Ascertaining the initiation of epidemic resurgences: an application to the COVID-19 second surges in Europe and the Northeast United States

Jose M. G. Vilar[1,2,*] and Leonor Saiz[3,*]

[1] Biofisika Institute (CSIC, UPV/EHU), University of the Basque Country (UPV/EHU), P.O. Box 644, 48080 Bilbao, Spain

[2] IKERBASQUE, Basque Foundation for Science, 48011 Bilbao, Spain

[3] Department of Biomedical Engineering, University of California, 451 E. Health Sciences Drive, Davis, CA 95616, USA

[*]Correspondence to: j.vilar@ikerbasque.org or lsaiz@ucdavis.edu

ABSTRACT

Assessing a potential resurgence of an epidemic outbreak with certainty is as important as challenging. The low number of infectious individuals after a long regression, and the randomness associated with it, makes it difficult to ascertain whether the infectious population is growing or just fluctuating. We have developed an approach to compute confidence intervals for the switching time from decay to growth and to compute the corresponding supra-location aggregated quantities to increase the precision of the determination. We estimated the aggregate prevalence over time for Europe and the Northeast United States to characterize the COVID-19 second surge in these regions during year 2020. We find a starting date as early as July 3 (95% Confidence Interval (CI): July 1–July 6) for Europe and August 19 (95% CI: August 16 – August 23) for the Northeast; subsequent infectious populations that, as of December 31 have always increased or remained stagnant; and the resurgences being the collective effect of each overall region with no location dominating the regional dynamics by itself.

INTRODUCTION

Identifying a potential resurgence of an epidemic outbreak is crucial to timely implementing measures for its mitigation and control. A major challenge, however, is the high uncertainty present because of the low prevalence values at which it typically happens after a long regression, as has been observed in many locations through the ongoing COVID-19 pandemic [1-4]. At the field level, direct characterization through randomized testing would need large population studies to provide significant results and using infection case data is dependent on varying testing rates [2, 3]. More robust approaches based on death counts are also affected by the extremely small number of random events on which they rely for inference [2, 4]. This uncertainty in assessing the state of the outbreak for a potential resurgence is a source of delays in the decision making and intervention implementation processes.

Here, we address two main computational needs to precisely characterize a resurgence. The first one is how to establish confidence intervals in the timing of the resurgence. These intervals range from the time it is certain that the infectious population has stopped decreasing to the time it is certain that the population has started to increase with a given confidence level. The second need is how to aggregate different local data into supra-local quantities to identify whether the resurgence is a collective regional effect and if so, to determine the initiation of the resurgence more confidently.

We focus explicitly on Europe and the Northeast United States (US), which have experienced a second surge of the COVID-19 outbreak after a similar initial outburst and subsequent regression. None of these resurgences was widely expected nor anticipated [5, 6]. Both regions display high mobility among their locations and broad independence among locations to enact measures to mitigate the propagation of the outbreak. In the case of Europe, Schengen Area countries allow for unrestricted border crossings among them.

Mobility restrictions, lockdowns, and other nonpharmaceutical interventions were able to achieve a major regression of the outbreaks, but the gradual lifting of restrictions has resulted in a resurgence across locations in these two regions [2, 7]. The characterization of the similarities and differences of the outbreak progression in these two areas is needed to provide insights into the effectiveness of the actions taken, to ascertain the extent their results can be extrapolated from one region to another, and to informedly mitigate the current and potentially forthcoming resurgences.

METHODS

Upper and lower bounds of the growth rate determine the confidence interval of the resurgences

We consider the estimated infectious population of the specific location at time t denoted by $n_I(t)$ and dynamics given by

$$\frac{d}{dt}n_I(t) = k_G(t)n_I(t),$$

where $k_G(t)$ is its per capita growth rate with upper and lower bounds of the confidence interval denoted by $k_G^U(t)$ and $k_G^L(t)$, respectively.

In epidemiology, it is customary to use the time-varying reproduction number R_t, which describes the expected number of infections arising from a single case in the population. It is related to the growth rate $k_G(t)$ through the Euler–Lotka equation,

$$R_t^{-1} = \int_0^\infty f_{GT}(\tau)e^{-k_G(t)\tau}d\tau,$$

where $f_{GT}(\tau)$ is the probability density function of the generation time. We consider the usual description of generation times through a gamma distribution

$$f_{GT}(\tau) = \frac{\beta^\alpha}{\Gamma(\alpha)}\tau^{\alpha-1}e^{-\beta\tau},$$

which leads to

$$R_t = \left(1 + \frac{k_G(t)}{\beta}\right)^{\alpha}$$

for $k_G(t) > -\beta$ and $R_t = 0$ for $k_G(t) \leq -\beta$. The values of the parameters are given by $\alpha = \tau_G^2/\sigma_G^2$ and $\beta = \tau_G/\sigma_G^2$, where τ_G and σ_G^2 are the mean and the variance of the generation time, respectively.

