Evaluating Covid-19 Mortality in Latin America: A Bayesian Perspective on Excess vs Official Death Counts

Timothy Lee

17 January, 2024

1. Introduction

Motivated by The Economist's analysis in “Tracking COVID-19 excess deaths across countries” (The Economist, 2021), this study aims to explore the discrepancy between official death counts and overall excess deaths attributed to the COVID-19 pandemic, with a focus on four Latin American countries: Chile, Ecuador, Mexico, and Peru. These countries are hypothesized to exhibit underreported COVID-19 mortality due to factors like underreporting, misclassification, and healthcare system overload (Litewka and Heitman, 2020).

The adoption of a Bayesian approach, utilizing the Integrated Nested Laplace Approximation (INLA) (Rue et al., 2009), was selected to address the limitations of linear models in The Economist's analysis. This method allows for a more flexible and nuanced analysis of mortality data, accommodating country-specific complexities and uncertainties. By applying Bayesian modeling, this study aims to deepen the understanding of COVID-19’s impact in Latin America and contribute to the global discussion on pandemic mortality reporting.

1.1 Definition of Excess Deaths

In this paper, “excess deaths” are defined as “the difference between the observed numbers of deaths in specific time periods and expected numbers of deaths in the same time periods” (CDC, 2022). This distinction is critical in the context of the COVID-19 pandemic, where attributing specific causes to deaths can be challenging. For example, some deaths due to COVID-19 may be misclassified, making all-cause mortality tracking a key indicator of potential undercounts in COVID-19 mortality (CDC, 2022).

2. Methodology and Modelling Assumptions

For simplicity and comparison purposes, I set the starting date of the COVID-19 pandemic for all countries to be on March 1st, 2020, considering the WHO’s declaration of COVID-19 as a pandemic on March 11, 2020 (WHO, 2020). For each country, I then fit a baseline model of total weekly mortality counts on all the pre-COVID years and generate out-of-sample forecasts for the COVID years (starting March 1st, 2020). The rationale is that this baseline model, based on pre-COVID mortality counts, should capture seasonal patterns and trends. Using this model, excess deaths are calculated by subtracting the actual number of mortality counts in the COVID years from this historical baseline. This approach accounts for people who died from COVID-19 but whose cause of death wasn’t correctly attributed to the official COVID mortality numbers for reasons previously mentioned. Finally, I compare these calculated excess deaths against the official reported COVID-19 deaths for each country.

3. Data and Sample

The total mortality counts and the official COVID-19 death counts for each country are retrieved from The Economist’s publicly available data repository on GitHub (https://github.com/TheEconomist/covid-19-excess-deaths-tracker/tree/master/output-data/historical-deaths). The original data source includes “the Human Mortality Database, a collaboration between UC Berkeley and the Max Planck Institute in Germany, and the World Mortality Dataset, created by Ariel Karlinsky and Dmitry Kobak.” (The Economist, 2021). I first plot the total deaths counts against the official covid-19 death counts for each Latin America country. Note that both death counts are reported in a weekly time frame.

