Estimation and worldwide monitoring of the effective reproductive number of SARS-CoV-2

Jana S. Huisman[1,2¶], Jérémie Scire[2,3¶*], Daniel C. Angst[1]
Richard A. Neher[2,4], Sebastian Bonhoeffer[1†], Tanja Stadler[2,3†*]

[1] Department of Environmental Systems Science, ETH Zurich, Swiss Federal Institute of Technology, Zurich, Switzerland

[2] Swiss Institute of Bioinformatics, Lausanne, Switzerland

[3] Department of Biosystems Science and Engineering, ETH Zurich, Swiss Federal Institute of Technology, Basel, Switzerland

[4] Biozentrum, University of Basel, Basel, Switzerland

[¶,†] These authors contributed equally

[*] Corresponding authors: jeremie.scire@bsse.ethz.ch, tanja.stadler@bsse.ethz.ch

Classification: Biological Sciences, Population Biology

Keywords: Reproductive number, SARS-CoV-2, infectious disease, public health surveillance

Abstract

The effective reproductive number R_e is a key indicator of the growth of an epidemic. Since the SARS-CoV-2 pandemic started, many methods and online dashboards have sprung up to monitor this number. However, these methods are not always thoroughly tested or are applied only to a limited geographic range. Here, we present a method for near real time monitoring of R_e, applied to epidemic data from 170 countries. We thoroughly validate the method on simulated data, and present an intuitive web interface for interactive data exploration. We show that in the majority of countries the estimated R_e dropped below 1 only after the introduction of major non-pharmaceutical interventions. For Europe, Asia, and North America we found that the implementation of non-pharmaceutical interventions was associated with reductions in the effective reproductive number. Globally, we found that relaxing non-pharmaceutical interventions did not fully revert R_e values to their original levels. Generally, our framework is useful both to inform governments and the general public on the status of the epidemic in their country, as well as a source for detailed comparison between countries and in relation to local public health policies and external covariates such as mobility, behavioural, or weather data.

Significance statement

During the SARS-CoV-2 pandemic, governments need a way to monitor the epidemiological situation in their country. A key indicator is the effective reproductive number R_e. It describes the average number of secondary infections caused by a primary infected individual. Here, we present a method to estimate R_e from case report data. We thoroughly validate the method on simulated data, and present R_e estimates for 170 countries on an interactive web interface. We then use this method to investigate the impact of non-pharmaceutical interventions on reducing R_e worldwide. We find that the estimated R_e was significantly above 1 prior to the introduction of major non-pharmaceutical interventions, and that relaxing these interventions does not fully revert R_e estimates to their prior levels.

1 Introduction

During an infectious-disease outbreak, such as the ongoing SARS-CoV-2 pandemic, accurate monitoring of the epidemic situation is critical to the decision-making process of governments and public health authorities. The magnitude of an epidemic, as well as its spatial and temporal infection dynamics determine the exposure risk posed to citizens in the near and long-term future, the pressure on critical infrastructure like hospitals, and the overall burden of disease to society.

The effective reproductive number R_e is a key indicator variable to describe how a pathogen spreads in a particular population [1, 2, 3]. It quantifies the average number of secondary infections caused by a primary infected individual. It also has a natural threshold value of 1, below which the epidemic reduces in size [1, 4]. R_e typically changes during the course of an epidemic as a result of the depletion of susceptible individuals, changed contact behaviour, seasonality of the pathogen, or the effect of pharmaceutical and non-pharmaceutical interventions (NPIs) [1, 5, 6, 7, 8].

Different methods have been developed to estimate the effective reproductive number. They broadly fall into two categories: those based on compartmental models, e.g. [5, 9, 10], and those that count the number of secondary infections per infected individual directly, based on a time series of infection incidence, e.g. [11, 12]. We focus on the latter methods, in particular the EpiEstim method of Cori et al. [12], as they rely on only few, simple assumptions, are less prone to model misspecifications, and well-suited for near real-time monitoring of the epidemic [13].

The infection incidence based methods face the difficulty that infection events cannot be observed directly [13]. These events can only be surmised with a certain time lag, e.g. when individuals show symptoms and are tested, via contact tracing, or via periodic testing of a cohort of individuals [4]. To use these methods, one must thus employ a proxy for infection events (e.g. the observed incidence of confirmed cases, hospitalisations, deaths). This proxy is either used directly in lieu of the infection incidence, or it is used as an indirect observation to infer past infections [13]. A benefit of the infection incidence based methods is that they can be applied to multiple proxies of infection events independently, allowing for direct comparison of the inferred results for the same epidemic [6]. However, depending on the method used to infer infections from the observed incidence time series, one can also introduce biases such as smoothing sudden changes in R_e [13, 14, 15].

Several methods, software packages, and online dashboards have been developed to monitor R_e during the SARS-CoV-2 pandemic (e.g. [16, 17, 18]). A pipeline for the continuous monitoring of R_e using infection incidence based methods should include four critical steps: (i) gathering and curation of observable proxy data of infection incidence, (ii) reconstruction of the unobserved infection events, (iii) R_e estimation, and (iv) communication of the results, including uncertainty and potential biases. These are four axes that also define the differences between existing methods. During the SARS-CoV-2 epidemic, many local public health authorities have made case data publicly available. Depending on the data sources used, R_e reports span different geographical scopes, from the scale of a city, region, country, or the entire globe [17, 19, 20]. The second step, i.e. going from a noisy time series of indirect observations to an infection incidence time series, is technically the most challenging. Biases can be introduced easily, and accurately assessing the uncertainty around the inferred infection incidence is a challenge in itself [13]. For the third step, i.e. to estimate R_e from a timeline of infection events, there are ready-to-use software packages [12, 21], which produce unbiased R_e estimates along with an estimate of the uncertainty resulting from this step. Finally, the communication of results to the general public and decision makers is essential, but is often overlooked.

We present a pipeline, together with an online dashboard, for near real-time monitoring of R_e. We use publicly available data gathered by different public health authorities. Wherever possible, we show results obtained from different types of case reports (confirmed cases, hospitalisations or death). This allows comparison across observation types and hence a better assessment of the validity of the estimates. Results are updated daily, and can be found on `https://ibz-shiny.ethz.ch/covid-19-re/`.

Because R_e estimates reflect changes in virus transmission dynamics, they can be used to assess the impact of public health interventions. Prior work on the relative impact of specific non-pharmaceutical interventions on R_e has shown conflicting results [8, 22, 23, 24, 25, 26]. These differences can be attributed mostly to different model formulations [24, 27], including differing assumptions on the independence of NPIs [27], differing timescales over which the effect of the NPI was analysed [8, 25], whether the time point of the NPI was assumed fixed or allowed to vary [26], and differing geographical scope.

There is a need to address whether the strength of measures and the speed of their implementation resulted in a larger and faster decrease in the R_e, and specifically whether highly restrictive lockdowns were necessary to reduce $R_e < 1$. Further, it remains unclear how the impact of interventions differed across time and geographical regions. We add to this debate by using our global R_e estimates across timescales that include the lifting of many NPIs. While we cannot determine causal relationships, we use our method to assess likely associations.

2 Results

A new pipeline to estimate the effective reproductive number of SARS-CoV-2. We have developed a pipeline to estimate the time-varying effective reproductive number of SARS-CoV-2 from observed COVID-19 incidence time series (see Materials and Methods). We build upon the existing EpiEstim method [12] to estimate $R_e(t)$ from a time series of infection incidence. To infer the infection incidence from a time series of (noisy) observations, we extended the deconvolution method by Goldstein *et al.* to deal with partially observed data and time-varying delay distributions [13, 14]. To reduce numerical artefacts resulting from the noisy nature of these observations, we smooth the data prior to deconvolution. We take into account uncertainty in the observation process using a bootstrap procedure, and in the R_e estimation using the 95% highest posterior density intervals from EpiEstim. As observed incidence data we use COVID-19 confirmed case data, hospital admissions, and deaths (with type specific delay distributions, see Materials and Methods).

Validation on simulated data. The method was validated with simulations of several epidemic scenarios (Materials and Methods section 4.5). For each scenario, we specified an R_e time-series, from which we simulated infection and observation incidence. Then, we used our method to infer the infection incidence and R_e from the observation incidence, and compared to the true underlying R_e values (Fig. 1A). The specified R_e trajectories were parametrised in a piecewise linear fashion, where we fixed the plateau values for R_e and the time-points at which the trajectory changed slope. To mimic the course of the epidemic observed in many European countries in spring and summer 2020 [28], we started with R_e values around 3, then dropped to a value below 1 (the 'initial decrease'), to subsequently rise slightly above 1 for some time (Fig. 1).

The results show that our method allows accurate monitoring of the effective reproductive number across the entire length of the time series (Fig. 1B; metrics described in Materials and Methods section 4.5). The low root mean square error (RMSE) indicates that our estimates closely track the true R_e value. In the simulated trajectories, the slope of the initial decrease can be correctly inferred, although the relative error is greater for steeper slopes (Slope error). More importantly, we correctly infer whether R_e is significantly above or below 1 for 95% of the time series (Correct $R > 1/R < 1$), and across all simulations we miss the date of the $R = 1$ crossing by at most 3 days ($R = 1$ difference; calculated between the R_e point estimate and the true R_e).

The minor misestimation of the slope is primarily due to the smoothing step included in our deconvolution algorithm. However, the inclusion of smoothing greatly improves our performance across more realistic scenarios with daily or weekly observation noise (Supplemental Fig. S1). With smoothing, the performance of our method is mostly independent of the amount of observation noise (Supplemental Fig. S2). When the mean of the delay distribution between infection and case observation is misspecified, we can still determine the shape of the R_e curve, but do misestimate when $R_e = 1$

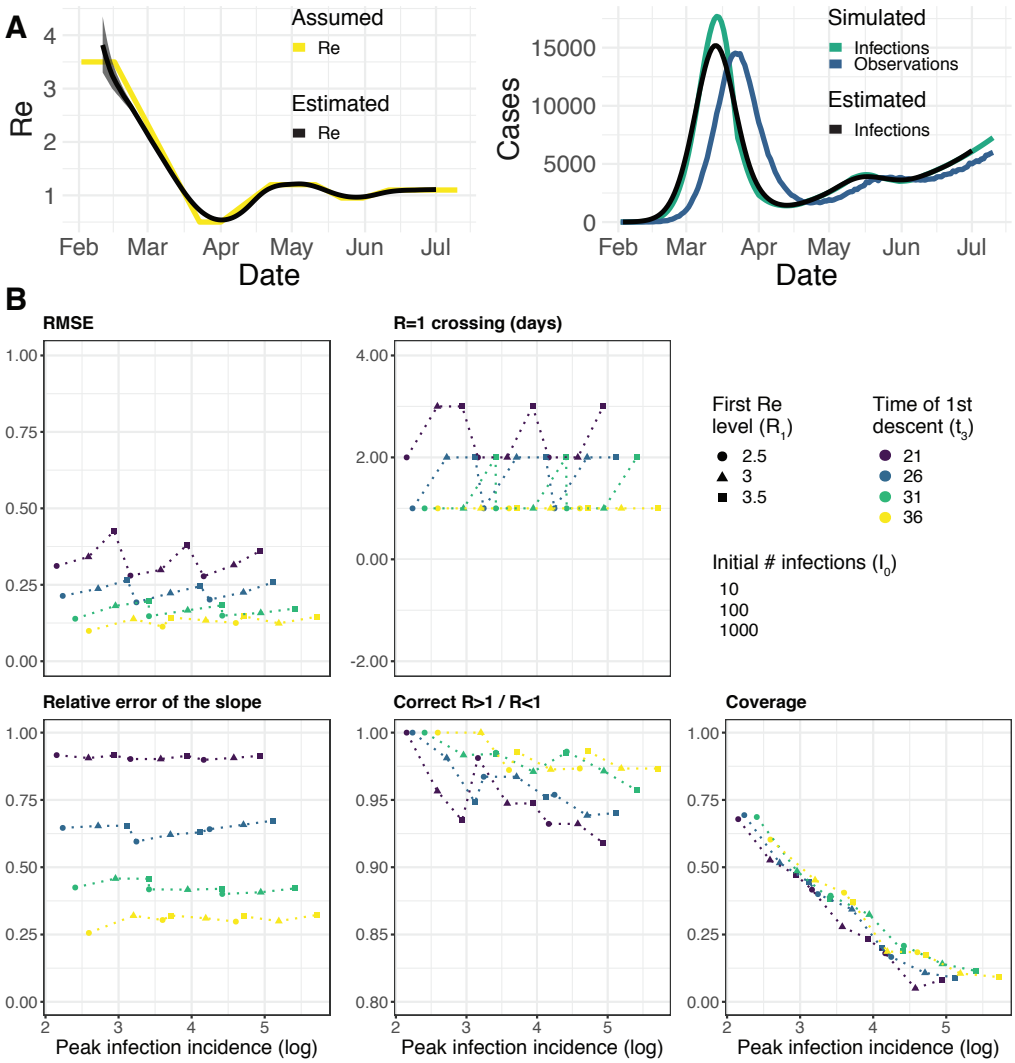


Figure 1: (A) Example simulation (coloured lines) and estimation (black lines). The left panel shows the specified R_e trajectory (yellow line) with five plateaus $R_e = (3.5, 0.5, 1.2, 0.95, 1.1)$ and the estimated R_e trajectory. The right panel shows the simulated infections (in green; initial infection incidence $I_0 = 10$), and the case observations (in blue) simulated from those infections. The infection incidence inferred from the case observations is shown in black. The trajectory shape mimics the course of the epidemic observed in many European countries, which started with an $R_e \sim 3$, then dropped to a low value below 1 (the 'initial decrease'), to subsequently oscillate around 1 for some time. **(B) Performance of our method on simulated scenarios.** From top to bottom, left to right, we show: (i) the normalised root mean squared error (RMSE), (ii) the error of the slope of the first decrease, (iii) the delay of the mean crossing $R_e = 1$, (iv) the fraction of time points for which we correctly infer that R_e is significantly above or below one, (v) the coverage of the true R_e value. These metrics are further described in Materials and Methods section 4.5. In the scenarios we varied the initial R_e level (R_1; point shape), the time of the first descent (t_3; point colour), and the number of initial infections ($I_0 \in \{10, 100, 1000\}$; not indicated separately as it determines the peak infection incidence directly). In these simulations a shorter R_e time series was used than in panel **(A)**, with three plateaus: $R_2 = 0.5, R_3 = 1.2$.

(Supplemental Fig. S3). Further model misspecifications, such as a wrong generation time interval, have been investigated by Gostic et al. [13].

The fraction of the R_e time-series where the true value of R_e falls within our estimated confidence intervals (the 'coverage'), decreases strongly for larger overall infection incidence. This indicates that the confidence intervals are too narrow for large case numbers. This result is consistent across

all scenarios, independent of the slope of the initial decrease (Fig. 1B) or the delay distribution used (data not shown). We are improving the bootstrapping method further to account for this variance.

Our method clearly outperforms the common approach of using a fixed delay to infer the infection incidence time series (Supplemental Fig. S5). Our ability to allow for empirical, time-varying distributions in the estimation also slightly improves the estimated R_e (Supplemental Figs. S4a, S4b), especially when strong time variation is present.

Monitoring R_e during the COVID-19 pandemic. We apply this R_e estimation method to COVID-19 case data from 170 countries (Fig. 2). All estimates are updated daily and made publicly available on an online dashboard `https://ibz-shiny.ethz.ch/covid-19-re/` (results available for download). For most countries, we include multiple observation sources, such as daily incidence of COVID-19 cases and deaths, and, when available, hospitalisation incidence.

The online app allows for comparison through time within a single country, between multiple observation traces, or between multiple countries. The data download further allows users to put these estimates in relation to external covariates such as mobility, weather, or behavioural data. The map view enables comparison across larger geographical areas and additionally reports the cases per 100'000 inhabitants per 14 days.

Fine-scale data allows fine-scale analysis: the example of Switzerland When detailed epidemiological data about individual cases (i.e. line lists) is available, we can increase the precision of our method by relaxing two assumptions: (i) distributions of delays between infection and observation do not change through time and (ii) outbreaks occur in a well-mixed homogeneous population at the country-level. In particular, we collaborated with the Federal Office of Public Health (FOPH) in Switzerland to further refine the monitoring of the Swiss SARS-CoV-2 epidemic.

The FOPH line list data contains information on the delays between onset of symptoms and reporting (of a positive test, hospitalisation or death) for a significant fraction of the reported cases. We estimate the time-varying empirical delay distribution from this data and use it as input to the deconvolution step, instead of estimates of these delays from the literature (for details see Materials and Methods section 4.2). The delay distribution is thus tailored to the specifics of the Swiss population and health system. Moreover, each distribution varies through time and thus reflects changes caused by e.g. improved contact tracing or overburdened health offices (see Fig S6; Supplementary Discussion). Whenever available in the FOPH line list, we use the symptom onset date of patients as the date of observation and thus only deconvolve the incubation period to obtain a time series of infection dates. The effect of these modifications is relatively minor in most parts of the estimated R_e curve (see Fig. S7), yet the difference between R_e point estimates for a particular day can be as big as 20%.

