Supplementary Materials
for
Excess deaths reveal the true spatial, temporal, and demographic impact of COVID-19 on mortality in Ecuador

Authors: Leticia Cuéllar[1], Irene Torres[2,*], Ethan Romero-Severson[3], Riya Mahesh[3,4], Nathaniel Ortega[3,5], Sarah Pungitore[3,6], Nicolas Hengartner[3,*] and Ruian Ke[3,*]

Affiliations:

[1] A-1 Information Systems and Modeling, Analytics, Intelligence and Technology Division, Los Alamos National Laboratory, NM 87545, USA.
[2] Fundación Octaedro, Quito, Ecuador.
[3] T-6 Theoretical Biology and Biophysics, Theoretical Division, Los Alamos National Laboratory, NM 87545, USA.
[4] Department of Biology, University of Texas, Austin, Texas, 78712, USA.
[5] University of California Santa Barbara, Santa Barbara, CA 93106, USA.
[6] Program in Applied Mathematics, University of Arizona, Tucson, Arizona 85721, USA.

Supplementary Methods

Delay in reporting

The 2015-2019 all-cause death records from the Ecuadorian National Institute of Statistics and Census included both the date of reporting and the date of death. There is a delay between the time of death and the time that death is recorded into government registries. According to Peralta[S1], "in Ecuador, by law, the declaration and registration of all deaths is mandatory. The mortality registry contains all death certificates completed in the country by health professionals or (in their absence): 1) police or civil authorities or 2) civil registration officials". That is, the reporting of deaths does not solely rely on the healthcare system, ensuring reasonable reporting during health crisis. Figure S1 shows the fraction of deaths reported as a function of days since death for the all-cause mortality during 2015-2019. The reporting delay compares favourably with that of other countries, see for example Weinberger[6] and supports our claim that the historical all-cause mortality from 2015-2019 are essentially complete and do not require adjustments for reporting delay. The death counts during COVID only contained date of death without any information about the reporting date. Therefore, it is not possible to estimate the delay in reporting during 2020. However, the summary statistics of deaths in 2020 obtained in Feburary 2021 suggests that the delay contributes to only less than 3% underestimates (as discussed in the main text).

Statistical Model

To estimate the number expected deaths in 2020 (without COVID-19), we model the number of weekly deaths as a Poisson distributed random variable with expected value λ_{Bijkl} and assume that the logarithm of its expected value can be expressed as a linear combination of a set of explanatory variables

$$\log(\lambda_{ijkl}) = c_0 + c_{1,i}Week_i + c_{2,j}\,Province_j + c_{3,k}Sex_k + c_{4,l}Age\_Group_l + c_{5,s}\,Sex_k * Age\_Group_l$$

where λ_{ijkl} is the expected number of deaths during week *i*, in province *j* for the sex k and age group *i*.

We estimate the parameters of this generalized linear model using the *glm* function in R (version 3.6.1), specifying family='Poisson'. We calculate the number of expected deaths per week, province, sex, and age group and aggregate the results at the province and national levels, per week, and for the different demographic groups.

Traditional models for death counts include sine and cosine terms into model for the logarithm of the expected value to capture seasonality. In contrast, our model estimates a separate effect for each week. This more flexible model also captures seasonality, but has the freedom to estimate the shape of the seasonal effect. This added flexibility mitigates the risk of model form uncertainty (bias) that may arise from mis-specifying the functional form of the seasonal effects in death patterns.

The root residual plot (see Figure S2) reveals that the death count data is over-dispersed. Indeed, for Poisson distributed random variables, the square-root transformation stabilises the standard deviation to 0.5. But the root-residuals exhibits more variability than is expected from random variables with standard deviation of 0.5. Standard approaches to account for such extra-Poisson variability is to use negative binomial or quasi-Poisson regression models. But estimation of the extra-Poisson variability is notoriously unreliable for small count data, and as a result, we use standard log-linear Poisson regression to fit the death counts. Indeed, we note that as standard linear regression estimates remain unbiased even if the response variable has heterogeneous variances, assuming a Poisson distribution for moderately over-dispersed data still yields consistent for conditional expectation. Formally, this follows from analysis of maximum likelihood estimators under model misspecification --- see Huber[S2] or White[S3] for example.

