ESTIMATION AND INTERACTIVE VISUALIZATION OF THE TIME-VARYING REPRODUCTION NUMBER R_t AND THE TIME-DELAY FROM INFECTION TO ESTIMATION

A PREPRINT

Fabian Valka
vektorraum
Vienna, Austria
fvalka@vektorraum.com

April 29, 2020

ABSTRACT

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Keywords First keyword · Second keyword · More

1 Introduction

An interactive web-based tool was developed for visualizing and exploring the time-varying reproduction number and time delays from infection to onset, to reporting and inclusion in the estimation of $R_{t,\tau}$ on a regional level for Austria. Users can choose the Austrian state, sliding time window length and overlay the AGES $R_{t,13}$ estimate [1].

The tool is available as free open-source software on GitHub:
`https://github.com/fvalka/r_estimate-methods`.

2 Estimation of the Time-Varying Reproduction Number R_t

In analogy to the method published by Richter, et al.[2] from AGES, the time-varying reproduction number R_t is estimated according to Cori, et al.[3] Sliding time windows of length τ days are used with the assumption that within the time window τ the reproduction number R_t is constant, defining $R_{t,\tau}$.

The R-package EpiEstim[4] provided by Cori, et al. is used for the implementation.

2.1 Serial Interval

Accuracy of the serial interval estimation plays a key role in the accuracy of the estimation of $R_{t,\tau}$.

Representing the uncertainties contained in the current serial interval estimates[5] was realized using the uncertain_si method in estimate_R. This samples the parameters for the serial intervals Gamma distribution Γ_s from a truncated normal distribution. The median of the $R_{t,\tau}$ estimate is only affected to a limited extend by this, but the credible interval

		95% ci		
Parameter	Mean	Lower	Upper	Std. Dev.
Serial interval mean μ_s	4.46	4.160	4.760	0.153
Serial interval standard deviation σ_s	2.63	2.369	2.891	0.133
		parameters for uncertain_si		

Table 1: Serial interval Gamma distribution Γ_s parameters based upon estimates by Richter, et al.[5]

will be increased around those median values[3], giving a more accurate representation based on our current believes about the serial interval.

Parameters used in the implementation are documented in Table 1.

Standard deviations of the mean serial interval and the standard deviation of the standard deviation were obtained based upon the assumption of normality from the 95% credible interval, see Equation 1.

$$\sigma_{\mu_x} = \frac{\mu_x - \mathrm{CI}_{lower}}{1.96} \tag{1}$$

Based upon these parameters multiple serial intervals are explored by estimate_R.[4] An example of explored serial interval distributions are shown in Figure 3.

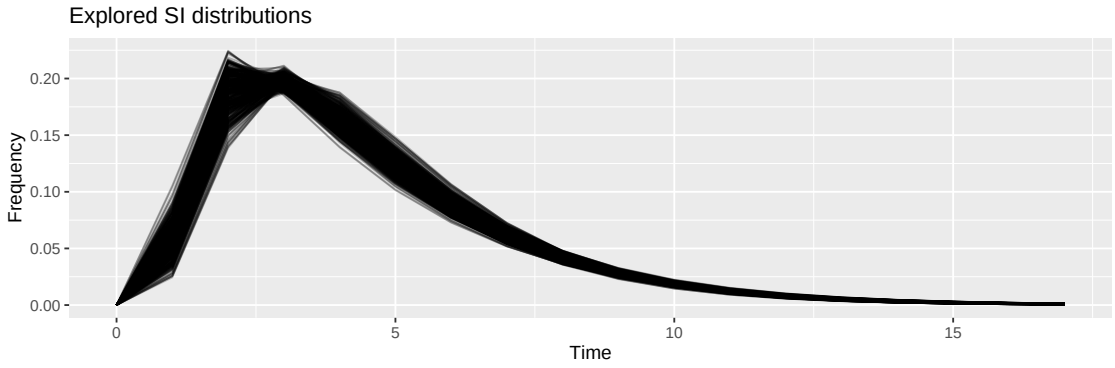


Figure 1: Example of serial interval distributions explored during $R_{t,\tau}$ estimation.

2.2 Sliding Time Window τ

Choosing τ is a trade-off between less noise and less delay. Smaller time windows τ lead to a more rapid detection of changes in $R_{t,\tau}$, but also to more statistical noise.[6]

Cori, et al. describe a method for choosing the time window size τ based upon the minimum number of cases included in each time window [6]. Based upon a chosen target CV of at least 0.3 a minium of 12 cases in each time window is required. With this in consideration and the usage of a time window of 7 days in Cori, et al.[3] a time window size τ of 7 days for the UI was chosen. AGES uses a 13 days time window for their estimation, no explanation is given for this choice and this τ was therefore only used for comparisons but can be chosen by the user on the interactive front-end.

The effect a change in τ has on the estimated $R_{t,\tau}$ is illustrated in Figure 2. Comparing the first date where the median of τ has fallen below 1 illustrates the trade-off discussed above. For $\tau = 7$ the condition $R_{t,7} < 1$ is first met on the 31. of March 2020, for $\tau = 13$ the condition $R_{t,13} < 1$ is only met 4 days later on the 4th of April 2020.