Using FOPH data on the fraction of cases infected abroad, we can correct our R_e estimate for imports to reflect only local transmission. This is especially important in phases during which the local epidemic is seeded from abroad, and local transmission occurs at a low rate relative to case importation (Fig. S8). Since we do not have data on the number of cases infected in Switzerland and then "exported" to other countries, we cannot correct for exports. Thus, the estimated R_e value corrected for imports is a lower bound for the R_e estimate which would be obtained if we could account for the location of infection of all cases detected in Switzerland or exported out of the country.

In the majority of countries the critical threshold R=1 was crossed only after the implementation of nationwide lockdowns. With our method, we can now assess the association between non-pharmaceutical interventions (NPIs) and the effective reproductive number R_e. We selected 20 European countries for which the reported data was free of major gaps or spikes (as these are

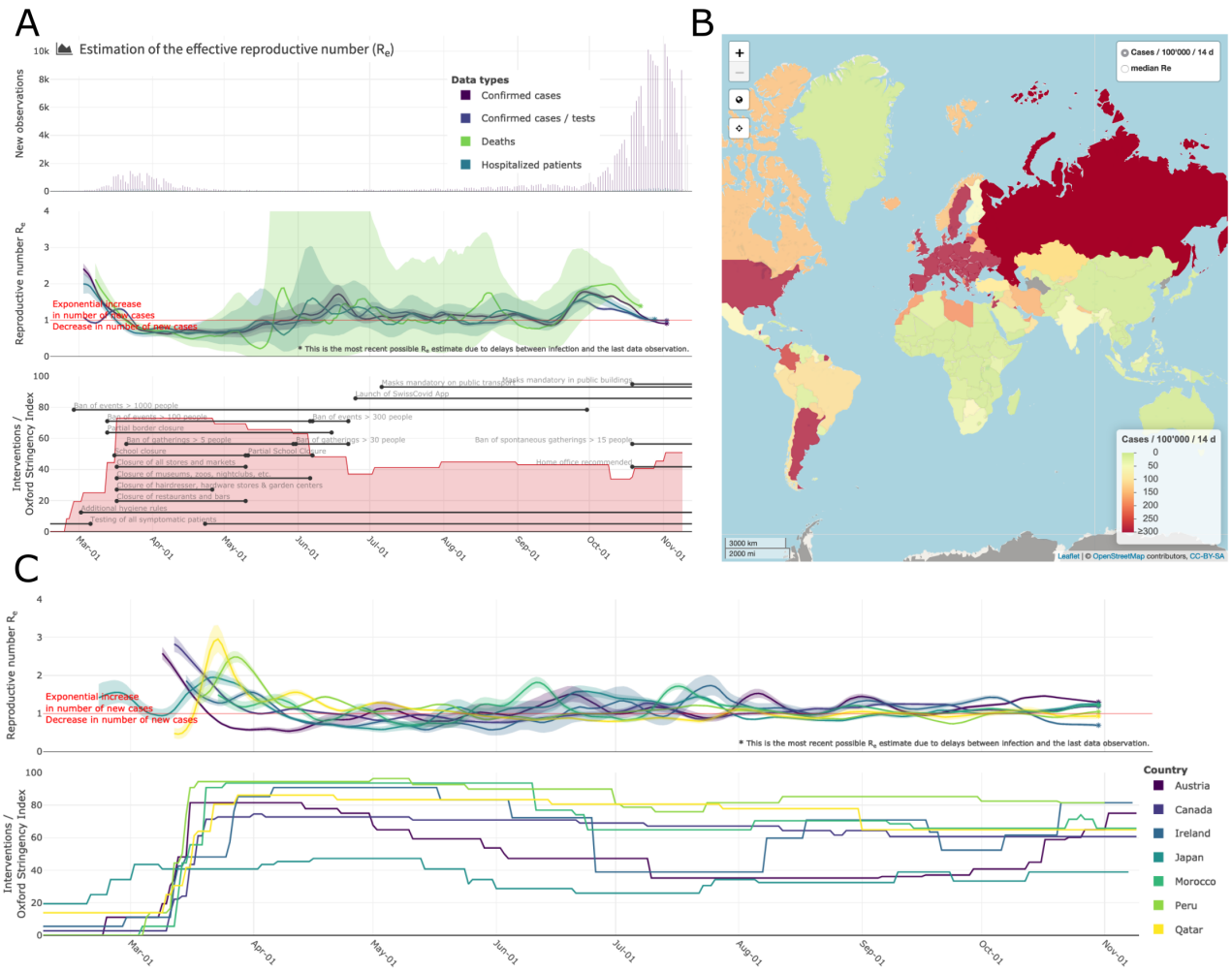


Figure 2: Example panels from the online dashboard. **(A)** Swiss cases, R_e estimates, timeline of non-pharmaceutical interventions (NPIs), and stringency index. **(B)** World map of incidence per 100'000 inhabitants over the last 14 days. **(C)** Comparison of R_e estimates across a handful of countries, with timelines of stringency indices. All panels were extracted on Nov. 13 2020. Dashboard url: `https://ibz-shiny.ethz.ch/covid-19-re`.

indicative of low-quality reporting), and for which we could estimate R_e prior to the nationwide implementation of a lockdown. The dates of interventions were extracted from news reports (sources listed in Supplementary Table S2), and 'lockdown' taken to refer to stay-at-home orders of differing intensity. Of the countries investigated, all except Sweden implemented a lockdown (19/20). We estimated that R_e was significantly above one prior to the lockdown measures in nearly all countries with a lockdown (17/19; Table 1). Only Denmark, which had a complex outbreak consisting of two initial waves, had an estimated R_e significantly below one prior to this date. We showed on simulated data that our method estimates the date of the $R_e = 1$ threshold crossing with up to 3 days delay. Accounting for these 3 days does not change our results (Table 1). The results are also remarkably consistent across the different observation types (Supplementary Table S1). However, the confidence intervals tend to be wider for the estimations based on the death incidence because the number of deaths is much smaller than the number of cases, and the relative noise in observations tends to be higher.

To consider this question for countries outside of Europe, we used the stringency index (SI) of the Blavatnik School of Government [29] to describe the public health response in different countries (Fig. 3a). This is a compound measure describing e.g. whether a state has closed borders, schools, or workplaces. For example, a country with widespread information campaigns, partially closed borders, closed schools, and a ban on public events and gatherings with more than 10 people would

Country	Lockdown	Date that $R < 1$	Time until $R < 1$
Austria	16-03	21-03 [21-03, 22-03]	5 days
Belgium	18-03	31-03 [30-03, 02-04]	13 days
Denmark	18-03	**$\leq$10-03 [$<$10-03, 11-03]**	-8 days
Finland	16-03	03-04 [01-04, 06-04]	18 days
France	17-03	27-03 [27-03, 28-03]	10 days
Germany	22-03	26-03 [26-03, 27-03]	4 days
Ireland	27-03	09-04 [08-04, 10-04]	13 days
Italy	10-03	18-03 [18-03, 19-03]	8 days
Netherlands	23-03	01-04 [26-03, 04-04]	9 days
Norway	14-03	23-03 [22-03, 24-03]	9 days
Poland	25-03	03-04 [02-04, 17-04]	9 days
Portugal	16-03	29-03 [28-03, 30-03]	13 days
Romania	16-03	06-04 [02-04, 30-04]	13 days
Russian Federation	30-03	05-05 [04-05, 05-05]	36 days
Slovenia	20-03	24-03 **[$\leq$14-03, 28-03]**	4 days
Spain	14-03	18-03 [02-02, 18-03]	4 days
Sweden		**19-04 [01-04, 21-04]**	
Switzerland	17-03	22-03 [21-03, 22-03]	5 days
Turkey	21-03	09-04 [08-04, 09-04]	19 days
United Kingdom	24-03	31-03 [30-03, 31-03]	7 days

Table 1: The date that $R_e < 1$ for the first time. Based on news reports, we report when a country implemented stay-at-home orders (a 'lockdown'). We determined when the R_e estimate and its confidence intervals first dropped below 1. Of the investigated countries that implemented a nationwide lockdown, only two (Denmark, Slovenia) had R_e estimates that included or were below one before a nationwide lockdown was implemented. 'Time until $R < 1$' indicates the number of days between the lockdown and the date that the mean $R_e < 1$.

have an SI slightly above 50. As reference date, we determined when a country first exceeded a stringency index of 50 (t_{SI50}). Then, we asked whether the estimated R_e was significantly above 1 prior to the reference date, where we again excluded countries without R_e estimates before the reference date. Out of 39 countries world-wide which fulfilled the criteria for inclusion (list in Supplement 8.3), 28/39 countries were significantly above 1 prior to t_{SI50}. As an additional analysis, for each day, we calculated the change of SI within the past 7 days. We used the day with the maximal change as the new reference date (t_{max}). This analysis yielded very similar results with 36/50 countries significantly above one before t_{max} (Supplement 8.3). Accounting for the 3 days of possible delay, 31/39 countries were still above 1 before t_{SI50} (significantly above for 25/39 countries), and 40/50 countries before t_{max} (significantly above for 33/50).

A strong government response is associated with a faster decrease in R_e. Next, we asked whether the slope of R_e on the reference date t_{SI50} (one-day change) is associated with the decisiveness of the government response, as measured by the change in the stringency index the week prior. In Europe, larger changes in the SI prior to t_{SI50} were significantly associated with stronger decreases in the estimated R_e on the reference date ($p = 0.04$, adjusted $R^2 = 0.18$, Fig. 3(b); 19 countries). The same trend was found at a global scale, though no longer significant (Fig. S9; 40 countries). However, this does not mean that the SI at lockdown is a predictor for the time until R_e is below 1 ($p = 0.9$, Fig. S10(a)). This time instead was better predicted by the estimated R_e on the day of lockdown ($p = 0.03$, adjusted $R^2 = 0.2$, Fig. S10(b)). Finally, there is no significant association between higher maximum SI during lockdown and a lower minimum R_e attained during that time ($p = 0.14$, Fig. S11).

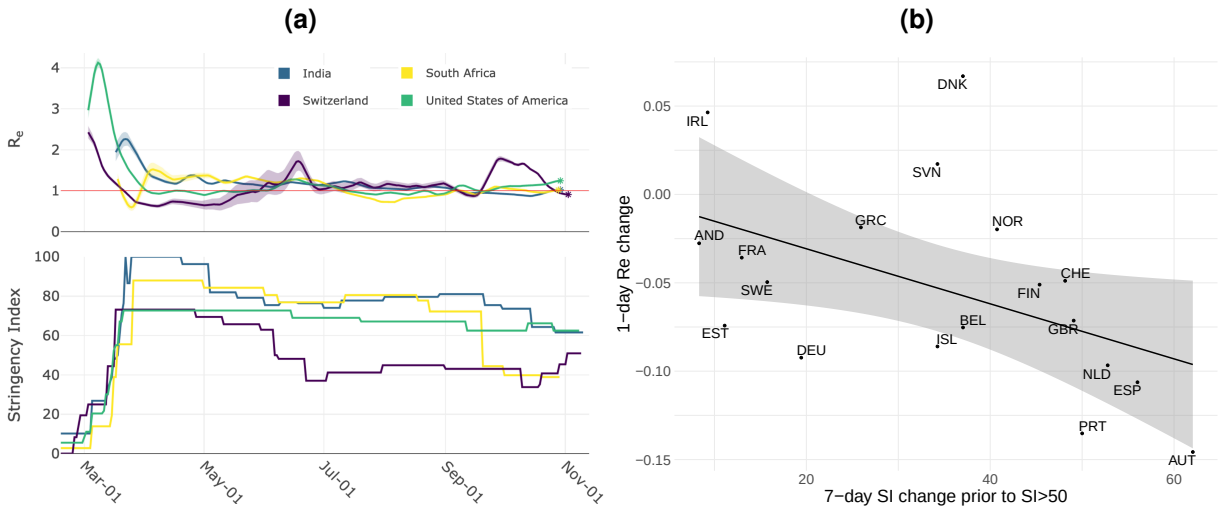


Figure 3: The association between non-pharmaceutical interventions and R_e during the first phase of the epidemic. (a) Estimated R_e (top) and government stringency index (bottom) for India, South Africa, Switzerland, and the United States of America. **(b)** Relation between the slope of the R_e estimates on t_{SI50} and the increase in the stringency index over the 7 days prior to t_{SI50} for 19 European countries. The countries are indicated by their ISO3 country code.

Insights into continent-specific responses to NPIs. To investigate the association between changes in government stringency and changes in R_e in more depth, we extended our analysis until November 10th and included both the implementation and lifting of NPIs (increases and decreases in stringency). For each week and country, we determined whether the stringency index changed, and if so, what the effect was on R_e after 7 days. If NPIs are working as expected, increases in stringency should be associated with a decrease in R_e and vice versa. We do find this for increases in stringency e.g in Europe (blue bars, Fig. 4a), but decreases in stringency have a more varied effect on R_e estimates on all continents (red bars, Fig. 4a). This may be modulated by temporal differences in when NPIs were implemented or lifted (Fig. 4b).

For North America, Asia, and Europe, an increase in stringency was significantly more associated with a subsequent reduction in the estimated effective reproductive number R_e than a reduction in stringency was (permutation test randomising SI increase/decrease associated with the estimated change in R_e, $\alpha = 0.05$, 6-way Bonferroni correction). For these continents, increases in stringency also resulted in a stronger absolute change than decreases in stringency (permutation test randomising SI increase/decrease associated with the estimated change in R_e, $\alpha = 0.05$, 6-way Bonferroni correction). This suggests that reversing non-pharmaceutical interventions had a very different effect than introducing them. For the continents Africa, South America, and Oceania no significant difference could be detected between the distributions of estimated R_e changes after implementation or lifting of NPIs.

We further repeated this analysis for 8 individual indices that make up the SI compound index separately. Continents differ substantially in which indices showed signal that an increase in stringency was significantly more associated with a subsequent reduction in R_e than a reduction in stringency. Four index-continent pairs showed a significant effect: school closing and closing public transport in Europe, cancelling public events in South America, and stay at home requirements in North America (as determined by permutation test randomising SI increase/decrease associated with the estimated change in R_e, $\alpha = 0.05$, 6*8-way Bonferroni correction; Supplementary Table S3). Similar, yet non-significant, trends were observed for some of the other index-continent pairs. We further tested whether there was a linear relationship between the change in stringency of the individual indices and the resulting estimated change in R_e (Supplementary Fig. S12, Table S4). The highest R^2 values were found for the index-continent pairs also identified in the previous analysis.

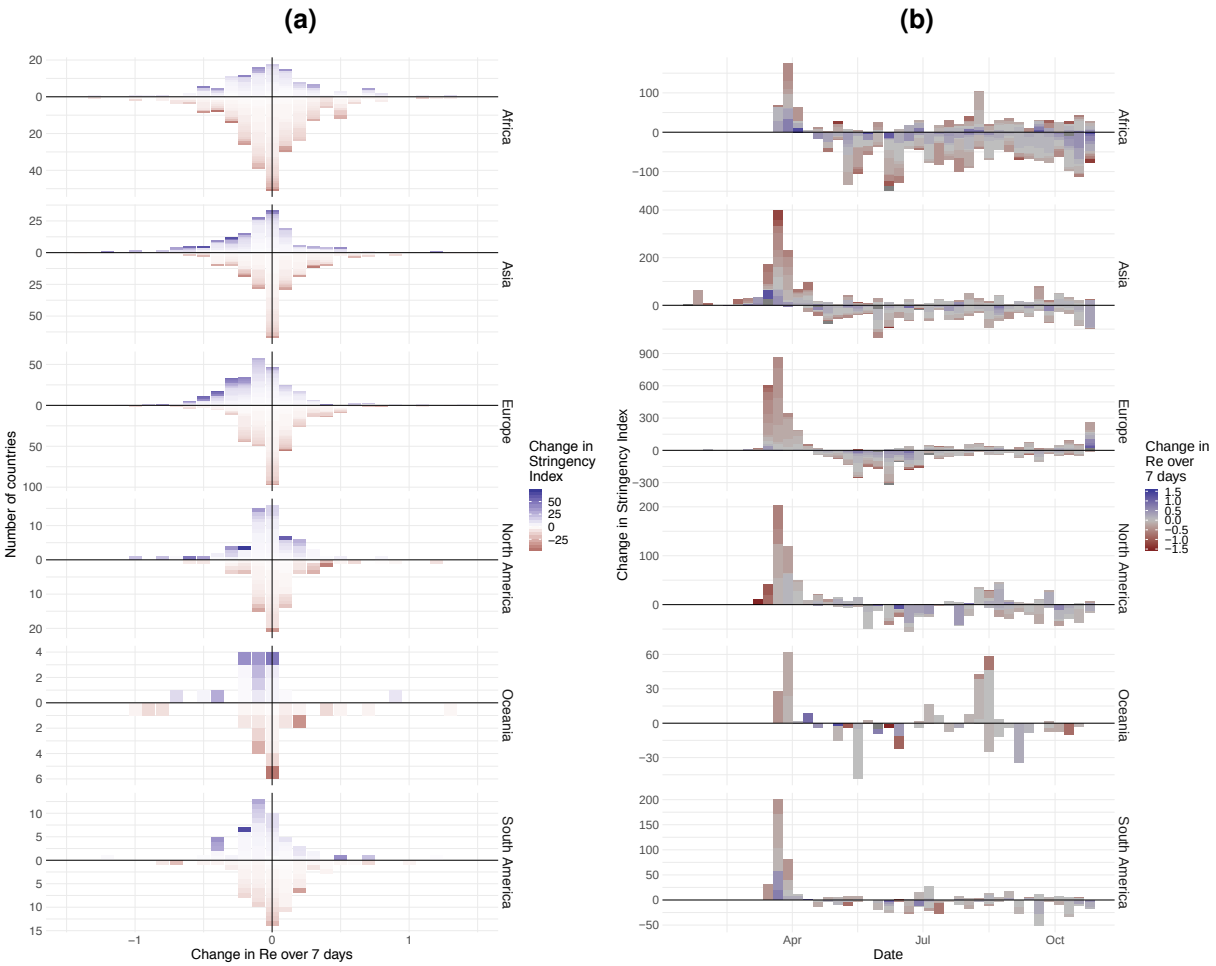


Figure 4: The association between non-pharmaceutical interventions and changes in R_e. (a) Histogram of the estimated 7-day change in R_e following the implementation (blue, above x-axis) or lifting (red, below x-axis) of NPIs in a given week. The height of the bars is given by the number of countries with this observation, coloured by the change in stringency index that week. If NPIs are working as expected, we would expect an increase in stringency (i.e. bars above the x-axis) to be more associated with a decrease in R_e (shifted to the left) and bars below the x-axis to be shifted more to the right. Note, scales differ between the continents. **(b)** Strength of NPI changes through time, coloured by their effect on R_e (increase is blue, a decrease red). Here, the height of bars is given by the summed change in the stringency index over all countries that week. If NPIs are working as expected, we would expect bars above the x-axis (i.e. an increase in stringency) to be more red (i.e. associated with a decrease in estimated R_e) and bars below the x-axis to be more blue.

