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

Jana S. Huisman[1,2,3¶*], Jérémie Scire[2,3¶*], Daniel C. Angst[1], Jinzhou Li[4],
Richard A. Neher[2,5], Marloes H. Maathuis[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] Department of Mathematics, ETH Zurich, Swiss Federal Institute of Technology, Zurich, Switzerland

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

[¶,†] These authors contributed equally

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

Abstract

The effective reproductive number R_e is a key indicator of the growth of an epidemic. Since the start of the SARS-CoV-2 pandemic, many methods and online dashboards have sprung up to monitor this number through time. However, these methods are not always thoroughly tested, correctly placed in time, or are overly confident during high incidence periods. Here, we present a method for near real time estimation of R_e, applied to epidemic data from 170 countries. We thoroughly evaluate 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 the implementation of non-pharmaceutical interventions was broadly associated with reductions in the estimated R_e. Globally though, relaxing non-pharmaceutical interventions had more varied effects on subsequent R_e estimates. Our framework is useful to inform governments and the general public on the status of the epidemic in their country, and is used as the official source of R_e estimates in Switzerland. It further allows detailed comparison between countries and in relation to covariates such as implemented public health policies, mobility, behaviour, or weather data.

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 to describe how a pathogen spreads in a given population at a given time [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 R_e. They broadly fall into two categories: those based on compartmental models, e.g. [5, 9, 10], and those that infer 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 class of methods as they rely on few, simple assumptions, are less prone to model misspecifications, and are well-suited for near real-time monitoring of the epidemic [13]. In particular, we consider the EpiEstim method of Cori et al. [12].

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, or 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]. It is important to relate R_e estimates to the timing of infection events because this allows multiple proxies of infection events, with differing delays, to be used independently to monitor the same epidemic [6]. In addition, any factors that may affect transmission dynamics will do so at the time infections occurred. If R_e is placed properly on this timescale, it can be compared directly to external covariates like weather and interventions [8, 14]. However, depending on the method used to infer the timing of infections from the observed incidence time series, one can also introduce biases such as smoothing sudden changes in R_e [13, 15, 16].

Several methods, software packages, and online dashboards have been developed to monitor developments in R_e during the SARS-CoV-2 pandemic (e.g. [17, 18, 19]). A pipeline for the continuous estimation 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 four axes also define the differences between existing methods. The first step dictates e.g. the geographical scope of the R_e estimates reported. During the SARS-CoV-2 epidemic, many local public health authorities have made case data publicly available. Depending on the data sources used, estimated R_e values span from the scale of a city, region, country, or the entire globe [18, 20, 21]. The second step, i.e. going from a noisy time series of indirect observations to an infection incidence time series, is technically 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, 22], which produce 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 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

deaths). This allows comparison across observation types and to balance the biases inherent in the different types. Results are updated daily, and can be found on `https://ibz-shiny.ethz.ch/covid-19-re-international/`. The results of this method are used directly in public health policy making in Switzerland, and are also communicated by the Federal Office of Public Health `https://www.covid19.admin.ch/en/overview`. Through continuous engagement with the public, scientific experts, and thorough evaluation on simulated scenarios, we have created a robust and transparent method of enduring relevance for the current and future epidemics.

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, 23, 24, 14, 25, 26]. These differences can be attributed mostly to different model formulations [14, 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 achieve $R_e < 1$. Further, it remains unclear how the impact of interventions differed across time and geographical regions. We add to this debate by using our R_e estimates across geographical regions and 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 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 case incidence time series (see Materials and Methods). The objective was to achieve stable estimates for multiple types of data, and with an adequate representation of uncertainty. To the best of our knowledge, no existing method fulfills this aim out-of-the box. At the core, we use the EpiEstim method [12] to estimate R_e 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, 15]. We smooth the data prior to deconvolution to reduce numerical artefacts resulting from their weekly patterns and overall noisy nature. We compute pointwise 95% confidence intervals for the true R_e values, using the union of a block bootstrap method, designed to account for variation in the case observations, and the credible 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).

Evaluation on simulated data. The method was evaluated with simulations of several epidemic scenarios (see Materials and Methods for more details). For each scenario, we specified an R_e time series, from which we simulated infection - and noisy observation incidence time series 100 times. Then we used our method to infer the infection incidence and R_e from the observation incidence, and compared these to the true underlying R_e values (Fig. 1). The specified R_e trajectories were parametrised in a piecewise linear fashion. To mimic the course of the epidemic observed in many European countries in 2020 [28], we started with R_e values around 3, then dropped to a value below 1 (the 'initial decrease'), stayed around 1 in summer and slightly above 1 (the 'second wave') in autumn (Fig. 1). We added additional noise to the observations, to mimic the observation noise in different countries around the world.

The results show that our method allows accurate estimation of the effective reproductive number (Fig. 1; metrics described in the Materials and Methods). Across most time points, our 95% confidence interval includes the true value of R_e (coverage; Fig. 1B). The low root mean squared error

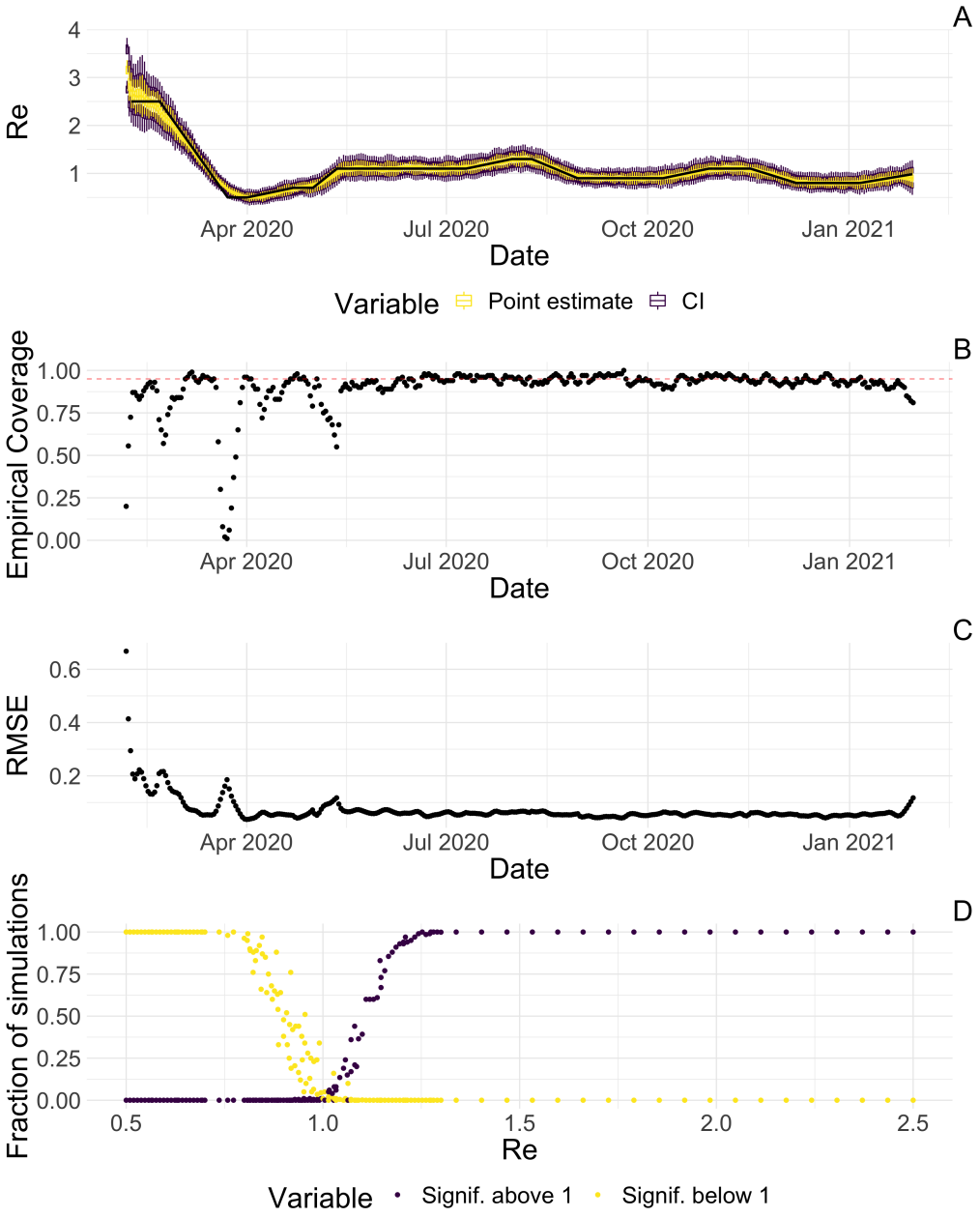


Figure 1: **A** The specified R_e trajectory (black line; described in Methods) was used to simulate a trajectory of reported cases (with noise) 100 times. From each trajectory we estimated R_e (yellow boxplots), and constructed a 95% confidence interval (purple boxplots of the lower/upper endpoint). **B** The fraction of simulations where the true R_e value was within the 95% confidence interval. The dashed red line indicates the nominal 95% coverage. **C** The root mean squared relative error for every time point. **D** The fraction of simulations where we estimate R_e to be significantly above or below one, depending on the true value of R_e.

(RMSE) indicates that our point estimates closely track the true R_e value (Fig. 1C). Importantly, we correctly infer whether R_e is significantly above or below 1 in this scenario: we never infer that R_e is significantly above 1 when the true value is below 1, and only for two time points the estimates are significantly below 1 for some simulations when the true value is a little above 1 (Fig. 1D). Due to the smoothing step prior to deconvolution, we slightly misestimate the R_e during steep slopes (see Supplementary Discussion 7.2, and Fig. S2 for more scenarios). However, the inclusion of smoothing greatly improves our performance across scenarios with different types of observation noise (Figs. S3, S4). For a wide range of infection incidences, our 95% confidence interval is informative and covers the true value of R_e (Fig. S5). This is the merit of our block bootstrap method,

and greatly improves the out-of-the-box EpiEstim method. Our method also clearly outperforms the common approach of using a fixed delay to infer the infection incidence time series (Supplementary Fig. S6).

We further tested the impact of model misspecification. Misspecifying the mean of the delay distribution between infection and case observation by up to 2 days does not have a strong effect on the R_e estimates (Supplementary Fig. S7). Correspondingly, our ability to allow for empirical, time-varying distributions in the estimation has a pronounced effect on the estimated R_e only for large changes in the mean of the delay distribution (Supplementary Fig. S8). Further model misspecifications, e.g. of the generation time interval, have been investigated by Gostic et al. [13].

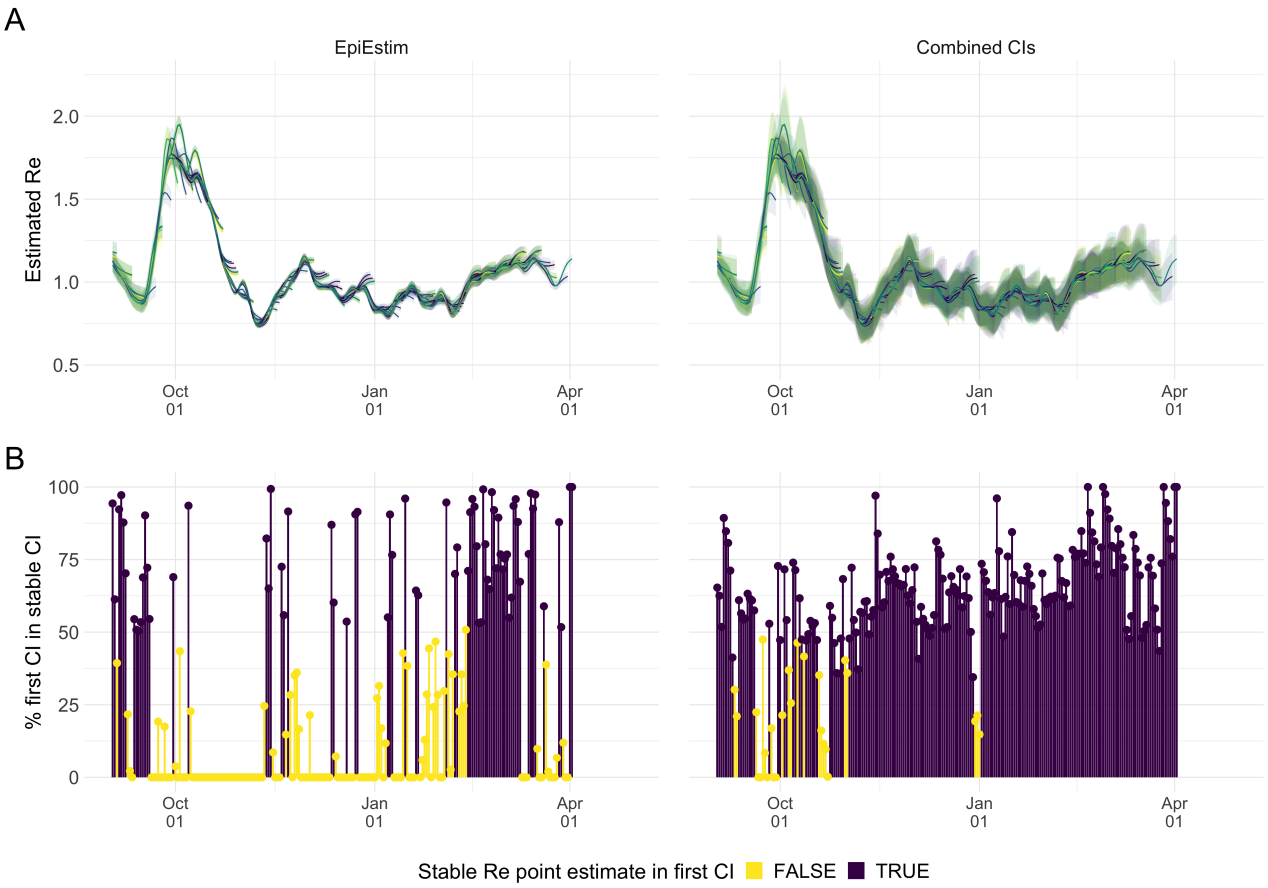


Figure 2: A Stability of Swiss R_e estimates upon adding additional days of observations. For each day we show R_e estimates and the corresponding uncertainty intervals, from the first possible estimate up to estimates including 3 additional weeks of data. Line segments correspond to 3 weeks of estimates made with the same input data (e.g. data up to December 1st) and were coloured for ease of distinction. The left of each segment corresponds to stable estimates, whereas the right endpoint will be unstable. **B** Percentage of the first estimated CI that is contained in the stable CI based on data from 30 April 2021. This percentage was calculated as the width of the intersection of both CIs, divided by the width of the first CI. The colour indicates whether the stable R_e estimate was contained in the first reported CI. In both rows, the left column shows uncertainty intervals from EpiEstim on the original data, and the right our improved 95% confidence intervals. Both columns include the same deconvolution steps.

