COVID19-Tracker: A shiny app to produce to produce comprehensive data visualization for SARS-CoV-2 epidemic in Spain

Aurelio Tobías[1], Joan Valls[2], Pau Satorra[3], Cristian Tebé[3]

(equal contribution)

1. Institute of Environmental Assessment and Water Research (IDEA), Spanish Council for Scientific Research (CSIC), Barcelona, Spain. (aurelio.tobias@idaea.csic.es)
2. Department of Mathematics, Universitat Autònoma de Barcelona (UAB), Bellaterra, Barcelona, Spain.
3. Biostatistics Unit, Institut d'Investigació Biomèdica de Bellvitge (IDIBELL), Barcelona, Spain.

(Date: March 31st, 2020)

Abstract

Data visualization is an important tool for exploring and communicating findings in medical research, and specially in epidemiological surveillance. The COVID19-Tracker app systematically produces daily updated data visualization and analysis of SARS-CoV-2 epidemic in Spain. It collects automatically daily data on COVID-19 diagnosed cases, intensive care unit admissions, and mortality, from February 24th, 2020 onwards. Two applications have already been developed; 1) to analyze data trends and estimating short-term projections; and 2) To assess the effect of the lockdown on the trend of incident data. We are currently planning to improve the app by uploading shortly new applications for data visualization and analysis, which may help for a better understanding of the SARS-CoV-2 epidemic data in Spain.

1. INTRODUCTION

The first confirmed cases of SARS-CoV-2 in Spain were identified in late February 2020 (1). Since then, Spain became, by the end of March, the third most affected country worldwide after the United States and Italy and recorded the second number of deaths due to the SARS-CoV-2 pandemic after Italy (2). Since March 16th, lockdown measures oriented on flattening the epidemic curve were in place in Spain, restricting social contact, reducing public transport, and closing businesses, except for those essential to the country's supply chains (3). However, this has not been enough to change the rising trend of the epidemic. For this reason, a more restrictive lockdown was suggested (4), and eventually undertaken by the Spanish Government on March 30th (5).

Data visualization is an important tool for exploring and communicating findings in medical research, and specially in epidemiological surveillance. It can help researchers and policy makers to identify and understand trends that could be overlooked if the data were reviewed in tabular form. We have developed a Shiny app allows users to evaluate daily time-series data from a statistical standpoint. The COVID19-Tracker app systematically produces daily updated data visualization and analysis of SARS-CoV-2 epidemic data in Spain. It is easy to use and fills a role in the tool space for visualization, analysis and exploration of epidemiological data during this particular scenario.

2. SOFTWARE AVAILABILITY AND REQUIREMENTS

The COVID19-Track app has been developed in RStudio using the Shiny package (6). Shiny offers the ability to develop a graphical user interface (GUI) that can be run locally or deployed online. Last is particularly beneficial to show and communicate updated findings to a broad audience.

The app has a friendly structure based on menus to shown data visualization for each of the analyses currently implemetned: projections, intervention, and methodology sections (Figure 1).

- **Projections** displays a plot with trends for daily ICU admissions and mortality since the epidemic began and estimates a 3-day projection.
- **Intervention** displayes and calculates the effect of the lockdown period on the trend of incident data on daily diagnosed cases, ICU admissions, and mortality.
- **Methodology** shows the statistical details on the analyses implemented.

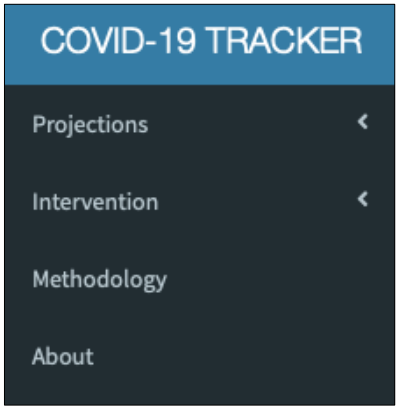

Figure 1. Menus currenlty available in the COVID19-Tracker app.

The app has an automated process to update data and all analyses every time a user connects to the app. It is available online at the following link: https://ubidi.shinyapps.io/covid19/ and shortly free available on GitHub as an R package.

3. DATA SOURCES

We collected daily data on COVID-19 diagnosed cases, intensive care unit (ICU) admissions, and mortality, from February 24th onwards. Data is collected automatically every day daily from Datadista github repository (7). This repository updates data according to the calendar and rate of publication of the Spanish Ministry of Health/Instituto de Salud Carlos III (8).

Data corresponding to the available number of ICU beds in Spain (year 2017) are also obtained from the Datadista github repository (7).

4. APPLICATIONS

4.1. Projections

For the evaluation of the observed trends of the accumulated number of cases, we used a classical quasi-Poisson regression model (9), allowing for over-dispersion and with a logarithmic link function, evaluating the existence of a quadratic effect. The two models are described as follows:

$$\text{Model 1: } \log(E(c_t))=\beta_0+\beta_1 t$$

$$\text{Model 2: } \log(E(c_t))=\beta_0+\beta_1 t+\beta_2 t^2$$

where t = 1, 2, …, T represents the time unit (from the first observed day until the last, T consecutive days in total), and it assumes that c_t, the observed cases, are distributed following a quasi-Poisson probability law. Estimated parameters and their standard error are used to obtain the predictions in the observed period of time but also the short-term projections, computing 95% confidence interval (95%CI)or the expected number of cases.