3 Discussion

We have developed a pipeline to monitor the effective reproductive number R_e of SARS-CoV-2 in near real-time, and validated our estimates with simulations. We showed that the inferred R_e curve can be slightly over-smoothed on simulated data, but that this is a necessary compromise given the inherent noisiness and sometimes low quality of real data. Overall, we show that the relative error in the R_e estimates is small. In particular, we can detect when R_e crosses the critical threshold of 1, which is important to an informed public health response.

During the ongoing SARS-CoV-2 pandemic, R_e estimates are of interest to health authorities, politicians, decision makers, the media and the general public. Because of this broad interest and the critical importance of R_e estimates, it is crucial to communicate both the results as well as the associated uncertainty and caveats in an open, transparent and accessible way. This is why we display daily up-

dated results on an online dashboard, accessible at `https://ibz-shiny.ethz.ch/covid-19-re/`. The dashboard shows R_e estimates in the form of time series for each included country or region, and a global map containing the latest R_e estimates and normalised incidence. For a number of European countries, we also display a timeline of the implementation and lifting of non-pharmaceutical interventions (NPIs). For all countries, we display a timeline of the stringency index of the Blavatnik School of Government.

A unique advantage of the monitoring method we have developed is the parallel use of different types of observation data, all reflecting the same underlying infection process [6]. Wherever we have data of sufficient quality, we estimate R_e separately based on confirmed cases, hospitalisations and death reports. The advantages and disadvantages of the different observation types are discussed in the Supplementary discussion 6.1. Comparing estimates from several types of data is a powerful way to evaluate the sensitivity of the results to the type of observations they were derived from. More generally, the method would be applicable to any other type of incidence data, such as admissions to intensive care units or excess death data. The potential limitations of our R_e estimation method are discussed in detail in the Supplementary Discussion 6.2.

The decision to implement, remove or otherwise adjust measures aimed at infection control will be informed by epidemiological, social and economic factors [30]. We can aid this decision making process by investigating the association between adjustments of public-health measures and the estimated R_e.

The merits of nation-wide lockdowns have been heavily discussed, both in the scientific literature as well as in the public sphere [8, 25, 23, 24, 31]. In particular, analyses showing that R_e estimates had dropped below 1 before the strictest measures were enforced were frequently used to claim that a lockdown was not necessary [31]. We showed that this argumentation is flawed: for 17 out of 20 European countries, we found that the estimated R_e was significantly above 1 prior to the lockdown in spring.

For this first epidemic wave, we further investigated the link between the strength of NPIs implemented and the concurrent decrease of R_e. There was a trend that countries with a strong increase in stringency prior to the reference date t_{SI50} saw a faster decrease in pathogen transmission on that day. However, this analysis focused only on the estimated one-day change in R_e around the onset of the lockdown. We did not find that the strength of the government response significantly determined the time it takes to bring $R_e < 1$, nor the value of R_e during a lockdown. To investigate these points conclusively, further analyses are needed that are beyond the scope of this manuscript.

Extending our analysis to data up until Nov. 10, we find differences between continents in the effect of NPIs on R_e. This could reflect differences in the speed with which lockdowns were put into practice [26], the de-facto lockdown stringency, or socio-cultural aspects [30, 32]. It is often argued that, especially in countries with a large informal business sector, there may be a difference between the official containment measures and those adhered to or implemented de-facto [32]. However, for continents where we find no significant effect of NPI implementation, this could also be because the majority of NPIs were implemented at a time for which we could not estimate changes in R_e. Many African countries implemented early and strict government responses, often prior to the first detected cases. These are thought to have delayed the virus in establishing a foothold on the continent [32]. Since we cannot estimate R_e without cases, such early response would not be seen to reduce R_e in our analyses.

Importantly, our analysis suggests that reversing non-pharmaceutical interventions had a very different effect than introducing them. This could be because the situation is not fully reverted: due to increased public awareness, testing, contact tracing, and quarantine measures still in place. In addition, the epidemic situation - in terms of number of infected individuals - is likely different when measures are implemented or lifted.

Our analysis could be confounded by other economic, social, and psychological factors motivating the implementation or release of measures. With the current stringency measures we can not ac-

count for diversity in adherence to NPIs across geographic regions and through time. Cultural norms, defiance towards public authorities, "lockdown fatigue", and economic pressures are all among the factors that may determine whether NPIs are in fact adhered to. In addition, there is increasing evidence that weather may be a factor influencing R_e through its effect on people's behaviour as well as properties of the virus [33]. In the future, our tools to quantify R_e could be used to explore associations of these many factors with R_e, with the aim of identifying minimal sets of factors ensuring an $R_e < 1$ for particular locations.

Acknowledgements

We thank the Federal Office of Public Health Switzerland for access to their line list data, and members of the modelling group of the Swiss National Covid-19 science task force for helpful discussions. We thank Marloes Maathuis for her feedback to the paper, discussions, and follow-up work on the bootstrapping methods together with Jinzhou Li. We thank Jūlija Pečerska for her help in finding governmental datasets on COVID-19. JSH was funded by the NRP72 SNF grant (407240-167121) awarded to SB and TS. SB and TS thank ETH Zurich for funding.

3.1 Author contributions

J.S.H, J.S, S.B, T.S designed research; J.S.H, J.S, D.C.A, R.A.N contributed new reagents or analytic tools; J.S.H, J.S performed research; J.S.H, J.S analyzed data; and J.S.H, J.S wrote the paper. All authors critically reviewed and approved the paper.

4 Materials and Methods

4.1 EpiEstim

The method presented here builds upon the R_e estimation method developed by Cori *et al.* [12], implemented in the EpiEstim R package. This method estimates $R_e(t)$ from a time series of infection incidence, we summarise its details below.

Disease transmission is modelled with a Poisson process. At time t, an individual infected at time $t - s$ causes new infections at a rate $R_e(t) \cdot w_s$, where w_s is the value of the infectivity profile s days after infection. The infectivity profile sums to 1, and can be approximated by the serial interval distribution [12]. The expected infection incidence I_t at time t is thus:

$$E(I_t) = R_e(t) \sum_{s=1}^{t} I_{t-s} w_s \tag{1}$$

and the likelihood of the incidence I_t is given by:

$$P(I_t | I_0, \ldots, I_{t-1}, w_t, R_e(t)) = \frac{(R_e(t)\Lambda_t)^{I_t} \, e^{-R_e(t)\Lambda_t}}{I_t!}\,, \tag{2}$$

$$\text{where} \qquad \Lambda_t = \sum_{s=1}^{t} I_{t-s} w_s\,. \tag{3}$$

The $R_e(t)$ inference is performed in a Bayesian framework, and an analytical solution can be derived for the posterior distribution of $R_e(t)$ (see [12]; Web Appendix 1). We choose a gamma distributed prior on $R_e(t)$ with mean 1, and standard deviation 5.

4.2 Deconvolution

To recover the non-observed time series of infection incidence, we deconvolve the observed time series of COVID-19 case incidence with a delay distribution specific to the type of case detection (case confirmation, hospital admission, death).

We extended the deconvolution method of Goldstein *et al.* [14], which is itself an adaptation of the Richardson-Lucy algorithm. Formally, the deconvolution infers an infection incidence time series $(\lambda_1, \ldots, \lambda_N)$ from a time series of observed cases $(D_K, \ldots, D_N)$, with $K \geq 1$. D_i indicates the number of observed cases on day i. Let m_l^j be the probability that an infection on day j takes l days to be detected. For any $k < 0$ and any l, $m_l^k = 0$. If no line list data is available $m_l^j = m_l$, and no time-variation of the delay distribution is assumed. Let q_j be the probability that an infection that occurred on day j is observed during the time-window of observations, between days K and N. Then:

$$q_j = \sum_{l=K-j+1}^{N-j} m_l^j \,. \tag{4}$$

Let E_i be the expected number of observed cases on day i, for a given infection incidence (λ_k):

$$E_i = \begin{cases} \sum_{j=1}^{i} \lambda_j \, m_{i-j}^j & \text{for} \quad i \geq K \\ 0 & \text{for} \quad 0 < i < K \,. \end{cases} \tag{5}$$

The Richardson-Lucy algorithm uses expectation maximisation to find a final infection incidence estimate, which has the highest likelihood of explaining the observed case time series. To do so, it starts from an initial guess of the infection incidence time series $\Lambda^0 = (\lambda_1^0, \ldots, \lambda_N^0)$, and updates the estimate in each iteration n according to the following formula:

$$\lambda_j^{n+1} = \frac{\lambda_j^n}{q_j} \cdot \sum_{i=K}^{N} \frac{m_{i-j}^j D_i}{E_i^n} \,. \tag{6}$$

The iteration proceeds until a termination criterion is reached. Here, we follow Goldstein *et al.* and iterate until the χ^2 statistic drops below 1 [14], or 100 iterations have been reached:

$$\chi^2 = \frac{1}{N-K+1} \sum_{i=K}^{N} \frac{(E_i^n - D_i)^2}{E_i^n} \,. \tag{7}$$

Convergence is typically fast and the stopping criterion based on the χ^2 statistic is reached in a few iterations. Due to the smoothing prior to deconvolution, this was the case for nearly all empirical data we analyzed. In some cases, e.g. when the observed incidence is very noisy, convergence can be slower and the threshold of 100 iterations can be reached. For 4/170 countries, convergence was not reached in 100 iterations: China, Ecuador, Equatorial Guinea and Peru all showed strong spikes in reporting which obstructed the deconvolution.

For the initial estimate of the incidence time series Λ^0, we shift the observation time series backwards in time by the mode of the delay distribution μ [14]. However, this leaves a gap of unspecified values at the start and end of the time series Λ_0. Contrary to Goldstein et al., we augment the shifted time series with the first observed value (D_K) on the left, and with the last observed value (D_N) on the right, to avoid initialising with a zero-value anywhere. If a day is initialised with zero incidence, it will also have zero incidence in the final estimate (compare equation 6), which would be a potential source of bias. We have compared several ways to pad the shifted observed time series for the initialisation step, and determined that augmentation with non-zero integers equal to the edge values is enough for the deconvolution to converge to the true distribution (see Supplementary Fig. S13).

We note that this Richardson-Lucy deconvolution algorithm accounts for 'right truncation', i.e. not all infections that have occurred are observed within the given observation time window (due to delay until symptoms/reporting), through the q_j indices.

Use of line list data When a line list is available, the time variation of delays between symptom onset and observation can be taken into account directly during the deconvolution step. This leads us to perform the deconvolution in two separate steps: first with the time-varying empirical onset-to-observation distributions, and then with the constant-through-time incubation period distribution. For those cases where symptom onset data is available, we only deconvolve with the incubation period distribution.

The $(m_0^j, \ldots, m_K^j)$ time-varying delay distributions from onset of symptoms to observation are determined as follows: for each date j, at least 300 of the most recent recorded delays between symptom onset and observation, with onset date before j, are taken into account. To avoid biases caused by the intensity of testing and reporting varying throughout the week, recorded delays are included in full weeks going in the past, until at least 300 delays are included.

As the incidence data is right-truncated, we have to fix the reporting delay distributions m_i^j at some point, or they would be downward biased for infection dates close to the present. Let $(m_0^0, \ldots, m_K^0)$ be the overall empirical delay distribution (aggregated over the entire window of observations) and n the 99^{th} percentile of this distribution (n is the smallest integer for which $\sum_{i=1}^{n} m_i^0 \geq 0.99$). For all infection dates z such that $M - z < n$, M being the index of the last available data point, we fix $(m_0^z, \ldots, m_K^z)$ to be equal to $(m_0^{M-n}, \ldots, m_K^{M-n})$.

4.3 Noise and uncertainty in the case observations

To reduce the influence of weekly patterns in case reporting data, as well as reporting irregularities, we smooth the observed incidence data prior to deconvolution. To smooth the incidence data, we use local polynomial regression fitting (LOESS) with 1st order polynomials and tricubic weights. The smoothing parameter alpha is set such that we always include 21 days of data in the local neighbourhood of each point. At the edges, the weights drop to 0 and less points are taken into account in total [34]. After smoothing, we normalise to the original total number of cases.

To reflect uncertainty in the observation process, we bootstrap the observed incidence data 50 times, prior to smoothing, deconvolution, and R_e estimation. The bootstrapping process here consists of sampling reported cases with replacement from the original case data (with equal probability per case), until a resampled time series with the same number of cases as the original one is obtained. This likely underestimates the total uncertainty because it does not take into account that some cases may be correlated, e.g. by belonging to the same transmission chain. A common way to circumvent this problem would be to use a (moving) block bootstrap method. We are working to implement this into our pipeline.

The R_e estimate reported for day T summarises the average estimated R_e over a 3-day period ending on day T. We report the median of the 50 R_e posterior distributions obtained, as well as the median of the 95% uncertainty interval boundaries. In addition, we provide step-wise estimates of R_e. In this step-wise analysis, R_e is assumed to be constant on a number of intervals spanning the entire epidemic time window. These intervals are bounded by dates at which public health interventions were implemented, altered, or lifted.

4.4 Data

We gathered case incidence data directly from public health authorities. Whenever accessible, we rely on data from local authorities. Otherwise, we default to the data of the European Centre for Disease Control (ECDC) [35]. A table summarising the incidence data sources is available in Supplementary File S1. Information on the start and end of interventions, or major changes in testing

policy, were obtained from media reports and the websites of public health authorities. The stringency index of the Blavatnik School of Government was accessed from their publicly available github repository [29].

We parametrised the discretised infectivity profile w_s using COVID-19 serial interval estimates from the literature [36], i.e. a gamma distribution with a mean of 4.8 days, and a standard deviation (SD) of 2.3 days. For a review of serial interval estimates published in the literature, see Griffin et al. [37]. The incubation period is parametrised by a gamma distribution with mean 5.3 days and SD 3.2 days-[38]. For countries for which we do not have access to line list data, i.e. all except Switzerland, Germany and Hong Kong at the time of writing, we assume delays from symptom onset to observation to be gamma-distributed, see table 2 for parameters.

Delay	Mean (days)	SD (days)	Reference
Onset of symptoms to case confirmation	5.5	3.8	[39]
Onset of symptoms to hospital admission	5.1	4.2	[40]
Onset of symptoms to death	15.0	6.9	[38]

Table 2: Delay distribution assumed for each observation type.

For Switzerland, Germany and Hong Kong, we use line lists to build time-varying empirical distributions on delays between symptom onset and case confirmation, hospitalisation or death. During the deconvolution step we use the empirical delay distribution of the last 300 recorded cases prior to the infection date. Moreover, for the fraction of cases for which the date of onset of symptoms is known, we use the onset date directly instead of deconvolving a delay from onset to reporting, allowing for more precise estimation of the infection date. For Switzerland, line lists contain information on which cases were infected abroad. By considering imported cases and locally-transmitted cases separately in the deconvolution step, we obtain two separate time series, one for local infections and one for imported infections.

4.5 Simulations

In the simulations, we first simulate a time series of infections and corresponding case observations from a specified piecewise linear R_e trajectory. Then, we estimate R_e from these observations using our method: deconvolution to infer the infection time series, followed by EpiEstim to estimate R_e.

To assess a range of scenarios, we parametrise R_e as a piecewise linear trajectory, where we fix the plateau values for R_e and the time-points at which the trajectory changes slope. Assuming I_0 infected individuals on the first day, the infection incidence is simulated forward in time, using the R_e time series and the discretised serial interval for SARS-CoV-2 [36] (see [12]; Web Appendix 11). These simulated infections are convolved with the observation type-specific delay distribution [38] to obtain the raw observation time series. In the case of time-varying delay distributions, we assume the mean of the delay distribution decreases by a fixed amount (1/20) each day, to a minimum of 2 days (e.g. for the confirmed cases this results in a range from 4.5 to 2). When estimating with a time-varying delay distribution, we draw observations from the true distributions, similar to line list information recorded by public health authorities. To assess the added value of the deconvolution method, we compare against a method where we estimate the infection time series by shifting the observations back by the mean of the delay distribution (termed 'fixed shift method').

