

# Estimación del número básico de reproducción temporal de la epidemia de la enfermedad COVID-19 en Colombia. v2

Jairo A. Diaz<sup>\*</sup>

División de Ciencias Básicas, Universidad del Norte, Barranquilla

Jairo J. Espinosa<sup>\*\*</sup>

Facultad de Minas, Universidad Nacional de Colombia, Medellín

Hector López<sup>\*\*\*</sup>

Gases del Caribe, Barranquilla

Jean Piero Suárez<sup>\*\*\*\*</sup>

División de Ciencias Básicas, Universidad del Norte, Barranquilla

Bernardo Uribe<sup>\*\*\*\*\*</sup>

División de Ciencias Básicas, Universidad del Norte, Barranquilla

28 de mayo de 2020

## Resumen

En este trabajo presentamos un modelo simple de estimación del número básico de reproducción temporal de la epidemia del virus SARS-CoV-2 en Colombia. El número básico de reproducción temporal  $\mathfrak{R}_t$  de una epidemia viral se refiere al número promedio de personas que contagia una persona infectada en un momento  $t$  del tiempo. Tengamos en cuenta que si este número  $\mathfrak{R}_t$  es mayor que 1, es porque habrá más infectados que antes y la epidemia crece. Por el contrario, si este número  $\mathfrak{R}_t$  es menor que 1, habrá menos infectados que antes y la epidemia decrece. La evolución del número de infectados depende del número básico de reproducción temporal  $\mathfrak{R}_t$  así como del tiempo promedio  $D$  en que una persona permanece infectada. Estos dos parámetros nos permiten comprender la evolución de la epidemia. En este trabajo se presenta un modelo de estimación de los mismos basado en los datos que provee día a día de la enfermedad COVID-19 el Ministerio de Salud y Protección Social de Colombia .

---

\*adjairo@uninorte.edu.co

\*\*jespinov@unal.edu.co

\*\*\*hlemus@uninorte.edu.co

\*\*\*\*solanopj@uninorte.edu.co

\*\*\*\*\*bjongbloed@uninorte.edu.co

## Introducción

La epidemia de la enfermedad COVID-19 que vivimos en el mundo nos ha llevado a tomar medidas nunca antes vistas. Con el fin de contrarrestar y de desaparecer la epidemia los gobiernos de los diferentes países han optado por obligar o recomendar diversos tipos de aislamiento social. En Colombia se dispuso de un aislamiento muy estricto de la población desde el día 24 de marzo de 2020. Las razones que justifican estas medidas extremas para contrarrestar la epidemia son la rapidez con que el virus se propaga, lo letal que puede ser en poblaciones vulnerables, el impacto que puede tener en la red hospitalaria y la falta de medicamentos o vacunas entre tantos otros.

Es importante conocer en tiempo real el impacto que tienen las medidas que adoptan los gobiernos y las personas en la evolución de la epidemia. De particular importancia es saber si dichas medidas han generado que la epidemia crezca o decrezca en términos de personas infectadas.

Este trabajo pretende hacer un pequeño aporte en la recopilación de información relativa a la evolución de la epidemia en Colombia. Nos hemos enfocado aquí en estimar el valor del número básico de reproducción temporal  $\mathfrak{R}_t$  y el tiempo promedio en días  $D$  que permanece una persona infectada, además de contabilizar el número de infectados nuevos por día, de fallecidos por día, el acumulado de infectados y el acumulado de fallecidos. Para esto no hemos basado en la información que provee el Ministerio de Salud y Protección Social en su página web: <https://coronaviruscolombia.gov.co/>. Esta estimación la hemos implementado para que se pueda visualizar por departamentos y por municipios.

Aclaremos que este trabajo sólo pretende mostrar una estimación del número básico de reproducción temporal  $\mathfrak{R}_t$  y como tal debe ser entendida. Esta estimación se basa en numerosas hipótesis, todas ellas dignas de ser tomadas con mucho cuidado, y sabemos que haría falta incorporar otras. Sin embargo nos hemos inclinado por presentar algo simple que cualquier persona sin entrenamiento matemático o epidemiológico puede comprender. La simplicidad del modelo lo hace sencillo de entender. Al mismo tiempo esta simplicidad debe generar en el lector una gran cautela sobre cómo interpreta y cómo usa la información aquí presentada.