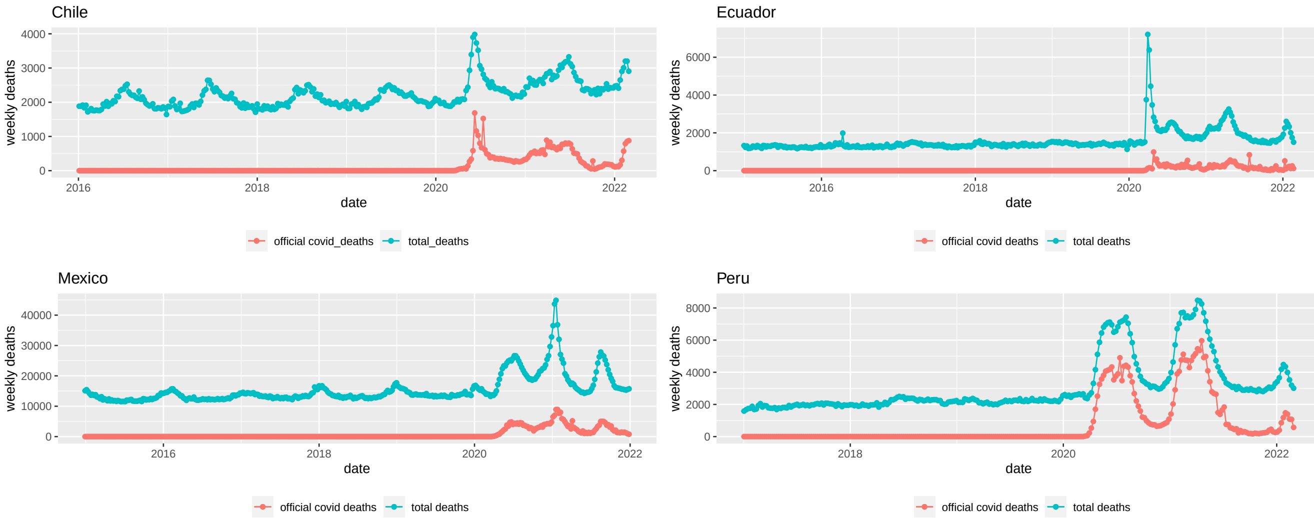

Figure 1: Plot of official covid-19 vs total deaths counts for each country

4. Bayesian Estimation of Historical Baselines and Excess Deaths

In the original analysis, The Economist utilized “a five-year average of deaths in a given region to calculate a baseline for excess deaths”, followed by “a statistical model for each region, which predicts the number of deaths we might normally have expected in 2020. The model fits a linear trend to years, to adjust from long-term increases or decreases in deaths, and a fixed effect for each week or month” (The Economist, 2021).

However, I believe that these models may lack the necessary complexity to accurately capture the underlying patterns in mortality data. The assumption of linearity in mortality trends overlooks potential seasonal/cyclical components and non-linear trends, which could significantly influence the data (Hyndman, 2018). Additionally, general mortality trends may exhibit non-stationarity, influenced by factors unrelated to COVID-19, such as the increasing elderly population (Cheng et al., 2020) and rising mortality from drug overdoses, alcohol, suicides, and cardiometabolic conditions in younger and middle-aged populations (National Academies of Sciences, 2021).

To address these limitations, I employ a Bayesian model using Integrated Nested Laplace Approximation (INLA) (Rue et al., 2009). Each country’s model is be fitted using data up to “2020-03-01”, serving as a historical baseline for calculating excess deaths during the pandemic. The model generates predicted values along with their corresponding 95% credible intervals (CI). Additionally, actual weekly deaths are overlaid for comparison.