To obtain the final observation time series we can add either weekly or daily sources of noise. In the case of weekly noise, we reduce the number of cases on the weekend to a fraction f of the simulated number, and add the subtracted cases to the following Monday and Tuesday instead. In the case of daily noise, we add multiplicative Gaussian noise with mean 1 and a set standard deviation on every day of the time series. If both sources of noise are chosen, the weekly noise is applied first.

Performance metrics To quantify the performance of our method on the simulated scenarios, we employ 5 metrics. The normalised root mean square error (RMSE) is given by:

$$1/\bar{R}\sqrt{\frac{1}{N}\sum_{j=1}^{N}(\hat{R}_j - R_j)^2}\,, \tag{8}$$

where $\bar{R}$ indicates the mean true R_e, N the length of the time series, $\hat{R}_j$ the estimated R_e and R_j the true R_e at time j. The relative error of the slope of the initial decrease (Slope error) is determined by comparing the slope of the true and estimated R_e between the time of the end of the first R_e plateau (t_e^1) and the start of the second plateau (t_s^2). The date of the $R = 1$ crossing ($R = 1$ difference) is determined by the number of days difference between when the true R_e crosses the threshold of $R = 1$ and our R_e point estimate does so. The fraction of correct above or below 1 estimation (Correct $R > 1/R < 1$) is determined as the fraction of the time series where we correctly infer that R_e is significantly above or below 1. Time points where the confidence interval includes values both above and below 1 are excluded from the calculation. The empirical coverage (Coverage) indicates the fraction of the time series for which our confidence interval includes the true R_e value.

4.6 Implementation and method availability

Daily updated results of our method on global COVID-19 data are available online on `https://ibz-shiny.ethz.ch/covid-19-re/`. The source code of this pipeline is openly accessible (see `https://github.com/covid-19-Re/dailyR`). We are also continuously updating our data sources, and welcome anyone who wishes to share quality data for a particular region or country (please contact the authors, or raise an issue on the Github repository of this project at `https://github.com/covid-19-Re/shiny-dailyRe`).

5 References

[1] Roy Malcolm Anderson and Robert M May. *Infectious diseases of humans : dynamics and control*. Oxford science publications. Oxford University Press, Oxford, 1991.

[2] Simon Cauchemez, Pierre-Yves Boëlle, Guy Thomas, and Alain-Jacques Valleron. Estimating in real time the efficacy of measures to control emerging communicable diseases. *American journal of epidemiology*, 164(6):591–597, 2006.

[3] Jacco Wallinga and Marc Lipsitch. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B: Biological Sciences*, 274(1609):599–604, 2007.

[4] Hiroshi Nishiura, Gerardo Chowell, James M. Hyman, Luís M. A. Bettencourt, and Carlos Castillo-Chavez. *The Effective Reproduction Number as a Prelude to Statistical Estimation of Time-Dependent Epidemic Trends*, pages 103–121. Springer Netherlands, Dordrecht, 2009.

[5] Paul L Delamater, Erica J Street, Timothy F Leslie, Y Tony Yang, and Kathryn H Jacobsen. Complexity of the basic reproduction number (r0). *Emerging infectious diseases*, 25(1):1, 2019.

[6] Jérémie Scire, Sarah Nadeau, Timothy Vaughan, Gavin Brupbacher, Simon Fuchs, Jürg Sommer, Katrin N. Koch, Reto Misteli, Lukas Mundorff, Thomas Götz, Tobias Eichenberger, Carlos Quinto, Miodrag Savic, Andrea Meienberg, Thilo Burkard, Michael Mayr, Christoph A. Meier, Andreas Widmer, Richard Kuehl, Adrian Egli, Hans H. Hirsch, Stefano Bassetti, Christian H. Nickel, Katharina S. Rentsch, Werner Kübler, Roland Bingisser, Manuel Battegay, Sarah Tschudin-Sutter, and Tanja Stadler. Reproductive number of the COVID-19 epidemic in Switzerland with a focus on the Cantons of Basel-Stadt and Basel-Landschaft. *Swiss medical weekly*, 150(February):w20271, 2020.

[7] Sheikh Taslim Ali, Lin Wang, Eric HY Lau, Xiao-Ke Xu, Zhanwei Du, Ye Wu, Gabriel M Leung, and Benjamin J Cowling. Serial interval of sars-cov-2 was shortened over time by non-pharmaceutical interventions. *Science*, 369(6507):1106–1109, 2020.

[8] Seth Flaxman, Swapnil Mishra, Axel Gandy, H. Juliette T. Unwin, Thomas A. Mellan, Helen Coupland, Charles Whittaker, Harrison Zhu, Tresnia Berah, Jeffrey W. Eaton, Mélodie Monod, Pablo N. Perez-Guzman, Nora Schmit, Lucia Cilloni, Kylie E. C. Ainslie, Marc Baguelin, Adhiratha Boonyasiri, Olivia Boyd, Lorenzo Cattarino, Laura V. Cooper, Zulma Cucunubá, Gina Cuomo-Dannenburg, Amy Dighe, Bimandra Djaafara, Ilaria Dorigatti, Sabine L. van Elsland, Richard G. FitzJohn, Katy A. M. Gaythorpe, Lily Geidelberg, Nicholas C. Grassly, William D. Green, Timothy Hallett, Arran Hamlet, Wes Hinsley, Ben Jeffrey, Edward Knock, Daniel J. Laydon, Gemma Nedjati-Gilani, Pierre Nouvellet, Kris V. Parag, Igor Siveroni, Hayley A. Thompson, Robert Verity, Erik Volz, Caroline E. Walters, Haowei Wang, Yuanrong Wang, Oliver J. Watson, Peter Winskill, Xiaoyue Xi, Patrick G. T. Walker, Azra C. Ghani, Christl A. Donnelly, Steven Riley, Michaela A. C. Vollmer, Neil M. Ferguson, Lucy C. Okell, Samir Bhatt, and Imperial College COVID-19 Response Team. Estimating the effects of non-pharmaceutical interventions on covid-19 in europe. *Nature*, 584(7820):257–261, 2020.

[9] Adam J Kucharski, Timothy W Russell, Charlie Diamond, Yang Liu, John Edmunds, Sebastian Funk, Rosalind M Eggo, Fiona Sun, Mark Jit, James D Munday, et al. Early dynamics of transmission and control of covid-19: a mathematical modelling study. *The lancet infectious diseases*, 2020.

[10] Tao Zhou, Quanhui Liu, Zimo Yang, Jingyi Liao, Kexin Yang, Wei Bai, Xin Lu, and Wei Zhang. Preliminary prediction of the basic reproduction number of the wuhan novel coronavirus 2019-ncov. *Journal of Evidence-Based Medicine*, 13(1):3–7, 2020.

[11] Jacco Wallinga and Peter Teunis. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of epidemiology*, 160(6):509–516, 2004.

[12] 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, 2013.

[13] Katelyn M Gostic, Lauren McGough, Edward Baskerville, Sam Abbott, Keya Joshi, Christine Tedijanto, Rebecca Kahn, Rene Niehus, James A Hay, Pablo M. De Salazar, Joel Hellewell, Sophie Meakin, James Munday, Nikos Bosse, Katharine Sherratt, Robin M Thompson, Laura F White, Jana Huisman, Jérémie Scire, Sebastian Bonhoeffer, Tanja Stadler, Jacco Wallinga, Sebastian Funk, Marc Lipsitch, and Sarah Cobey. Practical considerations for measuring the effective reproductive number, Rt. *medRxiv*, page 2020.06.18.20134858, 2020.

[14] Edward Goldstein, Jonathan Dushoff, Ma Junling, Joshua B. Plotkin, David J.D. Earn, and Marc Lipsitch. Reconstructing influenza incidence by deconvolution of daily mortality time series. *Proceedings of the National Academy of Sciences of the United States of America*, 106(51):21825–21829, 2009.

[15] Daniel Wyler and Markus Petermann. A pitfall in estimating the effective reproductive number Rt for COVID-19. *Swiss Medical Weekly*, 150(2930), 2020.

[16] S Abbott, J Hellewell, RN Thompson, K Sherratt, HP Gibbs, NI Bosse, JD Munday, S Meakin, EL Doughty, JY Chun, YWD Chan, F Finger, P Campbell, A Endo, CAB Pearson, A Gimma, T Russell, null null, S Flasche, AJ Kucharski, RM Eggo, and S Funk. Estimating the time-varying reproduction number of sars-cov-2 using national and subnational case counts [version 1; peer review: awaiting peer review]. *Wellcome Open Research*, 5(112), 2020.

[17] Kevin Systrom, Thomas Vladek, and Mike Krieger. Project title. `https://github.com/rtcovidlive/covid-model`, 2020.

[18] Cristian Tebé, Joan Valls, Pau Satorra, and Aurelio Tobías. Covid19-world: a shiny application to perform comprehensive country-specific data visualization for sars-cov-2 epidemic. *BMC Medical Research Methodology*, 20(1):1–7, 2020.

[19] An Pan, Li Liu, Chaolong Wang, Huan Guo, Xingjie Hao, Qi Wang, Jiao Huang, Na He, Hongjie Yu, Xihong Lin, et al. Association of public health interventions with the epidemiology of the covid-19 outbreak in wuhan, china. *Jama*, 323(19):1915–1923, 2020.

[20] Robert Koch Institut. *Täglicher Lagebericht des RKI zur Coronavirus-Krankheit-2019 (COVID-19)*, 2020 (accessed November 16, 2020).

[21] Thomas Obadia, Romana Haneef, and Pierre-Yves Boëlle. The r0 package: a toolbox to estimate reproduction numbers for epidemic outbreaks. *BMC medical informatics and decision making*, 12(1):1–9, 2012.

[22] Rachel T Esra, Lise Jamesion, Matthew P Fox, Daniel Letswalo, Nkosinathi Ngcobo, Sithabile Mngadi, Janne Global Estill, Gesine Meyer-Rath, and Olivia Keiser. Evaluating the impact of non-pharmaceutical interventions for sars-cov-2 on a global scale. *medRxiv*, 2020.

[23] Nicolas Banholzer, Eva van Weenen, Bernhard Kratzwald, Arne Seeliger, Daniel Tschernutter, Pierluigi Bottrighi, Alberto Cenedese, Joan Puig Salles, Stefan Feuerriegel, and Werner Vach. Estimating the impact of non-pharmaceutical interventions on documented infections with covid-19: A cross-country analysis. *medRxiv*, 2020.

[24] Kristian Soltesz, Fredrik Gustafsson, Toomas Timpka, Joakim Jaldén, Carl Jidling, Albin Heimerson, Thomas Schön, Armin Spreco, Joakim Ekberg, Örjan Dahlström, et al. Sensitivity analysis of the effects of non-pharmaceutical interventions on covid-19 in europe. *medRxiv*, 2020.

[25] Nils Haug, Lukas Geyrhofer, Alessandro Londei, Elma Dervic, Amelie Desvars-Larrive, Vittorio Loreto, Beate Pinior, Stefan Thurner, and Peter Klimek. Ranking the effectiveness of worldwide covid-19 government interventions. *MedRxiv*, 2020.

[26] Ilia Kohanovski, Uri Obolski, and Yoav Ram. Inferring the effective start dates of non-pharmaceutical interventions during covid-19 outbreaks. *medRxiv*, 2020.

[27] Mrinank Sharma, Sören Mindermann, Jan Markus Brauner, Gavin Leech, Anna B Stephenson, Tomáš Gavenčiak, Jan Kulveit, Yee Whye Teh, Leonid Chindelevitch, and Yarin Gal. On the robustness of effectiveness estimation of nonpharmaceutical interventions against covid-19 transmission. *arXiv preprint arXiv:2007.13454*, 2020.

[28] JC Lemaitre, J Perez-Saez, AS Azman, A Rinaldo, and J Fellay. Assessing the impact of non-pharmaceutical interventions on sars-cov-2 transmission in switzerland. *Swiss Medical Weekly*, 150:w20295–w20295, 2020.

[29] Thomas Hale, Sam Webster, Anna Petherick, Toby Phillips, and Beatriz Kira. *Oxford COVID-19 Government Response Tracker*. Blavatnik School of Government, 2020.

[30] Abiel Sebhatu, Karl Wennberg, Stefan Arora-Jonsson, and Staffan I Lindberg. Explaining the homogeneous diffusion of covid-19 nonpharmaceutical interventions across heterogeneous countries. *Proceedings of the National Academy of Sciences*, 117(35):21201–21208, 2020.

[31] Sascha Karberg. *Der "überflüssige" Lockdown?* Tagesspiegel, 2020 (accessed October 22, 2020).

[32] Moustapha Mbow, Bertrand Lell, Simon P Jochems, Badara Cisse, Souleymane Mboup, Benjamin G Dewals, Assan Jaye, Alioune Dieye, and Maria Yazdanbakhsh. Covid-19 in africa: Dampening the storm? *Science*, 369(6504):624–626, 2020.

[33] Dylan H Morris, Kwe Claude H Yinda, Amandine Gamble, Fernando W Rossine, Qishen Huang, Trenton Bushmaker, Robert J Fischer, M Jeremiah Matson, Neeltje van Doremalen, Peter J Vikesland, et al. The effect of temperature and humidity on the stability of sars-cov-2 and other enveloped viruses. *bioRxiv*, 2020.

[34] John M Chambers. *Statistical models in S*. Wadsworth & Brooks/Cole computer science series. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, California, 1992.

[35] European Centre for Disease Prevention and Control (ECDC). Daily number of new reported cases of covid-19 by country worldwide. `https://opendata.ecdc.europa.eu/covid19/casedistribution/csv`. Accessed: 2020-10-02.

[36] Hiroshi Nishiura, Natalie M. Linton, and Andrei R. Akhmetzhanov. Serial interval of novel coronavirus (COVID-19) infections. *International Journal of Infectious Diseases*, 93:284–286, apr 2020.

[37] John M Griffin, Aine B Collins, Kevin Hunt, David McEvoy, Miriam Casey, Andrew W Byrne, Conor G McAloon, Ann Barber, Elizabeth Ann Lane, and Simon J More. A rapid review of available evidence on the serial interval and generation time of covid-19. *medRxiv*, 2020.

[38] Natalie M. Linton, Tetsuro Kobayashi, Yichi Yang, Katsuma Hayashi, Andrei R. Akhmetzhanov, Sung-mok Jung, Baoyin Yuan, Ryo Kinoshita, and Hiroshi Nishiura. Incubation Period and Other Epidemiological Characteristics of 2019 Novel Coronavirus Infections with Right Truncation: A Statistical Analysis of Publicly Available Case Data. *Journal of Clinical Medicine*, 9(2):538, feb 2020.

[39] Qifang Bi, Yongsheng Wu, Shujiang Mei, Chenfei Ye, Xuan Zou, Zhen Zhang, Xiaojian Liu, Lan Wei, Shaun A Truelove, Tong Zhang, et al. Epidemiology and transmission of covid-19 in 391 cases and 1286 of their close contacts in shenzhen, china: a retrospective cohort study. *The Lancet Infectious Diseases*, 2020.

[40] Lorenzo Pellis, Francesca Scarabel, Helena B Stage, Christopher E Overton, Lauren HK Chappell, Katrina A Lythgoe, Elizabeth Fearon, Emma Bennett, Jacob Curran-Sebastian, Rajenki

Das, et al. Challenges in control of covid-19: short doubling time and long delay to effect of interventions. *arXiv preprint arXiv:2004.00117*, 2020.

[41] Esteban Ortiz-Ospina Max Roser, Hannah Ritchie and Joe Hasell. Coronavirus pandemic (covid-19). *Our World in Data*, 2020. https://ourworldindata.org/coronavirus.

[42] Gideon Meyerowitz-Katz and Lea Merone. A systematic review and meta-analysis of published research data on COVID-19 infection-fatality rates. *International Journal of Infectious Diseases*, Oct 2020.

[43] Anthony Hauser, Michel J. Counotte, Charles C. Margossian, Garyfallos Konstantinoudis, Nicola Low, Christian L. Althaus, and Julien Riou. Estimation of sars-cov-2 mortality during the early stages of an epidemic: A modeling study in hubei, china, and six regions in europe. *PLOS Medicine*, 17(7):1–17, 07 2020.

[44] Albert Esteve, Iñaki Permanyer, Diederik Boertien, and James W. Vaupel. National age and coresidence patterns shape covid-19 vulnerability. *Proceedings of the National Academy of Sciences*, 117(28):16118–16120, 2020.

[45] Wan Yang, Sasikiran Kandula, Mary Huynh, Sharon K Greene, Gretchen Van Wye, Wenhui Li, Hiu Tai Chan, Emily McGibbon, Alice Yeung, Don Olson, Anne Fine, and Jeffrey Shaman. Estimating the infection-fatality risk of SARS-CoV-2 in New York City during the spring 2020 pandemic wave: a model-based analysis. *The Lancet Infectious Diseases*, Oct 2020.

[46] Raphael Minder. *Counting Bodies and Pointing Fingers as Spain Tallies Coronavirus Dead*. New York Times, 2020 (accessed October 22, 2020).