However, over-dispersion impacts the construction of confidence intervals for the death counts, which in turn impact the confidence intervals of the excess death factor. Since our model does not explicitly accommodate the extra-Poisson variability, we use block bootstrap to compute confidence intervals for the expected death counts in “normal” years. Specifically, we resample with replacement observations in each cell --- combination of week, province, sex, and age group --- and refit the model using the bootstrap sample. Confidence intervals for provincial level expected values are obtained by computing 2.5% and 97.5% quantiles of the rolled-up bootstrap estimates.

The estimated model fitted with data from 2015 to 2019, fits fairly well as demonstrated by a closer inspection of the standardized residuals: square root of the observed deaths – the square root of the expected deaths. Taking the square root of the observed and the expected deaths is known as the variance stabilizing transformation that is commonly used for Poisson regression to alleviate heteroskedasticity. An analysis of deviance reveals that the all the factors have a p-values less than 0.00001.

Supplementary References

S1. Peralta, A., Benach, J., Borrell, C. *et al.* Evaluation of the mortality registry in Ecuador (2001–2013) – social and geographical inequalities in completeness and quality. *Popul Health Metrics* **17,** 3 (2019). https://doi.org/10.1186/s12963-019-0183-y

S2. Huber, P. The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 5.1 (1967), University of California press.

S3. White, H. (1982). Maximum Likelihood Estimation of Misspecified Models. Econometrica, 50(1), 1-25. doi:10.2307/1912526

Supplementary Figures

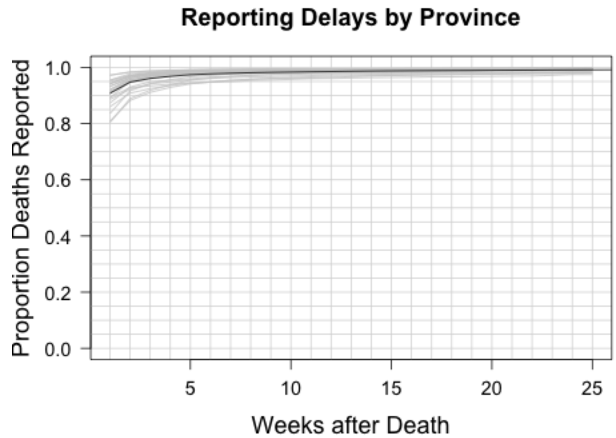

Figure S1. Reporting delay in death records for data between 2015 and 2019. Plot shows the percentage of deaths reported over time after deaths. Grey lines denote the percentage for each of the 24 Ecuadorian provinces and the black line shows the mean percentage.

