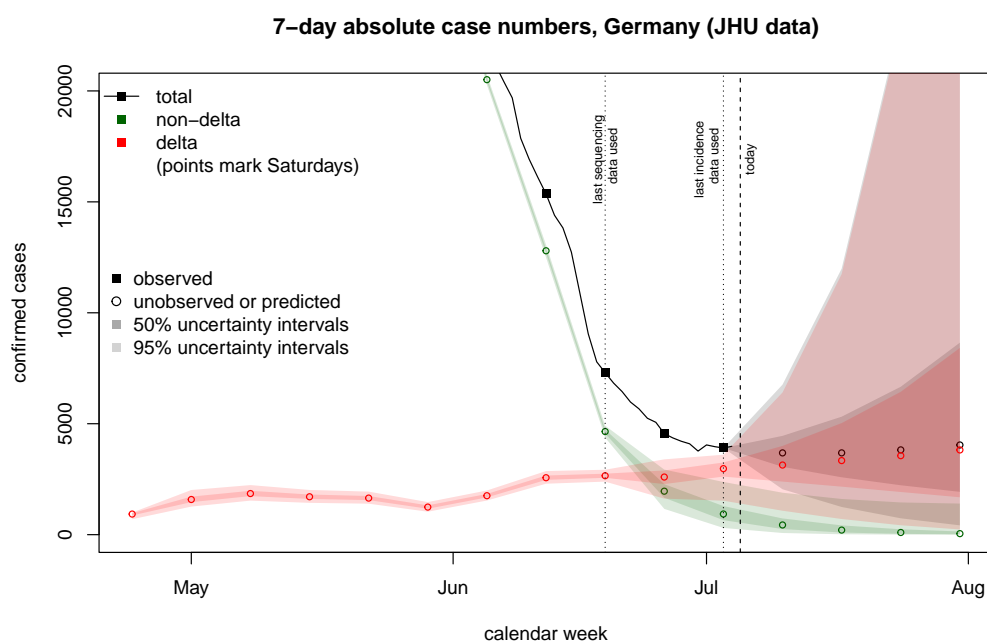


Figure 2: Results based on JHU data, 5 July 2021. Note that predictions have been generated up to 4 weeks ahead to conform with the Forecast Hub format, but I recommend focusing on one and two-week-ahead predictions only.