[47] James O Lloyd-Smith, Sebastian J Schreiber, P Ekkehard Kopp, and Wayne M Getz. Superspreading and the effect of individual variation on disease emergence. *Nature*, 438(7066):355–359, 2005.

6 Supplementary Discussion

6.1 Observation types and the influence of testing

Here, we briefly discuss the benefits and potential biases of the three types of observations we used. The most commonly used proxy for infection incidence is the incidence of confirmed cases. It is the least indirect way of observing infection events. However, it generally assumes that (i) the proportion of infected individuals that is tested, and (ii) the distribution of the delay between infection and testing are constant through time. Unfortunately, these assumptions do not generally hold.

As long as the sampling proportion is constant throughout the considered time period, the R_e estimates of EpiEstim are not affected by under-sampling [12]. During the SARS-CoV-2 epidemic, many countries initially restricted testing to only severe cases, before switching to a more extensive testing effort after curbing the first epidemic wave and ramping up testing capacity [41]. Changes in testing strategy as well as bottlenecks in testing capacity result in a varying fraction of infected individuals that are confirmed positive, but also in the delay between infection and test confirmation. This can bias the R_e estimate, as it will attribute an increase/decrease in case numbers between consecutive time points to a change in infection incidence, rather than a change in testing.

However, it is important to note that the 'memory' inherent in the R_e estimate is dictated by the serial interval w_s. An event at time t which changes the proportion of true infection incidence observed per day, e.g. a change in testing policy, will bias the R_e estimate for the number of days needed to reach the 95% of w_s (compare Materials and Methods, equation 1). We do not observe the infection incidence directly, but if the deconvolution is assumed to be perfect, the intuition for the number of days of biased R_e estimates still holds.

It is further possible to investigate the influence of testing intensity, by applying the R_e estimation method separately to a case incidence time series which is adjusted for the intensity of the testing effort. We have added this analysis to our online dashboard (where we show the number of confirmed cases / number of tests, normalised by the mean number of tests). However, one should note that such a normalisation does not take into account that the probability of test positivity might also change with the number of tests (e.g. by prioritising likely cases at low numbers of tests).

In contrast, the incidence of hospital admittance and deaths are likely based primarily on the severity of the symptoms, and mostly unaffected by changes in testing strategies, or the magnitude of the epidemic. This makes them valuable complementary observations of infection events [14]. However, also here biases can occur. First, only a small fraction of all infections results in hospitalisation or death (a recent meta-analysis found an average infection fatality ratio for SARS-CoV-2 of 0.68% [42]). This fraction varies with the risk group of the infected population [42, 43, 44, 45], introducing potential biases in R_e estimations when outbreaks occur in particularly age-stratified settings. Second, if a country's health infrastructure becomes overburdened and hospitals are forced to triage or delay admission, we expect the fraction of hospital admissions to decrease, and deaths to increase. Third, the likelihood to die from an infection may change through time as new treatment strategies are developed. Additionally, guidelines used to record COVID-19 as the cause of death have changed through time for some countries [46]. Lastly, the delay between infection and hospitalisation or death is expected to be longer than the delay until case confirmation, with the result that these R_e estimates are less timely. One should note that these observation type specific biases could also be seen as a source of information. The types simply describe a different epidemic if very structured population with highly different mortality rates are captured (e.g. elderly homes).

6.2 Method Limitations

The R_e estimation method we present in the main text relies on several assumptions. Here we highlight the limitations that occur when these assumptions are violated.

First, the geographical scale of the R_e estimates is determined by the incidence data itself. The R_e calculated for a country represents an average, summarised across multiple local epidemics unfolding in different regions. R_e values need not be identical in different local epidemics across a country or administrative region. In particular, in times of very low pathogen transmission, single super-spreading events can significantly increase the estimated R_e of the entire country [47].

Second, in our deconvolution step we account for an incubation period and a delay from symptom onset to case observation. Implicitly, we thus assume that all reported cases come from symptomatic individuals. This is certainly true for hospitalised and deceased patients, but does not have to hold for all confirmed cases. Similar to the testing intensity (discussed in Section 6.1), this would not bias our estimates as long as the fraction of asymptomatic or presymptomatic individuals is constant through time. However, the fraction of asymptomatic individuals could vary with the population structure and age-stratification. The fraction of tested presymptomatic individuals could vary with the testing strategy and the intensity of the testing effort.

Third, in our current analysis we assume a single serial interval distribution for all geographic locations and all times. However, behaviour, population contact structure, and cultural differences in dealing with infection symptoms, will cause geographic and temporal variations in the serial interval. In particular, the implementation of non-pharmaceutical interventions can significantly shorten the serial interval [7]. Misspecification of the generation interval will be most pronounced for R_e values further away from one [13].

Lastly, our estimates of the effective reproductive number R_e are subject to changes in data reporting. There are frequent changes in the way in which public health offices update their observed incidence data: the amount of variables shared (e.g. Brasil, the UK excluded testing information), their frequency (e.g. Swiss cantons moved to weekly data updates when daily numbers became low), the amount of consolidation (i.e. how much values reported for a given day change in subsequent days), and what constitutes a COVID-19 case [46]. These variables have all changed during the epidemic, frequently in response to political pressure or the magnitude of the local epidemic and the resulting workload at the public health offices [46]. This affects the timeliness of our estimates, and can cause the estimated R_e to change a bit between days.

7 Supplementary Methods

Discretisation of delay distributions When approximating delay distributions by gamma distributions, we discretise these in the following fashion:

$$m_l = \begin{cases} \int_0^{0.5} f(x)\,\mathrm{d}x & l = 0 \\ \int_{l-0.5}^{l+0.5} f(x)\,\mathrm{d}x & l \in \mathbb{N}^*\,, \end{cases} \tag{9}$$

where f is either the probability density function (p.d.f) of the gamma-distributed delay distribution, or the p.d.f of the sum of two independent gamma-distributed delay distributions. The former applies when line list data is available, and the observed data is deconvolved with the gamma-distributed incubation period separately from the empirical delay distribution of symptom onset to observation. The latter applies whenever the observed case data is jointly deconvolved with the incubation period and delay between symptom onset and observation.

Because the probability density function of a sum of two independent gamma distributions does not admit a simple form in the general case, in practice we approximate the p.d.f by drawing a million independent pairs of samples, one from each gamma distribution, summing the pairs, and computing the empirical cumulative distribution function of the sampled distribution.

8 Supplementary Materials

8.1 Supplementary Simulations

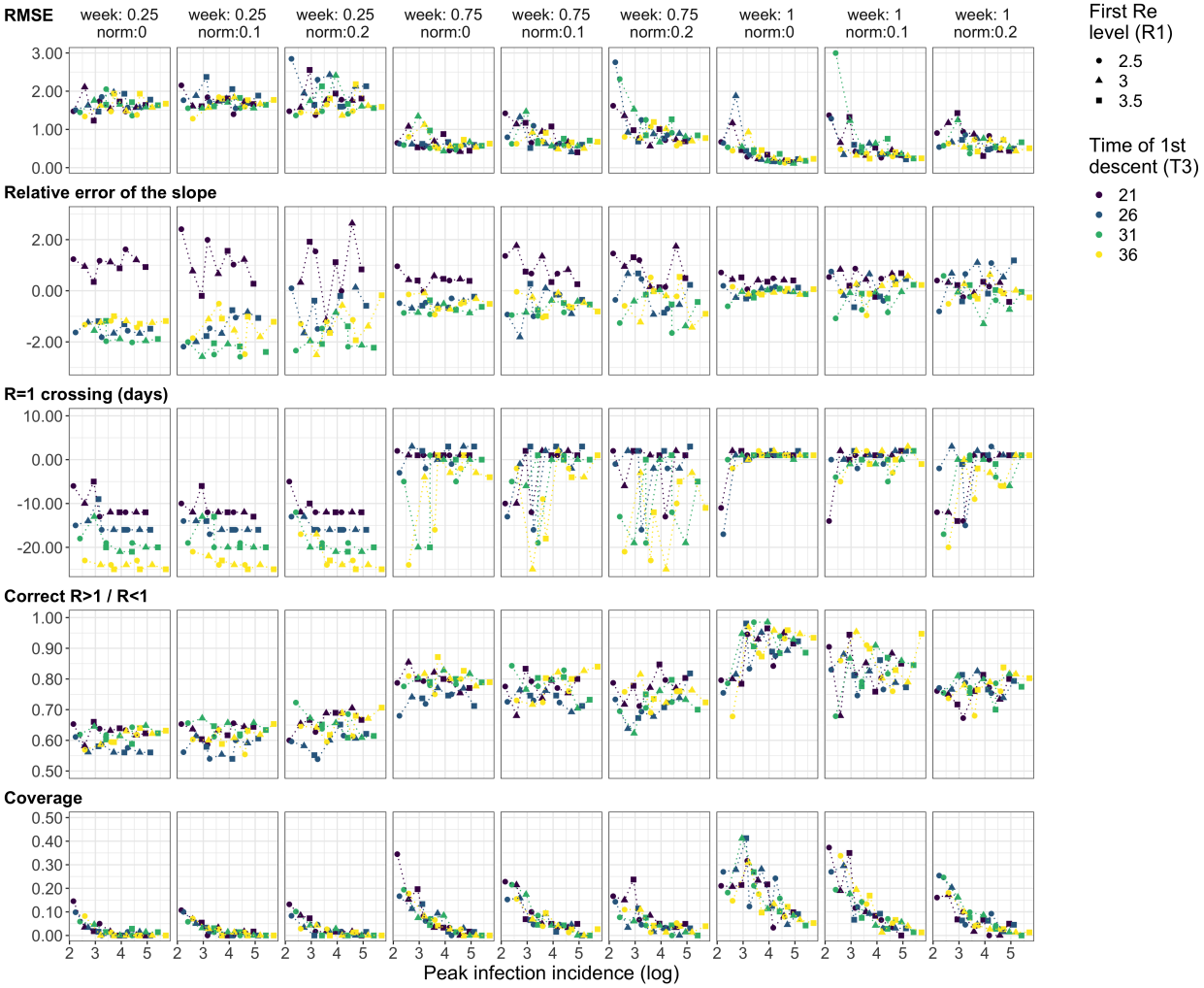


Figure S1: Performance of our method on simulated scenarios with noise and without smoothing. The R_1 (point shape), t_3 (point colour), and I_0 were varied, all other parameters kept constant (see Materials and Methods 4.5). Three plateaus were used, with $R_2 = 0.5, R_3 = 1.2$. The columns indicate different combinations noise: weekly noise ('week') reduces the number of cases on the weekend to a fraction f and redistributes these to Monday and Tuesday ($f = 1$ is the weakest noise), and daily noise ('norm') which refers to multiplicative Gaussian noise with mean 1 and a set standard deviation sd on every day of the time series ($sd = 0$ is weakest).

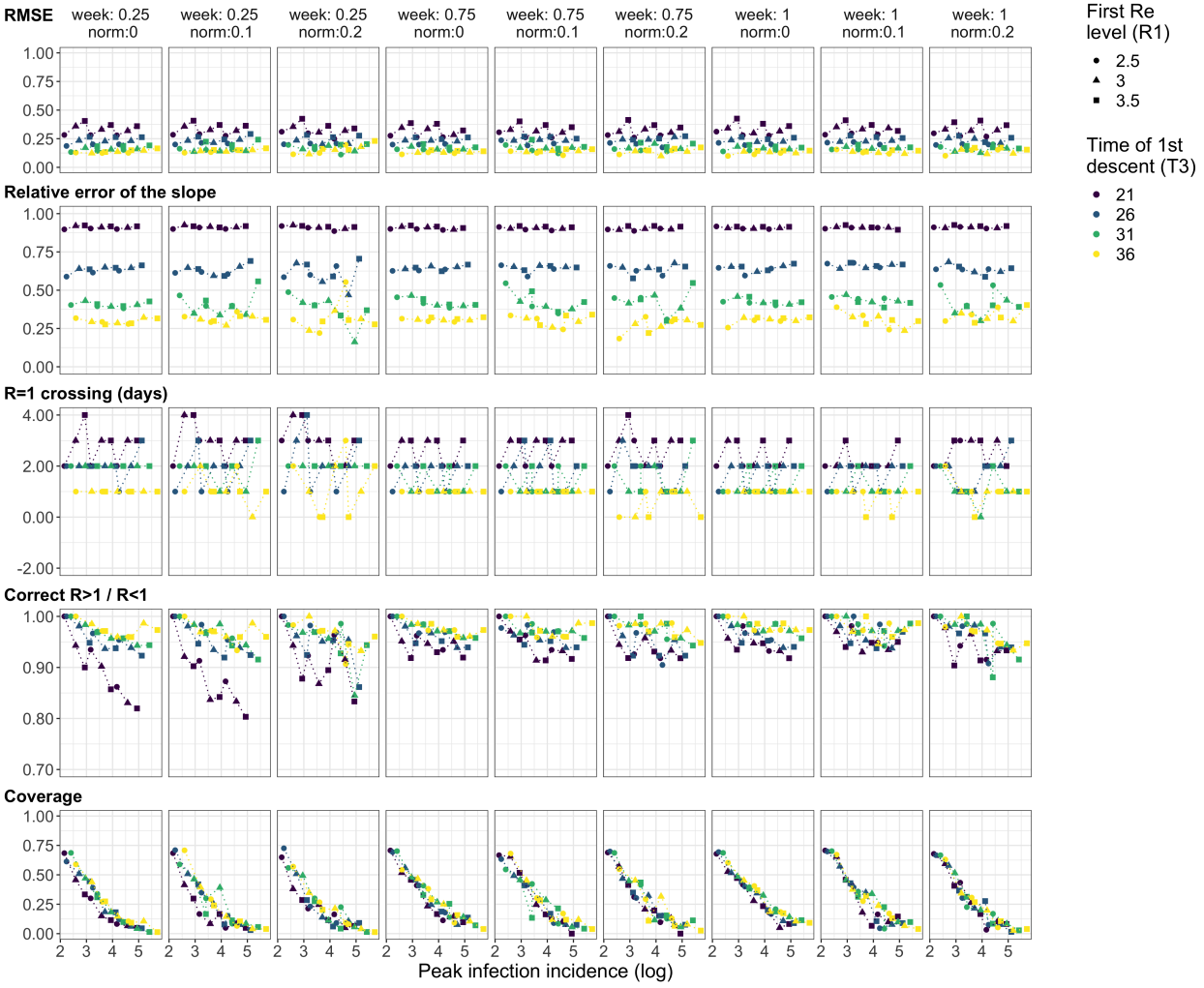
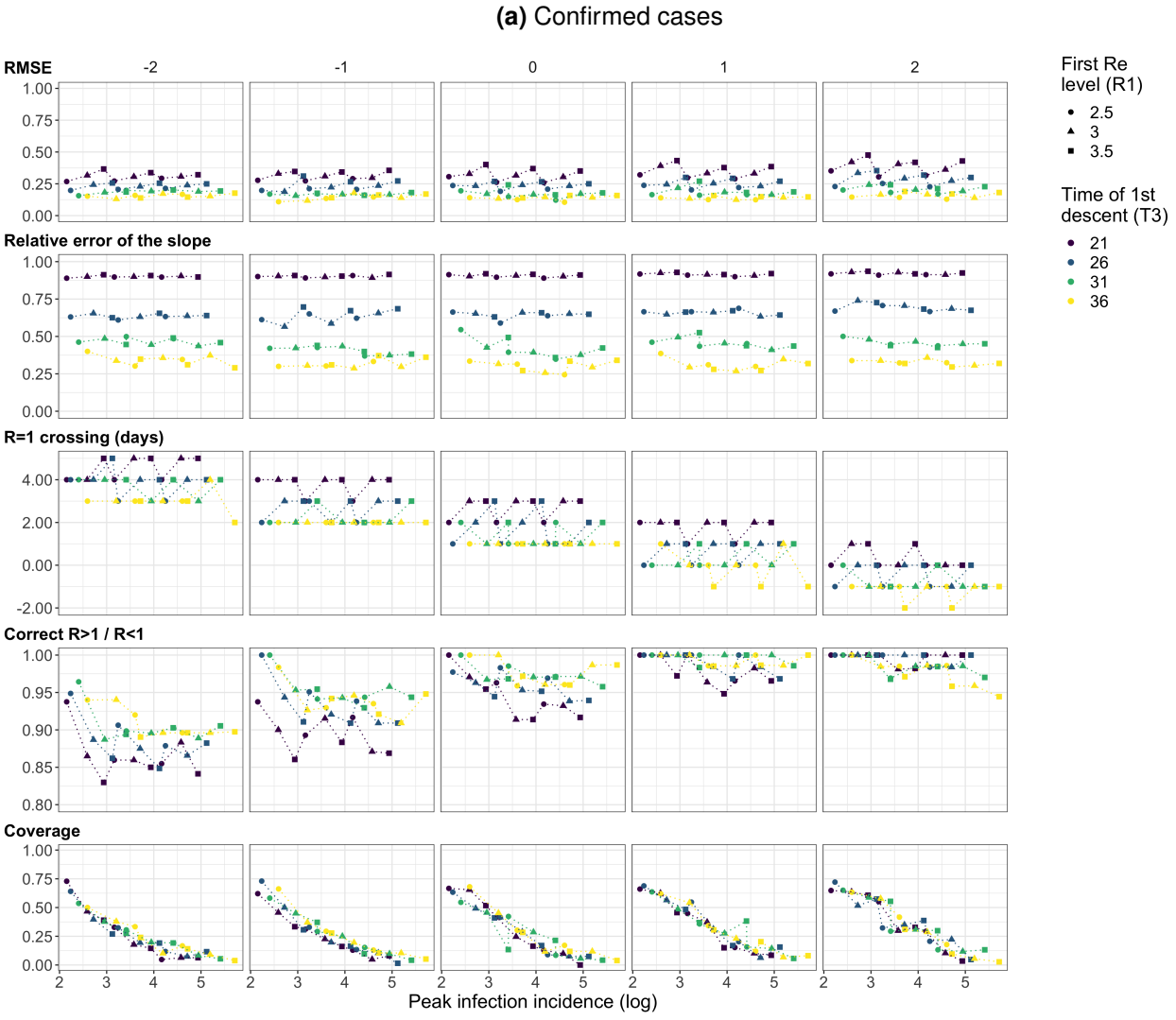


Figure S2: Performance of our method on simulated scenarios with noise and with smoothing. The R_1 (point shape), t_3 (point colour), and I_0 were varied, all other parameters kept constant (see Methods 4.5). Three plateaus were used, with $R_2 = 0.5, R_3 = 1.2$. The columns indicate different combinations noise: weekly noise ('week') reduces the number of cases on the weekend to a fraction f and redistributes these to Monday and Tuesday ($f = 1$ is the weakest noise), and daily noise ('norm') which refers to multiplicative Gaussian noise with mean 1 and a set standard deviation sd on every day of the time series ($sd = 0$ is weakest).



