

Comparison of methodologies for estimating the effective R associated with COVID-19 in Chile, by commune and week

Iván Gutiérrez¹, Luis Gutiérrez¹, Jessica Pavani¹, Leonardo Jofré¹, Inés Varas², Oscar Ortiz³, Jaime Cerda⁴, Danilo Garrido¹, and Gabriel Arriagada³

¹*Facultad de Matemáticas, Departamento de Estadística, Pontificia Universidad Católica de Chile*

²*Núcleo Milenio Centro para el Descubrimiento de Estructuras en Datos Complejos (MiDaS)*

³*Instituto de Ciencias Agroalimentarias, Animales y Ambientales – ICA3, Universidad de O’Higgins*

⁴*Facultad de Medicina, Departamento de Salud Pública, Pontificia Universidad Católica de Chile*

Resumen

This work review six methods for estimating the effective R of the COVID-19 pandemic in relatively small areas. The methods were illustrated with series of new cases in Chile. Effective R was estimated at the communal level with a weekly frequency. We found that all methods yielded noisy results, but some were more informative than others. In particular, we find that methods based on the Euler-Lotka equation work relatively well, even for relatively small Chilean communes, in the sense that the estimated effective R tracked the general trends of the pandemic. We then argue that while noisy, these estimates can still be helpful if employed to measure the general trend of the effective R over the communes or employed as part of a more extensive statistical model. For instance, we modeled the estimated effective R series using a Generalized Linear Model aimed to evaluate the effect of the Chilean government’s Step-by-Step plan –a gradual strategy to face the pandemic according to the health situation of each particular commune– on the spread of COVID-19 in Chile. We found that each step change towards tightening the measures reduced the value of the effective R.

1. Introduction

The recent COVID-19 pandemic has created a growing need for indicators of how fast the pandemic is spreading. Among these, one of the most popular indicators is the effective reproduction number or *effective R*. The effective R corresponds to the average number of people infected an infectious person, in a population where some people may have protection against the disease. In Chile, the effective R is periodically calculated by the Center for Mathematical Modeling (CMM) and can be viewed at the regional level¹ through a dashboard².

On the other hand, in July 2020, the Chilean government initiated the so-called *Step by Step* plan; a plan of dynamic sanitary restrictions (including lockdowns) at the communal level³. The official government website describes the plan as follows: ⁴

“The Step by Step Plan is a gradual strategy to face the pandemic according to the health situation of each particular area. These are 5 stages or gradual steps, ranging from Quarantine to Advanced Opening, with specific restrictions and obligations. The advance or retreat from one particular step to another is subject to epidemiological indicators, health care network and traceability.”

¹Chile is made up of 16 regions.

²Available in <https://covid-19vis.cmm.uchile.cl/geo>.

³Each region is made up of a variable number of communes, with 346 communes in total.

⁴Viewed at <https://www.gob.cl/coronavirus/pasoapaso/>.

Naturally, the implementation of this plan requires timely epidemiological information at a communal level. However, the disaggregation of indicators such as the effective R tends to stop at the regional level. This is because the higher the level of disaggregation, the higher the level of imprecision in the estimates of the effective R. The degree of disaggregation under discussion can be better understood with some figures: according to the projections of the National Institute of Statistics (INE) of Chile for 2020, the median population of the regions is 763,975 inhabitants, while the median population of the communes is only 20,071 inhabitants.

2. Objective

The main objective of this study is to calculate the effective R for COVID-19 in Chile at the communal/weekly level using different methodologies, and to evaluate which, despite their inherent volatility, can continue to provide useful and timely information. Once the effective R has been estimated using the most pertinent methodology, we illustrate how it can be used as a response variable in an observational study to estimate the effect of the different phases of the *Step by Step* plan on the effective R.

