

# Comparación de metodologías para la estimación del $R$ efectivo asociado al COVID-19 en Chile, según comuna y semana

Iván Gutiérrez<sup>1</sup>, Luis Gutiérrez<sup>1</sup>, Jessica Pavani<sup>1</sup>, Leonardo Jofré<sup>1</sup>, Inés Varas<sup>2</sup>, Oscar Ortiz<sup>3</sup>, Jaime Cerda<sup>4</sup>, and Gabriel Arriagada<sup>3</sup>

<sup>1</sup>*Facultad de Matemáticas, Departamento de Estadística, Pontificia Universidad Católica de Chile*

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

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

<sup>4</sup>*Facultad de Medicina, Departamento de Salud Pública, Pontificia Universidad Católica de Chile*

## Resumen

Este trabajo revisa seis métodos para estimar la  $R$  efectiva de la pandemia COVID-19 en áreas relativamente pequeñas. Los métodos se ilustran con una serie de casos nuevos en Chile. El  $R$  efectivo fue estimado a nivel comunal con una frecuencia semanal. Descubrimos que todos los métodos producen resultados ruidosos, pero algunos son más informativos que otros. En particular, encontramos que los métodos basados en la ecuación de Euler-Lotka funcionan relativamente bien, incluso para comunas chilenas relativamente pequeñas, en el sentido de que el  $R$  efectivo estimado sigue las tendencias generales de la pandemia. Luego argumentamos que, aunque son ruidosas, estas estimaciones pueden ser útiles si se emplean para medir la tendencia general del  $R$  efectivo en las comunas o se emplean como parte de un modelo estadístico más extenso. Por ejemplo, modelamos la serie del  $R$  efectivo estimado utilizando un Modelo Lineal Generalizado destinado a evaluar el efecto del plan Paso a Paso del gobierno chileno –una estrategia gradual para enfrentar la pandemia de acuerdo con la situación de salud de cada comuna en particular– sobre la propagación del COVID-19 en Chile. Descubrimos que cada cambio de paso hacia el endurecimiento de las medidas reduce el valor del  $R$  efectivo.

## 1. Introducción

La reciente pandemia de COVID-19 ha generado una necesidad cada vez mayor por indicadores de la velocidad de propagación de la pandemia. Entre estos, uno de los indicadores más populares es el número de reproducción efectiva o  $R$  efectiva. El  $R$  efectiva corresponde al número promedio de personas infectadas una persona infecciosa, en una población donde algunas personas pueden tener protección contra la enfermedad. En Chile, el  $R$  efectiva es periódicamente calculado por el Centro de Modelamiento Matemático (CMM) y puede ser visualizado, a nivel regional<sup>1</sup>, a través de un dashboard<sup>2</sup>.

Por otra parte, en julio del año 2020, el gobierno de Chile inició el plan *Paso a Paso*; una estrategia de restricciones sanitarias dinámica a nivel comunal<sup>3</sup>. La página oficial del gobierno describe el plan de la siguiente manera:<sup>4</sup>

*“El Plan Paso a Paso es una estrategia gradual para enfrentar la pandemia según la situación sanitaria de cada zona en particular. Se trata de 5 escenarios o pasos graduales, que van desde la Cuarentena hasta la Apertura Avanzada, con restricciones y obligaciones específicas. El avance o retroceso de un paso particular a otro está sujeta a indicadores epidemiológicos, red asistencial y trazabilidad.”*

---

<sup>1</sup>Chile se compone de 16 regiones.

<sup>2</sup>Disponible en <https://covid-19vis.cmm.uchile.cl/geo>.

<sup>3</sup>Cada región se compone de un número variable de comunas, habiendo 346 comunas en total.

<sup>4</sup>Consultado en <https://www.gob.cl/coronavirus/pasoapaso/>.

Naturalmente, la implementación de este plan requiere de información epidemiológica oportuna a un nivel comunal. Sin embargo, la desagregación de indicadores como el  $R$  efectivo suele detenerse en el nivel regional. Esto se debe a que a mayor nivel de desagregación, mayor nivel de imprecisión en las estimaciones del  $R$  efectivo. El grado de desagregación en discusión puede entenderse mejor con algunas cifras: según las proyecciones del Instituto Nacional de Estadísticas (INE) de Chile para 2020, la población mediana de las regiones es 763.975 habitantes, mientras que la población mediana de las comunas es de solo 20.071 habitantes.