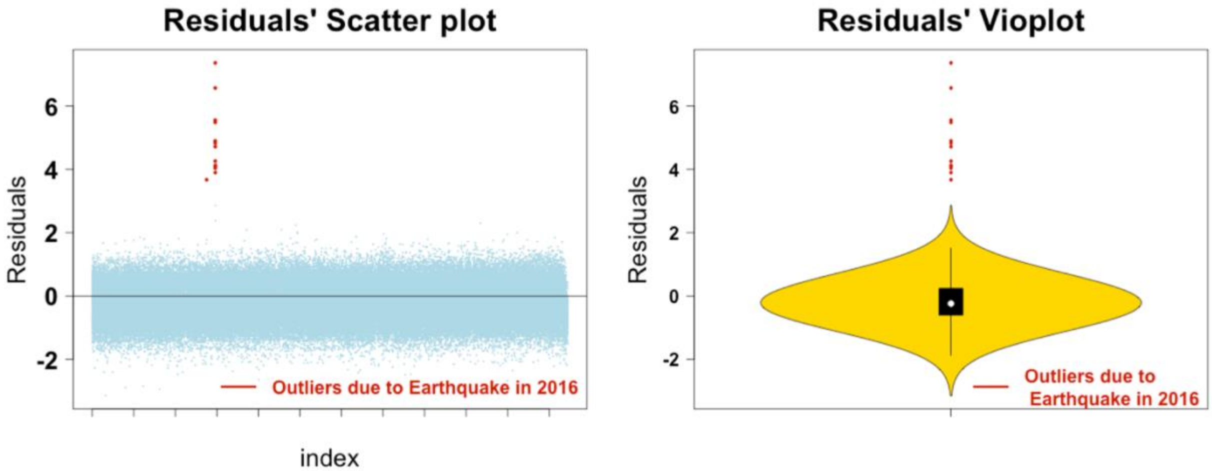

Figure S2. Plots of standardized residuals from the Poisson regression for excess deaths. Left plot shows scatter plot of the residuals and the right plot shows their distribution as depicted by a violin plot. Outliers (red), are cases that did not fit well the estimated model, correspond to deaths caused by the 2016 earthquake with epicentre in the province of Esmeraldas and a magnitude of 7.8 in the Mercalli intensity scale.

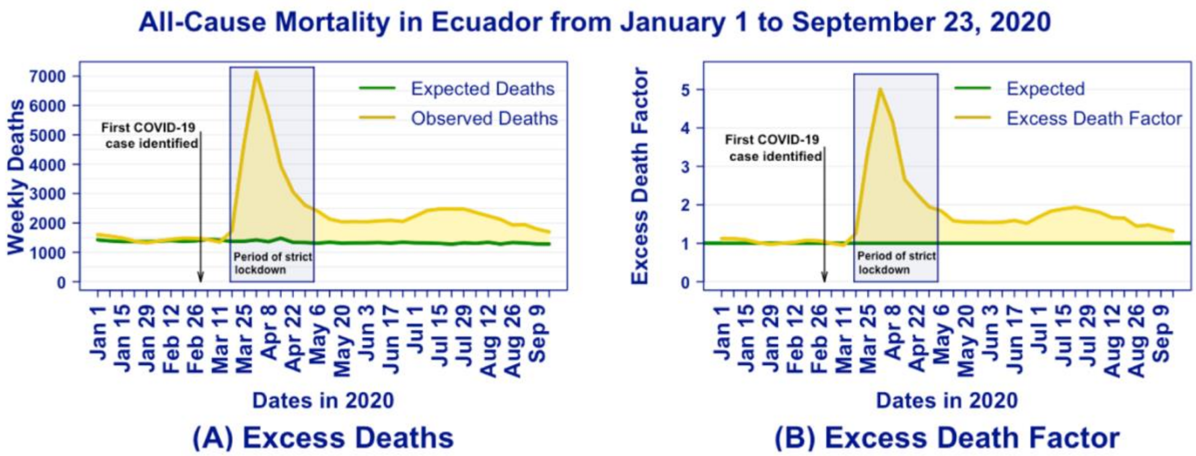

Figure S3. Time series for (A) weekly expected deaths and observed deaths and (B) corresponding time series of excess death factor. The period of strict national lockdown by the Ecuador government (shaded area) is taken from the Oxford COVID-19 Government Response Tracker[1].