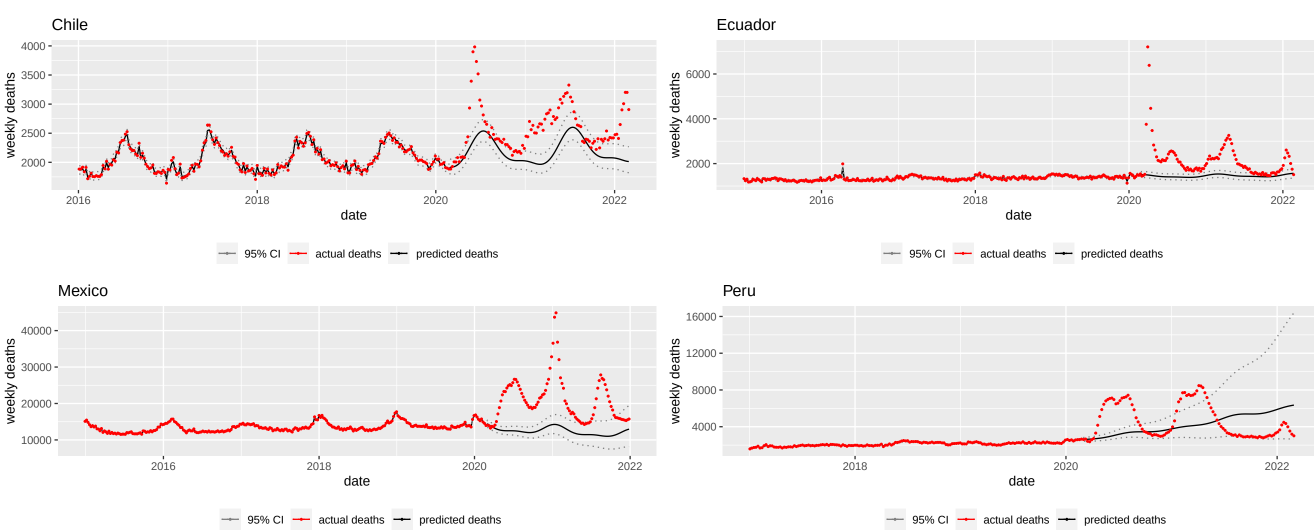

Figure 2: Plot of predicted vs actual deaths

4.1 Sampling of Predicted Deaths from the Posterior Distributions

Subsequently, I draw 10 simulations from the posterior distributions of each country's model. These simulations reflect varied scenarios of mortality data after incorporating the impacts of the COVID-19 pandemic. This approach allows for a comprehensive analysis of the potential ranges and uncertainties in actual death counts.

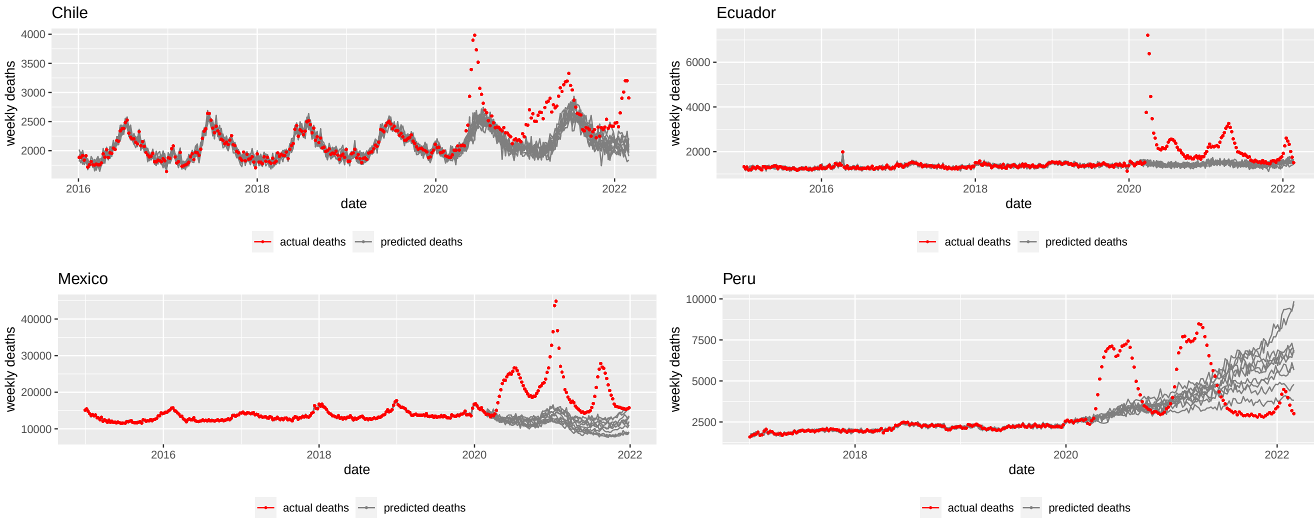

Figure 3: Plot of sampled predicted vs actual deaths from 10 simulations

4.2 Calculation of Excess Deaths and Reported COVID-19 Deaths

For the COVID-years (starting from March 1st, 2020), I calculate the excess deaths for each simulation by subtracting the estimated predicted deaths from the total deaths. Additionally, the official reported COVID-19 deaths for each country are overlaid on these calculations. This comparison provides a clear visualization of the discrepancy between the predicted and actual mortality figures in the context of the officially reported COVID-19 deaths.