## 2. Objetivo

El objetivo principal de este estudio es calcular el  $R$ -efectivo para el COVID-19 en Chile a nivel comunal/semanal usando distintas metodologías, y evaluar cuáles, pese a su inherente volatilidad, pueden seguir dando información útil y oportuna. Una vez estimado el  $R$  efectivo usando la metodología más pertinente, ilustramos cómo puede ser aprovechado utilizándolo como variable respuesta en un estudio observacional para estimar el efecto de las distintas fases del plan *Paso a Paso* sobre el  $R$  efectivo.

## 3. Metodología

Se compararon 6 metodologías, entre ellas las metodologías de Systrom (2020); Cislighi (2020); Asikainen and Annunziato (2020); Wallinga and Lipsitch (2007); Martinez-Beneito et al. (2020) y la metodología del Robert Koch Institut (2020). El método de Systrom (2020) se basa en un modelo de Estado-Espacio Bayesiano donde cada caso nuevo es una observación y cada  $R$ -efectivo es un estado, lo cual permite inferir los  $R$ -efectivos usando cualquier software de inferencia Bayesian (por ejemplo, Stan). El método de Cislighi (2020) asume que el  $R$ -efectivo puede ser razonablemente estimado por el crecimiento porcentual del número de casos activos, en la medida en que sean previamente suavizados. El método del Asikainen and Annunziato (2020) asume que el  $R$ -efectivo puede ser razonablemente aproximado como la solución de una ecuación diferencia que relaciona el  $R$ -efectivo con el número de casos nuevos, y luego resuelve dicha ecuación bajo el supuesto de que el  $R$ -efectivo no puede cambiar mucho en cortos períodos de tiempo. El método del Robert Koch Institut (2020) es similar al de Cislighi, diferenciándose solo en cómo ellos suavizan las series de casos nuevos. El método de Wallinga and Lipsitch (2007) asume que el  $R$ -efectivo puede ser razonablemente estimado como la solución de una ecuación en diferencia que relaciona al  $R$ -efectivo con el tiempo de generación (i.e., el tiempo promedio entre eventos de infección en una cadena de infecciones). Finalmente, el método de Martinez-Beneito et al. (2020) se basa en un modelo espacial Bayesiano donde los casos nuevos son las repuestas y el  $R$  efectivo es una función de varias variables latentes del mismo. En cada caso, las series del  $R$ -efectivo fueron calculadas a partir de los casos nuevos en cada semana y comuna (según el inicio de los síntomas), serie suministrada por el Ministerio de Ciencia, Tecnología, Conocimiento e Innovación (MINCIENCIA).<sup>5</sup>

Una vez implementados todos los métodos, se comparó su estabilidad y su capacidad para generar predicciones razonables (no tiene sentido, por ejemplo, que los valores del  $R$ -efectivo estimados sean superiores a 10 o negativos). Finalmente, la estimación del  $R$ -efectivo juzgada más razonable fue usada como variable respuesta en un modelo lineal generalizado de familia gamma con efectos eleatorios por comuna y región, usando la función inversa<sup>6</sup> como función de enlace. Esto es:

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

donde  $y_{rct}$  representa el  $R$ -efectivo para la  $c$ -ésima comuna dentro de la  $r$ -ésima región durante la  $t$ -ésima semana epidemiológica,  $u_{rc}$  es un efecto aleatorio por región y comuna,  $v_r$  es un efecto aleatorio

<sup>5</sup>Específicamente, el producto 15 en el repositorio <https://github.com/MinCiencia/Datos-COVID19>.

<sup>6</sup>Favor elaborar un poco más aquí, ¿A que te refieres con la función inversa?

por región, y  $x_{rct}$  es un vector de variables explicativas, entre ellas: la fase del plan paso en la semana pasada (como variable categórica), características sociodemográficas de cada comuna (e.g. un índice de desarrollo socioeconómico<sup>7</sup>, un índice de ruralidad<sup>8</sup>, porcentaje de la población en diversos tramos de edad), características relacionadas con la conectividad de cada comuna (e.g., si tiene un puerto o un aeropuerto), y el porcentaje de la población en las comunas vecinas que se encuentra en cuarentena (reconociendo el hecho de que los niveles de contagios no dependen solo de las medidas impuestas sobre las comunas, sino también sobre sus vecinos), y una base de B-splines (generados a partir de la semana epidemiológica y con un nodo por cada decil de esta variable).<sup>9</sup>.