The starting date of the second surge, t_2, is computed as the date the infectious population reached a minimum value after the maximum of the first surge at time denoted by t_1:

$$t_2 = \underset{t}{\operatorname{argmin}}\, n_I(t) \text{ subject to } t > t_1,$$

which in continuous time corresponds to a zero value of the growth rate (reproduction number equal to 1): $k_G(t_2) \simeq 0$. The lower bound, t_2^L, of its confidence interval is computed as the last day before the minimum in which the upper bound of the confidence interval of the growth rate is negative (reproduction number below 1):

$$t_2^L = \max_t t \text{ subject to } k_G^U(t) < 0 \text{ and } t < t_2.$$

Analogously, the upper bound, t_2^U, of the confidence interval is computed as the first day after reaching the minimum in which the lower bound of the confidence interval of the growth rate is positive (reproduction number above 1):

$$t_2^U = \min_t t \text{ subject to } k_G^L(t) > 0 \text{ and } t > t_2.$$

The approach is illustrated for Connecticut (Northeast US) and Austria (Europe's Schengen area) in Figure 1. The trajectories of the infectious populations, the growth rates, and the 95% confidence intervals (CI) for each location were downloaded on April 21, 2021, from https://github.com/Covid19Dynamics/trajectories. The data considers explicitly the age-specific infection fatality rates from Verity et al. [8], which are consistently similar among distinct locations [9, 10], to infer the local infectious population from reported death

counts [4]. Reproduction numbers were computed from growth rates considering a gamma-distributed generation interval with a mean of 6.5 days and a standard deviation of 4.2 days [11]. These two locations show that, in general, there is a high uncertainty in the timing that can be attributed to the starting date of the resurgence.

Aggregate values provide a potential avenue to increase the reliability of the estimates for low prevalence values

The aggregate infectious population for a region is expressed as

$$n_I(t) = \sum_j n_{I,j}(t),$$

where $n_{I,j}(t)$ is the infectious population of the specific location with index j. Using the method of variance estimates recovery [12], the corresponding upper and lower confidence intervals are computed as

$$n_I^U(t) = n_I(t) + \sqrt{\sum_j \left(n_{I,j}^U(t) - n_{I,j}(t)\right)^2}$$

and

$$n_I^L(t) = n_I(t) - \sqrt{\sum_j \left(n_{I,j}(t) - n_{I,j}^L(t)\right)^2}$$

from the upper, $n_{I,j}^U(t)$, and lower, $n_{I,j}^L(t)$, confidence intervals for each location.

The method of variance estimates recovery cannot be used directly to compute the confidence intervals for the aggregate growth rate. We derive the expressions for the upper and lower bounds by considering that the overall time-dependent growth rate is given by

$$k_G(t) = \sum_j k_{G,j}(t) \frac{n_{I,j}(t)}{n_I(t)},$$

where $k_{G,j}(t)$ is the growth rate of the infectious population of the specific location with index j. This expression follows from

$$\frac{d}{dt}n_I(t) = \sum_j \frac{d}{dt}n_{I,j}(t) = \sum_j k_{G,j}(t)\, n_{I,j}(t) = k_G(t)n_I(t).$$

The corresponding upper and lower confidence intervals are computed as

$$k_G^U(t) = k_G(t) + \frac{1}{n_I(t)}\sqrt{\sum_j \left(k_{G,j}^U(t) - k_{G,j}(t)\right)^2 n_{I,j}(t)^2}$$

and

$$k_G^L(t) = k_G(t) - \frac{1}{n_I(t)}\sqrt{\sum_j \left(k_{G,j}(t) - k_{G,j}^L(t)\right)^2 n_{I,j}(t)^2}$$

from the upper, $k_{G,j}^U(t)$, and lower, $k_{G,j}^L(t)$, confidence intervals for each location. These expressions explicitly consider that the uncertainty in the infectious populations times the corresponding growth rate is much smaller than the uncertainty in the growth rates times the corresponding infectious population.