Stability of the estimates used for situation monitoring. In addition, we have assessed the stability of our estimates. Since our estimates are directly policy relevant in Switzerland, we specifically investigated the stability of the $R_e(t)$, estimated for the most recent possible time point t, when adding additional days of data (up to 21 additional days; Fig. 2). When more observations become available, one can estimate further values of $R_e(s)$, where $s > t$, but also the estimates for $R_e(t)$

are updated. This means that the $R_e(t)$ estimates initially change with each new day of data, before settling on a long-term, stable value. In the analysis in Fig. 2B, estimates for Sept 1 to April 1 based on data up to April 30 are referred to as 'stable R_e estimates'. Especially during rapid changes in R_e, the initial estimates for $R_e(t)$ occasionally under- or overshoot the long-term trend. However, with our improved 95% confidence intervals (CI), the percentage of the first estimated CI for $R_e(t)$ that is contained within the stable CI is much improved compared to purely EpiEstim-based uncertainty intervals (Fig. 2). This difference is particularly striking during periods of high case incidence (e.g. October 2020), where the EpiEstim uncertainty interval is very narrow.

Detailed data allows more precise analysis: the example of Switzerland. When detailed epidemiological data about individual cases (i.e. line list data) is available, we can increase the precision of our method by relaxing the assumptions that (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 using fixed delay distributions from the literature (for details see Materials and Methods section 4.3). 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 S9; 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 (Fig. S10), yet the difference between R_e point estimates for a particular day can be as big as 20%. The difference between R_e estimates has dwindled since early 2021 as the fraction of cases for which the date of onset of symptoms was collected has been very low.

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. S11). Since we do not have data on the number of cases infected in Switzerland that are "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.

Monitoring R_e during the COVID-19 pandemic. We developed an online dashboard (`https://ibz-shiny.ethz.ch/covid-19-re-international/`) on which we present daily-updated results of this R_e estimation method applied to COVID-19 case data from 170 countries (Fig. 3). For most countries, we include multiple observation sources, such as daily incidence of COVID-19 cases and deaths, and, when available, hospitalisation incidence. We make these estimates available for download, as resource for other researchers and the public alike.

The online app allows for comparison through time within a single country, between multiple observation traces, and 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. We additionally show the Oxford Stringency Index and vaccination coverage for context [29, 30].

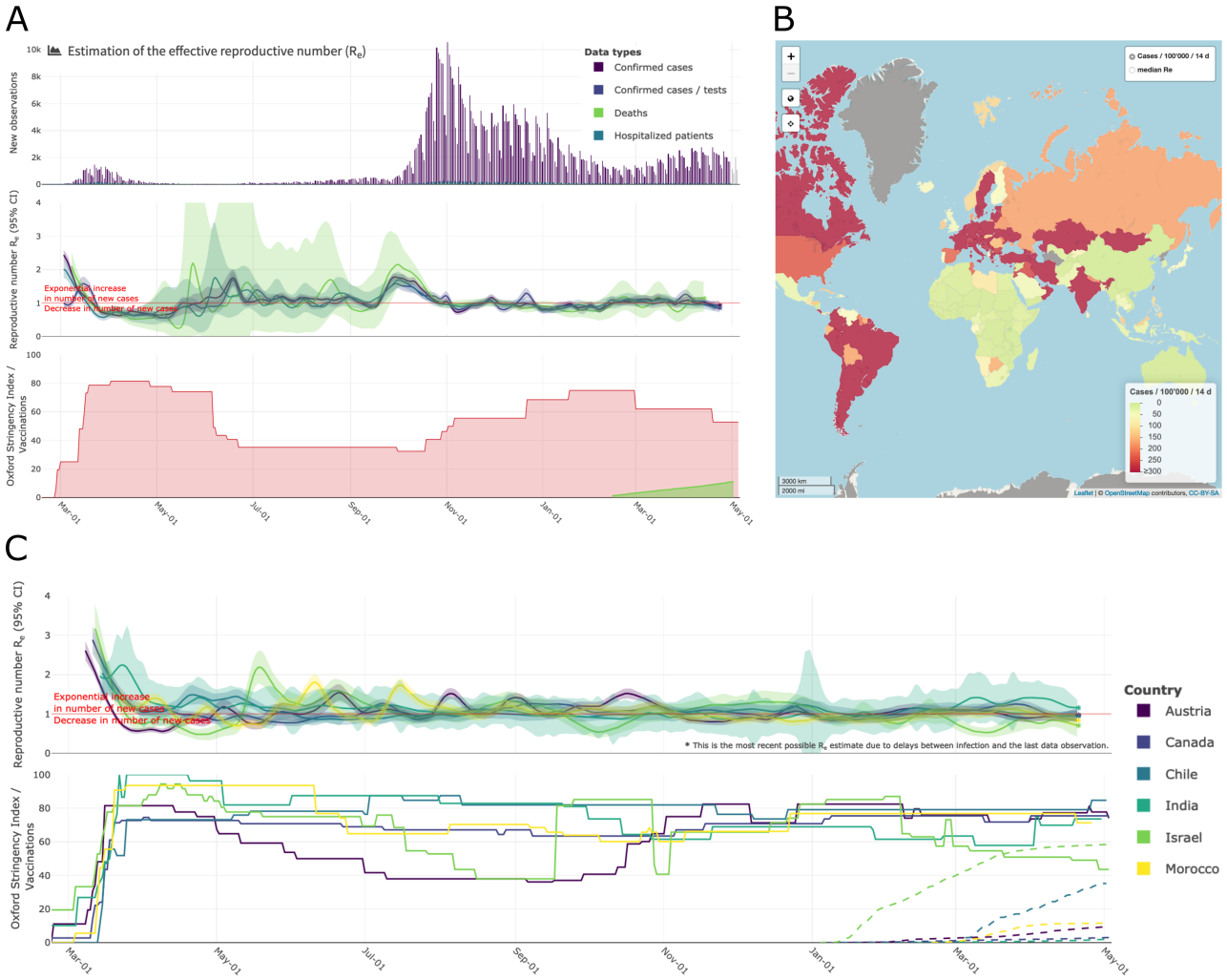


Figure 3: Example panels from the online dashboard. (A) Swiss case incidence with evidence of weekly testing patterns (top row), R_e estimates from four types of observation data (middle row), and timeline of stringency index and vaccination coverage (bottom row). **(B)** World map of incidence per 100'000 inhabitants over the last 14 days. One can also display the worldwide R_e estimates instead. **(C)** Comparison of R_e estimates across a handful of countries, with timelines of stringency indices and vaccination coverage. All panels were extracted on May 5. 2021. Dashboard url: `https://ibz-shiny.ethz.ch/covid-19-re-international`.

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 estimated effective reproductive number R_e. We selected 20 European countries for which the reported data was free of major gaps or spikes, and for which we could estimate R_e prior to the nationwide implementation of a lockdown in spring 2020. The dates of interventions were extracted from news reports (sources listed in Supplementary Table S3), 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). Using case data, we inferred that R_e was significantly above one prior to the lockdown measures in nearly all countries with a lockdown (15/19; Table 1). Denmark, which had a complex outbreak consisting of two initial waves, and Germany, which experienced a cluster of early cases, had an estimated R_e significantly below one prior to this date. For countries with very short delays between the lockdown and the estimated date that $R_e < 1$ (e.g. Austria, Switzerland) we can not exclude the possibility that the 'true' R_e may have been below 1 prior to the lockdown since our pipeline introduces smoothing to the estimates (see Supplementary Discussion 7.2). The results are remarkably consistent across the different observation types (Supplementary Table S2). However, the 95% confidence intervals tend to be wider for the

estimates based on death incidence data because the number of deaths is much smaller than the number of cases, and the relative noise in observations tends to be higher.

Table 1: Investigating the relation between the date of 'lockdown' and the date that the R_e estimated from case reports dropped below 1. Based on news reports, we report when a country implemented stay-at-home orders (a 'lockdown'). The column '$\hat{R}_e < 1$' indicates when the R_e point estimate first dropped below 1. The column 'CI includes 1' details the corresponding time interval where the 95% confidence interval included 1. Of the investigated countries that implemented a nationwide lockdown, four (Denmark, Germany, the Netherlands, Slovenia) had 95% confidence intervals that included 1 or were below before a nationwide lockdown was implemented. The column 'Time until $\hat{R}_e < 1$' indicates the number of days between the lockdown and the date that the R_e point estimate dropped below 1.

Country	Lockdown	$\hat{R}_e < 1$	CI includes 1	Time until $\hat{R}_e < 1$
Austria	16-03	20-03	[20-03, 20-03]	4 days
Belgium	18-03	30-03	[25-03, 03-04]	12 days
Denmark	18-03	**$\leq$10-03**	**[$\leq$10-03, 20-06]**	-8 days
Finland	16-03	02-04	[29-03, 30-04]	17 days
France	17-03	27-03	[23-03, 07-04]	10 days
Germany	22-03	**18-03**	**[17-03, 19-03]**	-4 days
Ireland	27-03	08-04	[04-04, 15-04]	12 days
Italy	10-03	18-03	[17-03, 19-03]	8 days
Netherlands	23-03	05-04	**[22-03, 10-04]**	13 days
Norway	14-03	21-03	[17-03, 19-03]	7 days
Poland	25-03	02-04	[31-03, 17-04]	8 days
Portugal	16-03	28-03	[23-03, 15-04]	12 days
Romania	24-03	06-04	[31-03, 29-04]	13 days
Russian Federation	30-03	04-05	[01-05, 08-05]	35 days
Slovenia	20-03	23-03	**[$\leq$13-03, 26-03]**	3 days
Spain	14-03	26-03	[25-03, 26-03]	12 days
Sweden		**01-04**	**[06-03, $\geq$03-05-2021]**	
Switzerland	17-03	22-03	[20-03, 22-03]	5 days
Turkey	21-03	08-04	[01-04, 13-04]	18 days
United Kingdom	24-03	30-03	[28-03, 20-04]	6 days

To consider the association between NPIs and the estimated R_e for countries outside of Europe, we used the stringency index (SI) of the Blavatnik School of Government [31] to describe the public health response in different countries (Fig. 3C). 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 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 investigated whether the estimated R_e was significantly above 1 prior to the reference date (i.e. the lower bound of the 95% confidence interval was above 1), where we excluded countries without R_e estimates before the reference date t_{SI50}. We found that this was the case for 35 out of the 42 countries world-wide which fulfilled the criteria for inclusion (list in Supplementary Section 8.3). As an additional analysis we calculated, for each day, 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 38/45 countries significantly above one before t_{max} (Supplementary Section 8.3).

Insights into continent-specific responses to NPIs. To investigate the association between changes in the stringency of measures and changes in R_e in more depth, we extended our analysis beyond the first wave and included both the implementation and lifting of NPIs (increases and decreases in stringency; data until May 3rd 2021). For each week and country, we determined the change in

stringency index over the past week ($\Delta SI_{-7} = SI(t) - SI(t-7)$) and the change in the estimated R_e the following week ($\Delta \hat{R}_{e,+7} = \hat{R}_e(t+7) - \hat{R}_e(t)$). If NPIs are working as expected, increases in stringency should be associated with a decrease in the estimated R_e and vice versa. We do find this for increases in stringency e.g. in Europe, but decreases in stringency have a more varied effect on R_e estimates on all continents (Fig. 4).