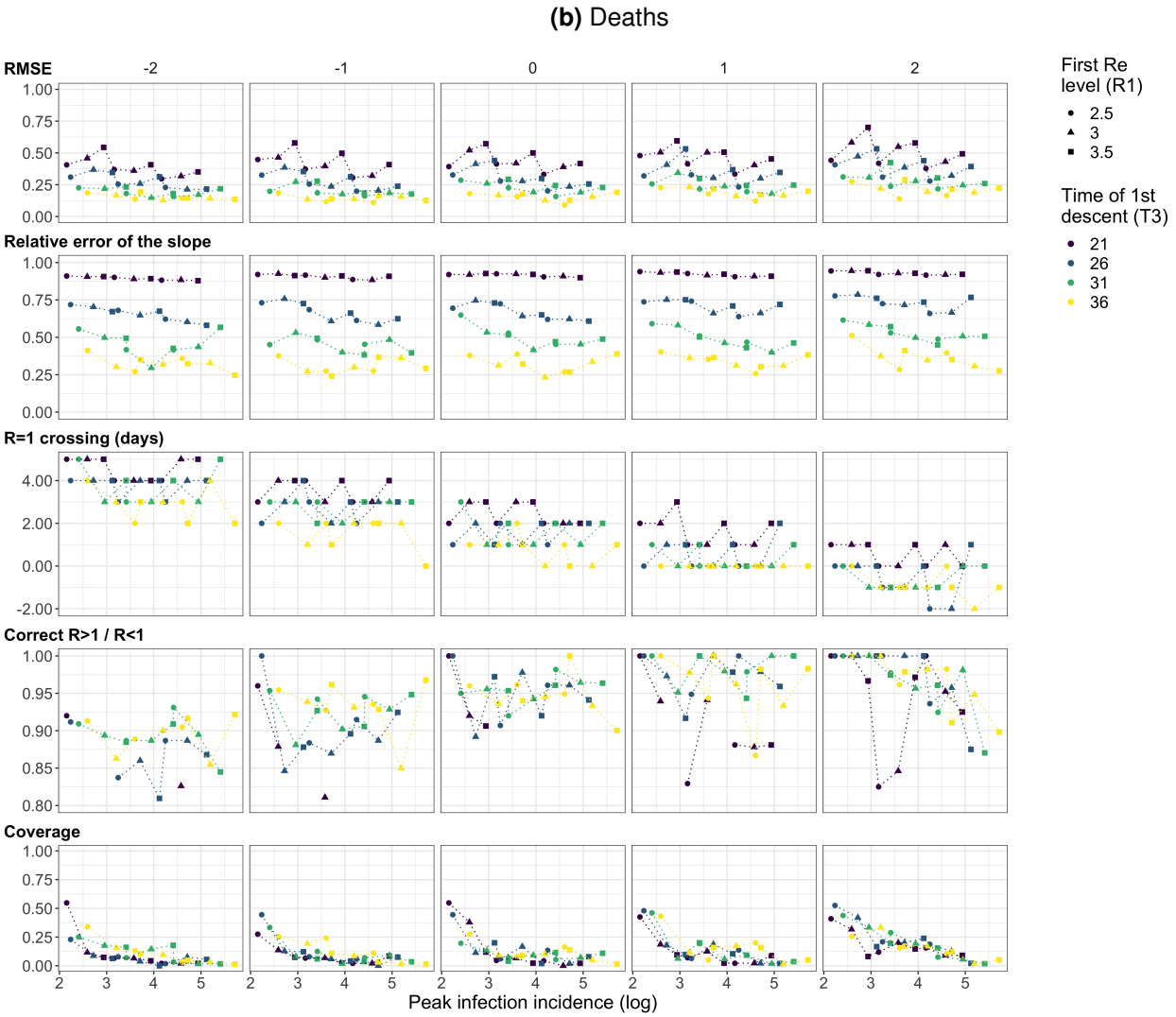
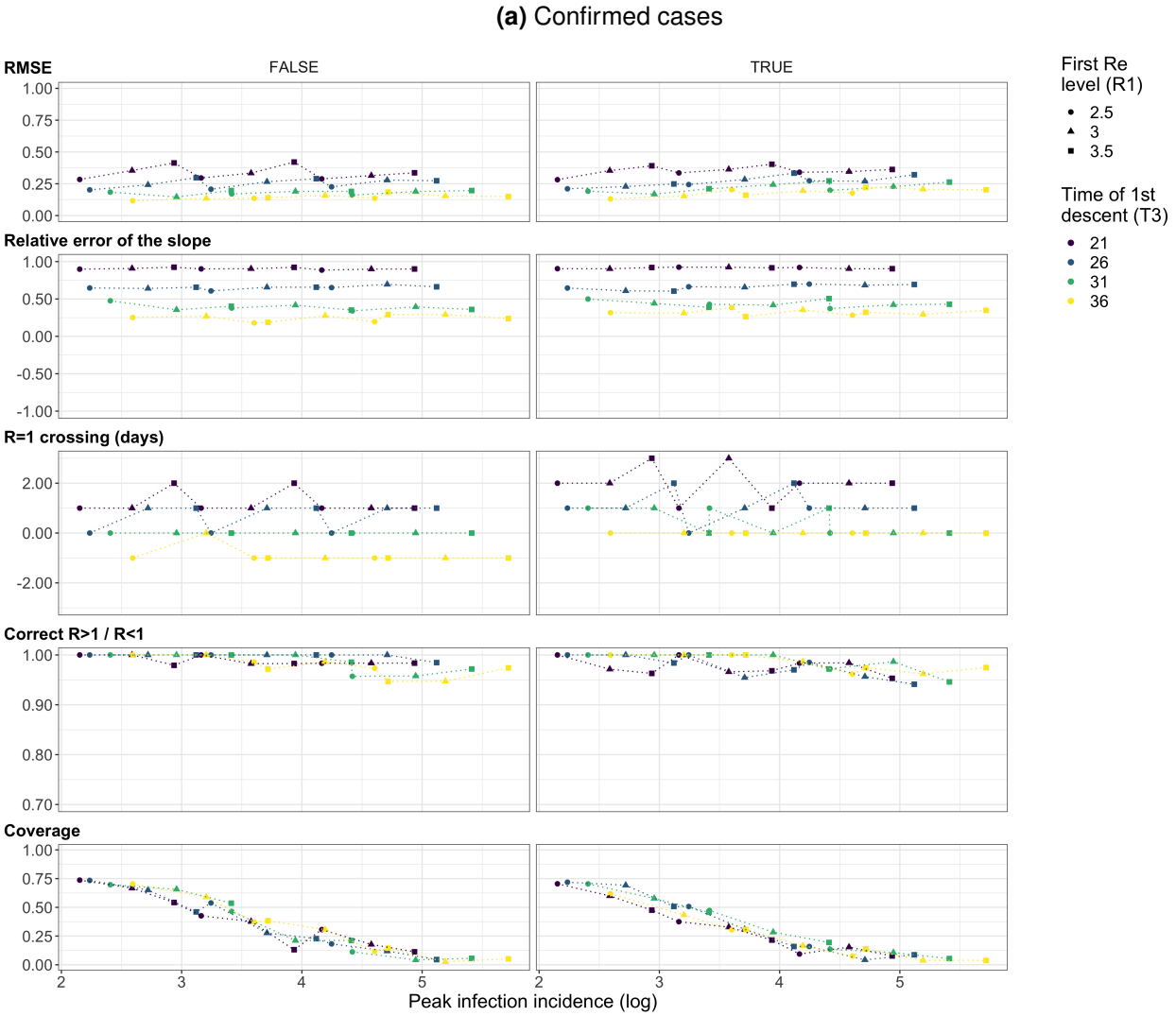


Figure S3: Performance of our method on simulated scenarios with misspecified delay distributions. We changed the mean of the delay distribution (5.5 for symptom-onset to case confirmation, 15.0 for symptom-onset to death) by the numbers above the columns. The R_1 (point shape), t_3 (point colour), and I_0 were varied, all other parameters kept constant (see Materials and Methods 4.5). The observations were smoothed and no noise was added. Three plateaus were used, with $R_2 = 0.5, R_3 = 1.2$. No noise was added to the observations.



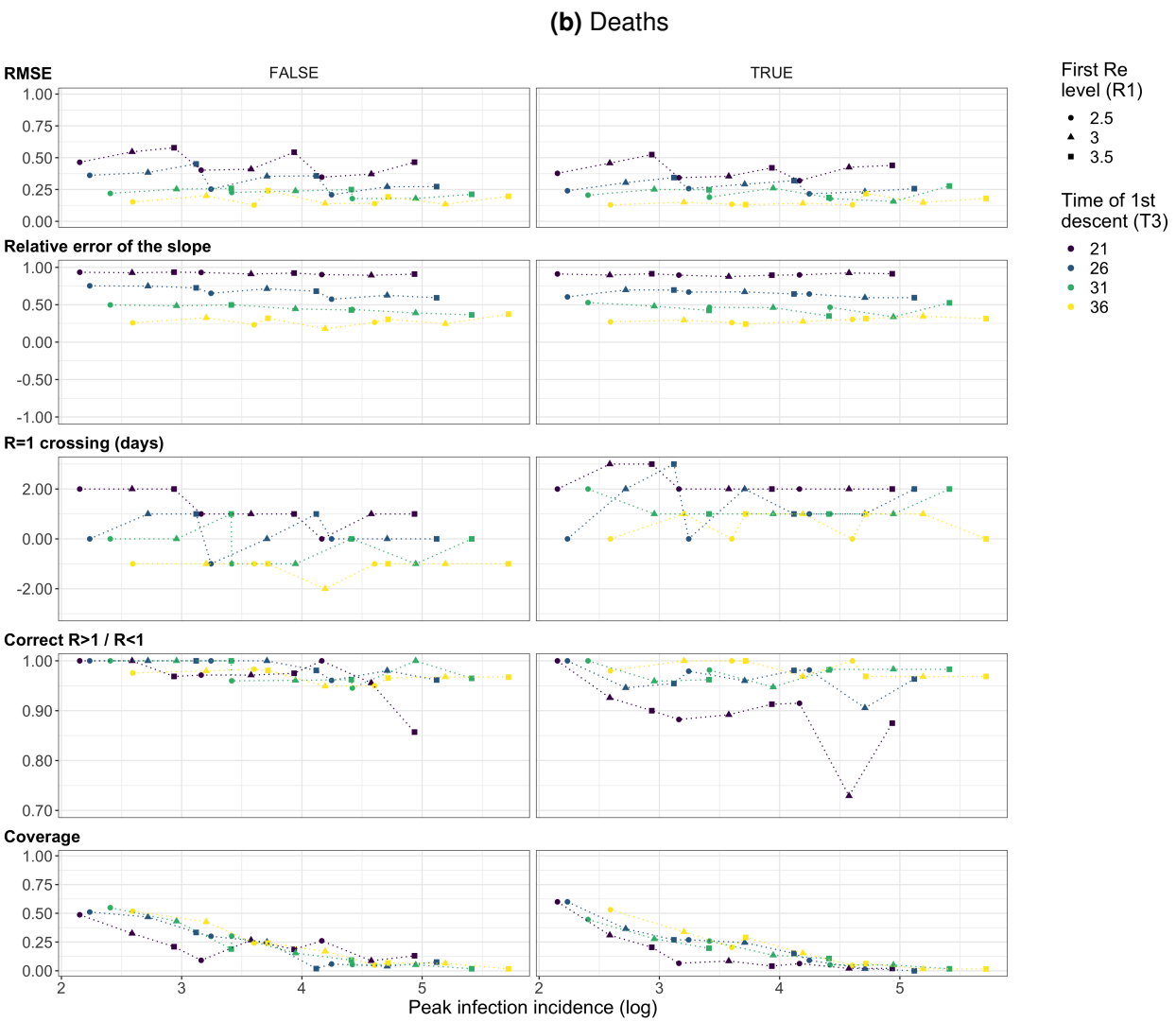


Figure S4: Performance of our method on simulated scenarios with time-varying delay distributions. The observations were simulated with time-varying delay distributions (see Materials and Methods, section on simulations), and then estimated with (right column) or without (left column) taking the time-varying distributions into account. The R_1 (point shape), t_3 (point colour), and I_0 were varied, all other parameters kept constant. The observations were smoothed and no noise was added. Three plateaus were used, with $R_2 = 0.5, R_3 = 1.2$.

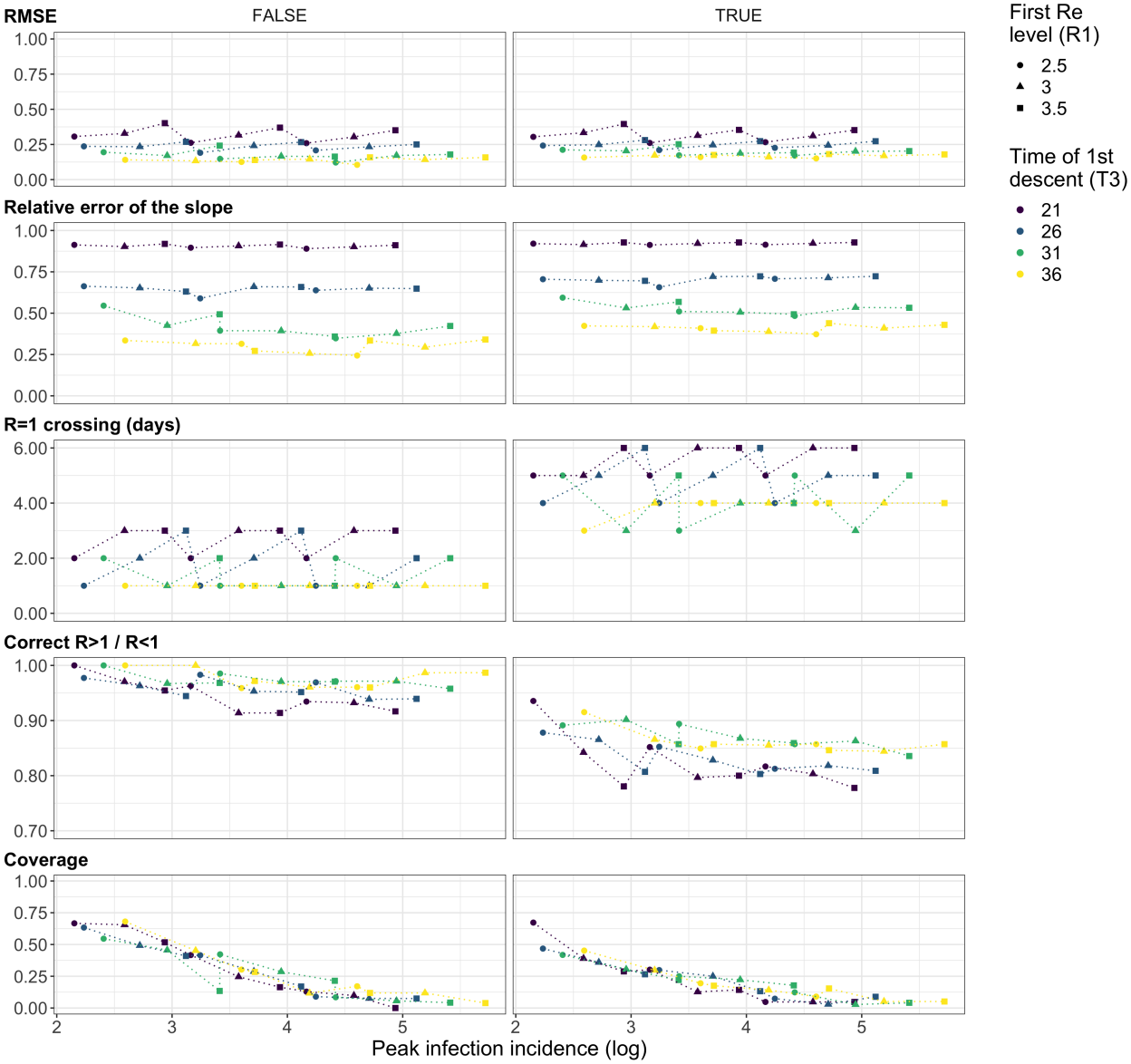


Figure S5: Performance of our method on simulated scenarios using the fixed shift (TRUE) instead of the deconvolution (FALSE) to infer infection incidence. In case of the fixed shift method the observations are shifted back by the mean of the delay distribution. For both methods the observations are bootstrapped in the same way, leading to similar width of the confidence intervals. The R_1 (point shape), t_3 (point colour), and I_0 were varied, all other parameters kept constant (see Materials and Methods 4.5). The observations were smoothed and no noise was added. Three plateaus were used, with $R_2 = 0.5, R_3 = 1.2$.