3. Methodology

Six methodologies were compared, among them, the [Systrom \(2020\)](#); [Cislaghi \(2020\)](#); [Asikainen and Annunziato \(2020\)](#); [Wallinga and Lipsitch \(2007\)](#); [Martinez-Beneito et al. \(2020\)](#) methodologies and the [Robert Koch Institut \(2020\)](#) methodology. The [Systrom \(2020\)](#) method is based on a Bayesian State-Space model where each new case is an observation and each effective R is a state, which allows to infer the effective R using any Bayesian inference software. The [Cislaghi \(2020\)](#) method assumes that the effective R can be reasonably estimated by the percentage growth of the number of active cases, to the extent that they are previously smoothed. The [Asikainen and Annunziato \(2020\)](#) method assumes that the effective R can be reasonably approximated as the solution of a difference equation that relates the effective R to the number of new cases, and then solves that equation under the assumption that the R-cash may not change much in short periods of time. The [Robert Koch Institut \(2020\)](#) method is similar to Cislaghi's, differing only in how they smooth out the new case series. The [Wallinga and Lipsitch \(2007\)](#) method assumes that effective R can be reasonably estimated as the solution of a difference equation that relates effective R to generation time (i.e., the average time between infection events in a chain of infections). Finally, the [Martinez-Beneito et al. \(2020\)](#) method is based on a Bayesian spatial model where the new cases are the responses and the effective R is a function of several latent variables. In each case, the effective R series were calculated from the new cases in each week and commune (according to the onset of symptoms), a series provided by the Ministry of Science, Technology, Knowledge and Innovation (MINCIENCIA).⁵

Once all the methods were implemented, their stability and their ability to generate reasonable predictions were compared (it does not make sense, for example, an effective R greater than 10 or lower than 0). Finally, the estimation of the effective effective R judged more reasonable was used as the response variable in a generalized linear model of the gamma family with random effects by commune and region, using the inverse function as the link function. Specifically, we consider the following model:

$$\begin{aligned} y_{rct} | x_{rct}, u_{rc}, v_r &\stackrel{ind}{\sim} \text{Gamma}(\nu, \text{rate} = \nu / \mu_{rct}), \\ \mu_{rct} &= (x'_{rct} \beta + u_{rc} + v_r)^{-1}, \\ u_{rc} &\stackrel{iid}{\sim} N(0, \sigma_u^2), \\ v_r &\stackrel{iid}{\sim} N(0, \sigma_v^2), \end{aligned} \tag{1}$$

where y_{rct} represents the effective R for the c -th commune within the r -th region during the t -th epidemiological week, u_{rc} is a random effect by region and commune, v_r is a random effect by region, and x_{rct} is a vector of explanatory variables, among them: the phase of the plan passed in the past week (as

⁵Specifically, we product 15 from the repository <https://github.com/MinCiencia/Datos-COVID19>.

a categorical variable), sociodemographic characteristics of each commune (e.g. a socioeconomic development index⁶, a rurality index⁷, percentage of the population in various age groups), characteristics related to connectivity of each commune (e.g. if it has a port or an airport), and the percentage of the population in neighboring communes that is in quarantine (acknowledging the fact that the levels of infections not only depend on the measures imposed on the communes, but also on their neighbors), and a base of B-splines (generated from the epidemiological week and with a node for each decile of this variable).⁸.

4. Results

Figure 1 presents the 50th percentile of the effective R, according to commune, in 6 contiguous maps (one per method). As can be seen, the medians estimated by all the methods take reasonable values. However, not all methods are equally stable or have the same dynamics. Figure 2 gives a good account of this. In it, it is possible to appreciate how some methods (such as Systrom's) give predictions out of range (specifically, negative), while other methods are either too volatile (such as Cislaghi's) or too stable (such as Martínez-Beneito's). In our opinion, the best performing method in general is Wallinga and Lipsitch (2007). In particular, it is the only method that generates predictions within reasonable ranges while its evolution over time is in line with the main milestones related to COVID-19. Meanwhile, Table 1 presents the estimate of β when using the previously estimated effective R (with the Wallinga method) as the response variable in the mixed generalized linear model described in Equation (1) (the coefficients associated with the splines base were omitted from the Table, but all were significant). In particular, we obtained that all the dummies associated with the Step by Step plan phase have associated negative and significant coefficients (see Table 1) which, due to the nature of the link, implies that each increase in the plan phase increases the effective R. Indeed, under this model, $E[y_{rct}|x_{rct}, u_{rc}, v_r] = (x'_{rct}\beta + u_{rc} + v_r)^{-1}$, so that $\partial E[y_{rct}|x_{rct}, u_{rc}, v_r]/\partial x_{rct,k} = -\beta_k/(x'_{rct}\beta + u_{rc} + v_r)^2$. Therefore, a $\beta_k < 0$ implies a positive marginal effect, which in pandemic terms implies an increase in the speed of contagion.