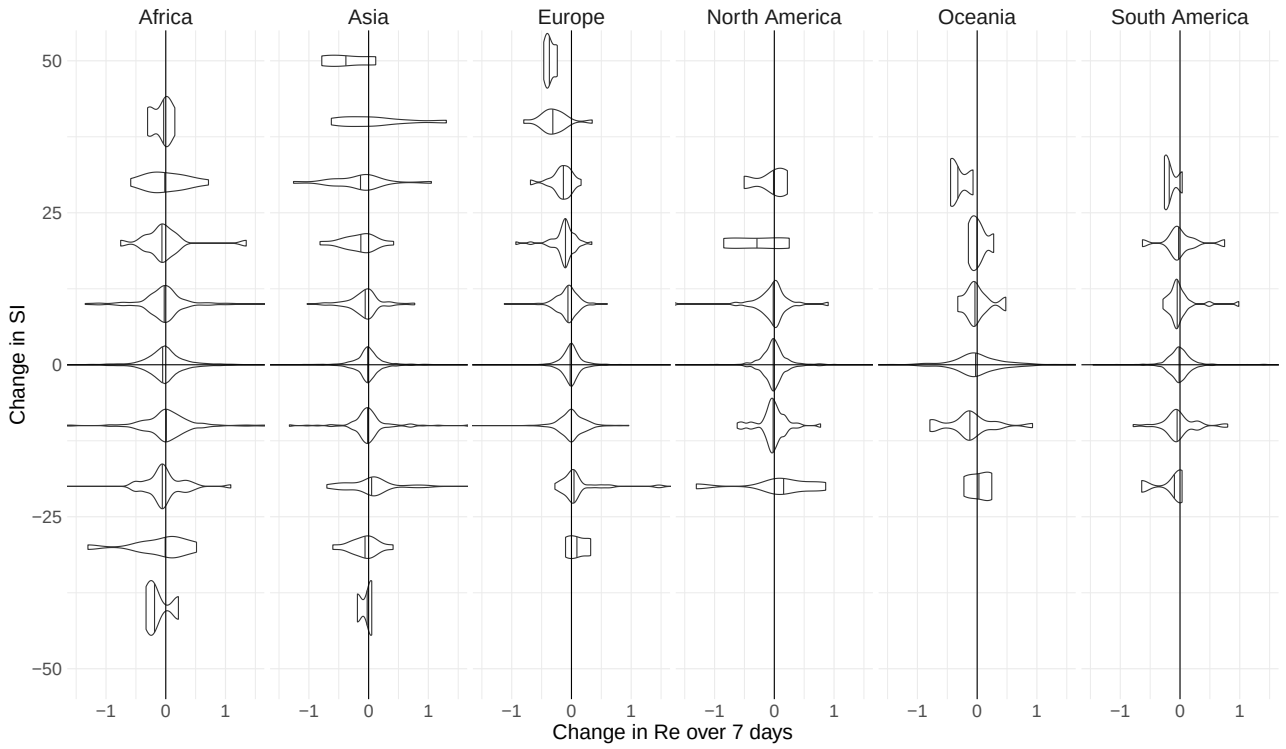


Figure 4: The association between the implementation or lifting of non-pharmaceutical interventions and changes in R_e. Violin plot of the estimated 7-day change in R_e following the implementation (above x-axis) or lifting (below x-axis) of NPIs in a given week. If NPIs are working as expected, we would expect an increase in stringency (i.e. above the x-axis) to be more associated with a decrease in R_e (shifted to the left) and a decrease in stringency (below the x-axis) to be associated with an increase in R_e (shifted to the right).

3 Discussion

We have developed a pipeline to estimate the effective reproductive number R_e of SARS-CoV-2 in near real-time, and evaluated our estimates using simulations. We showed that the inferred R_e curve can be over-smoothed on simulated data, but that this disadvantage is outweighed by the increased stability of the estimates. Overall, we show that the relative error in the R_e estimates is small.

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 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 updated results on an online dashboard, accessible at `https://ibz-shiny.ethz.ch/covid-19-re-international/`. 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 all countries, we further display a timeline of the stringency index of the Blavatnik School of Government [29], and current vaccination coverage.

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 [32]. We can aid this decision making process by investigating the association between adjustments of public-health measures and the estimated R_e. In particular, the merits of nation-wide lockdowns have been heavily discussed, both in the scientific literature and the public sphere [8, 25, 24, 14, 33]. 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 [33]. We showed that this argumentation cannot be applied universally: for 15 out of 20 European countries, we found that the estimated R_e was significantly above 1 prior to the lockdown in spring of 2020. Interestingly, the result we obtain for Germany critically depends on whether we use symptom onset data, or more widely available case reports.

Extending our analysis beyond the first wave, we find differences between continents in the correlation of changes in the stringency of NPIs and changes in R_e. This could reflect differences in the speed with which lockdowns were put into practice [26], the de facto lockdown stringency, or sociocultural aspects [32, 34]. 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 [34]. However, for continents where we find no significant correlation, this could also be because a large fraction of NPIs were implemented at a time for which we could not estimate changes in R_e. Many African countries had 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 [34].

Importantly, our analysis suggests that reversing non-pharmaceutical interventions may have 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 economic, social, and psychological factors motivating the implementation or release of measures. With the current stringency measures we cannot account 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 and on properties of the virus [35]. In the future, our tools to estimate R_e could be used to explore associations of these many factors with R_e estimates, with the aim of identifying minimal sets of factors that may ensure an $R_e < 1$ for a particular location.

Acknowledgements

We thank the Federal Office of Public Health Switzerland for access to their line list data, and Jūlija Pečerska for help in finding governmental datasets on COVID-19. We thank members of the modelling group of the Swiss National Covid-19 science task force for helpful discussions, and Markus Petermann and Daniel Wyler for their comments and suggestions. This work was supported by the Swiss National Science Foundation (SNSF) through grant number 31CA30_196267 (to TS), 200021_172603 (to MHM), 310030B_176401 (to SB), and NRP72 grant 407240-167121 (to SB and TS).

Author contributions

J.S.H, J.S, M.H.M, S.B, T.S designed research; J.S.H, J.S, D.C.A, J.L, R.A.N, M.H.M contributed new reagents or analytic tools; J.S.H, J.S, J.L 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 Overview of the pipeline

The pipeline allows the estimation of R_e from different proxies for the infection incidence, such as the time series of confirmed cases, hospitalisations or deaths. In a first step, we smooth the case observations and deconvolve by the type-specific delay distribution to obtain an estimate of the infection incidence time series. Second, we used the package EpiEstim to estimate the effective reproductive number Re from this infection incidence. We assess the uncertainty in the estimates using the union of a block bootstrap method, designed to account for variation in the case observations, and the credible intervals from EpiEstim.

4.2 Smoothing 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 (LOESS) with 1st order polynomials and tricubic weights. The smoothing parameter alpha is set such that we include 21 days of data in the local neighbourhood of each point. After smoothing, we normalise to the original total number of cases.

4.3 Estimating the infection incidence through deconvolution

To recover the non-observed time series of infection incidence, we deconvolve the smoothed observed time series of COVID-19 case incidence with a delay distribution specific to the type of case detection (case confirmation, hospital admission, death). To this end we extended the deconvolution method of Goldstein *et al.* [15], which is itself an adaptation of the Richardson-Lucy algorithm [36, 37], to deal with zero-incidence case observations and time-varying delay distributions.

Formally, the method infers a deconvolved output time series $(\lambda_1, \ldots, \lambda_N)$ from an input time series $(\bar{D}_K, \ldots, \bar{D}_N)$, where $K \geq 1$ and $\bar{D}_i$ indicates the smoothed number of observations on day i (e.g. confirmed cases, hospitalisations, or deaths). Let m_l^j be the probability that an infection on day j takes $l \geq 0$ days to be observed. 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, i.e. is counted towards $(\bar{D}_K, \ldots, \bar{D}_N)$. Then:

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

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

$$E_i = \begin{cases} \sum_{j=1}^{i} \lambda_j \, m_{i-j}^j & \text{for} \quad K \geq i \geq N \\ 0 & \text{for} \quad 0 < i < K \,. \end{cases} \tag{2}$$

The Richardson-Lucy algorithm uses expectation maximisation [38] to find a final infection incidence estimate, which has the highest likelihood of explaining the observed input time series. To do so, it starts from an initial guess of the infection incidence time series $\Lambda^0 = (\lambda_1^0, \ldots, \lambda_N^0)$, used to compute

E_i^0 according to equation 2, and updates the estimate in each iteration n according to the following formula:

$$\lambda_j^{n+1} = \frac{\lambda_j^n}{q_j} \cdot \sum_{i=K}^{N} \frac{m_{i-j}^j \bar{D}_i}{E_i^n} \,. \tag{3}$$

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 [15]:

$$\chi^2 = \frac{1}{N-K+1} \sum_{i=K}^{N} \frac{(E_i^n - \bar{D}_i)^2}{E_i^n} \,, \tag{4}$$

or 100 iterations have been reached.

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 is the case for the vast majority of the empirical data we analyzed. In some cases, e.g. when the observed incidence is especially noisy, convergence is slower and the threshold of 100 iterations is reached. In 22 of the 170 countries analyzed, the iteration threshold was reached for at least 5 bootstrap replicates (out of 100).

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 [15]. 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 ($\bar{D}_K$) on the left, and with the last observed value ($\bar{D}_N$) 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 (3)), which would be a potential source of bias.

We note that the Richardson-Lucy deconvolution algorithm accounts for 'right truncation', i.e. that not all infections 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 information on the time variation of delays between symptom onset and observation is available (e.g. through a line list), this can be taken into account directly during the deconvolution step. In this case, we 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 distribution for the reporting delay (m_l^j) after a certain day j, or the distributions would be downward biased for infection dates close to the present. Let $(\bar{m}_0, \ldots, \bar{m}_K)$ 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 $\sum_{i=1}^{n} \bar{m}_i \geq 0.99$). For infection dates z that are closer to the present than n (i.e. $N - z < n$, where N is the index of the last available data point), we fix $(m_0^z, \ldots, m_K^z)$ to be equal to $(m_0^{N-n}, \ldots, m_K^{N-n})$.

4.4 Estimating the effective reproductive number R_e

Once we have obtained an estimate for the time series of infection incidence, we use the method developed by Cori *et al.* [12], implemented in the EpiEstim R package, to estimate R_e.

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 (discretised) serial interval distribution [12]. The likelihood of the incidence I_t at time t is thus given by:

$$P(I_t|I_0,\ldots,I_{t-1},R_e(t)) = \frac{(R_e(t)\Lambda_t)^{I_t}\, e^{-R_e(t)\Lambda_t}}{I_t!}\,, \tag{5}$$

$$\text{where} \qquad \Lambda_t = \sum_{s=1}^{t} I_{t-s} w_s\,. \tag{6}$$

The R_e 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. This is a conservative assumption, which means the R_e estimates will tend to 1 during periods of low case incidence (i.e. when the posterior is dominated by the prior).

For the smooth R_e estimates, we assume R_e is constant over a sliding window of 3 days ($\tau = 3$ in EpiEstim), i.e. the reported R_e estimate for day T summarises the average R_e over a 3-day period ending on day T. In addition, we provide step-wise estimates of R_e on our dashboard. 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 determined by dates at which public health interventions were implemented, altered, or lifted. All results in this paper are based on the smooth R_e estimate. In both cases, we use the mean of the posterior distribution of R_e as the point estimate.

4.5 Estimating the uncertainty intervals

To account for the uncertainty in the case observations, we construct 95% bootstrap confidence intervals for R_e. We first re-sample case observations as follows: given the original case observations $D_t, t = K,\ldots,N$, we apply the LOESS with smoothing parameter 21 days on the log-transformed data $log(D_t+1)$ to obtain the smoothed value $\hat{\mu}_t$ and additive residuals e_t. Then we re-sample the residuals to get e_t^*. We obtain the bootstrap case observations by

$$D_t^* = \max(\exp(\hat{\mu}_t + e_t^*) - 1, 0). \tag{7}$$

We use overlapping block bootstrap in re-sampling the residuals to account for the time series nature of the data. Specifically, given the original residuals $(e_K,\ldots,e_N)$, we first sample a block $(e_1^{*1},\ldots,e_b^{*1})$ with block length $b = 10$. To account for weekly patterns in the case observations, we make sure that the sampled block starts on the same day of the week as the original case observations D_K (e.g. Tuesday). That is, we keep the longest possible $(e_m^{*1},\ldots,e_b^{*1})$ such that e_m^{*1} has the same day of the week as D_K. Then, we sample a new block $(e_1^{*2},\ldots,e_b^{*2})$ and keep the longest possible $(e_m^{*2},\ldots,e_b^{*2})$ such that the weekday of e_m^{*2} follows on e_b^{*1} (i.e. has the next day of the week). We glue these two sampled blocks together to get the temporal re-sampled residuals $(e_m^{*1},\ldots,e_b^{*1},e_m^{*2},\ldots,e_b^{*2})$. We repeat this process of adding blocks until the length of the re-sampled residuals is equal to or larger than the original residuals. In the latter case, we cut the last part of the re-sampled residuals to make sure its length is the same as the original residuals.

Given the bootstrap case observations, we apply our method to obtain an estimate for $R_e(t)$, and denote it by $\hat{\theta}^*(t)$. By repeating the above steps 100 times, we obtain $\hat{\theta}_1^*(t),\ldots,\hat{\theta}_{100}^*(t)$. Then, we construct a Normal based bootstrap confidence interval for each time point t by:

$$[\hat{\theta}(t) - q_z(1-\frac{\alpha}{2})\widehat{sd}(\hat{\theta}^*(t)),\ \hat{\theta}(t) + q_z(1-\frac{\alpha}{2})\widehat{sd}(\hat{\theta}^*(t))], \tag{8}$$

where $\hat{\theta}(t)$ denotes the estimated $R_e(t)$ based on the original case observations, $q_z(1-\frac{\alpha}{2})$ denotes the $1-\frac{\alpha}{2}$ quantile of the standard normal distribution, and $\widehat{sd}(\hat{\theta}^*)$ denote the empirical standard

deviation of $\hat{\theta}^*_1(t), \ldots, \hat{\theta}^*_{100}(t)$. In this paper, we aim at confidence interval level 95%, so $\alpha = 0.05$ is used.

An implicit assumption for the above bootstrap confidence interval to be reasonable, is that the variance of the residuals e_t is a constant over time t and does not depend on the value of the log-transformed data $log(D_t + 1)$. This assumption roughly holds when the case incidence is high. During periods of low case incidence (e.g. deaths or regional data in summer 2020 in Switzerland), however, this assumption is no longer appropriate. Therefore, to be conservative and rather err on the side of too large uncertainty intervals, we also consider the credible interval of R_e which is obtained by taking the 0.025 and 0.975 quantiles from the posterior distribution of R_e using EpiEstim based on the original data D_t. The final reported interval is then the union of the credible interval and the 95 % bootstrap confidence interval. Based on our experience, the above credible interval will be reported during periods of very low case incidence. But at high case numbers, the bootstrap confidence interval will be much wider than the credible interval and so will be reported.

4.6 Data

We gather case incidence data directly from public health authorities. Whenever accessible, we rely on data from local authorities. Otherwise, we use data from 'Our World in Data' since the European Centre for Disease Control (ECDC) has stopped its daily updates (December 2020) [30, 39]. 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, are obtained from media reports and the websites of public health authorities. The stringency index of the Blavatnik School of Government is accessed from their publicly available github repository [31]. The vaccination coverage is taken from 'Our World in Data' [30].

