Reconstructing the past true incidence of the Covid-19 epidemic

Arturo del Cerro[1], Sergio Villar[2], Jorge Piqueras[3], Saioa Extxebarria[4], ശാരിക വിജയകുമാർ (Sharika Vijayakumar)[5]. Supervised by Xavier Mora[6]

[1]arturo.delcerro@ e-campus.uab.cat, [2]sergio.villar@uab.cat, [3]jorge.piqueras@e-campus.uab.cat,
[4]saioa.etxebarria@e-campus.uab.cat,[5]sharika.vijayakumar@e-campus.uab.cat, [6]xmora@mat.uab.cat

Universitat Autònoma de Barcelona

Master in Modelling for Science and Engineering

Abstract

We reconstruct the past true incidence of the Covid-19 epidemics in contrast to the reported one. We base ourselves on a simple variation of the SIR model with an unknown time-varying contagion rate. We rely on the number of deaths together with assumed known values for two clinical parameters of the disease, namely the infection fatality ratio and the average time of illness. The procedure is applied to data for Catalonia and Spain in order to confirm and quantify the suspicion of underreported incidence and mortality of the COVID-19.

Introduction

Measuring the COVID-19 pandemic has been a tough challenge in all the fields it involves, such us epidemiology and medicine.

The lack of resources like PCRs to check for real cases in the first months means that the real cases must be higher than the reported ones.

What's more, considering that a significative part of the population has been asymptomatic, it is not trivial to estimate the total number of people who have undergone COVID-19.

Following a SIR model, we were able to reconstruct the behavior of COVID-19 pandemic obtaining an estimated count of the number of recoveries and the incidence.

In order to reconstruct those variables, we rely on the number of deaths.

Methodology

Our focus is to reconstruct the **I(t)** curve about infected population. To perform this work, we consider the daily deaths, recoveries and incidence, which we denote respectively by we used **ΔM**, **ΔH** and **ΔK**. We apply 7 days averages in order to drop the weekend underreporting effects.

All of them are proportional to **I** and related to it by **κ**, **γ** and **f**, the contagion rate, average duration of illness and infection fatality ratio, respectively. In this work, we started with the data about daily deaths **ΔM,** then computed **ΔH.** Up to this point **I** is computed by

$$I = \Delta M / \gamma f$$

and from there, **ΔI.** More specifically, the equations for our model are

$$\Delta I = \Delta K - \Delta H - \Delta M$$

where:

$$\Delta K = \kappa I, \quad \Delta H = \gamma (1-f) I, \quad \Delta M = \gamma f I$$

We can also reconstruct the reproduction number R, that is the average number of infections that derive from one infected individual, regardless of when do the contagions take place. So, **R** is a dimensionless multiplication factor.

For the epidemic to die out, it is essential to have **R** < 1, the lower the better. From the definition of **R,** we have that **R = κ / γ,** and from the equations above **R** can be expressed in the form of:

$$R = \Delta K / (\Delta H + \Delta M)$$

Where the right-side quantities are known to us.

A more refined model assumes

$$\Delta M(t) = \gamma f I(t- \delta)$$

where the delay **δ** will be assumed to coincide with the average duration of illness, namely

$$\delta = 1 / \gamma$$

Materials

The data used to the present work was extracted from:

- [1] MoMo (Daily Mortality Monitoring System) reports from *Instituto de Salud Carlos III*. These data give the excess mortality, i.e. the number of deaths from all causes minus its average in the last years.
- *[2] Departament de Salut de la Generalitat de Catalunya*. https://dadescovid.cat/
- [3] Centro Nacional de Epidemiología. https://cnecovid.isciii.es/
- [4] Escovid19data project. https://github.com/montera34/escovid19data

The deaths data for Spain and Catalonia were taken from [1] until April 17. After that date, the direct data from [2, 3, 4] seem to be reliable enough.

Results

Figure 1 shows the **ΔM** data where we can highlight that in the first wave more than 45000 people died due to the COVID-19 in Spain, accordingly to the INE***1** estimation results and more than 20000 died in Catalonia during the whole pandemic.