8.2 Switzerland specific results

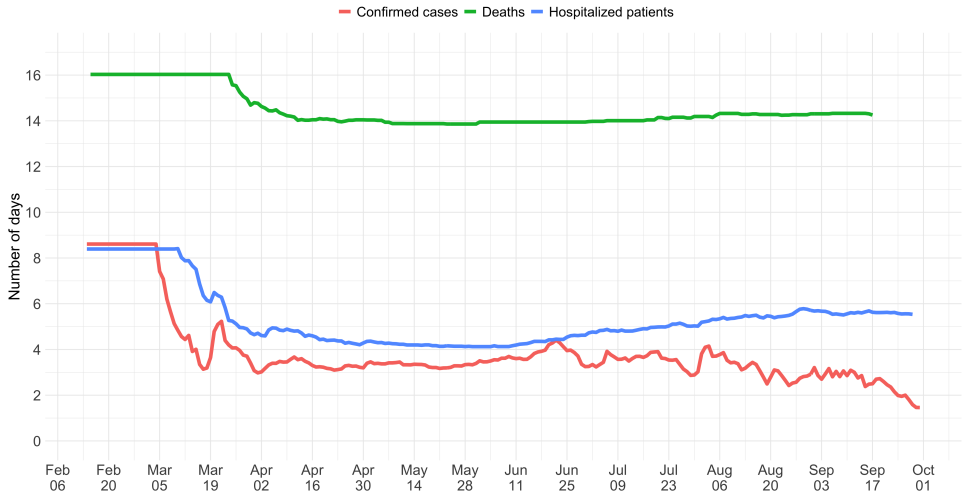


Figure S6: Mean delay in Switzerland between onset of symptoms and reporting. For each date, the mean is taken over the last 300 reports with known symptom onset date, based on line list data from the FOPH. For early dates, before 300 reports are available, the mean is taken over the first 300 reports.

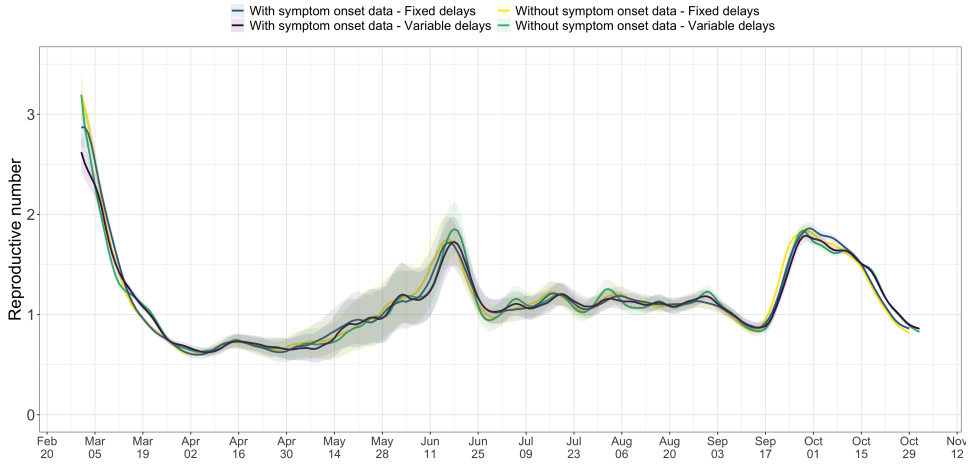


Figure S7: Comparison of effective reproductive number estimates with or without accounting for known symptom onset dates and for time-variability on reporting delays. The comparison is based on time series of confirmed cases in Switzerland, from line list data provided by the FOPH. Both the inclusion of known symptom onset dates and of the time-variability of reporting delay distributions have an effect on the Re estimates, in particular for early estimates in this case.

8.3 R = 1 crossings

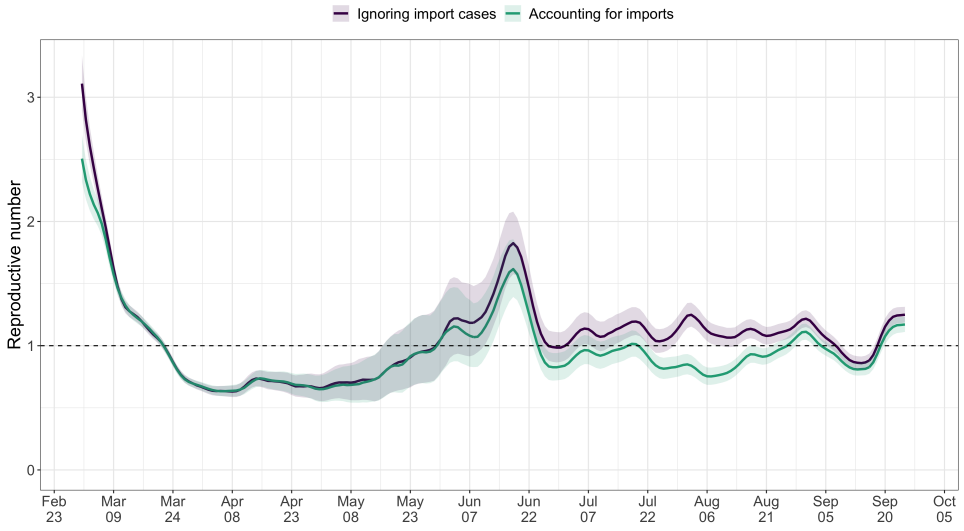


Figure S8: Effective reproductive number estimates with or without accounting for known imports. The comparison is based on time series of confirmed cases in Switzerland, from line list data provided by the FOPH. The analysis without imports is unbiased if the number of imports equals the number of exports. Since the analysis accounting for imports is not accounting for exports, the results are a lower limit for the effective reproductive number.

Country	Lockdown	R=1 based on Confirmed cases	R=1 based on Deaths	R=1 based on Hospitalisations
Austria	16-03	21-03 [21-03, 22-03]		
Belgium	18-03	31-03 [30-03, 02-04]	26-03 [25-03, 28-03]	25-03 [24-03, 26-03]
Denmark	18-03	$\leq$**10-03 [$\leq$10-03, 11-03]**	23-03 [20-03, $\geq$19-09]	
Finland	16-03	03-04 [01-04, 06-04]	08-04 [26-03, 13-04]	
France	17-03	27-03 [27-03, 28-03]	24-03 [23-03, 24-03]	26-03 [26-03, 27-03]
Germany	22-03	26-03 [26-03, 27-03]	02-04 [29-03, 04-04]	
Ireland	27-03	09-04 [08-04, 10-04]	09-04 [07-04, 12-04]	
Italy	10-03	18-03 [18-03, 19-03]	14-03 [13-03, 15-03]	
Netherlands	23-03	01-04 [26-03, 04-04]	**20-03 [18-03, 23-03]**	
Norway	14-03	23-03 [22-03, 24-03]	28-03 [17-03, 11-04]	
Poland	25-03	03-04 [02-04, 17-04]	10-04 **[25-03, $\geq$19-09]**	
Portugal	16-03	29-03 [28-03, 30-03]	29-03 [23-03, 11-04]	
Romania	24-03	06-04 [02-04, 30-04]	30-03 [26-03, 01-05]	
Russian Federation	30-03	05-05 [04-05, 05-05]	19-05 [17-04, 04-06]	
Slovenia	20-03	24-03 **[$\leq$14-03, 28-03]**	26-03 [22-03, $\geq$19-09]	
Spain	14-03	18-03 [02-02, 18-03]		
Sweden	-	**19-04 [01-04, 21-04]**	**29-03 [26-03, 06-04]**	
Switzerland	17-03	22-03 [21-03, 22-03]	21-03 [19-03, 24-03]	17-03 **[16-03, 19-03]**
Turkey	21-03	09-04 [08-04, 09-04]	05-04 [01-04, 08-04]	
United Kingdom	24-03	31-03 [30-03, 31-03]	25-03 [25-03, 26-03]	29-03 [28-03, 29-03]

Table S1: The date that $R_e < 1$ for the first time. Based on news reports, we determined when a country implemented stay-at-home orders (a 'lockdown'). Using our method we determined when the R_e estimate and its confidence intervals first dropped below 1. Based on our R_e estimates for confirmed cases, only two countries that implemented a nationwide lockdown (Denmark, Slovenia) had confidence intervals that included or were below one before a nationwide lockdown was implemented. For R_e estimates based on COVID-19 deaths, there are also two, but different ones (the Netherlands, Poland).

Country	First Measure	Lockdown	URL
Austria	10-03	16-03	https://mrc-ide.github.io/covid19estimates/#/interventions
Belgium	10-03	18-03	https://mrc-ide.github.io/covid19estimates/#/interventions
Denmark	12-03	18-03	https://mrc-ide.github.io/covid19estimates/#/interventions
Finland	16-03	16-03	https://en.wikipedia.org/wiki/COVID-19_pandemic_in_Finland#Response_by_sector
France	29-02	17-03	https://www.politico.eu/article/europes-coronavirus-lockdown-measures-compared/
Germany	06-03	22-03	https://www.bundesregierung.de/breg-de/themen/coronavirus/besprechung-der-bundeskanzlerin-mit-den-regierungschefinnen-und-regierungschefs-de
Ireland	12-03	27-03	https://en.wikipedia.org/wiki/COVID-19_pandemic_in_the_Republic_of_Ireland#Containment_phase
Italy	22-02	10-03	https://www.politico.eu/article/europes-coronavirus-lockdown-measures-compared/
Netherlands	10-03	23-03	https://www.volkskrant.nl/nieuws-achtergrond/bijeenkomsten-tot-1-juni-verboden-burgemeesters-mogen-handhaven-met-forse-boetes~a4411
Norway	12-03	14-03	https://www.euractiv.com/section/coronavirus/short_news/norway-update-covid-19/
Poland	09-03	25-03	https://www.politico.eu/article/europes-coronavirus-lockdown-measures-compared/
Portugal	11-03	16-03	https://www.politico.eu/article/europes-coronavirus-lockdown-measures-compared/
Romania	21-02	24-03	https://en.wikipedia.org/wiki/COVID-19_pandemic_in_Romania
Russian Federation	25-03	30-03	https://en.wikipedia.org/wiki/COVID-19_pandemic_in_Russia
Slovenia	09-03	20-03	https://en.wikipedia.org/wiki/COVID-19_pandemic_in_Slovenia
Spain	10-03	14-03	https://www.elmundo.es/espana/2020/03/13/5e6b844e21efa0dd258b45a5.html
Sweden	11-03		
Switzerland	28-02	17-03	'Verordnung 2 über Massnahmen zur Bekämpfung des Coronavirus (COVID-19)'
Turkey	12-03	21-03	https://en.wikipedia.org/wiki/COVID-19_pandemic_in_Turkey#Government_response
United Kingdom	12-03	24-03	https://www.bbc.com/news/uk-52012432

Table S2: News and public resources used to determine when a country implemented the first non-pharmaceutical interventions, and a nationwide lockdown.

SI 50 analysis Reference date: first day the stringency index exceeded 50 ($SI > 50$).

The 39 included countries: Algeria, Andorra(*), Australia, Austria, Bahrain, Belgium, Brazil, Canada, Chile, China, Denmark, Egypt, Estonia, Finland, France, Germany, Greece, Iceland, India, Indonesia, Iran, Ireland(*), South Korea(*), Malaysia, Mexico, Netherlands, Norway, Portugal, Qatar, Singapore, Slovenia, Spain, Switzerland, Tajikistan, Thailand, United Arab Emirates, United Kingdom, United States of America, Vietnam.

The star indicates the country was not included in the ΔSI analysis.

For 34/39 countries the R_e estimate was above one prior to the reference date, and significantly so for 28/39. The countries that reached $R_e < 1$ prior to the reference date were Andorra (15 days prior), Australia (1 day prior), Denmark (3 days prior), Qatar (5 days prior), and Vietnam (2 days prior).

ΔSI analysis Reference date: date of the biggest 7-day increase in the SI.

The 50 included countries: Algeria, Australia, Austria, Bahrain, Belarus(*), Belgium, Brazil, Canada, Chile, China, Colombia(*), Croatia(*), Czech Republic(*), Denmark, Egypt, Estonia, Finland, France, Germany, Greece, Iceland, India, Indonesia, Iran, Israel(*), Japan(*), Lebanon(*), Malaysia, Mexico, Netherlands, New Zealand(*), Norway, Pakistan(*), Philippines(*), Portugal, Qatar, Russia(*), Serbia(*), Singapore, Slovenia, Spain, Sweden(*), Switzerland, Tajikistan, Thailand, Turkey(*), United Arab Emirates, United Kingdom, United States of America, Vietnam.

The star indicates the country was not included in the $SI50$ analysis.

For 44/50 countries the R_e estimate was above one prior to the reference date, and significantly so for 36/50. The countries that reached $R_e < 1$ prior to the reference date were Australia (2 days prior), Denmark (4 days prior), New Zealand (2 days prior), Qatar (6 days prior), Tajikistan (17 days prior), and Vietnam (9 days prior).

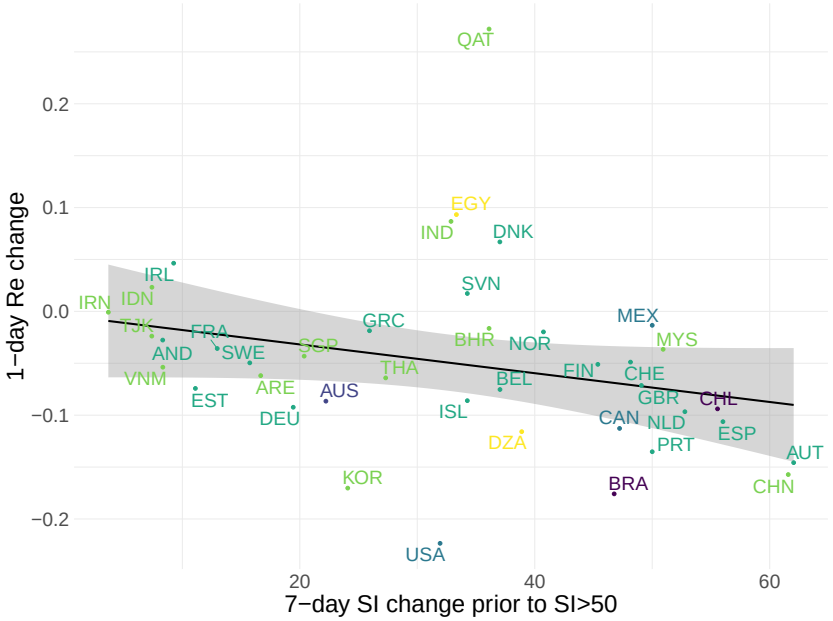


Figure S9: The association between non-pharmaceutical interventions and R_e. Relation between the slope of the R_e estimates on t_{SI50} and the increase in the stringency index in the 7 days prior to t_{SI50}. Countries are indicated by their ISO3 country code, colours represent continents.

8.4 Implementation/Lifting of individual NPIs

8.5 Sensitivity of the deconvolution to initial conditions

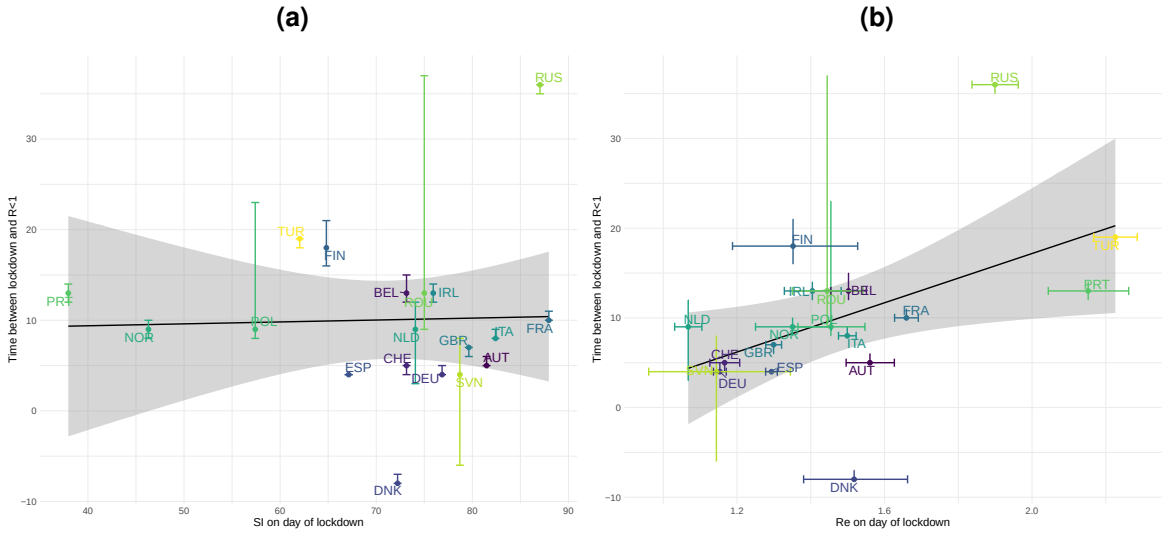


Figure S10: The time between the start of lockdown and $R < 1$. (a) The stringency index on the day of lockdown is not a good predictor for the number of days it will take for $R_e < 1$ ($p = 0.9$). **(b)** The R_e estimate on the day of lockdown is significantly associated with the number of days it will take for $R_e < 1$ ($p = 0.03$, adjusted $R^2 = 0.2$). The regression analysis used the R_e point estimate for each country. The uncertainties shown in the plot were not used for this analysis. Countries are the same as in table 1, and indicated by their ISO3 country code.