The analyses have been carried out using R version 3.6.3.

This analysis is accessible on the *Projections* menu, displaying short-term projections up to 3 days for COVID19 diagnosed cases, ICUs, and mortality in a time-series plot (Figure 2). Results are available nationwide by default, but also at the regional level, allowing a dropdown menu for this purpose. In addition, the produced graph is mouse-sensitive, showing the exact number of observed and predicted/projected cases for both models through the time-series.

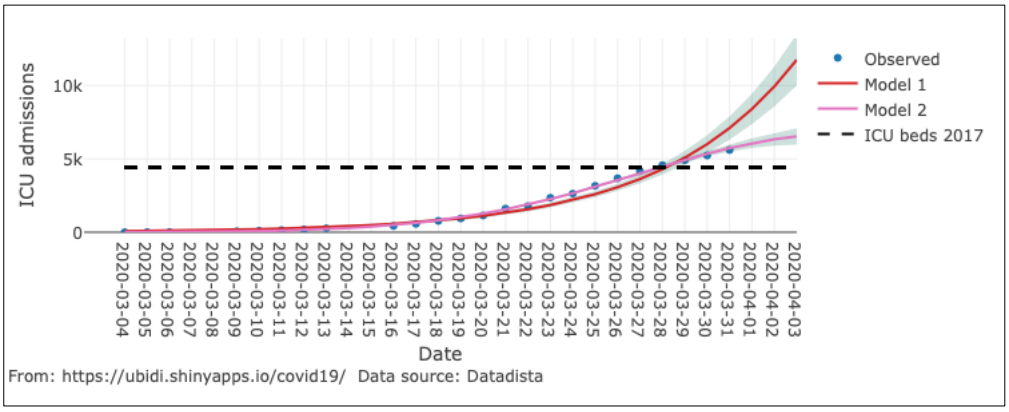

Figure 2. Standard output for the projections analysis.

4.2. Intervention analysis

To assess the effect of the lockdown on the trend of incident cases, admissions in ICU intensive care units, and mortality, we used an interrupted time-series design (10). The data is analyzed with quasi-Poisson regression with an interaction model to estimate the change in trend:

$$\log(E(c_t))=\beta_0+\beta_1 t+\beta_2 lockdown+\beta_3 t*lockdown$$

where *t* =1, 2, …, T represents the time unit (from the first observed day until the last, T consecutive days in total); and *lockdown* is a binary variable that identifies the periods before after the alarm status decree (0 = before Mar 15th, 2020; 1 = after Mar 16th, 2020).

The analyses have been carried out using R, version 3.6.3.

We should acknowledge that this is a descriptive analysis without predictive purposes. For an easy interpretation, and comparison of the effectiveness of lockdown measures between countries, a linear trend is assumed before and after the lockdown. The changes in the definition of diagnosed cases have not been taken into account, nor has the reduction in the susceptible population because of the lockdown. Therefore, the incident cases are modelled directly instead of the incidence rate, assuming that the entire population is at risk. Although not accounted for residual autocorrelation, the estimates are unbiased but possibly inefficient.

This analysis is accessible on the *Intervention* menu, displaying trends in a time-series plot before and after the lockdown for COVID19 diagnosed cases, ICUs, and mortality (Figure 3). The daily percentage (%) mean increase and its 95%CI are also reported. Results are available nationwide.

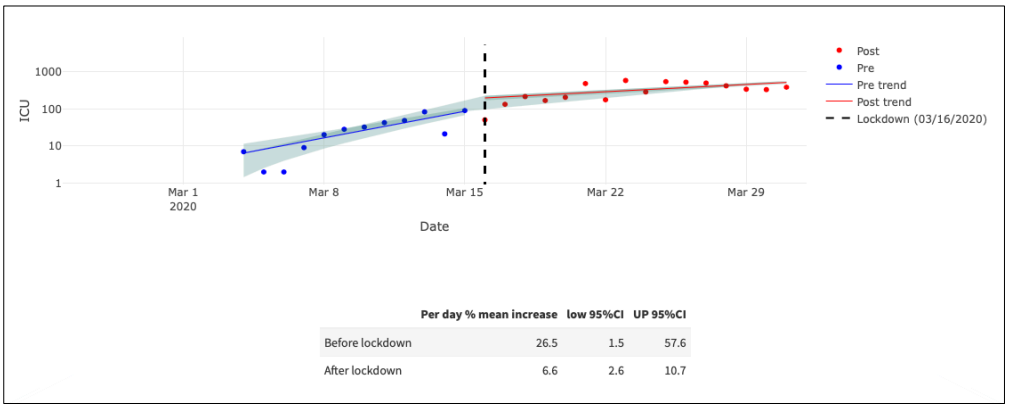

Figure 3. Standard output for the projections analysis.