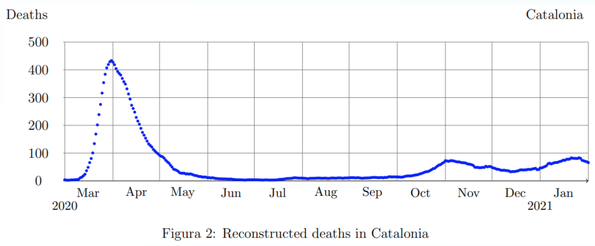

Figura 2: Reconstructed deaths in Catalonia

Taking values for δ = 1 / γ = 7 days and f = 1.15% ***2** , we computed the reconstructed daily incidence ΔK shown in figure 2 compared with the reported daily incidence.

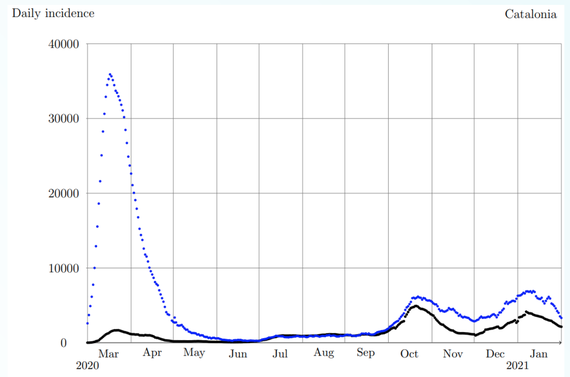

Figura 1: Comparison between reconstructed daily incidence (blue) and reported daily incidence (black) in Catalonia

Results show how first wave had 20 times more people infected in Catalonia. Some other works, based on seroprevalence data, estimate in Spain this number as 10 ***3.**.

Finally, figure 3 shows the resulting temporal evolution of the reproduction number R in comparison with that which is obtained from the reported daily incidence by means of the Cislaghi***4** method.

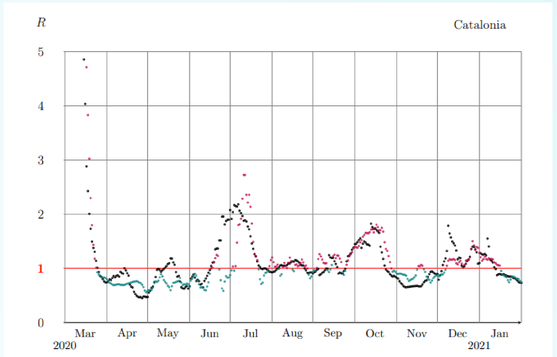

Figure 3: Reproduction number in Catalonia. In black, Cislaghi method from the reported cases. In green and red, the method presented in this paper.

The differences between both curves of figure 3 mean that policy decisions based on the deaths-based reconstructed incidence could have been different from those that have been based on the reported incidence.

Conclusion

Assuming a known value for a couple of clinical parameters of the disease, namely the infection fatality rate and the average duration of illness, the temporal evolution of the Covid epidemic can be reconstructed from the deaths data.

In so doing, it is advisable to rely not only on the reported Covid deaths but also on the excess deaths above the average of the preceding years. This is especially suitable for the starting phase of the epidemic, where Covid cases where not always recognized as such.

In the cases of Catalonia and Spain, the reconstructed incidence curve reaches much larger values than the reported one, especially in the first wave of the epidemic, when grave cases were the only reported ones. In that situation the reconstructed values differ from the reported ones by a factor of 20 in Catalonia and 16 in Spain.

The reconstructed temporal evolution of the epidemic may result in different values of the time-dependent reproduction number, which might lead to different policy decisions.

1: https://www.eldiario.es/datos/45-000-personas-murieron-causa-covid-primera-ola-pandemia-espana_1_6494110.html
2: Infection fatality rate of COVID-19 inferred from seroprevalence data. John P. Ioannidis. Bulletin of the World Health Organization
3: https://www.datadista.com/coronavirus/estimacion-diagnostico-segunda-ola-covid19/
4: Effective Reproduction Number Estimation from Data Series. A. Annunziato, T. Asikainen. Joint Research Centre.