A 7-day window has the added advantage that it always contains exactly 2 weekend days.

2.3 Crossvalidation with AGES estimates for Austria

The output of the estimation described in this section was compared to the $R_{t,\tau}$ estimates obtained by AGES on the 24th of April 2020[1].

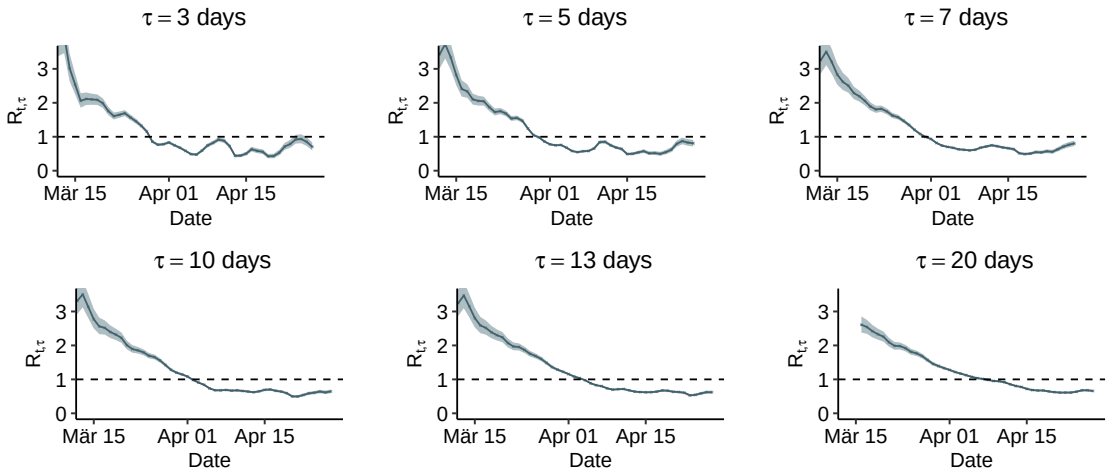
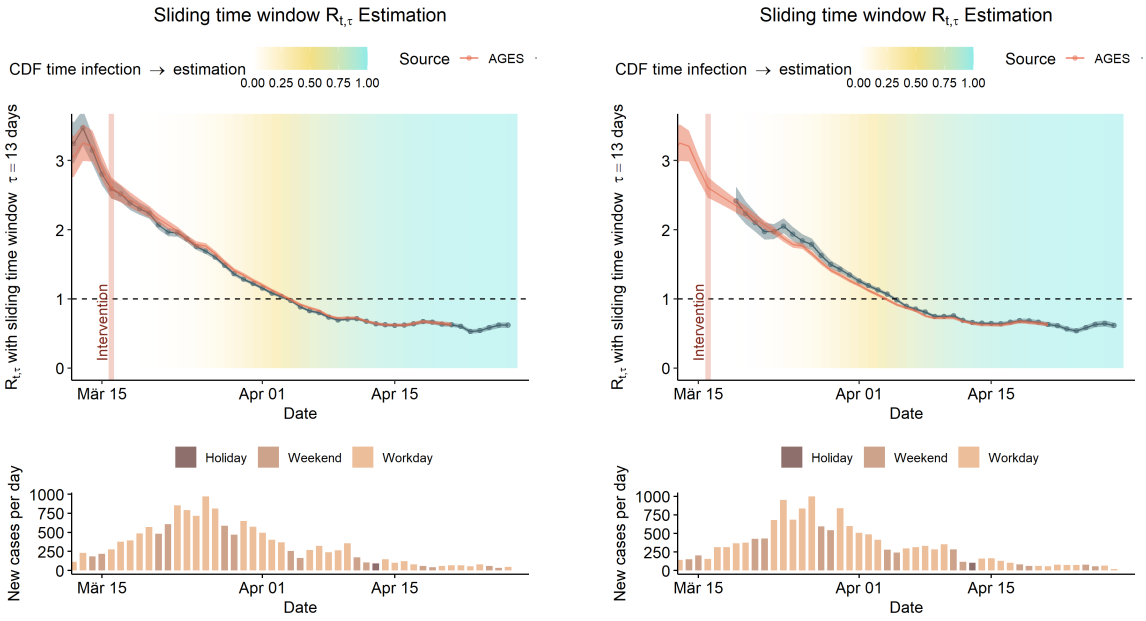


Figure 2: Effect of different time window, τ, choices on the $R_{t,\tau}$ estimate for Austria, as of 2020-04-27. Left censored at first entry in data-set $t_0 + \tau$ Data source: data.gv.at.[7]



(a) Closest reproduction using the data.gv.at dataset[7] and parametric_si .

(b) Dataset used in actual visualization and using uncertain_si in estimate_R.

Figure 3: Comparison between own estimates and $R_{t,\tau}$ estimates published by AGES.[8]

Two different methods were chosen for the comparison, one where the dataset which allows the closest reproduction of the estimates obtained by AGES was used as shown in figure Figure 3a. In the actual interactive visualization a different dataset is used. The comparison against this dataset is shown in Figure 3b. For a further discussion on the datasets see section 4.