We parametrise the discretised infectivity profile w_s using COVID-19 serial interval estimates from the literature [40]. For a review of published serial interval estimates, see Griffin et al. [41]. The incubation period is parametrised by a gamma distribution with mean 5.3 days and SD 3.2 days [42]. 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, with parameters taken from the literature. Table 2 summarises the distributions used in our pipeline.

Table 2: Gamma distributions used in the pipeline: serial interval, incubation period, and the delay distributions assumed for each observation type.

Distribution	Mean (days)	SD (days)	Reference
Serial interval	4.8	2.3	[40]
Infection to onset of symptoms	5.3	3.2	[42]
Onset of symptoms to case confirmation	5.5	3.8	[43]
Onset of symptoms to hospital admission	5.1	4.2	[44]
Onset of symptoms to death	15.0	6.9	[42]

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.

4.7 Simulations

In the simulations, we first specify a piecewise linear R_e trajectory, and simulate 100 time series of infections and corresponding case observations from it. Then, we estimate R_e from these observations using our pipeline. 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. The infection incidence on day t is drawn from a Poisson distribution, corresponding to equation (6), using the specified R_e time series and the discretised serial interval for SARS-CoV-2 [40] as the infectivity profile (see [12]; Web Appendix 11).

These simulated infections are convolved with the observation type-specific delay distribution [42] to obtain the raw observation time series $\tilde{D}_t$. The final observation time series is generated based on these raw observations and a noise model. To obtain a realistic noise model, we apply the LOESS smoother with smoothing parameter 21 days on the log-transformed confirmed case data from some country (e.g. Switzerland) to obtain additive residuals, and then fit an ARIMA model on these residuals. The final observation time series $D_t = \tilde{D}_t \cdot \exp(e_t)$, where e_t is simulated from the fitted ARIMA model.

In the case of time-varying delay distributions, we assume that 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 5.5 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 further 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 quantify the performance of our method on the simulated scenarios, we compute the root mean squared error (RMSE) at time point j:

$$RMSE(j) = \sqrt{\frac{1}{M}\sum_{m=1}^{M}\left(\hat{R}_e(j,m) - R_e(j)\right)^2}, \tag{9}$$

where M is the total number of simulations, $\hat{R}_e(j,m)$ the estimated R_e and $R_e(j)$ the true R_e at time j, for simulation m.

For each simulation we also compute the 95% confidence interval (CI) of our estimates across 100 bootstrap replicates. The empirical coverage indicates the fraction of simulations for which our CI includes the true R_e value.

4.8 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-international/`. The source code of this pipeline is openly accessible at `https://github.com/covid-19-Re/shiny-dailyRe`, and the code necessary to reproduce the figures in this paper is at `https://github.com/covid-19-Re/paper-code`. 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).

5 References

[1] Roy Malcolm Anderson and Robert M May. *Infectious diseases of humans: dynamics and control*. 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. *Mathematical and Statistical Estimation Approaches in Epidemiology*, chapter 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. *PLoS Computational Biology*, 16(12):e1008409, 2020.

[14] Kristian Soltesz, Fredrik Gustafsson, Toomas Timpka, Joakim Jaldén, Carl Jidling, Albin Heimerson, Thomas B Schön, Armin Spreco, Joakim Ekberg, Örjan Dahlström, et al. The effect of interventions on COVID-19. *Nature*, 588(7839):E26–E28, 2020.

[15] 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.

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

[17] 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. *Wellcome Open Research*, 5(112), 2020.

[18] Kevin Systrom, Thomas Vladek, and Mike Krieger. Rt covid live. `https://github.com/rtcovidlive/covid-model`, 2020.

[19] 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.

[20] 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.

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

[22] 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.

[23] 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. *https://www.medrxiv.org/content/early/2020/08/05/2020.07.30.20164939*, 2020.

[24] Nicolas Banholzer, Eva Van Weenen, Adrian Lison, Alberto Cenedese, Arne Seeliger, Bernhard Kratzwald, Daniel Tschernutter, Joan Puig Salles, Pierluigi Bottrighi, Sonja Lehtinen, et al. Estimating the effects of non-pharmaceutical interventions on the number of new infections with COVID-19 during the first epidemic wave. *PLoS one*, 16(6):e0252827, 2021.

[25] Nils Haug, Lukas Geyrhofer, Alessandro Londei, Elma Dervic, Amélie Desvars-Larrive, Vittorio Loreto, Beate Pinior, Stefan Thurner, and Peter Klimek. Ranking the effectiveness of worldwide COVID-19 government interventions. *Nature Human Behaviour*, 4(12):1303–1312, 2020.

[26] Ilia Kohanovski, Uri Obolski, and Yoav Ram. Inferring the effective start dates of non-pharmaceutical interventions during COVID-19 outbreaks. *https://www.medrxiv.org/content/10.1101/2020.05.24.20092817v2*, 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: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, Noam Angrist, Rafael Goldszmidt, Beatriz Kira, Anna Petherick, Toby Phillips, Samuel Webster, Emily Cameron-Blake, Laura Hallas, Saptarshi Majumdar, and Helen Tatlow. A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nature Human Behaviour*, 5(4):529–538, April 2021.

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

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

[32] 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.

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

[34] 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.

[35] Dylan H Morris, Kwe Claude Yinda, Amandine Gamble, Fernando W Rossine, Qishen Huang, Trenton Bushmaker, Robert J Fischer, M Jeremiah Matson, Neeltje Van Doremalen, Peter J Vikesland, et al. Mechanistic theory predicts the effects of temperature and humidity on inactivation of SARS-CoV-2 and other enveloped viruses. *ELife*, 10:e65902, 2021.

[36] William Hadley Richardson. Bayesian-based iterative method of image restoration. *Journal of the Optical Society of America*, 62(1):55–59, 1972.

[37] Leon B Lucy. An iterative technique for the rectification of observed distributions. *The Astronomical Journal*, 79:745, 1974.

[38] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, 1977.

[39] 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.

[40] 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.

[41] John Griffin, Miriam Casey, Áine Collins, Kevin Hunt, David McEvoy, Andrew Byrne, Conor McAloon, Ann Barber, Elizabeth Ann Lane, and SImon More. Rapid review of available evidence on the serial interval and generation time of COVID-19. *BMJ Open*, 10(11):e040263, 2020.

[42] 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.

[43] 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.

[44] 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:2004.00117*, 2020.

[45] 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*, 2020.

[46] 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.

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

[48] 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.

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

[50] 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.

[51] Menekse Tokyay. *Turkey hits record COVID-19 cases after change in reporting*. Arab News, 2020 (accessed April 26, 2021).

[52] Apple. COVID-19 Mobility Trends Reports. `https://covid19.apple.com/mobility`. Accessed: 2021-05-06.

[53] Google. COVID-19 Community Mobility Reports. `https://www.google.com/covid19/mobility/`. Accessed: 2021-05-06.

[54] Yousef Alimohamadi, Maryam Taghdir, and Mojtaba Sepandi. Estimate of the basic reproduction number for COVID-19: A systematic review and meta-analysis. *Journal of Preventive Medicine and Public Health*, 53(3):151–157, 2020.

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 COVID-19 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 [30]. Changes in testing strategy as well as bottlenecks in testing capacity result in a varying fraction of infected individuals that are confirmed positive, and also variation 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 infectivity profile 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 a number of days given by w_s (compare Materials and Methods, equation 6). For SARS-CoV-2 the time needed to reach the 95% quantile of w_s is 9 days. 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 [15]. However, also here biases can occur. First, only a small fraction of all infections results in hospitalisation or death (a meta-analysis found an average infection fatality ratio for SARS-CoV-2 of 0.68% [45]). This fraction varies with the risk group of the infected population [45, 46, 47, 48], introducing potential biases in R_e estimations when outbreaks occur in particularly age-stratified settings. Also, new variants may result in different hospitalisation or fatality rates. 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 or if hospitals are overburdened. Additionally, guidelines used to record COVID-19 as the cause of death have changed through time for some countries [49]. 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 populations with highly different mortality rates are captured (e.g. elderly homes).

It is important to note that all analyses here are focused on the period before vaccination mediated immunity became widespread. Since vaccinations change the fraction of infections that eventually become hospitalised or die, they may introduce temporary biases for the R_e estimated from hospitalisation and death incidence. We have added the metric of vaccination coverage to the online

dashboard, so one can estimate when these effects start to become important.

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 [50].

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 serial interval will lead to larger errors in R_e estimates 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 number 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 data consolidation (i.e. to which extent values reported for a given day change in subsequent days), and what constitutes a COVID-19 case [49, 51]. 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 [49]. This affects the timeliness of our estimates, and can cause the estimated R_e to change a bit between days.

7 Supplementary Methods

7.1 Discretisation of delay distributions

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

$$m_l = \begin{cases} \int_0^{0.5} f(x)\,\mathrm{d}x & l = 0 \\ \int_{l-0.5}^{l+0.5} f(x)\,\mathrm{d}x & l \in \{1, 2, \dots\}, \end{cases} \tag{10}$$

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 the 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, 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.

7.2 The effect of smoothing on our ability to infer when $R_e = 1$

Our LOESS smoothing roughly spreads sudden changes in R_e over 20 days in the estimated R_e. Whether this is a substantial problem depends on the smoothness of the true R_e that we are trying to estimate. Direct observations of behavioural changes, specifically changes in mobility, suggest the true R_e is quite smooth: for instance it took 2-3 weeks for mobility to drop to its lowest level in response to government interventions in Switzerland [52, 53].

However, to get a rough feeling for the impact smoothing would have on our estimates and downstream analysis in case the true R_e does change abruptly, we can use a simple analysis using a linear approximation. In the case of a step-wise change from R_0 to R_1 (with $R_0 > R_1$) at time t_0, the estimated smooth R_e will start decreasing about 10 days prior to t_0, and take another 10 days after to reach the terminal value (Fig. S1). When inferring the day that a certain threshold value was reached (e.g. $R_e = 1$) we will be off by a number of days s, dictated by R_0 and R_1. Specifically, the delay s is greater if the turning point $R_{tp} = \frac{R_1+R_0}{2}$ is further above 1, or the slope $a_{tp} = \frac{R_1-R_0}{20}$ is closer to 0:

$$s = (R_{tp} - 1)\frac{20}{R_0 - R_1}. \tag{11}$$

In Table S1 we have listed some possible delay values, using R_0 values spanning the range of values reported for SARS-CoV-2 [54]. The delay is positive if $R_0 > R_1$ and $R_{tp} > 1$, which was the case for most countries around the 1st lockdown. In general, these numbers can be considered a 'worst-case' scenario: when the true underlying R_e changes more gradually than considered here, the smoothing introduced by our pipeline will have a smaller effect.

Note that these calculations specifically refer to the point estimate. The estimates may stop being significantly above the threshold already earlier, especially when the confidence interval is wide and the slope is close to 0.

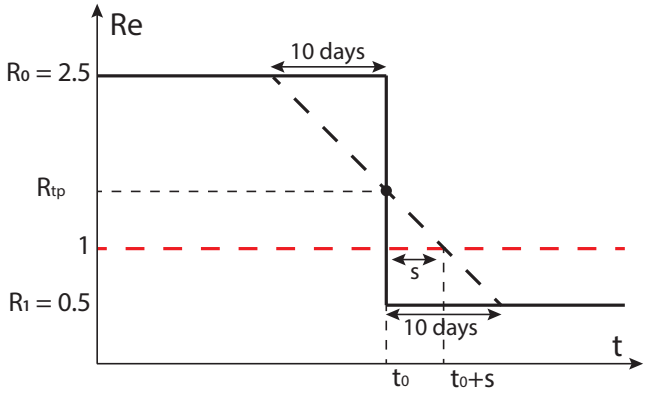


Figure S1: Schematic of the effect of smoothing on the ability to estimate when $R_e = 1$. The true R_e is indicated by the black solid line, the black dashed line shows a linear approximation of the smoothed R_e. Instead of crossing 1 at t_0, this line crosses 1 at $t_0 + s$.

Table S1: The effect of smoothing on the ability to estimate when $R_e = 1$. These values were calculated using equation 11.