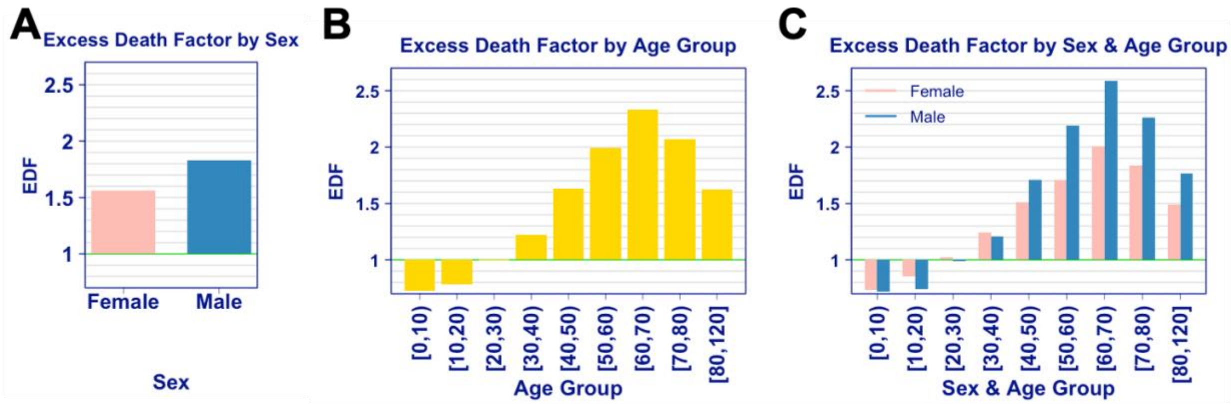

Figure S4. The excess death factor by demographics, i.e. the number of all-cause deaths over the number of expected deaths: (A) by sex, (B) by age group, and (C) by sex and age group. Note that the age group [60, 70) has the largest excess death factor.

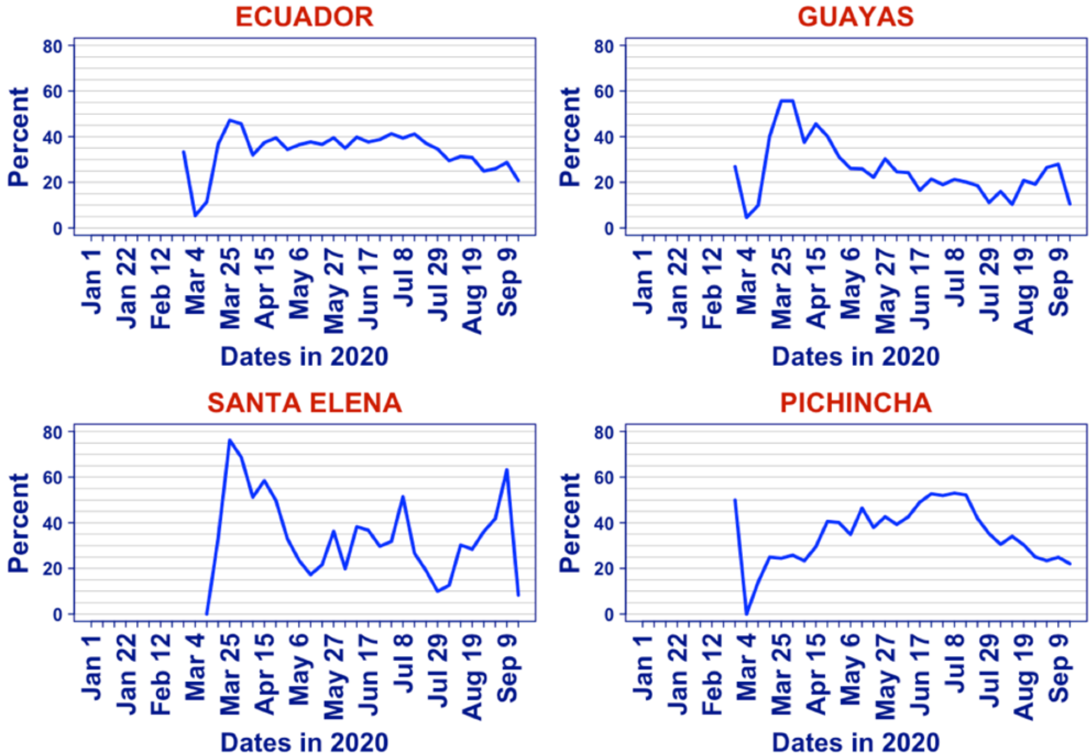

Figure S5. Time series of COVID-19 test positivity rate, i.e. percent of positive COVID-19 tests for Ecuador (national level), and for the provinces of Guayas, Santa Elena, and Pichincha.