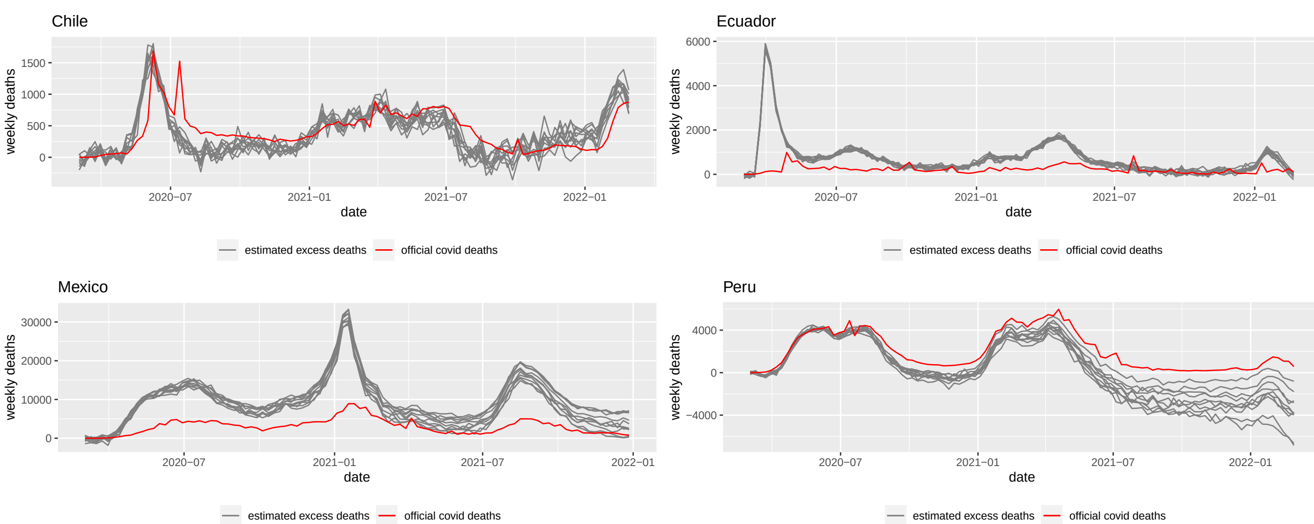

Figure 4: Plot of sampled predicted vs official covid-19 deaths

5 Discussion

5.1 Evaluating Model Performance and Predictions

Figure 2 demonstrates that INLA effectively captured pre-pandemic death trends, including seasonal patterns. Predictions for the pandemic period, based solely on pre-COVID data, appear reasonable. Seasonality is predominantly driven by a linear combination of 6-month and 12-month sine and cosine cycles, as indicated by sin12, sin6, cos12, and cos6 parameters.

5.2 Model Priors and Implications

The selected pc.prec priors (Precision Penalized Complexity prior) for both random walk 2 and independent/iid models effectively penalize deviations from the base model, supporting Occam's razor and invariant to reparameterizations (Virgilio, 2020; Daniel et al., 2017). With a standard deviation of 0.01 and a median quantile of 0.5, these priors suggest a possible weekly change of 0.01 in death rates, though this is an estimate due to varying death rates across countries.

The widening of the 95% C.I. in predictions over time, particularly for Peru, indicates increasing uncertainty. This is further evidenced by Figure 3's posterior distribution samples, revealing consistent results for Chile and Ecuador but variations for Mexico and Peru, suggesting the need for more sensitive priors or hyperparameter adjustments for these countries.

5.3 Comparing Excess Deaths with Official COVID-19 Deaths

In Figure 4, the calculated excess deaths, obtained by subtracting predicted deaths from actual deaths, are compared with official COVID-19 death counts. While Chile's data aligns, indicating that excess deaths correspond with reported COVID-19 deaths, Ecuador and Mexico show higher excess deaths than reported. Conversely, Peru presents an underfitting model with lower estimated excess deaths than official counts, implying significant noise and variance in the data.

6 Conclusion

6.1 Summary of Findings

In this study, I find mixed evidence regarding the discrepancy between official COVID-19 death counts and estimated excess deaths in Latin America. Among the four countries analyzed (Chile, Ecuador, Mexico, and Peru), only Ecuador and Mexico displayed significant disparities, with excess deaths exceeding reported COVID-19 fatalities. Chile's data showed parity between excess and reported deaths. These findings align with those of The Economist (2021), which employed a linear fixed effects model. However, the model's performance on Peru's data was suboptimal, yielding inconclusive results.