Consideramos de suma importancia que el lector contraste la información que se provee en este trabajo con con otras mediciones y estimaciones que han realizado otros grupos de trabajo. No pretendemos que esta estimación sea la única. Sólo esperamos que este trabajo aporte su grano de arena en la comprensión de la evolución de la epidemia del virus SARS-CoV-2 en Colombia.<sup>1</sup>

---

<sup>1</sup>Los autores agradecen al Grupo Asesor SCM COVID-19 de la Sociedad Colombiana de Matemáticas por el apoyo en este proyecto. J. Díaz y B. Uribe agradecen a la Universidad del Norte y J. Espinosa a la Universidad Nacional de Colombia por su apoyo y confianza.

## Descripción

El número básico de reproducción temporal  $\mathfrak{R}_t$  de una epidemia viral lo definimos como el número promedio de personas que contagia una persona infectada promedio en el lapso de tiempo que permaneció infecciosa. Este número es importante porque nos permite tener en tiempo real un indicador simple que valora la evolución de una epidemia. Si  $\mathfrak{R}_t$  es mayor que 1 la epidemia crece en términos de infectados, si es menor que 1 la epidemia decrece. Vale la pena resaltar que este valor  $\mathfrak{R}_t$  no es constante y en él se ven reflejadas las medidas de distanciamiento social que toman las comunidades para enfrentar la epidemia. El valor de este número cuando  $t = 0$  es el conocido  $R_0$  y refleja la evolución de la epidemia al inicio de la misma. Transcribimos textualmente una descripción de  $R_0$ <sup>2</sup>:

*El número básico de reproducción ( $R_0$ ), pretende ser un indicador del contagio o transmisibilidad de agentes infecciosos y parasitarios.  $R_0$  se encuentra a menudo en la bibliografía de epidemiología y salud pública y también se puede encontrar en la prensa popular.  $R_0$  se ha descrito como una de las métricas fundamentales y más utilizadas para el estudio de la dinámica de las enfermedades infecciosas. Un  $R_0$  para un evento de enfermedad infecciosa generalmente se informa como un valor numérico único o rango bajo-alto, y la interpretación generalmente se presenta como directa; se espera que continúe un brote si  $R_0$  tiene un valor mayor que 1 y que termine si  $R_0$  es menor que 1. El tamaño potencial de un brote o epidemia a menudo se basa en la magnitud del valor  $R_0$  para ese evento, y  $R_0$  puede usarse para estimar la proporción de la población que debe vacunarse para eliminar una infección de esa población.*

El número básico de reproducción temporal  $\mathfrak{R}_t$  debe venir acompañado del periodo de tiempo  $D$  en días que una persona permanece infecciosa. Los dos parámetros juntos nos permiten modelar la evolución de la epidemia y por lo tanto siempre deben ir acompañados el uno del otro.

La idea de este trabajo es estimar estos números  $\mathfrak{R}_t$  y  $D$  basándonos en la información relativa a la evolución de casos de la COVID-19 que provee el Ministerio de Salud y Protección Social. De este reporte diario se conoce la información relativa a los infectados detectados por día, los que se reportan como curados y los que se reportan como fallecidos.

Entendamos entonces las ecuaciones que nos permitirán relacionar los números  $\mathfrak{R}_t$  y  $D$  con la evolución de la epidemia.

Denotemos por  $I(t)$  a los infectados por el virus en el día  $t$ , esto es, el número de personas que en un tiempo específico son infecciosas debido al virus SARS-CoV-2, y por  $R(t)$  al número de personas que ya sea se han curado de la enfermedad o que desgraciadamente han fallecido debido al virus. La letra  $R$  en  $R(t)$  denota los recuperados, sin embargo, en este modelo simple  $R(t)$  denotará los que estuvieron infectados y dejaron de estarlo, ya sea porque se curaron o porque fallecieron.

---

<sup>2</sup> Delamater PL, Street EJ, Leslie TF, Yang Y, Jacobsen KH. Complexity of the Basic Reproduction Number ( $R_0$ ). Emerg Infect Dis. 2019;25(1):1-4.

Denotemos con la letra  $D$  el tiempo medio en días en que una persona permanece infectada con el virus. Después de este tiempo la persona infectada pasará al grupo de las recuperadas, ya sea porque se curó o porque falleció.