RESULTS

The second surges started in early-mid summer with the Northeast US trailing Europe

To assess the properties of the resurgences with increased confidence, we computed the aggregate values of the infectious populations and the corresponding time-varying reproduction numbers for Europe's Schengen Area and the Northeast US from the individual values of the locations of each region [4]. We considered overall region values and overall region values excluding one location. Exclusion of one location provides an avenue to reliably infer the effects of the location in the overall region.

The initial progression of the overall infectious populations for both regions consisted of exponential growth followed by exponential decay (Figures 2A and 2B). Subsequently, there was a sharp transition to fast exponential growth in Europe on July 3 (95% Confidence Interval (CI): July 1–July 6), 2020, from an estimated infectious population of 3.0×10^4 (95 % CI: 2.4×10^4 – 3.5×10^4) individuals and a stagnant overall infectious population in the Northeast US, which started to grow slowly but with increasing speed on August 19 (95% CI: August 16 – August 23), 2020, from an estimated infectious population of 3.3×10^4 (95 % CI: 2.8×10^4 – 3.8×10^4) individuals.

The resurgence has been more abrupt and intense in Europe than in the Northeast US

Concomitantly, the time-varying reproduction numbers crossed above one on the resurgence dates less abruptly in the Northeast US than in Europe (Figure 2C), reaching maximum values of 1.50 (95% CI: 1.48–1.51) in Europe and 1.30 (95% CI: 1.27–1.34) in the Northeast US. The sharp resurgence to exponential growth in Europe is coincidental with lifting major nonpharmaceutical interventions that curved the outbreak [13], including the coordinated end of travel bans in Schengen Area's countries on July 1, 2020 [14].

No substantial decreases in the overall infectious population, nor corresponding reproduction numbers below one, were observed for any of the two regions over three months after the starting dates of the second surges (Figure 2). The estimated infectious population just stooped growing in the Northeast US in late December (Figure 2B) and entered a prolonged stagnant state in Europe in early November (Figure 2A).

Aggregate values are highly reliable compared to location-specific data

The low prevalence at the specific-location level leads to broad confidence intervals for both the infectious population and the time-varying reproduction numbers, which makes ascertaining the local growth properties of the outbreak unreliable over prolonged periods of time (Figure 1, Figure 3, and Supplementary Figure S1). The aggregated values for each region provide precise evidence of sustained growth of the outbreaks already over the

summer, despite the uncertainty and variability present in each of the locations independently (Figure 3 and Supplementary Figure S1).

Our results also provide robust evidence that the resurgence was not driven by a unique location since any aggregate value of the starting date for each region leaving one of their locations out is within the confidence limits of that of the overall region (Figure 3 and Supplementary Figure S1). Therefore, the resurgences were the collective effect of each overall region.

DISCUSSION

COVID-19 second surges in Europe and the Northeast US exemplify the difficulties of ascertaining the presence of an incipient epidemic resurgence and to determine whether the infectious population is growing or just fluctuating. We have provided an avenue to quantify the uncertainty present and the methodology to increase the reliability of the assessment by aggregating location-specific data in regional quantities.

Our results highlight the need to implement policies and surveillance approaches that include data at a supra-location level when there is high mobility among locations. In this regard, the Northeast US, as a region, closely trailed Europe in the second surge of the outbreak, but with a markedly smaller growth and evidence of slowing down earlier in the growth phase than Europe. Key differences in the actions taken included more gradual lifting and swifter progressive reimplementation of measures in the Northeast US than in Europe [13]. The progression over time of the aggregate prevalence of Europe's Schengen Area countries shows, with high certainty, that Europe's initial acting upon the second surge in mid-late October [5] took place well after a three-month-long period of sustained growth of the COVID-19 infectious population in the overall region, which has resulted in a second surge deadlier than the first one [15]. Such a high death toll has not been reached in the Northeast US [13].

ACKNOWLEDGMENTS

J.M.G.V. acknowledges support from Ministerio de Ciencia e Innovación under grant PGC2018-101282-B-I00 (MCI/AEI/FEDER, UE). L.S. acknowledges support from the University of California, Davis.