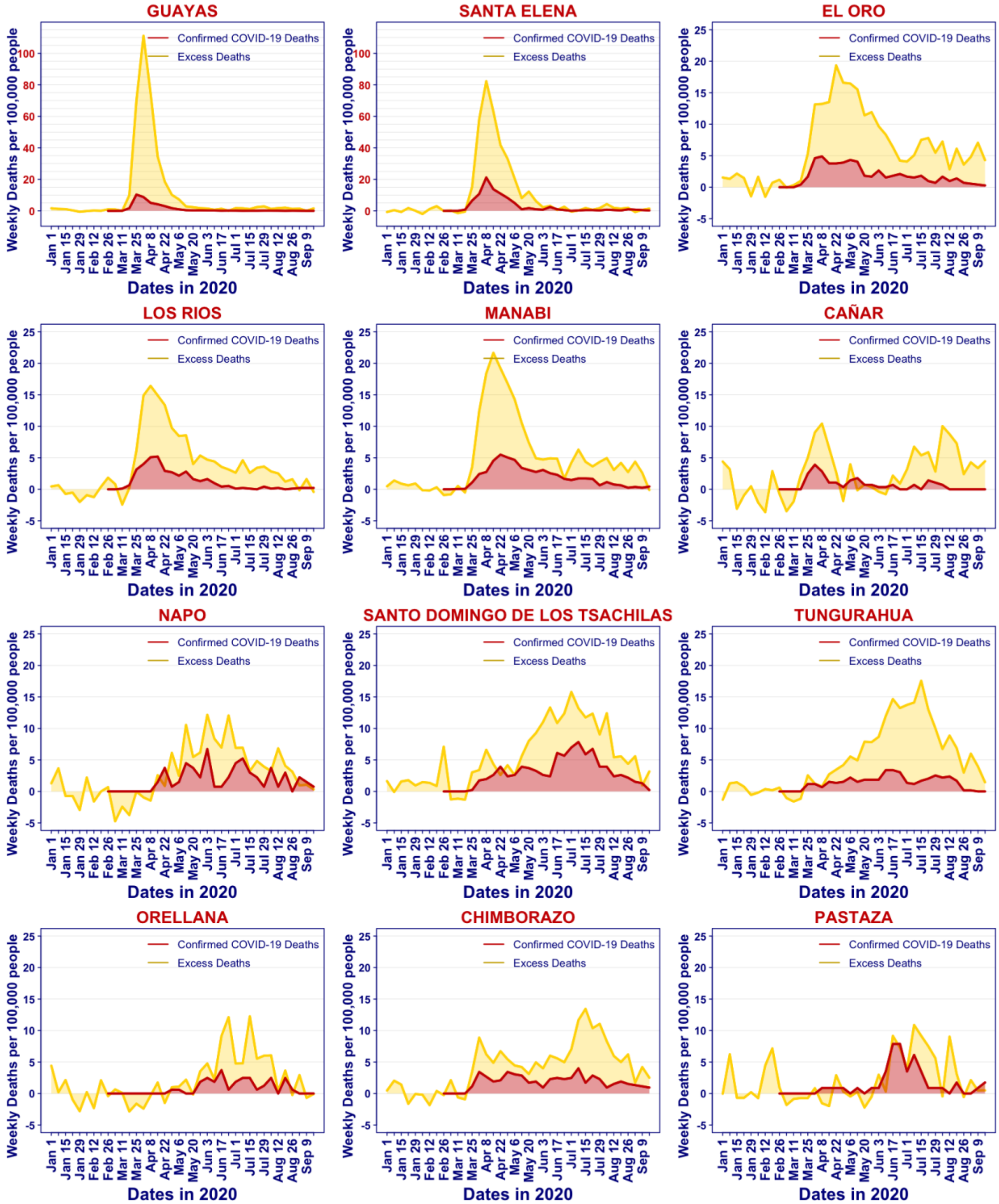

Figure S6. Provincial time series of excess deaths per 100,000 people (yellow) and for documented COVID-19 deaths per 100,000 people (red). Plots ordered according to the order in which provinces reached 10 excess deaths per 100,000 people. The first 12 provinces are shown.