Consideremos el tiempo  $t$  medido en días y veamos cómo cambia el número de infectados con respecto al tiempo. El número de infectados aumenta debido a los casos nuevos de contagio y disminuye debido a los que se curaron o fallecieron. Teniendo en cuenta la ventana de tiempo promedio en que una persona permanece infectada, tendremos que en ese periodo de tiempo se habrán contagiado  $\mathfrak{R}_t$  personas y al final del periodo esa persona se habrá curado o habrá fallecido. Esto quiere decir que cada infectado, en la ventana de tiempo  $D$ , genera un cambio en la cantidad de infectados de  $\mathfrak{R}_t - 1$  personas.

Por lo tanto, el cambio en el número de infectados por unidad de tiempo  $\frac{\Delta I}{\Delta t}$  sería  $((\mathfrak{R}_t - 1)/D)I(t)$ . Recordemos que la unidad de tiempo es el día y que por lo tanto tenemos que dividir por el lapso de tiempo en que una persona está enferma. La ecuación queda entonces

$$\frac{dI(t)}{dt} = \frac{\mathfrak{R}_t - 1}{D} I(t). \quad (1)$$

Ahora, siguiendo el mismo argumento anterior tenemos que el cambio en el número de recuperados (que nunca es negativo porque el número de recuperados no puede disminuir) sería:

$$\frac{dR(t)}{dt} = \frac{1}{D} I(t). \quad (2)$$

Es decir, tomamos el número de infectados y dividimos entre el periodo de infección para obtener el número de personas promedio que se recuperan en el día  $t$ .

Notemos que lo que queremos comprender es la evolución del número de infectados. Por lo tanto de la ecuación (1) podemos deducir que si  $\mathfrak{R}_t - 1$  es mayor que cero es porque el número de infectados aumenta con respecto al tiempo, y si  $\mathfrak{R}_t - 1$  es menor que cero es porque el número de infectados disminuye.

La evolución de la epidemia depende tanto de  $\mathfrak{R}_t$  como de  $D$ . Como vimos, cada nuevo día habrá  $((\mathfrak{R}_t - 1)/D) \cdot I(t)$  más infectados que el anterior. De ahí la importancia de estimar los valores de  $\mathfrak{R}_t$  y de  $D$  de los datos en tiempo real.

## Variables y parámetros del modelo

Las ecuaciones que hemos presentado en (1) y (2) sólo involucran el número de infectados activos y de recuperados (más fallecidos) en un tiempo  $t$ , el periodo promedio de infección  $D$  y el número básico de reproducción temporal  $\mathfrak{R}_t$ . Esta información la estimaremos de los datos de los casos reportados que se puede obtener diariamente de la página del Ministerio de Salud y Protección Social:

<https://www.datos.gov.co/Salud-y-Proteccion-Social/Casos-positivos-de-COVID-19-en-Colombia/gt2j-8ykr/data>

Las variables relevantes para este trabajo que se obtienen de la base de datos son las siguientes:

- `fecha_sintomas`: Fecha del inicio de síntomas
- `fecha_muerte`: Fecha de fallecimiento
- `fecha_recuperacion`: Fecha en la cual ya no es portador del virus
- `fecha_reporte`: Fecha en la que se reporta en la página web

Tengamos en cuenta que en los datos reportados hay casos de personas asintomáticas. A estos casos se les asigna una fecha de reporte en la página web pero no se les asigna fecha de inicio de síntomas. Más adelante veremos cómo le asignaremos fecha de inicio de síntomas a los asintomáticos teniendo en cuenta el valor medio del retraso en reportar los casos.

## Estimación del periodo de infecciosidad

La duración del periodo en que una persona se encuentra enferma  $D$  se podría asumir que es un parámetro externo. Sin embargo las estimaciones que se han hecho de este periodo han variado. Se ha dicho en un principio que es de dos semanas<sup>3</sup>, se ha estimado en 20.8 días<sup>4</sup> y se ha estimado en 11 días<sup>5</sup>

Como este valor de  $D$  no ha sido completamente determinado, consideramos pertinente estimar el valor de  $D$  desde los datos.