R_0	R_1	R_{tp}	a_{tp} (per day)	Delay s (days)
6.0	0.0	3.0	-0.30	6.7
3.0	0.0	1.5	-0.15	3.3
3.5	0.5	2.0	-0.15	6.7
2.5	0.5	1.5	-0.10	5.0
3.3	0.9	2.1	-0.12	9.2
1.8	0.8	1.3	-0.05	6.0

8 Supplementary Materials

8.1 Supplementary Simulations

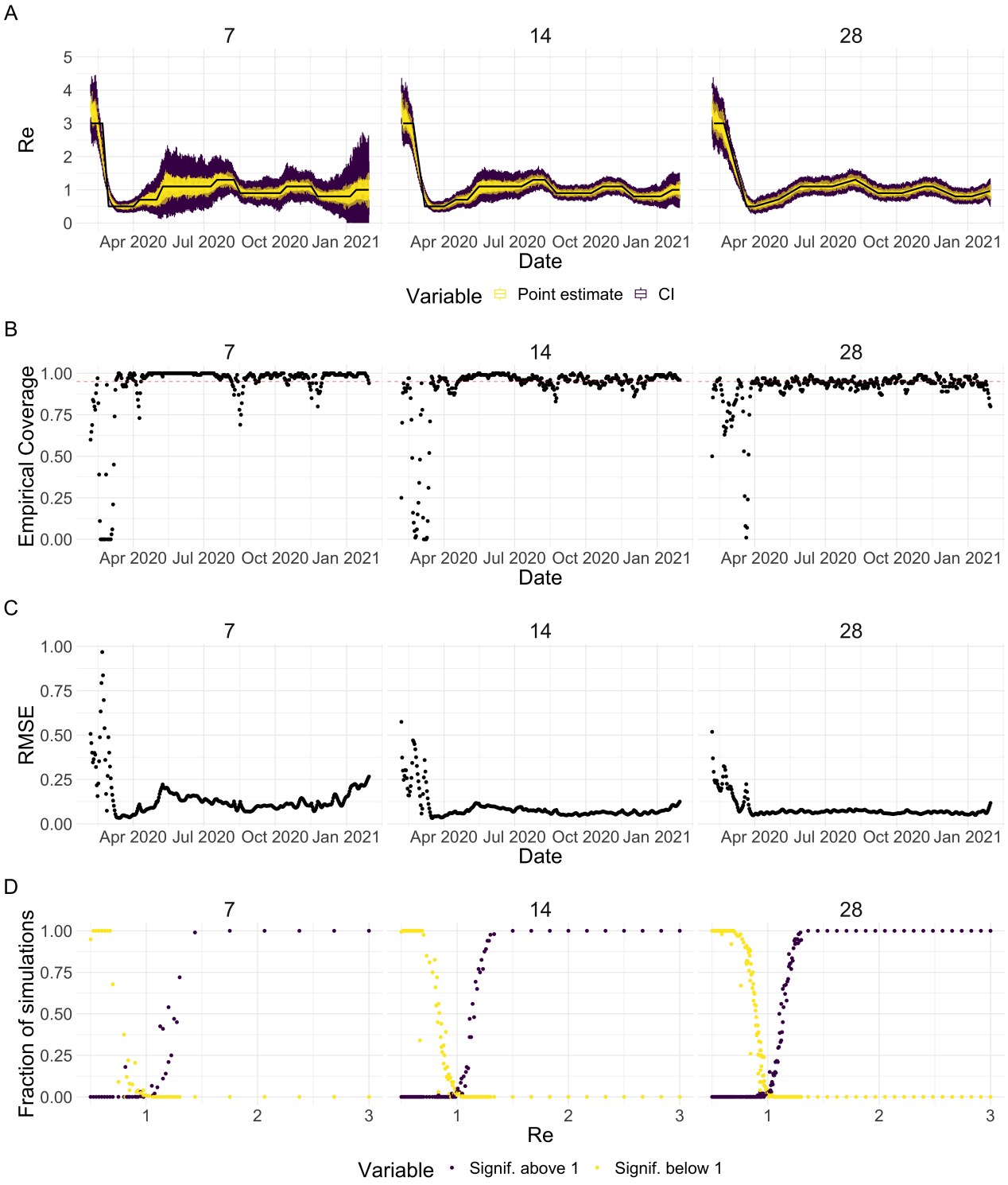


Figure S2: Performance of our method on simulated scenarios with differing slopes. **A** The specified R_e trajectory (black line; see Methods) was used to simulate a trajectory of reported cases (with Swiss case observation noise) 100 times. From each trajectory we estimated R_e (yellow boxplots), and constructed a 95% confidence interval (purple boxplots of the lower/upper endpoint). We varied the time it took to change from one R_e value to the next, $t \in \{7, 14, 28\}$ (columns). Larger values of t correspond to less abrupt changes. **B** The fraction of simulations where the true R_e value was within the 95% confidence interval. The dashed red line indicates the nominal 95% coverage. **C** The root mean squared relative error for every time point. **D** The fraction of simulations where we estimate R_e is significantly above or below one, depending on the true value of R_e. We see that the method closely tracks the true R_e in all scenarios, although the error is greater for steeper slopes. In the case of steeper changes in R_e the overall size of the epidemic is also smaller, which explains the larger confidence intervals.

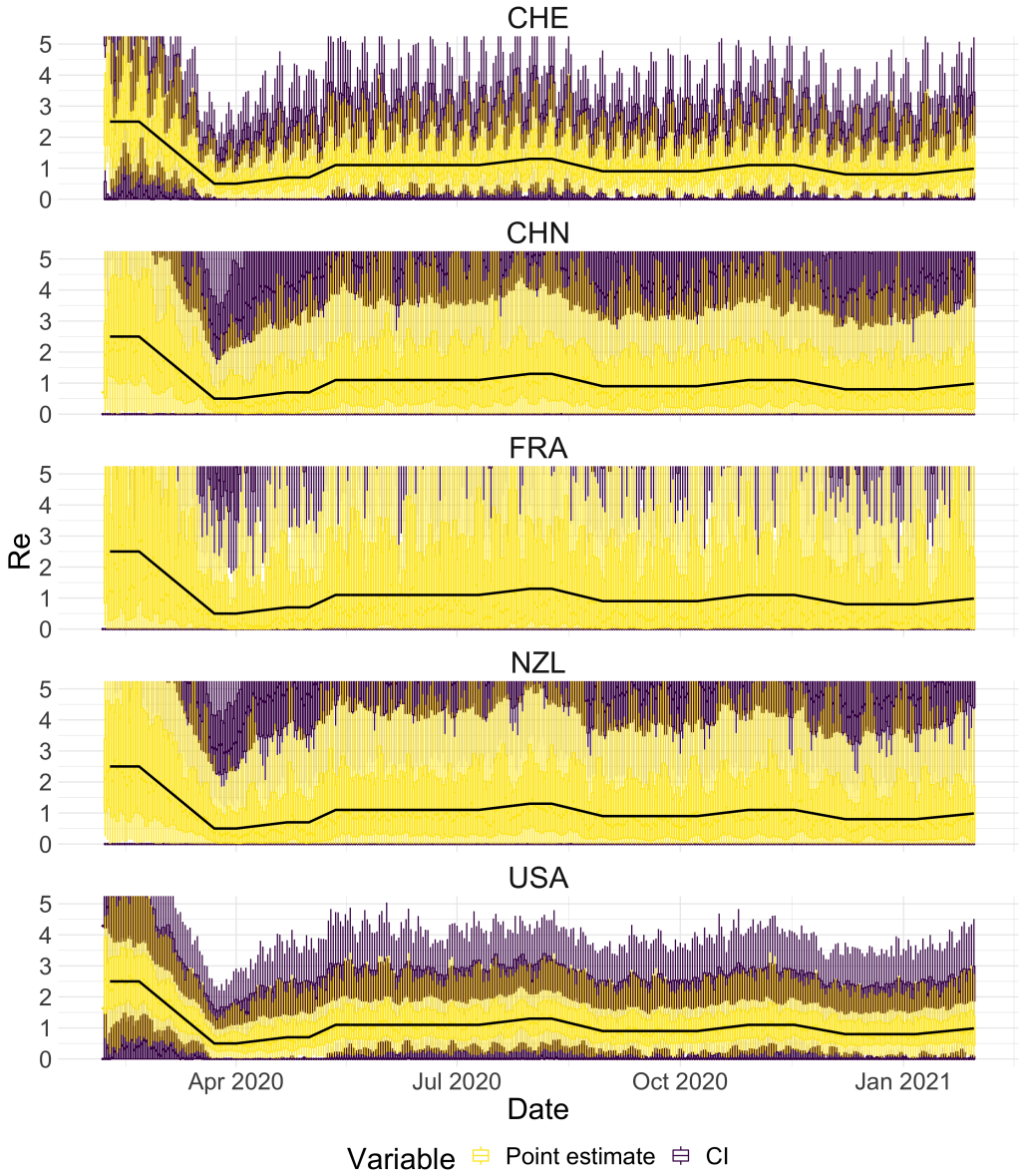


Figure S3: Performance of our method, modified to skip the smoothing step in the pipeline, on simulated scenarios with observation noise. The specified R_e trajectory (black line; see Methods) was used to simulate a trajectory of reported cases (with varying country-specific noise profiles; rows) 100 times. From each trajectory we estimated R_e (yellow boxplots), and constructed a 95% confidence interval (purple boxplots of the lower/upper endpoint). Contrary to our normal pipeline, the observations were not smoothed prior to the deconvolution and R_e estimation. We see that the estimates are highly variable.

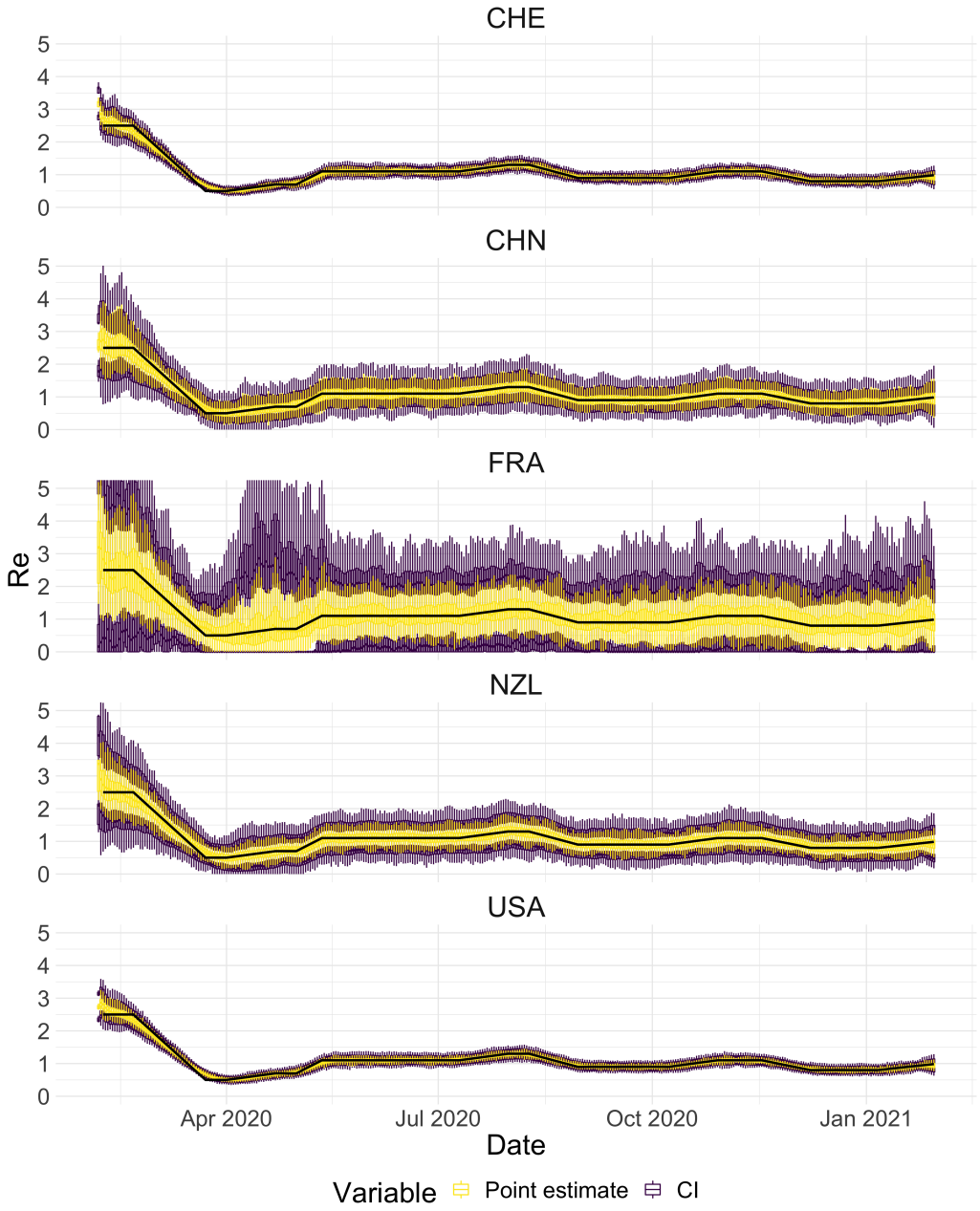
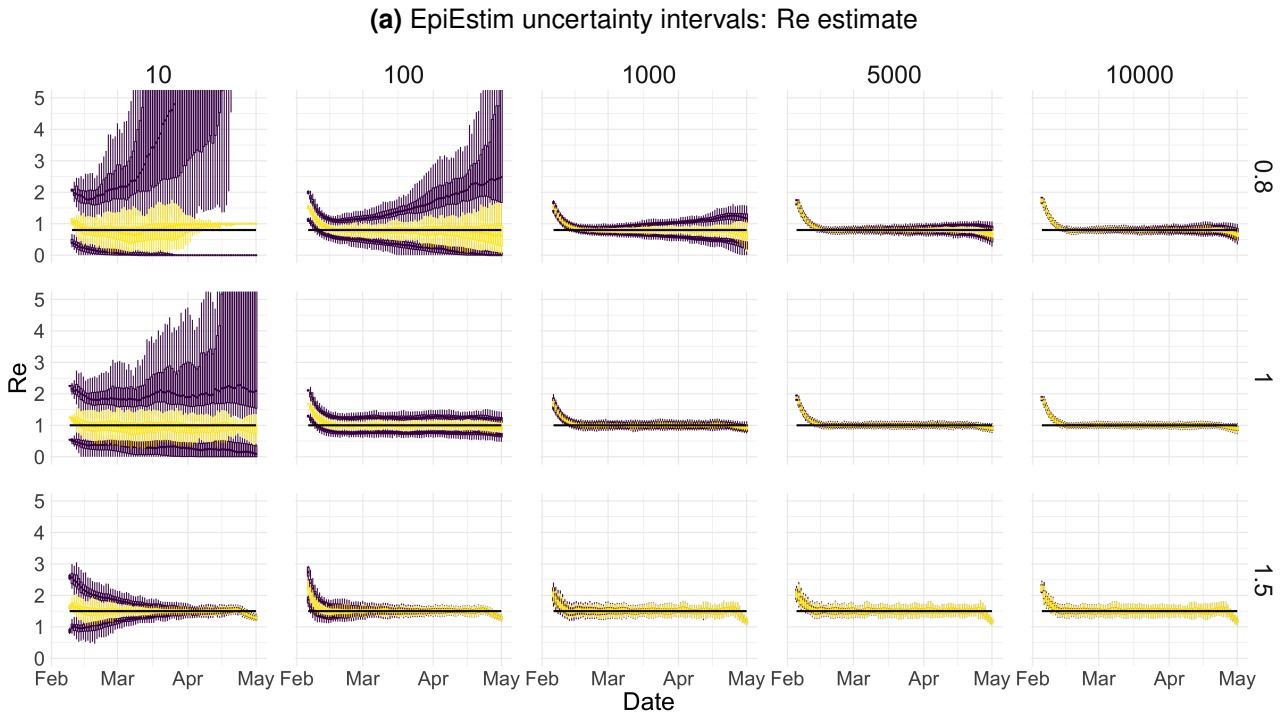
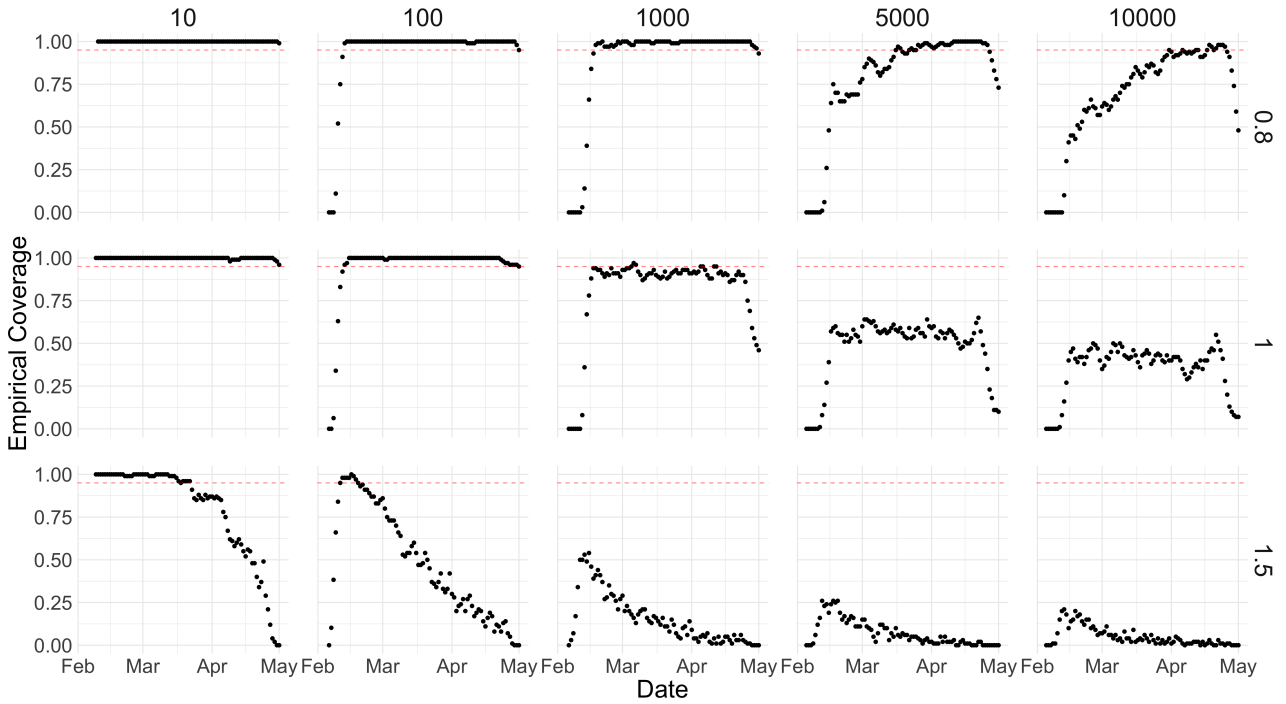