3 Estimation of the time delay from infection to inclusion in the $R_{t,\tau}$ estimation

The total time delay from infection to inclusion in the estimate $t_{infection,estimation}$ will be split up into three different time delays. The time delay from infection to onset $t_{infection,onset}$, the incubtion period. The time delay from onset to

reporting $t_{onset,reporting}$ and the time delay from reporting to inclusion in the estimation $t_{reporting,estimation}$, which is caused by the $R_{t,\tau}$ estimation method of assuming a constant R_t within the time window τ.

3.1 Time delay from infection to reporting $t_{infection,reporting}$

Both $t_{infection,onset}$ and $t_{onset,reporting}$ are assumed to be independent random variables and a Gamma distribution was used for the estimation. Estimates of the Gamma distribution parameters for $t_{infection,onset}$ were taken from the supplementary appendix of the study by Zhang, et al.[9] based upon case data from Mainland China, excluding Hubei province. The parameters used are $\alpha_{infection,onset} = 4.23(\text{SD } 1.28)$ and $\beta_{infection,onset} = 0.81(\text{SD } 0.24)$ with a mean of 5.2days and a 95% CI of the mean of (95% CI: $1.5 - 11.3$) days.

Onset to reporting Gamma distribution parameters were also taken from the estimates from Mainland China, excluding Hubei[9] from the second period (Jan 28 – Feb 17) were used. Those parameters were obtained from 2079 observations and are $\alpha_{onset,reporting} = 3.18$ and $\beta_{onset,reporting} = 0.59$ giving a mean of 5.3 (95% CI: $1.2 - 13.1$) days.

In order to obtain the empirical CDF of $t_{infection,reporting}$ random samples were drawn from the Gamma distributions of $t_{infection,onset}$ and $t_{onset,reporting}$ using the R-function **rgamma**[10]. Random samples were added up pair-wise as shown in Equation 2.

$$t^i_{infection,reporting} = t^i_{infection,onset} + t^i_{onset,reporting} \tag{2}$$

A total of 10^6 samples were drawn from each distribution and the resulting emperical distribution function $\widehat{F}_{infection,reporting}$ was obtained by applying the ecdf function[10] to the list of $t^i_{infection,reporting}$.

3.2 Time delay from reporting to inclusion in the $R_{t,\tau}$ estimation $t_{reporting,estimation}$

For modeling the effect of the assumption that R_t is constant within the time window τ[3] the assumption was made that this time delay can be modeled by applying a simple moving average to the ecdf $\widehat{F}_{infection,reporting}$ with a backwards time-window of length τ.

Combining the ecdf for $t_{infection,reporting}$ with the simple moving average we obtain the modeled ecdf for $t_{infection,estimation}$ as shown in Equation 3.

$$\widehat{F}_{infection,reporting}(t) = \text{SMA}\left(\widehat{F}_{infection,reporting}\right)(t) \tag{3}$$

4 Data Sources

References

[1] Lukas Richter, Daniela Schmid, Ali Chakeri, Sabine Maritschnik, Sabine Pfeiffer, and Ernst Stadlober. Epidemiologische Parameter des COVID19 Ausbruchs - Update 24.04.2020, Österreich, 2020. page 5, 2020.

[2] Lukas Richter, Daniela Schmid, and Ernst Stadlober. Methodenbeschreibung für die Schätzung von epidemiologischen Parametern des COVID19 Ausbruchs, Österreich. page 4.

[3] Anne Cori, Neil M. Ferguson, Christophe Fraser, and Simon Cauchemez. A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics. *American Journal of Epidemiology*, 178(9):1505–1512, November 2013.

[4] Anne Cori. *EpiEstim: Estimate Time Varying Reproduction Numbers from Epidemic Curves*, 2019. R package version 2.2-1.

[5] Lukas Richter, Daniela Schmid, Ali Chakeri, Sabine Maritschnik, Sabine Pfeiffer, and Ernst Stadlober. Schätzung des seriellen Intervalles von COVID19, Österreich. page 3, 2020.

[6] Anne Cori, Neil M. Ferguson, Christophe Fraser, and Simon Cauchemez. A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics: Web material. *American Journal of Epidemiology*, 178(9):1505–1512, November 2013.

[7] Sozialministerium. COVID-19: Alle verfügbaren Dateien (gezippt) - data.gv.at. Library Catalog: www.data.gv.at.

[8] Wissen Aktuell: Epidemiologische Parameter des COVID19 Ausbruchs, Österreich, 2020.

[9] Juanjuan Zhang, Maria Litvinova, Wei Wang, Yan Wang, Xiaowei Deng, Xinghui Chen, Mei Li, Wen Zheng, Lan Yi, Xinhua Chen, Qianhui Wu, Yuxia Liang, Xiling Wang, Juan Yang, Kaiyuan Sun, Ira M Longini, M Elizabeth Halloran, Peng Wu, Benjamin J Cowling, Stefano Merler, Cecile Viboud, Alessandro Vespignani, Marco Ajelli, and Hongjie Yu. Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. *The Lancet Infectious Diseases*, page S1473309920302309, April 2020.

[10] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020.