5. Further developing

So far, the COVID19-Tracker app has been very well received online, with a large number of connections generating an outsized memory usage on our server (Figure 4).

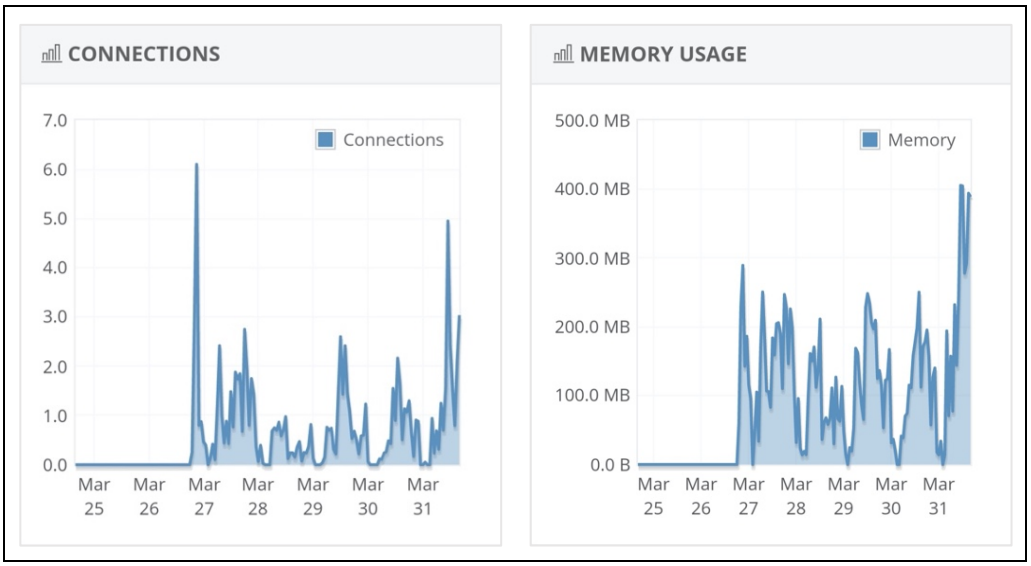

Figure 4. Number of connections and memory usage since March 27th, 2020.

We are currently planning to improve the app by uploading shortly new applications for data visualization and analysis, which may help for a better understanding of the SARS-CoV-2 epidemic data in Spain. Moreover, the COVID19-Tracker app could also be extensible to data visualizations across other countries and geographical regions.

Funding

None

Acknowledgements

None

Conflict of interest

None

References

1. Saglietto A, D'Ascenzo F, Zoccai GB, De Ferrari GM. COVID-19 in Europe: the Italian lesson. The Lancet 2020 (Published online March 24th). doi: 10.1016/S0140-6736(20)30690-5.
2. Ourworld in Data. Coronavirus Disease (COVID-19) Statistics and Research. Oxford Martin School, The University of Oxford, Global Change Data Lab; 2020. [Accessed March 30th, 2020]. Available from: https://ourworldindata.org/coronavirus/
3. Ministerio de la Presidencia, Relaciones con las Cortes y Memoria Democrática. Real Decreto 463/2020, de 14 de marzo, por el que se declara el estado de alarma para la gestión de la situación de crisis sanitaria ocasionada por el COVID-19. [Accessed March 17th, 2020]. Available from: https://www.boe.es/eli/es/rd/2020/03/14/463/con
4. Mitjà O, Arenas À, Rodó X, Tobias A, Brew J, Benlloch JM. Experts' request to the Spanish Government: move Spain towards complete lockdown. The Lancet 2020 (Published online March 26th). doi: 10.1016/S0140-6736(20)30753-4.
5. Ministerio de la Presidencia, Relaciones con las Cortes y Memoria Democrática. Real Decreto-ley 10/2020, de 29 de marzo, por el que se regula un permiso retribuido recuperable para las personas trabajadoras por cuenta ajena que no presten servicios esenciales, con el fin de reducir la movilidad de la población en el contexto de la lucha contra el COVID-19. [Accessed Mar 31st, 2020]. Available from: https://www.boe.es/buscar/doc.php?id=BOE-A-2020-4166
6. Team R. RStudio: Integrated Development for R. RStudio, Inc. Boston, MA; 2015.
7. Datadista. Extracción, limpieza y normalización de las tablas de la situación diaria acumulada de la enfermedad por el coronavirus SARS-CoV-2 (COVID-19) en España en un formato accesible y reutilizable 2020. [Accessed Mar 30th, 2020]. Available from: https://github.com/datadista/datasets/tree/master/COVID%2019
8. Instituto de Salud Carlos III. Situación de COVID-19 en España 2020. [Accessed Mar 30th, 2020]. Available from: https://covid19.isciii.es/
9. Dyba T, Hakulinen T. Comparison of different approaches to incidence prediction based on simple interpolation techniques. Statistics in Medicine. 2000;19(13):1741-52. doi:10.1002/1097-0258(20000715)19:13<1741::aid-sim496>3.0.co;2-o.
10. Bernal JL, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. Int J Epidemiol. 2017;46(1):348-55. doi: 10.1093/ije/dyw098.