El periodo  $D$  podemos estimarlo como el valor medio del tiempo que una persona se reporta como infectada, a este periodo lo denominamos  $\hat{D}_T$  ya que depende de la información recopilada hasta el día  $T$ :

$$\hat{D}_T = \text{Valor medio del tiempo que una persona se reporta como infectada.}$$

Esta estimación se hace como el valor medio del tiempo de infecciosidad de los recuperados y los fallecidos reportados. Para los recuperados este periodo se calcula como

$$\text{fecha\_recuperacion} - \text{fecha\_sintomas}$$

y para los fallecidos

$$\text{fecha\_muerte} - \text{fecha\_sintomas.}$$

En particular tenemos que el periodo de infecciosidad de la mitad de los casos que cuentan con inicio de síntomas y que se han recuperado o fallecido es menor a  $\hat{D}_T$ .

<sup>3</sup>Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19)

<sup>4</sup>Qifang Bi, et al. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. The Lancet Infectious Diseases. April 27 2020

<sup>5</sup>Position Statement from the National Centre for Infectious Diseases and the Chapter of Infectious Disease Physicians, Academy of Medicine, Singapore – 23 May 2020

En el modelo SIR de ecuaciones diferenciales se usa la variable  $\gamma = 1/D$  definida como el inverso del periodo de infección. Hay que tener en cuenta que esta es la variable que a veces se reporta en algunas modelaciones.

## Retraso medio

Consideramos el *retraso* en reportar los datos como la cantidad de días entre la fecha de inicio de síntomas y la fecha en que se reporta en la página web:

$$\text{retraso} = \text{fecha\_reporte} - \text{fecha\_sintomas}.$$

Denotamos por  $\widehat{W}_T$  al valor medio del retraso (mediana) de los casos reportados con fecha de inicio de síntomas al tiempo  $T$ :

$$\widehat{W}_T = \text{Valor medio del retraso en el reporte.}$$

Este valor medio del retraso nos dice que la mitad de los casos se tardaron menos que  $\widehat{W}_T$  días en reportarse.

## Asignación de inicio de síntomas a los asintomáticos

Para poder tener en cuenta a las personas asintomáticas en el conteo de infectados activos diarios es necesario asignarles una fecha de inicio de síntomas. Suponiendo que el reporte de los casos asintomáticos tienen un retraso estimado similar al retraso medio  $\widehat{W}_T$ , les asignamos a todos los casos asintomáticos una fecha de inicio de síntomas de  $\widehat{W}_T$  días antes de la fecha de reporte:

$$\text{fecha\_sintomas\_asintomaticos} = \text{fecha\_reporte} - \widehat{W}_T.$$

Esta asignación de fecha de inicio de síntomas es la que hace el Instituto Robert Koch de Alemania<sup>6</sup> en sus análisis de la evolución de la epidemia.

## Conteo de infectados, recuperados y fallecidos

Con la información antes definida podemos proceder a definir las variables de interés. Sea  $T$  el día en que queremos hacer los cálculos y  $t \leq T$ .

- $NI_{sin}(t, T)$ : número de nuevos infectados en el día  $t$  que no son asintomáticos reportados hasta el día  $T$ .
- $NI_{asin}(t, T)$ : número de nuevos infectados asignados al día  $t$  que son asintomáticos reportados hasta el día  $T$ .
- $NI(t, T) = NI_{sin}(t, T) + NI_{asin}(t, T)$ : número total de nuevos infectados en el día  $t$  reportados hasta el día  $T$ .

<sup>6</sup>an der Heiden M, Hamouda O: Schätzung der aktuellen Entwicklung der SARS-CoV-2-Epidemie in Deutschland - Nowcasting. Epid. Bull. 2020;17:10, 15 — DOI 10.25646/6692.2

- $I_{tot}(t, T)$ : acumulado de infectados al día  $t$  reportados hasta el día  $T$ .
- $NR(t)$ : número de recuperados en el día  $t$ .
- $NM(t)$ : número de fallecidos en el día  $t$ .
- $R(t)$ : acumulado de recuperados al día  $t$ .
- $M(t)$ : acumulado de fallecidos al día  $t$ .
- $I(t, T) = I_{tot}(t, T) - R(t) - M(t)$ : número de infectados activos en el día  $t$  reportados hasta el día  $T$ .

Notemos la fecha de reporte de los fallecidos y recuperados es muy cercana a la fecha de su reporte en la página web. Por ese motivo no incorporamos la variable  $T$  es un conteo.