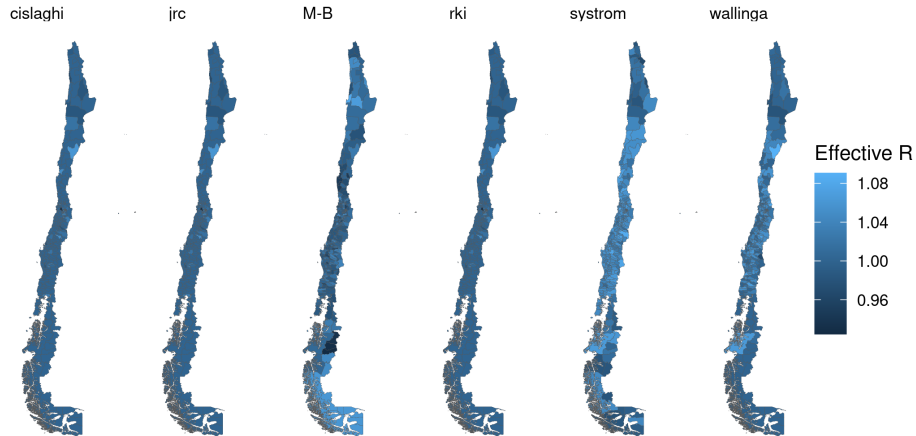


Figura 1: 50th percentile of the effective R, by commune and method.

⁶Developed by the Chilean Public Health Observatory.

⁷Developed by the Ministry of Social Development.

⁸Most of these variables are available in the repository <https://github.com/MinCiencia/Datos-COVID19>. In particular, we use products 7 (# of PCR tests / communal population, according to region and epidemiological week), 15 (new cases, according to commune and epidemiological week), 74 (phase in step-by-step plan, according to commune and week epidemiological), 29 (quarantine status, according to commune and epidemiological week). In the periods leading up to the start of the step-by-step plan, we treat a quarantine as a phase 1 and a non-quarantine as a phase 4. All details can be found in the long version of this report, available at , and all the codes necessary to reproduce the results are available in the repository <https://github.com/COVID0248/covid-re>