Figure S4: Performance of our method on simulated scenarios with observation noise. The specified R_e trajectory (black line; see Methods) was used to simulate a trajectory of reported cases (with varying country-specific noise profiles; rows) 100 times. From each trajectory we estimated R_e (yellow boxplots), and constructed a 95% confidence interval (purple boxplots of the lower/upper endpoint). Compared to Fig. S3, we see that the estimates are much more stable.





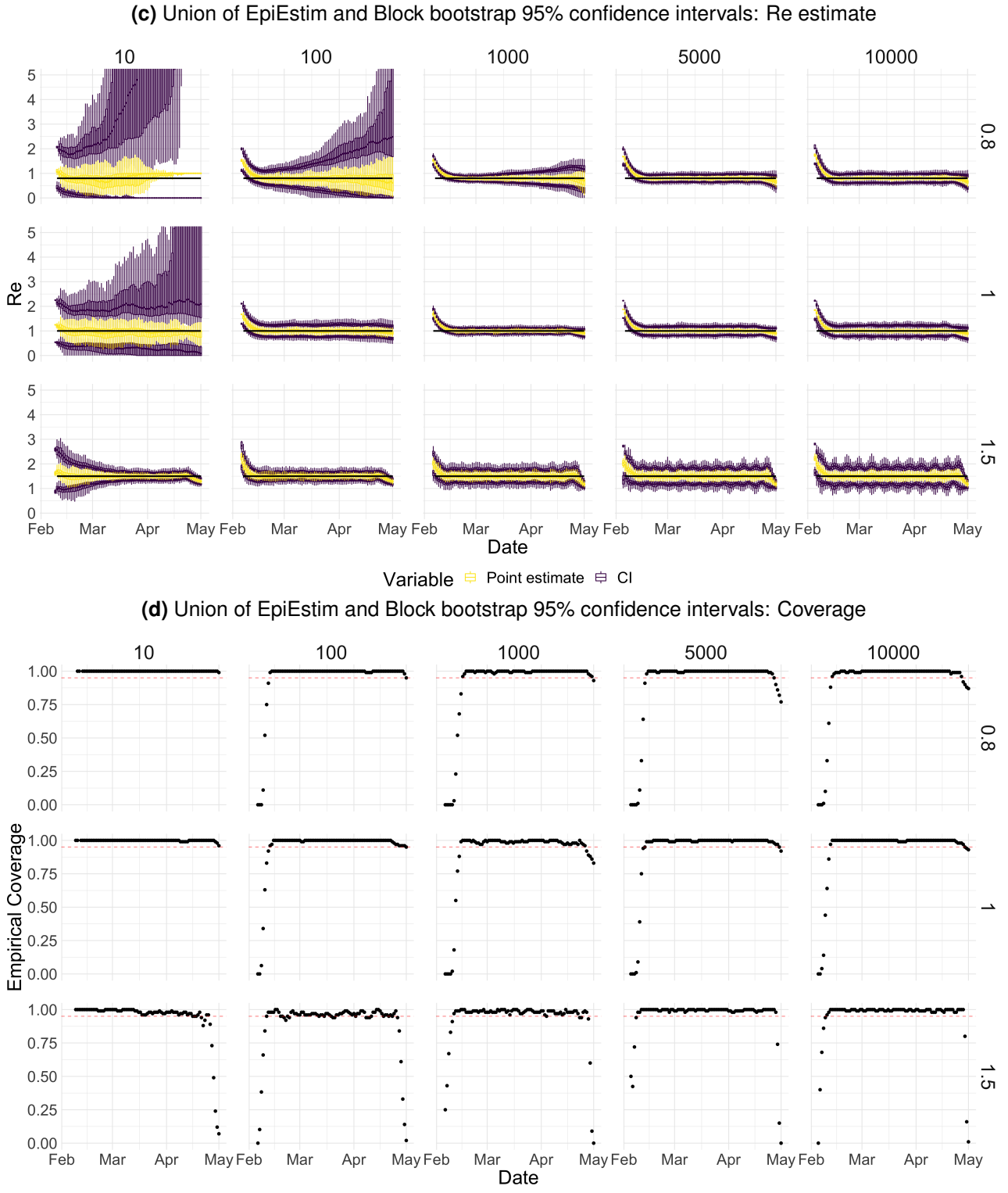


Figure S5: Performance of our method on simulated scenarios with varying population size. **(a), (c)** We specified a constant $R_e \in \{0.8, 1, 1.5\}$ value (black line; rows) to simulate a trajectory of reported cases (with Swiss case observation noise) 100 times. From each trajectory we estimated R_e (yellow boxplots), and constructed a 95% confidence interval (purple boxplots of the lower/upper endpoint). The simulated scenarios had differing initial incidence of $I_0 \in \{10, 100, 1000, 5000, 10000\}$ infections per day (columns). In the top row, $R_e < 1$ so the epidemic is decreasing. In the middle row, $R_e = 1$, the epidemic is constant, and in the bottom row, $R_e > 1$, the epidemic is increasing. The bias at the start is due to the initialisation of the simulation. **(b), (d)** The fraction of simulations where the true R_e value was within the 95% confidence interval. The dashed red line indicates the nominal 95% coverage. We see that for a wide range of infection incidences, our 95% confidence interval is informative and covers the true value of R_e, whereas the EpiEstim coverage strongly declines with increased epidemic size.

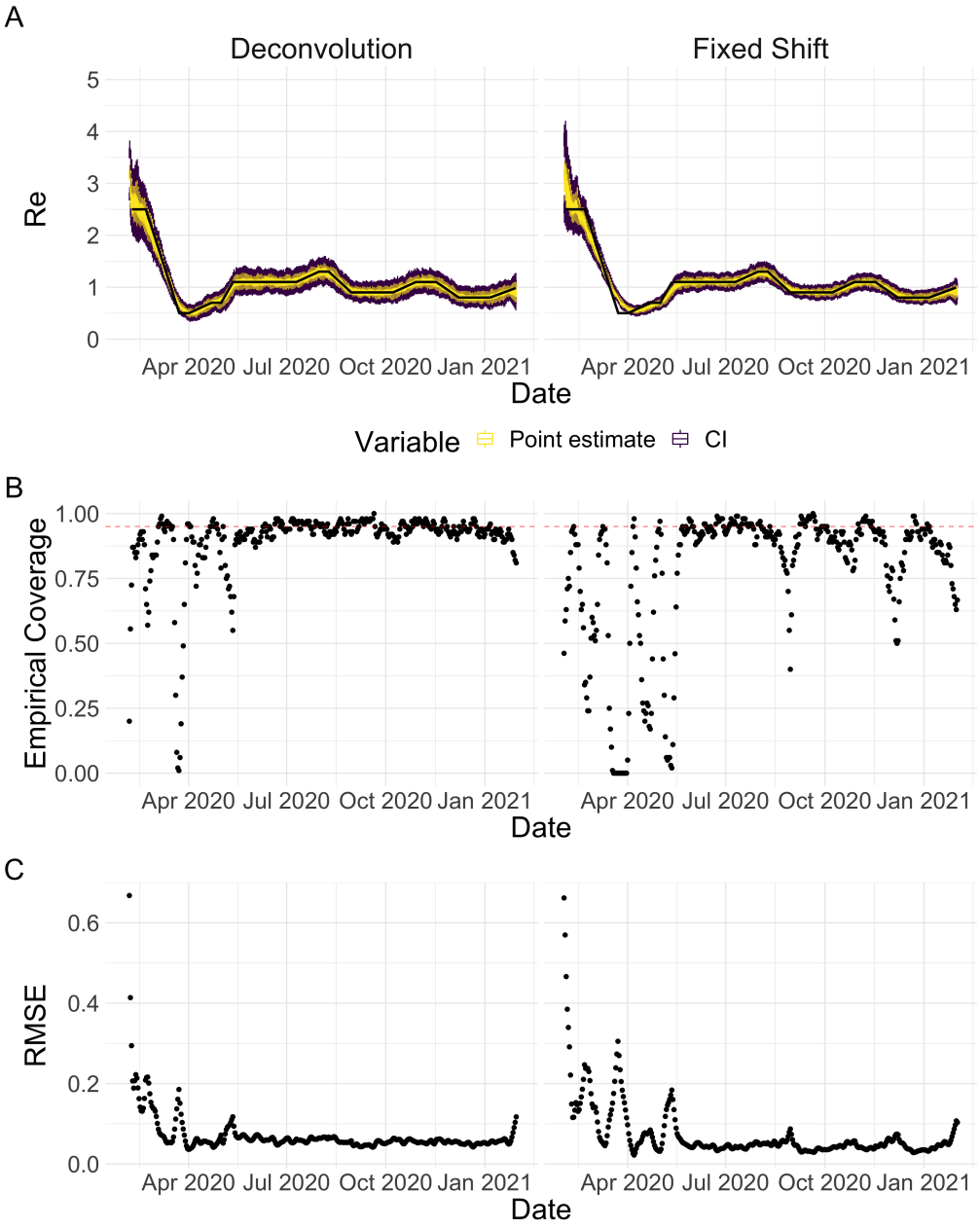


Figure S6: Performance of our method on simulated scenarios using a fixed shift versus the deconvolution to infer infection incidence. The fixed shift method shifts the observations back by the mean of the delay distribution (here assumed to correspond to confirmed cases). **A** The specified R_e trajectory (black line; see Methods) was used to simulate a trajectory of reported cases (with Swiss case observation noise) 100 times. From each trajectory we estimated R_e (yellow boxplots), and constructed a 95% confidence interval (purple boxplots of the lower/upper endpoint). **B** The fraction of simulations where the true R_e value was within the 95% confidence interval. The dashed red line indicates the nominal 95% coverage. **C** The root mean squared relative error for every time point. We see that our method (left) outperforms the fixed-shift method.