REFERENCES

1. Li, R., et al., *Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2).* Science, 2020. **368**(6490): p. 489-493.
2. Flaxman, S., et al., *Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe.* Nature, 2020. **584**(7820): p. 257-261.
3. Vilar, J.M.G. and J.M. Rubi, *Determinants of population responses to environmental fluctuations.* Sci Rep, 2018. **8**(1): p. 887.
4. Vilar, J.M.G. and L. Saiz, *The evolving worldwide dynamic state of the COVID-19 outbreak.* medRxiv, 2020: p. 2020.11.26.20239434.
5. Looi, M.-K., *Covid-19: Is a second wave hitting Europe?* BMJ, 2020. **371**: p. m4113.
6. Kim, J., *Is a Second Wave Starting in N.Y.C.?* The New York Times 2020: p. www.nytimes.com/2020/11/10/nyregion/second-coronavirus-wave-.html.
7. Li, Y., et al., *The temporal association of introducing and lifting non-pharmaceutical interventions with the time-varying reproduction number (R) of SARS-CoV-2: a modelling study across 131 countries.* The Lancet Infectious Diseases 2020: p. DOI:10.1016/s1473-3099(20)30785-4.
8. Verity, R., et al., *Estimates of the severity of coronavirus disease 2019: a model-based analysis.* Lancet Infect Dis, 2020. **20**(6): p. 669-677.
9. Levin, A.T., et al., *Assessing the age specificity of infection fatality rates for COVID-19: systematic review, meta-analysis, and public policy implications.* Eur J Epidemiol, 2020. **35**(12): p. 1123-1138.
10. O'Driscoll, M., et al., *Age-specific mortality and immunity patterns of SARS-CoV-2.* Nature, 2020: p. DOI:10.1038/s41586-020-2918-0.
11. Bi, Q., 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. **20**(8): p. 911-919.

12. Donner, A. and G.Y. Zou, *Closed-form confidence intervals for functions of the normal mean and standard deviation.* Stat Methods Med Res, 2012. **21**(4): p. 347-59.

13. Hale, T., et al. *Oxford COVID-19 Government Response Tracker, Blavatnik School of Government*. 2020. www.bsg.ox.ac.uk/covidtracker.

14. *Council of the European Union. Council Recommendation on the temporary restriction on non-essential travel into the EU and the possible lifting of such restriction*. 2020. https://data.consilium.europa.eu/doc/document/ST-9208-2020-INIT/en/pdf.

15. *JHU CSSE COVID-19 Data*. December 31, 2020]; Available from: https://github.com/CSSEGISandData/COVID-19.

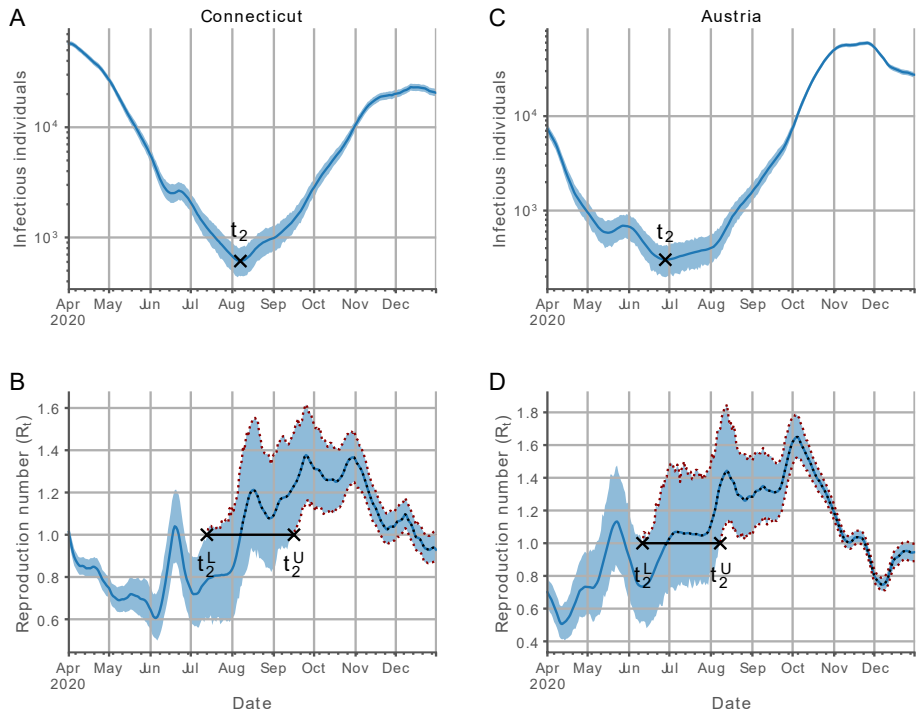

Figure 1. Location-specific characterization of the COVID-19 resurgence. The approach to locate the time of the resurgence and its confidence intervals is illustrated with data for Connecticut (A, B) and Austria (C, D). The top panels (A, C) show the temporal evolution of the infectious population with the shaded blue region indicating the 95% confidence intervals (CI). The bottom panels (B, D) show the temporal evolution of the reproduction number (blue line) with the shaded blue region indicating the 95% CI. The dotted lines highlight the span of the second wave and its confidence interval over the reproduction number data. Black markers indicate the starting date of the resurgence (t_2) and the lower (t_2^L) and upper (t_2^U) bound of its 95% CI. Reproduction numbers were computed from growth rates considering a gamma-distributed generation interval with a mean of 6.5 days and a standard deviation of 4.2 days.