## 4. Resultados

La Figura 1 presenta el percentil 50 del R efectivo, según comuna, en 6 mapas contiguos (uno por método). Como se puede apreciar, las medianas estimadas por todos los métodos toman valores razonables. Sin embargo, no todos los métodos son igualmente estables o presentan las mismas dinámicas. La Figura 2 da buena cuenta de ello. En ella, es posible apreciar cómo algunos métodos (como el de Systrom) arrojan predicciones fuera de rango (específicamente, negativas), mientras que otros métodos son o demasiado volátiles (como el de Cislighi) o bien demasiado estables (como el de Martínez-Beneito). En nuestra opinión, el método que mejor se comporta en general es el de Wallinga and Lipsitch (2007). En particular, es el único método que genera predicciones dentro de rangos razonables al tiempo que su evolución en el tiempo va en línea con los principales hitos relativos al COVID-19. En tanto, la Tabla 1 presenta el estimado de  $\beta$  al usar el R-efectivo previamente estimado (con el método de Wallinga) como variable respuesta en el modelo lineal generalizado mixto descrito en la Ecuación (1) (los coeficientes asociados a la base de splines fueron omitidos de la Tabla, pero todos resultaron significativos). En particular, obtuvimos que todas las dummies asociadas a la fase plan Paso a Paso tienen asociados coeficientes negativos y significativos (ver Tabla 1) lo cual, por la naturaleza del link, implica que cada aumento en la fase del plan aumenta el R-efectivo. En efecto, bajo este modelo,  $E[y_{rct}|x_{rct}, u_{rc}, v_r] = (x'_{rct}\beta + u_{rc} + v_r)^{-1}$ , de modo que  $\partial E[y_{rct}|x_{rct}, u_{rc}, v_r] / \partial x_{rct,k} = -\beta_k / (x'_{rct}\beta + u_{rc} + v_r)^2$ . Por lo tanto, un  $\beta_k < 0$  implica un efecto marginal positivo, lo que en términos de la pandemia implica un aumento en la velocidad de contagio.

term	estimate	std.error	statistic	p.value
(Intercepto)	0.47	0.04	12.56	0.00
Índice de desarrollo socio-económico index	0.06	0.03	2.30	0.02
Índice de ruralidad	0.12	0.02	6.88	0.00
1(fase 2 del plan Paso a Paso) (lag 1)	-0.07	0.00	-18.72	0.00
1(fase 3 del plan Paso a Paso) (lag 1)	-0.16	0.00	-36.04	0.00
1(fase 4 del plan Paso a Paso) (lag 1)	-0.22	0.01	-40.55	0.00
% de la población vecina en cuarentena (lag 1)	0.04	0.00	10.15	0.00
# de tests PCR / población comuna (lag 2)	1.00	0.34	2.96	0.00

Cuadro 1: Resumen de la estimación para los efectos fijos del modelo utilizado para estimar el efecto de las fases del plan *Paso a Paso* sobre el R efectivo.

## 5. Conclusiones preliminares

Mientras mayor sea el nivel de desagregación, más volátiles se volverán las estimaciones del R efectivo. Pese a lo anterior, algunos de los métodos examinados aún pueden arrojar estimaciones sensibles

<sup>7</sup>Desarrollado por el Observatorio Chileno de Salud Pública.

<sup>8</sup>Desarrollado por el Ministerio de Desarrollo Social.

<sup>9</sup>La mayoría de estas variables están disponibles en el repositorio <https://github.com/MinCiencia/Datos-COVID19>. En particular, usamos los productos 7 ((n. exámenes pcr / población comunal), según región y semana epidemiológica), 15 (casos nuevos, según comuna y semana epidemiológica), 74 (fase en plan paso a paso, según comuna y semana epidemiológica), 29 (status de cuarentena, según comuna y semana epidemiológica). En los períodos previos al inicio del plan paso a paso, tratamos una cuarentena como una fase 1 y una no-cuarentena como una fase 4. Todos los detalles pueden encontrarse en la versión extensa de este informe, disponible en [este enlace](#), y todos los códigos necesarios para reproducir los resultados están disponibles en el repositorio <https://github.com/COVID0248/covid-re>