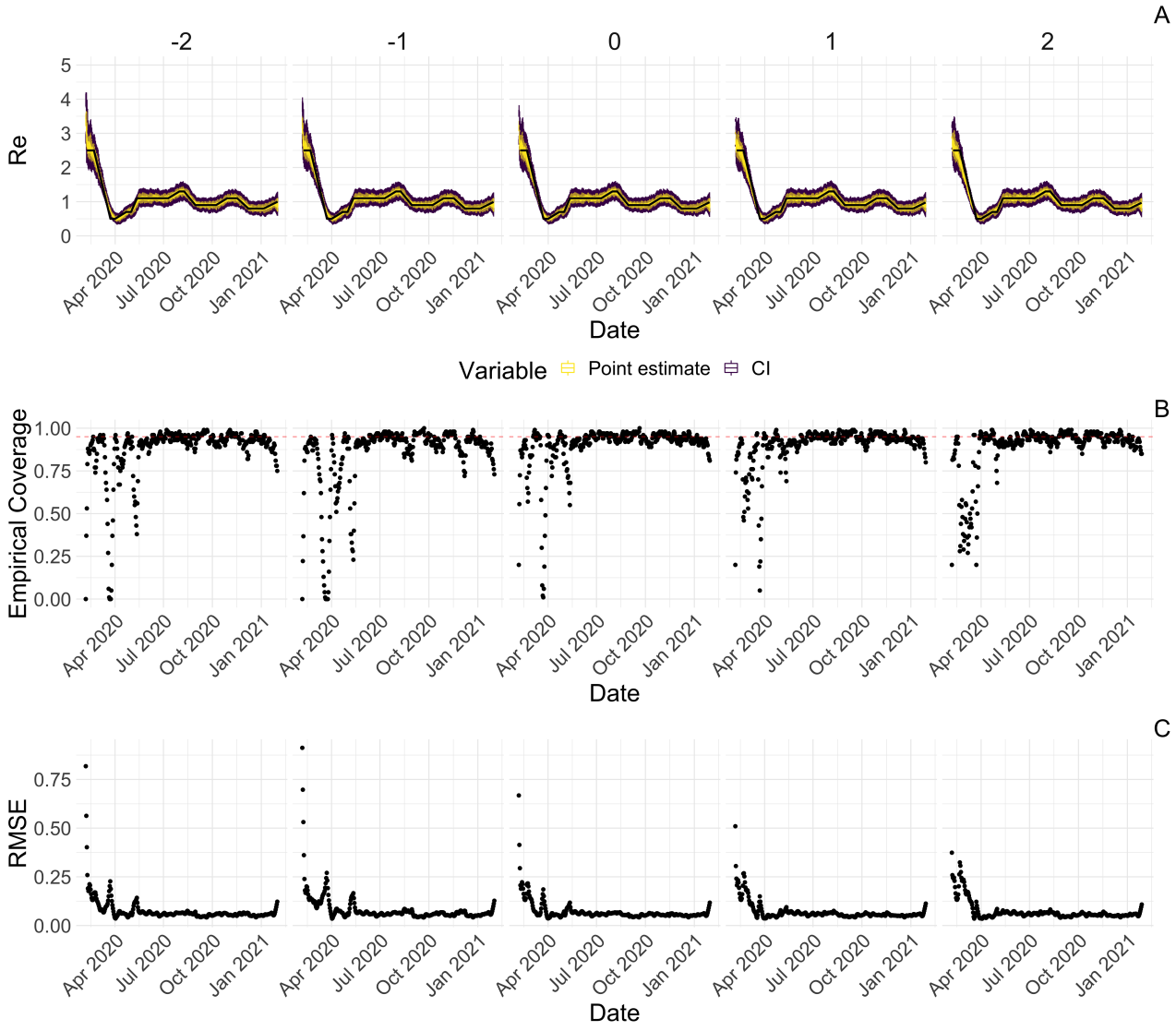
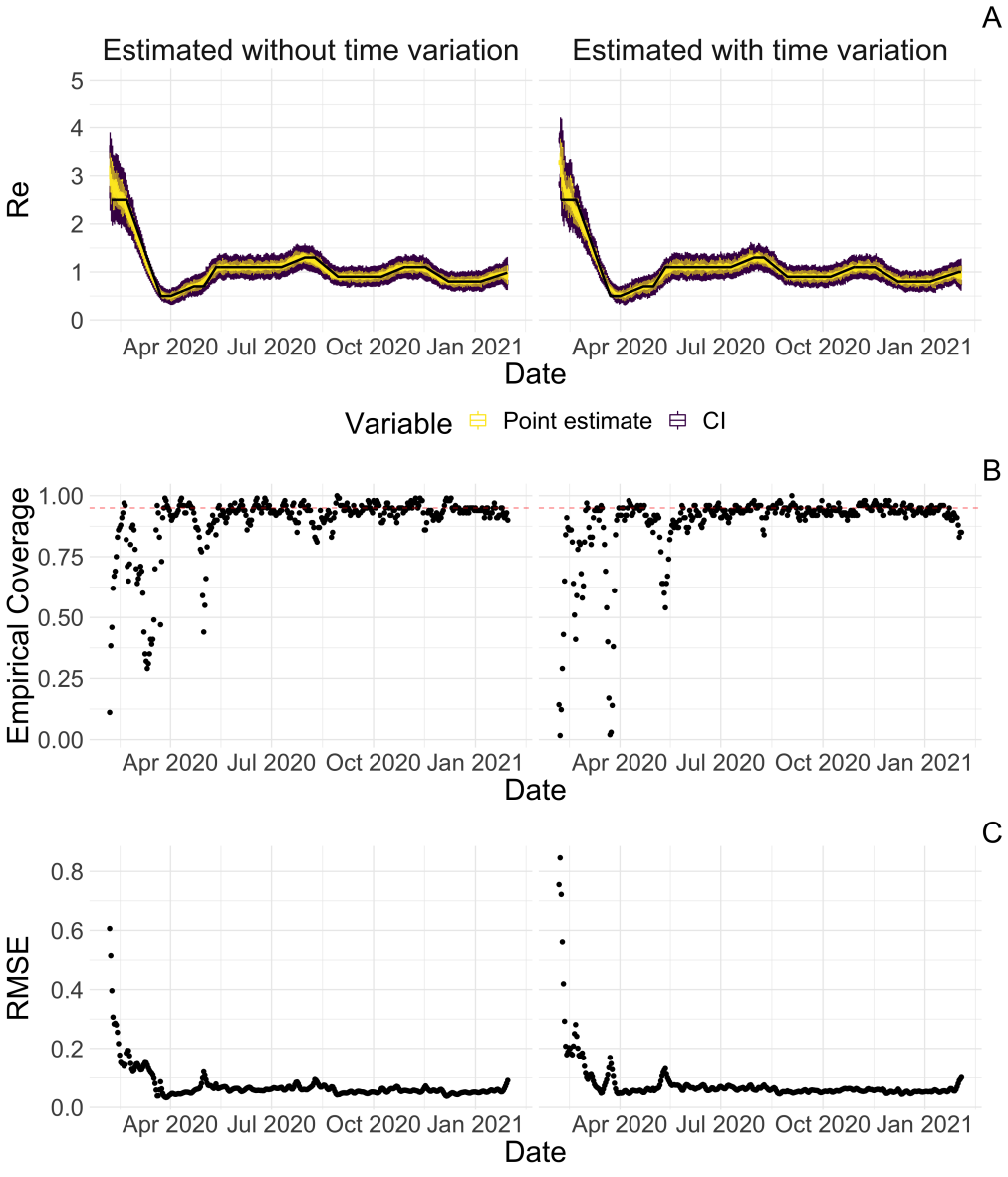


Figure S7: Performance of our method on simulated scenarios with misspecified delay distributions. When estimating R_e, we misspecified the mean of the delay distribution (5.5 for symptom-onset to case confirmation) by the numbers above the columns. **A** The specified R_e trajectory (black line; see Methods) was used to simulate a trajectory of reported cases (with Swiss case observation noise) 100 times. From each trajectory we estimated R_e (yellow boxplots), and a 95% confidence interval (purple boxplots of the lower/upper endpoint). **B** The fraction of simulations where the true R_e value was within the 95% confidence interval. The dashed red line indicates the nominal 95% coverage. **C** The root mean squared relative error for every time point.

(a) Confirmed cases: mean from 5.5 to 2



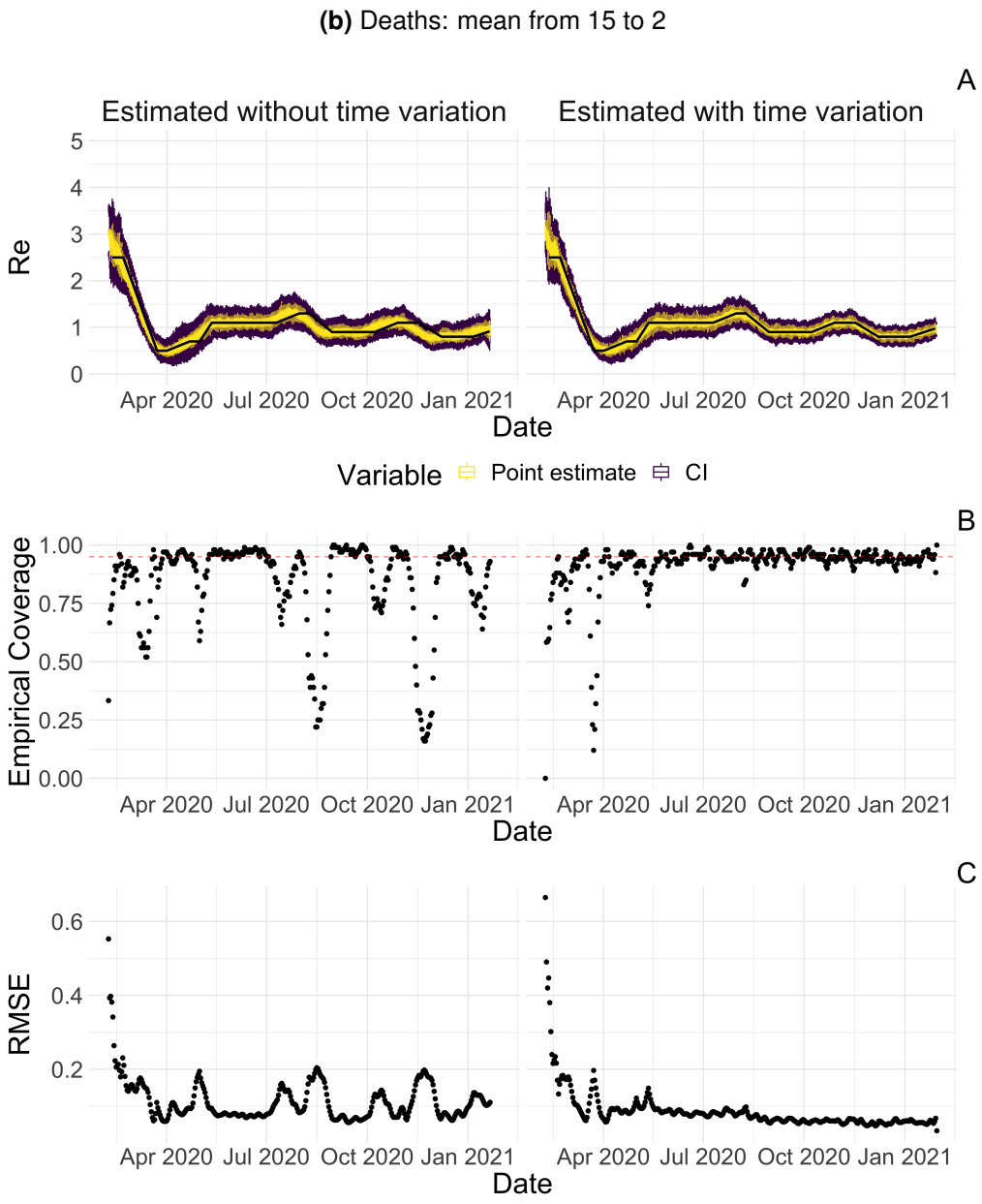


Figure S8: Performance of our method on simulated scenarios with time-varying delay distributions. The observations were simulated with a time-varying delay distribution for (a) confirmed cases, or (b) deaths (see Methods), and then estimated with (right column) or without (left column) taking the time-varying distributions into account. **A** The specified R_e trajectory (black line; see Methods) was used to simulate a trajectory of reported cases (with Swiss case observation noise) 100 times. From each trajectory we estimated R_e (yellow boxplots), and a 95% confidence interval (purple boxplots of the lower/upper endpoint). **B** The fraction of simulations where the true R_e value was within the 95% confidence interval. The dashed red line indicates the nominal 95% coverage. **C** The root mean squared relative error for every time point.

8.2 Switzerland specific results

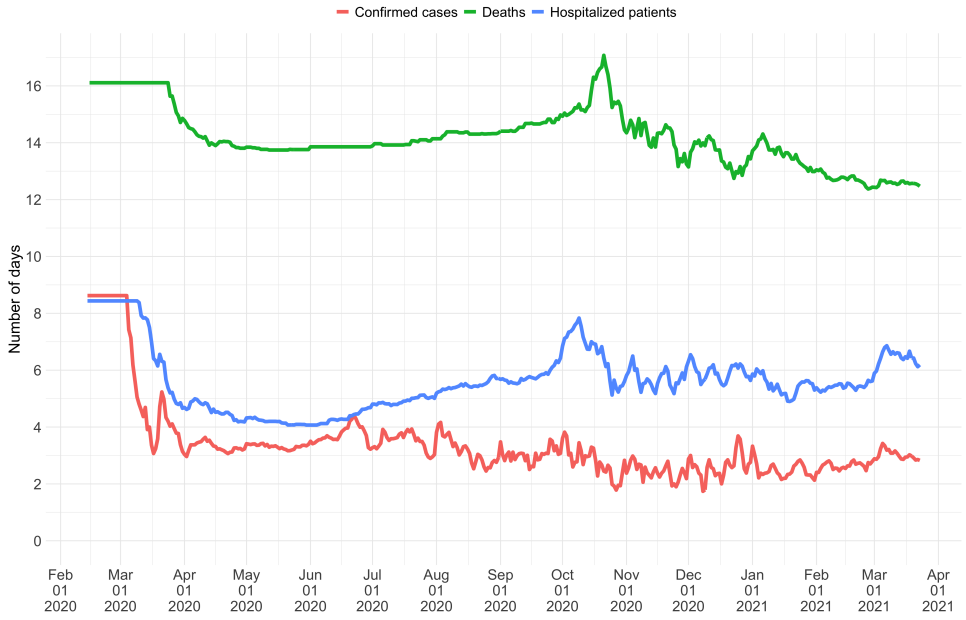


Figure S9: Mean delay in Switzerland between onset of symptoms and reporting. For each date, the mean is taken over the last 300 reports with known symptom onset date, based on line list data from the FOPH. For early dates, before 300 reports are available, the mean is taken over the first 300 reports.