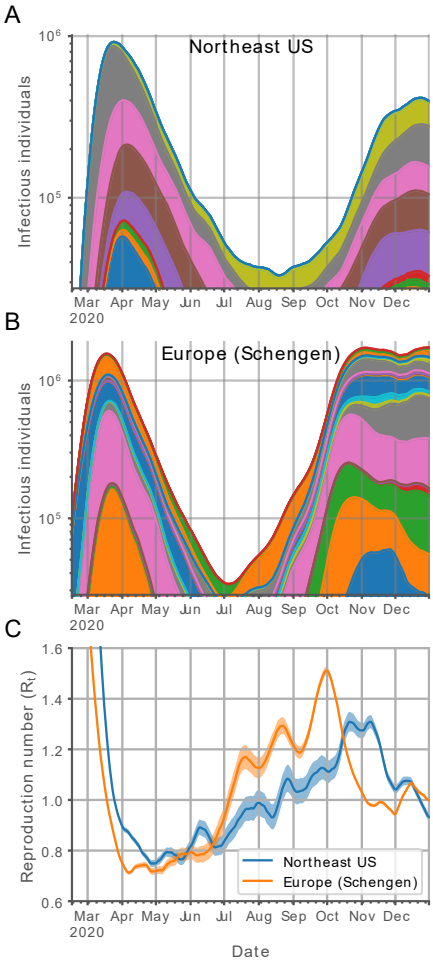

Figure 2. Temporal evolution of the COVID-19 outbreak in Europe and the Northeast US. Progression over time of the infectious population for countries in Europe's Schengen Area (A) and locations in the Northeast US (B) and of their corresponding reproduction numbers (C). Each colored section in the area plots represents the contribution of a country (A) or state (B) to the overall infectious populations. Countries and states are arranged in alphabetical order from bottom to top. The infectious populations are plotted on a logarithmic scale to highlight the triphasic behavior (growth-decay-growth) of the outbreak. The shaded regions in the reproduction number plots (C) represent the 95% CI. Locations with fewer than 30 reported COVID-19 deaths were not considered in the analysis.

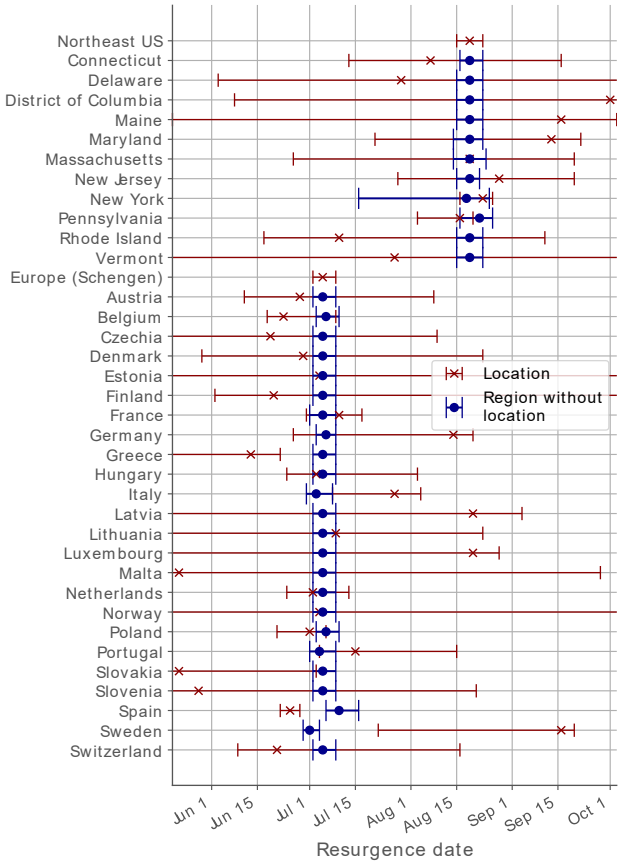

Figure 3. Timing of the COVID-19 outbreak resurgence in Europe and the Northeast US. Dates of the minimum infectious population reached are shown for each location (red crosses) and for the whole region without the population of the location indicated (blue circles). The intervals represent the 95% CI. Locations with fewer than 30 reported COVID-19 deaths were not considered in the analysis.