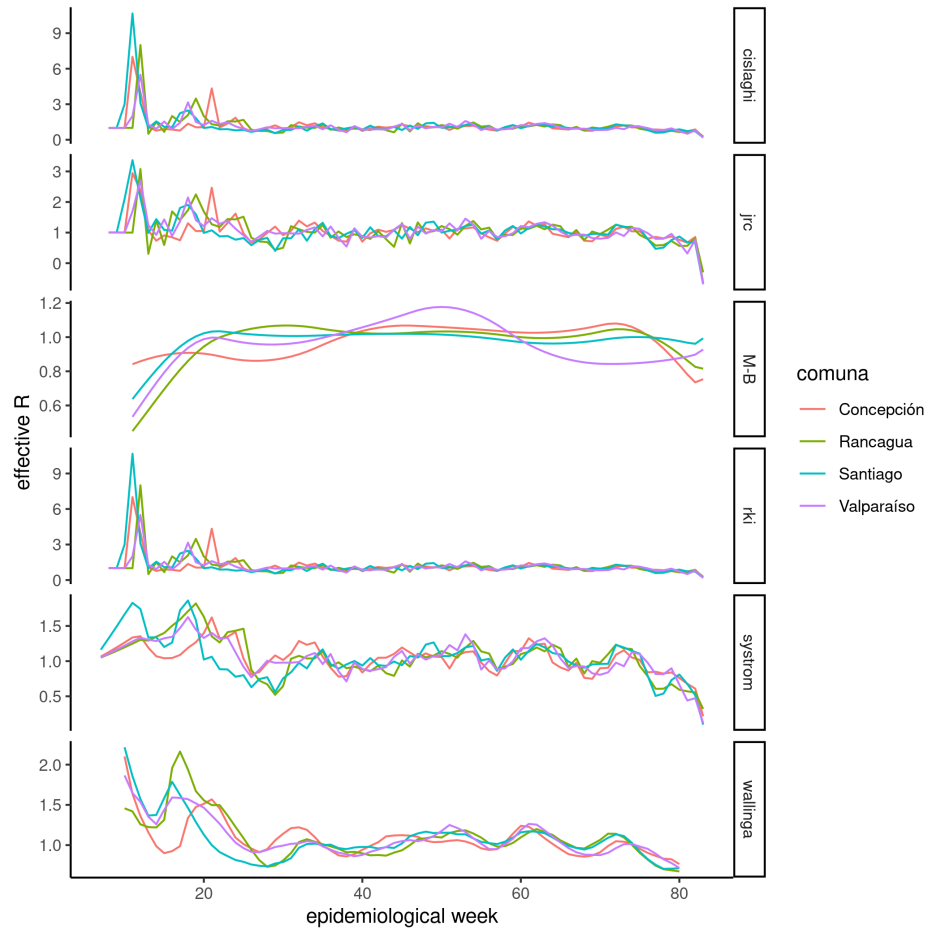


Figura 2: Effective R for 4 selected communes, by epidemiological week and method.

5. Preliminary conclusions

The higher the level of disaggregation, the more volatile the estimates of effective R become. Despite this, some of the methods examined may still yield sensitive estimates that timely reflect changes in trends within the pandemic. In particular, using the method proposed by Wallinga and Lipsitch (2007), it is possible to verify that the changes in the phases within the Step by Step plan do have effects on the effective R. In particular, each step change towards tightening the measures reduces the value of the communal effective R.

Referencias

- Asikainen, T. and Annunziato, A. (2020). Effective reproduction number estimation from data series. Technical report, Joint Research Centre (European Commission).
- Cislighi, C. (2020). Un cruscotto per monitorare l'evoluzione dell'epidemia.
- Martinez-Beneito, M. A., Mateu, J., and Botella-Rocamora, P. (2020). Spatio-temporal small area surveillance of the covid-19 pandemics. *arXiv e-prints*, page arXiv:2011.03938.
- Robert Koch Institut (2020). Epidemiologisches bulletin 17.
- Systrom, K. (2020). The metric we need to manage covid-19.

term	estimate	std.error	statistic	p.value
(Intercept)	0.47	0.04	12.56	0.00
socio-economic development index	0.06	0.03	2.30	0.02
rurality index	0.12	0.02	6.88	0.00
1(Step-by-Step plan phase 2) (lag 1)	-0.07	0.00	-18.72	0.00
1(Step-by-Step plan phase 3) (lag 1)	-0.16	0.00	-36.04	0.00
1(Step-by-Step plan phase 4) (lag 1)	-0.22	0.01	-40.55	0.00
% of neighbor population in quarantine (lag 1)	0.04	0.00	10.15	0.00
# of PCR tests / population size (lag 2)	1.00	0.34	2.96	0.00

Cuadro 1: Estimation summary for the fixed effects of the model employed for estimating the effects of Step-by-Step plan's phases on the effective R. The coefficients associated with the B-spline basis were omitted, but they were all significant.

Wallinga, J. and Lipsitch, M. (2007). How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B: Biological Sciences*, 274(1609):599–604.