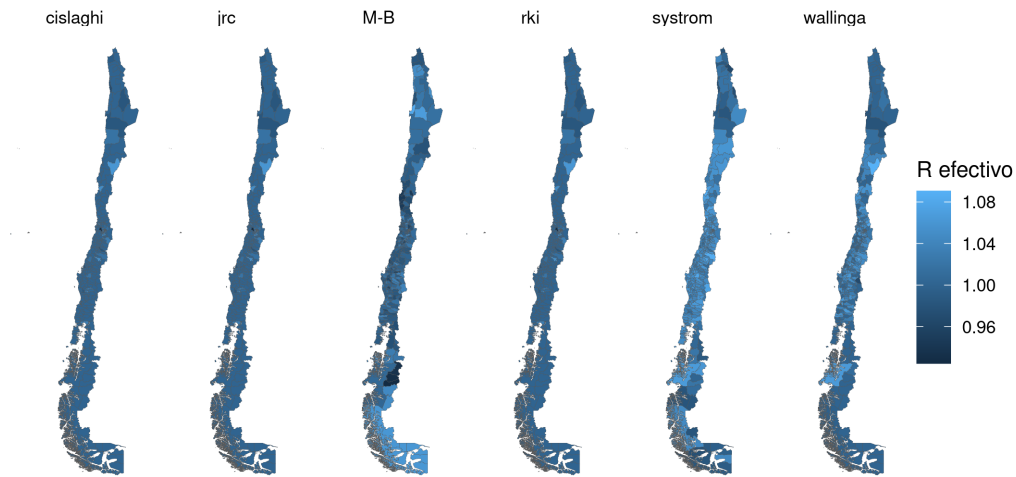


Figura 1: Percentil 50 del R efectivo, según comuna y método.

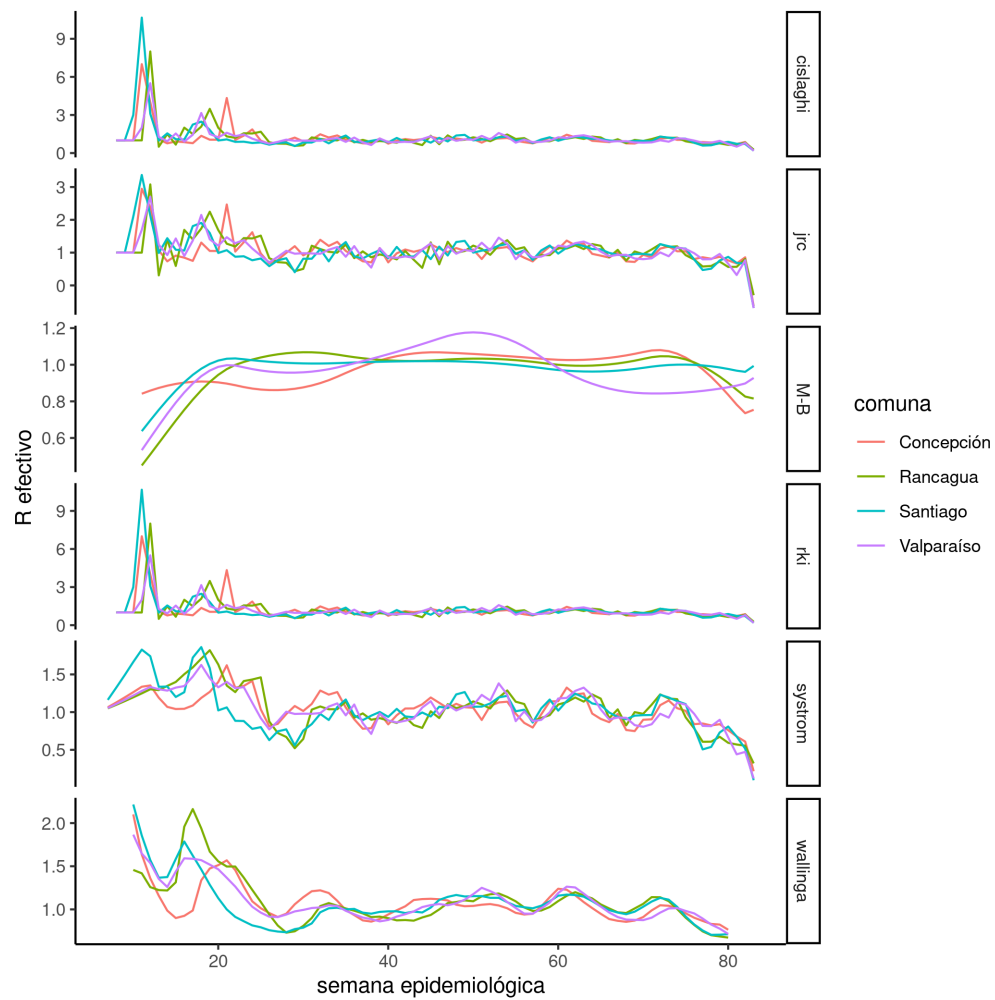


Figura 2: R efectivo para 4 comunas seleccionadas, según semana epidemiológica y método.

y que reflejan oportunamente los cambios en las tendencias dentro de la pandemia. En particular, usando el método de [Wallinga and Lipsitch \(2007\)](#), es posible verificar que los cambios en las fases dentro del plan Paso a Paso sí tienen efectos sobre el efectivo. En particular, cada cambio de paso hacia endurecer las medidas reduce el valor del R-efectivo comunal.

## Referencias

- Asikainen, T. and Annunziato, A. (2020). Effective reproduction number estimation from data series. Technical report, Joint Research Centre (European Commission).
- Cislaghi, 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.
- 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.