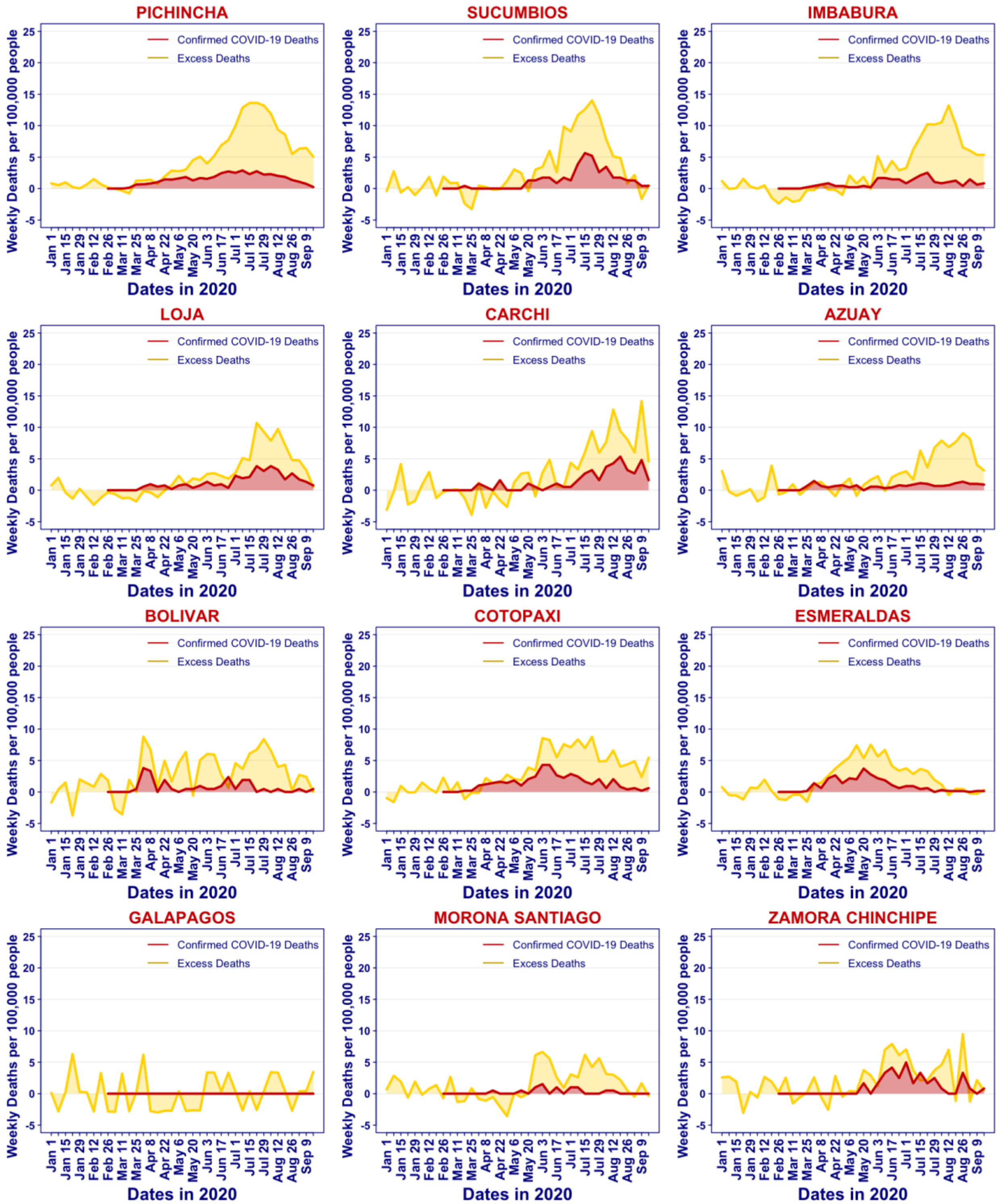

Figure S7. Provincial time series of excess deaths per 100,000 people (yellow) and for documented COVID-19 deaths per 100,000 people (red). Plots ordered according to the order in which provinces reached 10 excess deaths per 100,000 people. The last 12 provinces are shown.