6.2 Acknowledging Limitations and Recommendations for Future Research

A primary limitation of this study lies in not treating each Latin American country as an independent entity, each with its distinct epidemiological and socio-economic context. Future research could be enhanced by incorporating country-specific Bayesian priors, reflecting factors such as GDP per capita, healthcare infrastructure quality, and corruption indices. This would improve the model's predictive accuracy and offer a nuanced analysis beyond broad regional generalizations.

Additionally, exploring alternative modeling approaches, such as Generalized Additive Models (GAMs), ARIMA, or exponential smoothing, could yield valuable comparative insights. Comparing the model's findings on excess deaths with reported COVID-19 deaths in OECD countries like the U.S. and Canada might also serve as a useful benchmark for evaluating the model's accuracy and generalizability.

References

- CDC, Centers for Disease Control and Prevention (2022) *Excess Deaths Associated with COVID-19.* Retrieved: https://www.cdc.gov/nchs/nvss/vsrr/covid19/excess_deaths.htm#references
- Cheng X, Yang Y, Schwebel DC, Liu Z, Li L, Cheng P, et al. (2020) *Population ageing and mortality during 1990–2017: A global decomposition analysis.* PLoS Med 17(6): e1003138. https://doi.org/10.1371/journal.pmed.1003138
- Daniel Simpson, Håvard Rue, Andrea Riebler, Thiago G. Martins, Sigrunn H. Sørbye. "Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors." Statistical Science, 32(1) 1-28 February 2017. https://doi.org/10.1214/16-STS576
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice.* (2nd ed.) OTexts. https://otexts.org/fpp2/
- National Academies of Sciences, Engineering, and Medicine. (2021). *High and Rising Mortality Rates Among Working-Age Adults.* Washington, DC: The National Academies Press.https://doi.org/10.17226/25976.
- H. Rue, S. Martino, and N. Chopin. *Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). Journal of the Royal Statistical Society, Series B, 71(2):319{392, 2009.*
- Litewka, S. G., & Heitman, E. (2020). *Latin American healthcare systems in times of pandemic. Developing world bioethics, 20(2), 69–73.* https://doi.org/10.1111/dewb.12262
- The Economist (2021) *Tracking covid-19 excess deaths across countries.* Retrieved: https://www.economist.com/graphic-detail/coronavirus-excess-deaths-tracker
- Gómez-Rubio, Virgilio (2020). Bayesian Inference with INLA. Chapman & Hall/CRC Press. Boca Raton, FL. Retrieved: https://becarioprecario.bitbucket.io/inla-gitbook/index.html
- WHO (2020) *WHO Timeline - COVID-19.* Retrieved: https://www.who.int/news/item/27-04-2020-who-timeline---covid-19