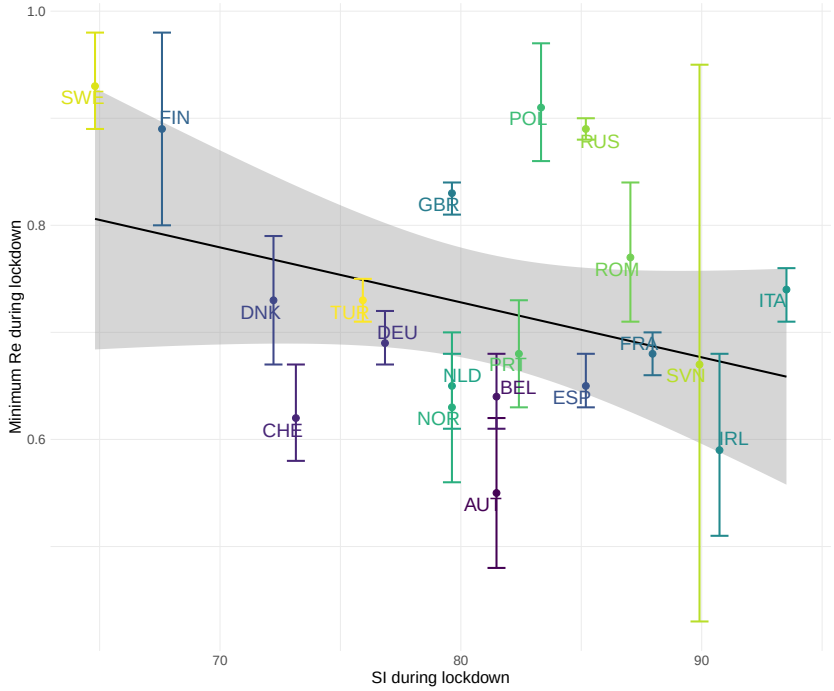


Figure S11: The relation between the stringency and minimum R_e attained during lockdown in Europe. This trend is not significant ($p = 0.14$). Lockdown here refers to the period of constant, maximum stringency during the first wave in Europe (April 2020). There is substantial variation in the duration of the lockdown, and the number of days after the start of the lockdown that the minimum R_e was reached. The regression analysis used the R_e point estimate for each country. The uncertainties shown in the plot were not used for this analysis. Countries are the same as in table 1, and indicated by their ISO3 country code.

Intervention index	P-value $< 0.05/48$	P-value $0.05/48 < p < 0.05$
C1: School closing	Europe	Africa, Asia, Oceania
C2: Workplace closing		Asia, Europe
C3: Cancel public events	South America	Asia
C4: Restriction on gatherings		Asia, Europe
C5: Close public transport	Europe	North America
C6: Stay at home requirements	North America	Europe, Oceania
C7: Restrictions on internal movement		North America, Asia, Europe
C8: International travel controls		Europe

Table S3: Continents for which an increase in stringency was significantly more associated with a subsequent reduction in R_e than a reduction in stringency. Significance was determined by permutation test for each index separately, with significance threshold $\alpha = 0.05$, and corrected for 48-way hypothesis testing.

Intervention index	P-value $< 0.05/48$	P-value $0.05/48 < p < 0.05$
C1: School closing	Europe ($R^2 = 0.13$)	Asia ($R^2 < 0.1$)
C2: Workplace closing	Europe ($R^2 < 0.1$)	Asia ($R^2 < 0.1$)
C3: Cancel public events	South America ($R^2 = 0.77$)	Asia ($R^2 < 0.1$)
C4: Restriction on gatherings	Europe ($R^2 < 0.1$)	
C5: Close public transport	Europe ($R^2 = 0.17$)	
C6: Stay at home requirements	Europe ($R^2 < 0.1$)	North America ($R^2 = 0.15$)
C7: Restrictions on internal movement	Europe ($R^2 < 0.1$)	North America ($R^2 = 0.13$), Asia ($R^2 < 0.1$)
C8: International travel controls		Europe ($R^2 < 0.1$)

Table S4: Continents where the estimated one-week change in R_e is significantly determined by changes in individual stringency indices. Significance was determined with significance threshold $\alpha = 0.05$, and corrected for 48-way hypothesis testing.

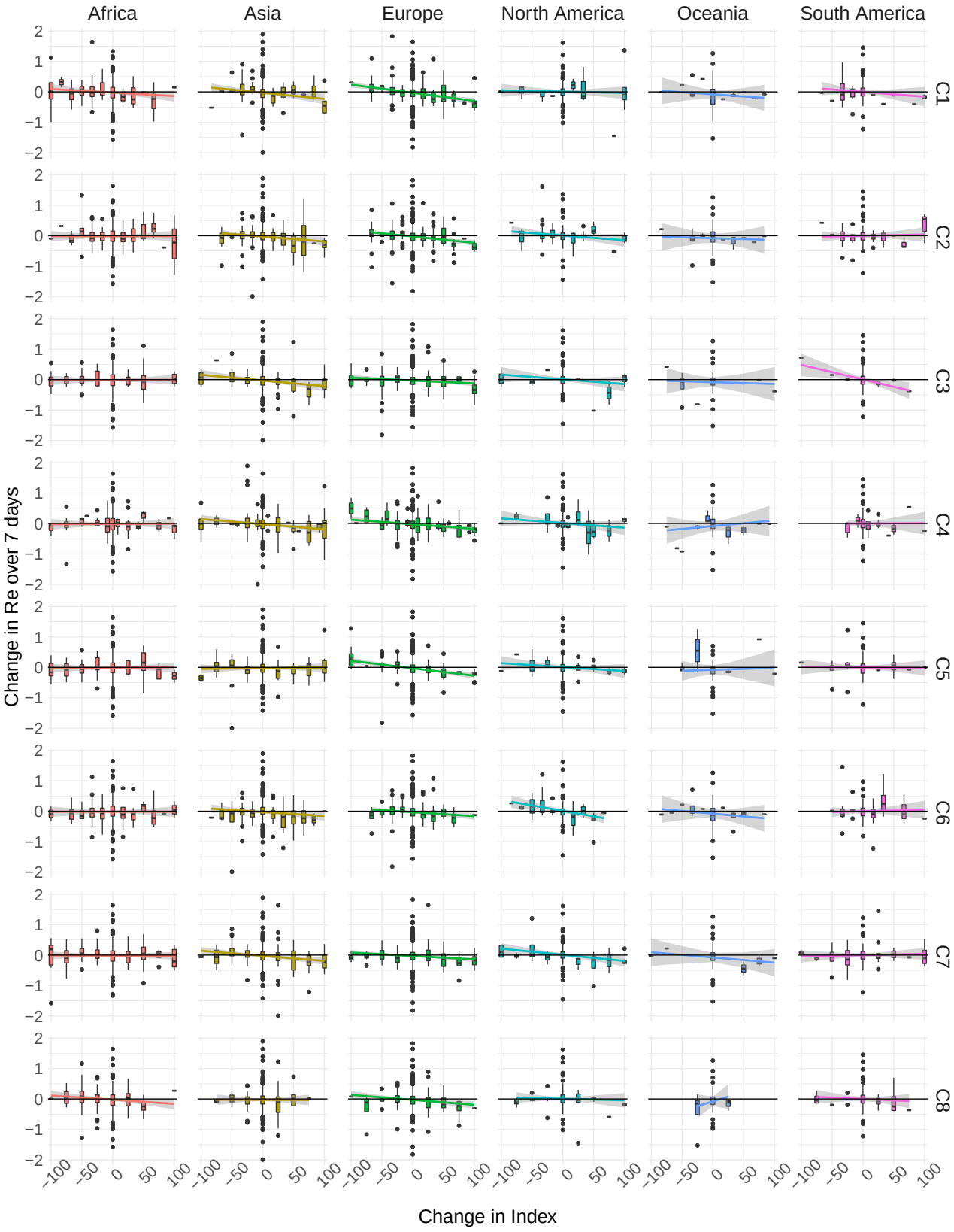
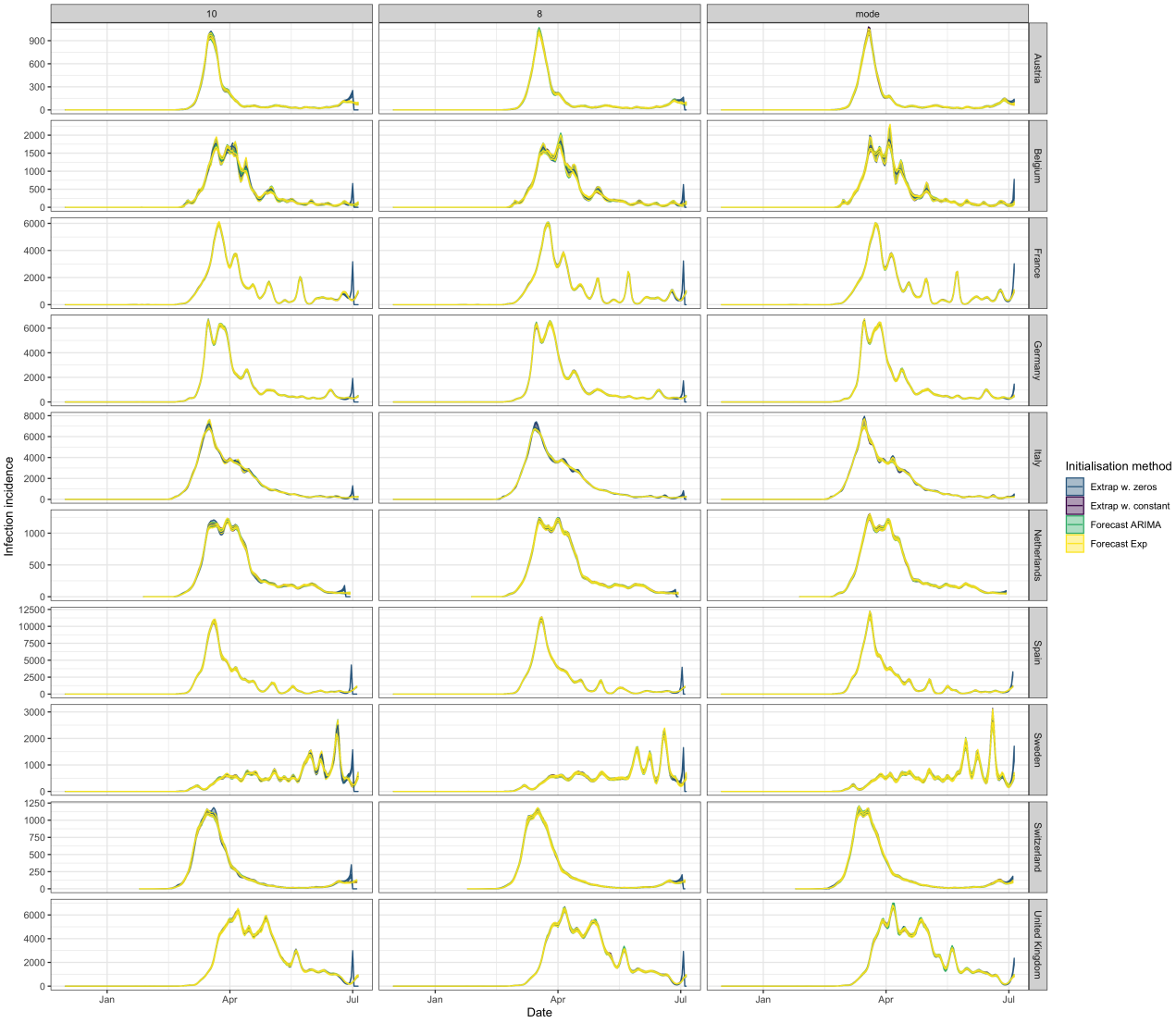


Figure S12: The one-week change in R_e following changes in government stringency. The rows C1-C8 refer to different stringency indices (see table S4).

(a) Whole timeseries



(b) Last Fortnight

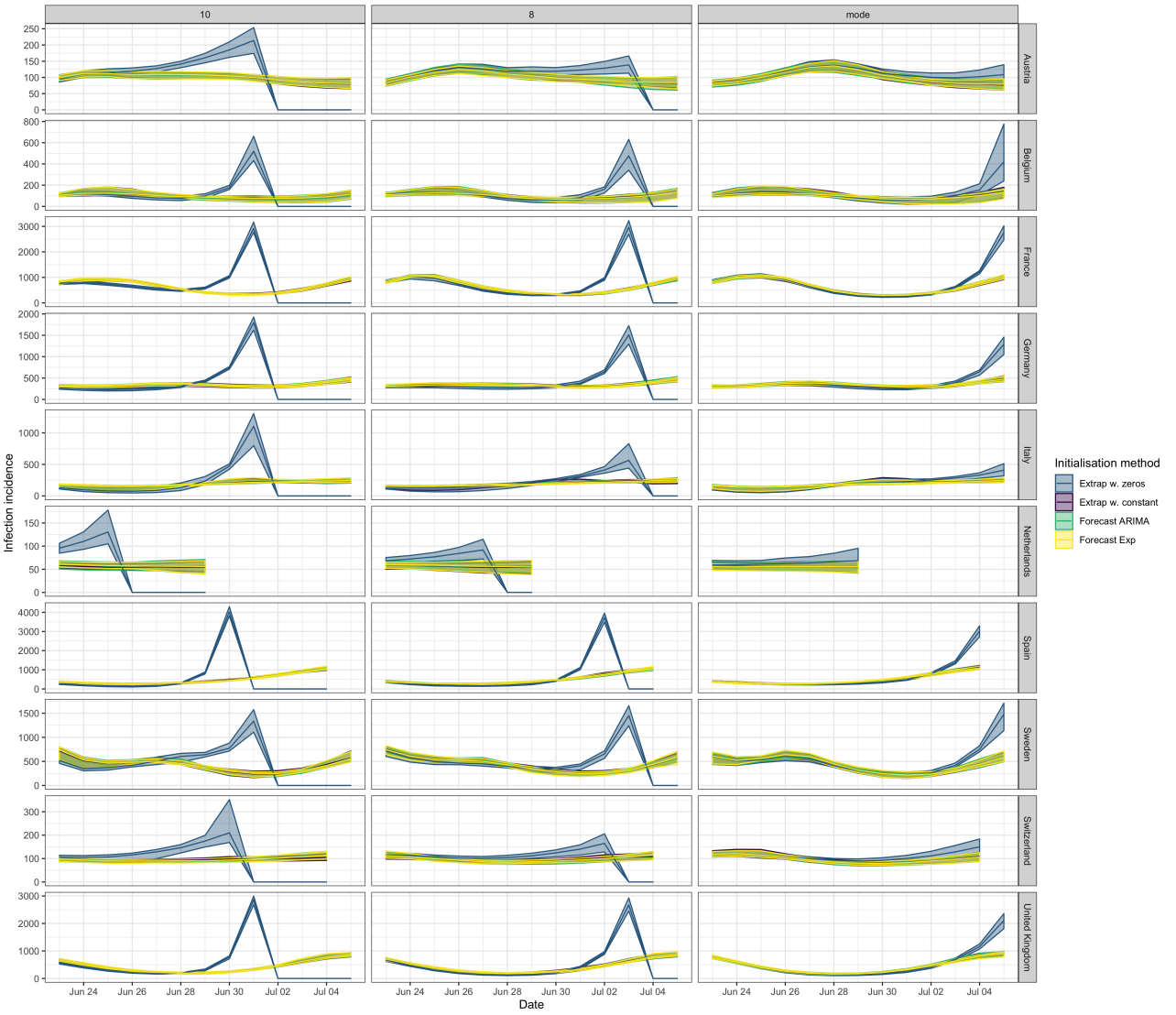


Figure S13: Assessing the sensitivity of the deconvolution to different initial estimates. Panel **(b)** shows the last fortnight of panel **(a)**. Confidence intervals are computed on 20 bootstraps of the original time series. The columns represent a shift by 10 days, 8 days, or the mode of the delay distribution (7 days). The rows show ten countries in our dataset. It is clear that augmenting both ends with zeros can lead to spurious increases in incidence towards the present (blue curves). Augmenting with a constant non-zero integer (purple), forecasting with an ARIMA model (green), and forecasting with an exponential model (yellow), all perform similarly.