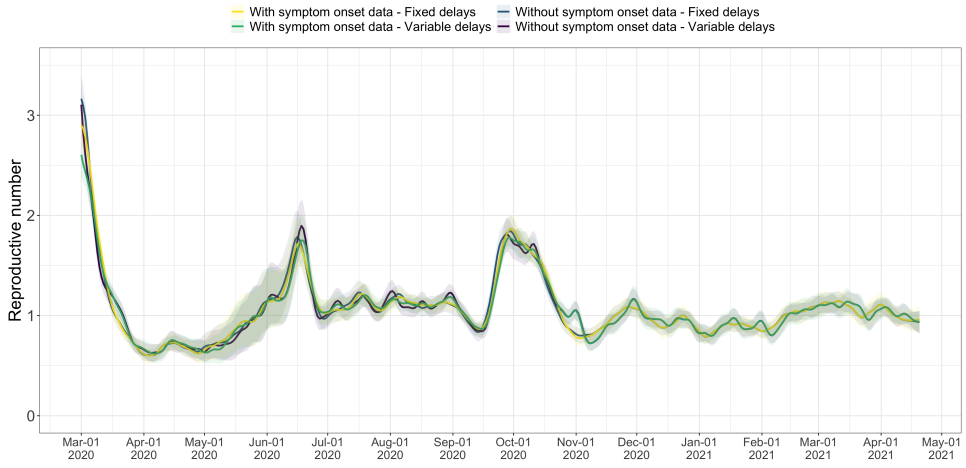


Figure S10: Comparison of the R_e estimates with or without accounting for known symptom onset dates and for time-variability on reporting delays. The comparison is based on time series of confirmed cases in Switzerland, from line list data provided by the FOPH. Both the inclusion of known symptom onset 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. The fraction of cases with known symptom onset date has drastically reduced since November 2020, hence the overlap in curves with and without symptom onset data for later dates.

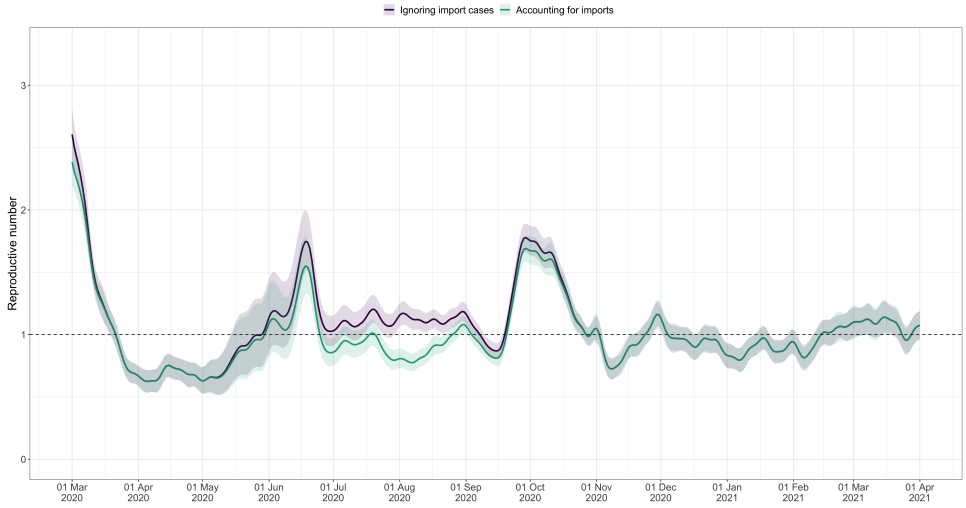


Figure S11: Effective reproductive number estimates with or without accounting for known imports. The comparison is based on time series of confirmed cases in Switzerland, from line list data provided by the FOPH. The analysis ignoring imports is unbiased if the number of imports equals the number of exports. Since the analysis accounting for imports is not accounting for exports, the results are a lower limit for the effective reproductive number. Very few imported cases were reported since November 2020, hence the complete overlap in the curves after that date.

8.3 R = 1 crossings

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

The 42 included countries: Algeria, Andorra(*), Australia, Austria, Belgium, Canada, Chile, Croatia, Czech Republic, Denmark, Egypt, Estonia, Finland, France, Germany, Greece, Iceland, Indonesia, Iran, Ireland, Israel, Japan, South Korea(*), Lebanon, Malaysia, Mexico, Netherlands, Norway, Philippines, Poland, Portugal, Saudi Arabia, Singapore, Slovenia, Spain, Switzerland, Tajikistan(*), Thailand, United Arab Emirates, United Kingdom, United States, Vietnam.

A star indicates the country was not included in the ΔSI analysis (e.g. because the biggest jump in SI took place prior to the first possible R_e estimate).

For 37/42 countries the R_e estimate was above one prior to the reference date, and significantly so for 35/42. The countries that reached $R_e < 1$ prior to the reference date were Andorra (17 days prior), Australia (2 day prior), Denmark (3 days prior), Japan (359 days prior), and Vietnam (3 days prior).

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

The 45 included countries: Algeria, Australia, Austria, Belarus(*), Belgium, Canada, Chile, Colombia(*), Croatia, Czech Republic, Denmark, Egypt, Estonia, Finland, France, Germany, Greece, Iceland, Indonesia, Iran, Ireland, Israel, Japan, Lebanon, Malaysia, Mexico, Netherlands, New Zealand(*), Norway, Philippines, Poland, Portugal, Russia(*), Saudi Arabia, Serbia(*), Singapore, Slovenia, Spain, Switzerland, Thailand, Turkey(*), United Arab Emirates, United Kingdom, United States, Vietnam.

A star indicates the country was not included in the $SI50$ analysis (e.g. because $SI = 50$ was never reached).

For 41/45 countries the R_e estimate was above one prior to the reference date, and significantly so for 38/45. The countries that reached $R_e < 1$ prior to the reference date were Australia (3 days prior), Denmark (4 days prior), Germany (4 days prior), and Vietnam (10 days prior).

Table S2: Investigating the relation between the date of 'lockdown' and the date that the estimated R_e dropped below 1. The first three columns contain the same information as the first four columns of the main text table 1. The last two columns are analogous to the third ('$\hat{R}_e < 1$ based on confirmed cases') but are based on R_e estimates for hospitalisations and deaths respectively. For each of these observation types, we used our method to determine when the R_e estimate first dropped below 1, and for which dates the corresponding 95% confidence interval contained 1. Further, we used news reports to determine when a country implemented stay-at-home orders (a 'lockdown'). Based on our R_e estimates for confirmed cases, Denmark, Germany, the Netherlands, and Slovenia had 95% 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 four: Denmark, the Netherlands, Poland, and the United Kingdom. See Supplementary Discussion 7.2 for smoothing related caveats.

Country	Lockdown	$\hat{R}_e < 1$ based on Confirmed cases	$\hat{R}_e < 1$ based on Deaths	$\hat{R}_e < 1$ based on Hospitalisations
Austria	16-03	20-03 [20-03, 20-03]		
Belgium	18-03	30-03 [25-03, 03-04]	26-03 [24-03, 26-03]	25-03 [24-03, 25-03]
Denmark	18-03	**10-03 [10-03, 20-06]**	22-03 **[18-03, 07-01]**	
Finland	16-03	02-04 [29-03, 30-04]	07-04 [25-03, 11-04]	
France	17-03	27-03 [23-03, 07-04]	24-03 [22-03, 26-03]	27-03 [25-03, 26-03]
Germany	22-03	**18-03 [17-03, 19-03]**	31-03 [23-03, 04-04]	
Ireland	27-03	08-04 [04-04, 15-04]	05-04 [31-03, 09-04]	06-04 [06-04, 26-04]
Italy	10-03	18-03 [17-03, 19-03]	14-03 [01-03, 29-05]	
Netherlands	23-03	05-04 **[22-03, 10-04]**	**22-03 [19-03, 02-04]**	26-03 [24-03, 26-03]
Norway	14-03	21-03 [17-03, 24-03]	25-03 [18-03, 08-04]	
Poland	25-03	02-04 [31-03, 17-04]	09-04 **[24-03, 01-12]**	
Portugal	16-03	28-03 [23-03, 15-04]	28-03 [21-03, 12-04]	
Romania	24-03	06-04 [31-03, 29-04]	17-04 [25-03, 28-04]	
Russian Federation	30-03	04-05 [01-05, 08-05]	18-05 [14-05, 12-12]	
Slovenia	20-03	23-03 **[13-03, 26-03]**	26-03 [20-03, $\geq$03-05-2021]	
Spain	14-03	26-03 [25-03, 26-03]		
Sweden		**01-04 [06-03, $\geq$03-05-2021]**	**05-04 [13-03, $\geq$03-05-2021]**	
Switzerland	17-03	22-03 [20-03, 22-03]	21-03 [18-03, 23-03]	18-03 **[16-03, 18-03]**
Turkey	21-03	08-04 [01-04, 13-04]	04-04 [31-03, 06-04]	
United Kingdom	24-03	30-03 [28-03, 20-04]	25-03 **[24-03, 25-03]**	29-03 [27-03, 29-03]

Table S3: News and public resources used to determine when a country implemented the first non-pharmaceutical interventions, and a nationwide lockdown.

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-1
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~b4
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

8.4 Implementation/Lifting of individual NPIs

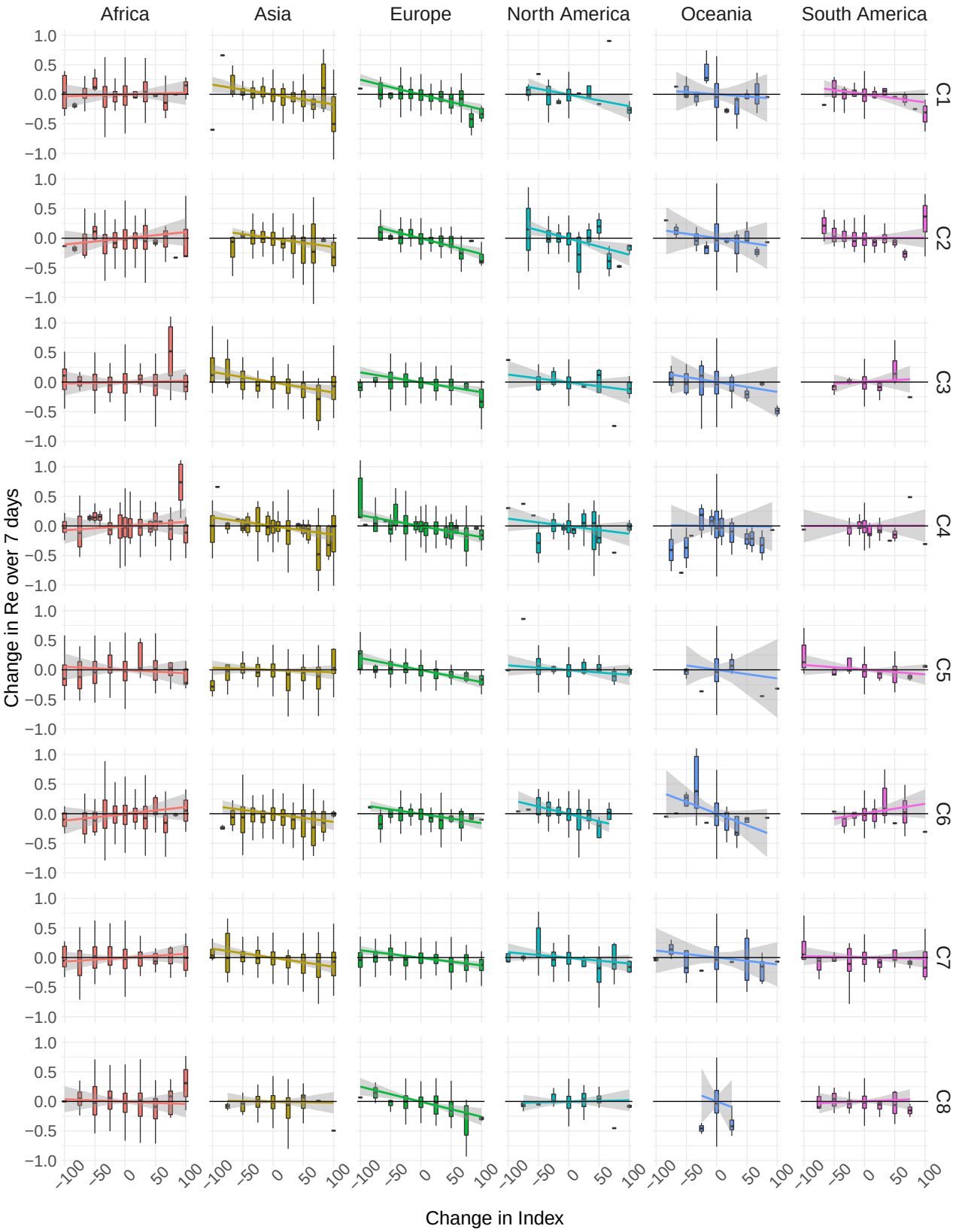


Figure S12: The one-week change in R_e following changes in government stringency. The rows C1-C8 refer to different stringency indices: C1 School closing; C2 Workplace closing; C3 Cancelling public events; C4 Restrictions on gatherings; C5 Closing public transport; C6 Stay at home requirements; C7 Restrictions on internal movement; C8 International travel controls.