Appendix

```r
# importing libraries
library(tidyverse)
library(INLA)
library(Pmisc)

chile <- read.csv("https://raw.githubusercontent.com/TheEconomist/covid-19-excess-deaths-tracker/master/ou
ecuador <- read.csv("https://raw.githubusercontent.com/TheEconomist/covid-19-excess-deaths-tracker/master/
peru <- read.csv("https://raw.githubusercontent.com/TheEconomist/covid-19-excess-deaths-tracker/master/out
mexico <- read.csv("https://raw.githubusercontent.com/TheEconomist/covid-19-excess-deaths-tracker/master/o

chile_plt <- chile %>% ggplot(aes(x= as.Date(start_date)) ) +
          geom_point(aes(y= total_deaths, color="total_deaths")) +
          geom_line(aes(y= total_deaths, color="total_deaths")) +
          geom_point(aes(y= covid_deaths, color="official covid_deaths")) +
          geom_line(aes(y= covid_deaths, color="official covid_deaths")) +
          xlab("date") + ylab("weekly deaths") + labs(title="Chile", color="") +
          theme(legend.position="bottom")

ecuador_plt <- ecuador %>% ggplot(aes(x= as.Date(start_date)) ) +
          geom_point(aes(y= total_deaths, color="total deaths")) +
          geom_line(aes(y= total_deaths, color="total deaths")) +
          geom_point(aes(y= covid_deaths, color="official covid deaths")) +
          geom_line(aes(y= covid_deaths, color="official covid deaths")) +
          xlab("date") + ylab("weekly deaths") + labs(title="Ecuador", color="") +
          theme(legend.position="bottom")

mexico_plt <- mexico %>% ggplot(aes(x= as.Date(start_date)) ) +
          geom_point(aes(y= total_deaths, color="total deaths")) +
          geom_line(aes(y= total_deaths, color="total deaths")) +
          geom_point(aes(y= covid_deaths, color="official covid deaths")) +
          geom_line(aes(y= covid_deaths, color="official covid deaths")) +
          xlab("date") + ylab("weekly deaths") + labs(title="Mexico", color="")+
          theme(legend.position="bottom")

peru_plt <- peru %>% ggplot(aes(x= as.Date(start_date)) ) +
          geom_point(aes(y= total_deaths, color="total deaths")) +
          geom_line(aes(y= total_deaths, color="total deaths")) +
          geom_point(aes(y= covid_deaths, color="official covid deaths")) +
          geom_line(aes(y= covid_deaths, color="official covid deaths")) +
          xlab("date") + ylab("weekly deaths") + labs(title="Peru", color="")+
          theme(legend.position="bottom")

cowplot::plot_grid(chile_plt, ecuador_plt, mexico_plt, peru_plt)

generate_predicted_deaths <- function(x){
  x$time <- as.Date(x$start_date)
  x$dead <- x$total_deaths

  dateCutoff = as.Date('2020/3/1')
  xPreCovid = x[x$time < dateCutoff, ]
  xPostCovid = x[x$time >= dateCutoff, ]
  toForecast = expand.grid(time = unique(xPostCovid$time), dead = NA)

  xForInla = rbind(xPreCovid[,colnames(toForecast)], toForecast)
  xForInla= xForInla[order(xForInla$time), ]
```

```r
  xForInla$timeNumeric = as.numeric(xForInla$time)
  xForInla$timeForInla = (xForInla$timeNumeric)/365.25
  xForInla$timeIid = xForInla$timeNumeric
  xForInla$sin12 = sin(2*pi*xForInla$timeNumeric/365.25)
  xForInla$sin6 = sin(2*pi*xForInla$timeNumeric*2/365.25)
  xForInla$cos12 = cos(2*pi*xForInla$timeNumeric/365.25)
  xForInla$cos6 = cos(2*pi*xForInla$timeNumeric*2/365.25)

  res = inla(dead ~ sin12 + sin6 + cos12 + cos6 +
      f(timeIid, model='iid', prior='pc.prec', param= c(0.01, 0.5)) +
      f(timeForInla, model = 'rw2', scale.model=FALSE,
        prior='pc.prec', param= c(0.01, 0.5)),
    data=xForInla,
    control.predictor = list(compute=TRUE, link=1),
    control.compute = list(config=TRUE),
  #  control.inla = list(fast=FALSE, strategy='laplace'),
    family='poisson')
  qCols = paste0(c('0.5', '0.025', '0.975'), 'quant')
  res_bind = rbind(res$summary.fixed[,qCols], Pmisc::priorPostSd(res)$summary[,qCols])
  res_bind_inla = cbind(xForInla, res$summary.fitted.values[,qCols],
                        total_deaths=x$total_deaths, covid_deaths = x$covid_deaths)
  return(list(res, res_bind_inla, res_bind, xForInla))
}

chile_predicted_deaths = generate_predicted_deaths(chile)
ecuador_predicted_deaths = generate_predicted_deaths(ecuador)
mexico_predicted_deaths = generate_predicted_deaths(mexico)
peru_predicted_deaths = generate_predicted_deaths(peru)


chile_inla_plt = chile_predicted_deaths[[2]] %>%
  ggplot(aes(x=time)) +
  geom_line(aes(y=`0.5quant`, color='predicted deaths')) +
  geom_line(aes(y=`0.025quant`, color='95% CI'), linetype="dotted") +
  geom_line(aes(y=`0.975quant`, color='95% CI'), linetype="dotted") +
  geom_point(aes(y=total_deaths, color='actual deaths'), size=0.5) +
  scale_color_manual(values=c("grey50", "red", "black"))+
  xlab("date") + ylab("weekly deaths") + labs(title="Chile", color="")+
  theme(legend.position="bottom")

ecuador_inla_plt = ecuador_predicted_deaths[[2]] %>%
  ggplot(aes(x=time)) +
  geom_line(aes(y=`0.5quant`, color='predicted deaths')) +
  geom_line(aes(y=`0.025quant`, color='95% CI'), linetype="dotted") +
  geom_line(aes(y=`0.975quant`, color='95% CI'), linetype="dotted") +
  geom_point(aes(y=total_deaths, color='actual deaths'), size=0.5) +
  scale_color_manual(values=c("grey50", "red", "black"))+
  xlab("date") + ylab("weekly deaths") + labs(title="Ecuador", color="")+
  theme(legend.position="bottom")

mexico_inla_plt = mexico_predicted_deaths[[2]] %>%
  ggplot(aes(x=time)) +
  geom_line(aes(y=`0.5quant`, color='predicted deaths')) +
  geom_line(aes(y=`0.025quant`, color='95% CI'), linetype="dotted") +
  geom_line(aes(y=`0.975quant`, color='95% CI'), linetype="dotted") +
  geom_point(aes(y=total_deaths, color='actual deaths'), size=0.5) +
  scale_color_manual(values=c("grey50", "red", "black"))+
  xlab("date") + ylab("weekly deaths") + labs(title="Mexico", color="")+
```

```r
  theme(legend.position="bottom")

peru_inla_plt = peru_predicted_deaths[[2]] %>%
  ggplot(aes(x=time)) +
  geom_line(aes(y=`0.5quant`, color='predicted deaths')) +
  geom_line(aes(y=`0.025quant`, color='95% CI'), linetype="dotted") +
  geom_line(aes(y=`0.975quant`, color='95% CI'), linetype="dotted") +
  geom_point(aes(y=total_deaths, color='actual deaths'), size=0.5) +
  scale_color_manual(values=c("grey50", "red", "black"))+
  xlab("date") + ylab("weekly deaths") + labs(title="Peru", color="")+
  theme(legend.position="bottom")
cowplot::plot_grid(chile_inla_plt, ecuador_inla_plt, mexico_inla_plt, peru_inla_plt)

sample_deaths_from_posterior <- function(res, res_bind_inla, n_sim=10){
  sampleList = INLA::inla.posterior.sample(n_sim, res, selection = list(Predictor=0))
  sampleIntensity = exp(do.call(cbind,
    Biobase::subListExtract(sampleList, 'latent')))
  sampleDeaths = matrix(rpois(length(sampleIntensity), sampleIntensity),
    nrow(sampleIntensity), ncol(sampleIntensity))
  sample_deaths_df = cbind(sampleDeaths, res_bind_inla)
  sample_deaths_df_wide = sample_deaths_df %>%
    pivot_longer(1:n_sim, names_to = "simulation", values_to = "predicted_deaths")
  return(sample_deaths_df_wide)
}

sample_deaths_chile = sample_deaths_from_posterior(chile_predicted_deaths[[1]], chile_predicted_deaths[[2]
sample_deaths_ecuador = sample_deaths_from_posterior(ecuador_predicted_deaths[[1]], ecuador_predicted_deat
sample_deaths_mexico = sample_deaths_from_posterior(mexico_predicted_deaths[[1]], mexico_predicted_deaths[
sample_deaths_peru = sample_deaths_from_posterior(peru_predicted_deaths[[1]], peru_predicted_deaths[[2]])

chile_posterior_plt = sample_deaths_chile %>%
  ggplot(aes(x=time)) +
  geom_line(aes(y=predicted_deaths, group=simulation, color="predicted deaths"))  +
  geom_point(aes(y=total_deaths, color='actual deaths'), size=0.5) +
  scale_color_manual(values=c("red", "grey50"))+
  xlab("date") + ylab("weekly deaths") + labs(title="Chile", color="")+
  theme(legend.position="bottom")

ecuador_posterior_plt = sample_deaths_ecuador %>%
  ggplot(aes(x=time)) +
  geom_line(aes(y=predicted_deaths, group=simulation, color="predicted deaths"))  +
  geom_point(aes(y=total_deaths, color='actual deaths'), size=0.5) +
  scale_color_manual(values=c("red", "grey50"))+
  xlab("date") + ylab("weekly deaths") + labs(title="Ecuador", color="")+
  theme(legend.position="bottom")

mexico_posterior_plt = sample_deaths_mexico %>%
  ggplot(aes(x=time)) +
  geom_line(aes(y=predicted_deaths, group=simulation, color="predicted deaths"))  +
  geom_point(aes(y=total_deaths, color='actual deaths'), size=0.5) +
  scale_color_manual(values=c("red", "grey50"))+
  xlab("date") + ylab("weekly deaths") + labs(title="Mexico", color="")+
  theme(legend.position="bottom")

peru_posterior_plt = sample_deaths_peru %>%
  ggplot(aes(x=time)) +
  geom_line(aes(y=predicted_deaths, group=simulation, color="predicted deaths"))  +
  geom_point(aes(y=total_deaths, color='actual deaths'), size=0.5) +
```

```r
  scale_color_manual(values=c("red", "grey50"))+
  xlab("date") + ylab("weekly deaths") + labs(title="Peru", color="")+
  theme(legend.position="bottom")
cowplot::plot_grid(chile_posterior_plt, ecuador_posterior_plt, mexico_posterior_plt, peru_posterior_plt)

generate_excess_deaths_from_posterior <- function(sample_deaths){
  excess_deaths = sample_deaths %>% filter(time > as.Date('2020/3/1')) %>%
  mutate(excess_deaths=total_deaths - predicted_deaths)
  return(excess_deaths)
}

excess_deaths_chile = generate_excess_deaths_from_posterior(sample_deaths_chile)
excess_deaths_ecuador = generate_excess_deaths_from_posterior(sample_deaths_ecuador)
excess_deaths_mexico = generate_excess_deaths_from_posterior(sample_deaths_mexico)
excess_deaths_peru = generate_excess_deaths_from_posterior(sample_deaths_peru)

excess_deaths_chile_plt = excess_deaths_chile %>%
  ggplot(aes(x=time)) +
  geom_line(aes(y=excess_deaths, group=simulation, color="estimated excess deaths"))  +
  geom_line(aes(y=covid_deaths, color="official covid deaths"))  +
  scale_color_manual(values=c("grey50", "red"))+
  xlab("date") + ylab("weekly deaths") + labs(title="Chile", color="")+
  theme(legend.position="bottom")

excess_deaths_ecuador_plt = excess_deaths_ecuador %>%
  ggplot(aes(x=time)) +
  geom_line(aes(y=excess_deaths, group=simulation, color="estimated excess deaths"))  +
  geom_line(aes(y=covid_deaths, color="official covid deaths"))  +
  scale_color_manual(values=c("grey50", "red"))+
  xlab("date") + ylab("weekly deaths") + labs(title="Ecuador", color="")+
  theme(legend.position="bottom")

excess_deaths_mexico_plt = excess_deaths_mexico %>%
  ggplot(aes(x=time)) +
  geom_line(aes(y=excess_deaths, group=simulation, color="estimated excess deaths"))  +
  geom_line(aes(y=covid_deaths, color="official covid deaths"))  +
  scale_color_manual(values=c("grey50", "red"))+
  xlab("date") + ylab("weekly deaths") + labs(title="Mexico", color="")+
  theme(legend.position="bottom")

excess_deaths_peru_plt = excess_deaths_peru %>%
  ggplot(aes(x=time)) +
  geom_line(aes(y=excess_deaths, group=simulation, color="estimated excess deaths"))  +
  geom_line(aes(y=covid_deaths, color="official covid deaths"))  +
  scale_color_manual(values=c("grey50", "red"))+
  xlab("date") + ylab("weekly deaths") + labs(title="Peru", color="")+
  theme(legend.position="bottom")

cowplot::plot_grid(excess_deaths_chile_plt, excess_deaths_ecuador_plt,
                   excess_deaths_mexico_plt, excess_deaths_peru_plt)
```