### Cálculo de $\mathfrak{R}_t$

Para calcular  $\mathfrak{R}_t$  en el día  $T$  vamos a hacer uso de la variable  $D$  que calculamos antes. Tomando el logaritmo natural del número de infectados y derivando con respecto a  $t$  obtenemos la ecuación

$$\frac{d \log(I(t))}{dt} = \frac{1}{I(t)} \frac{dI(t)}{dt} = \frac{\mathfrak{R}_t - 1}{D}.$$

Por lo tanto tenemos que

$$\mathfrak{R}_t = D \cdot \frac{d \ln(I(t))}{dt} + 1$$

y de esta manera obtenemos el primer cálculo del número básico de reproducción temporal en el día  $T$  con  $t \leq T$ :

$$\boxed{\mathfrak{R}_t = \hat{D}_t \cdot [\log(I(t, T)) - \log(I(t-1, T))] + 1}.$$

Es importante notar aquí que el cálculo de  $\mathfrak{R}_t$  para el día  $t$  incorpora el tiempo medio de recuperación del día  $t$ . Así tenemos que para valores diferentes de  $T$ , el valor de  $\mathfrak{R}_t$  siempre está calculado con el mismo periodo medio de recuperación.

El cálculo de  $\mathfrak{R}_t^s$  suavizado es una media móvil de 3 puntos:

$$\boxed{\mathfrak{R}_t^s = \frac{\mathfrak{R}_t + \mathfrak{R}_{t-1} + \mathfrak{R}_{t-2}}{3}}.$$

## Estimación de los nuevos infectados en los últimos días (nowcast)

Los casos positivos de infección son reportados con retraso. Esto quiere decir que en el conteo diario de infectados nuevos  $NI(t, T)$  en el día  $t$ , basado en la información recibida hasta el día  $T$ , no se han contado los infectados nuevos que se reportarán después del día  $T$ . La idea es estimar el valor de los infectados totales del día  $t$  a partir de la información que conocemos hasta el día  $T$ . Procedamos a definir las siguientes variables:

- $\widehat{NI}(t, T; T_1)$ : número de casos totales de nuevos infectados en el día  $t$  que se deberían haber reportado hasta el día  $T_1 \geq T$  a partir de la información que conocemos hasta el día  $T$ .

El valor de  $\widehat{NI}(t, T; \infty)$  es el que nos interesa estimar, esto es, el número total de nuevos infectados en el día  $t$  conociendo la información de los nuevos infectados hasta el día  $T$ .

En esta estimación vamos a seguir el análisis de la información presentada en [Lawless-1994]<sup>7</sup>.

La distribución del retraso en los casos reportados no es aleatoria y vamos a suponer que es casi-estacionaria. Esto quiere decir que asumiremos que se estabiliza en el tiempo la probabilidad de que un caso se reporte con retraso de  $d$  días.

Para esto denotemos las variables de conteo de retraso de la siguiente manera. Sea  $d \geq 0$  el retraso en días y  $T$  el día en que queremos estimar dicha probabilidad. Considere la variable

- $N(d, T)$ : número de casos reportados hasta el día  $T$  con retraso menor o igual a  $d$ .

De esta manera la probabilidad de que un caso sea reportado en el transcurso de  $d$  días es

$$\mathbb{P}(\text{retraso} \leq d) = \frac{N(d, T)}{N(T, T)}$$

donde  $N(T, T)$  consiste de todos los casos reportados con fecha de inicio de síntomas.

Notemos que el retraso medio  $\widehat{W}_T$  que habíamos precisamente nos da una probabilidad de un medio sobre los casos totales:

$$\mathbb{P}(\text{retraso} \leq \widehat{W}_T) \approx \frac{1}{2}.$$

Estimaremos entonces el número de nuevos infectados en el día  $t$  reportados hasta el día  $T_1 \geq T$  a partir de la información del día  $T$  como:

$$\widehat{NI}(t, T; T_1) = \frac{\mathbb{P}(\text{retraso} \leq T_1 - t)}{\mathbb{P}(\text{retraso} \leq T - t)} NI(t, T).$$

---

<sup>7</sup>J. F. Lawless, The Canadian Journal of Statistics, La Revue Canadienne de Statistique Vol. 22, No. 1 (Mar., 1994), pp. 15-31



Por lo tanto la estimación del número total de nuevos infectados en el tiempo  $t$  conociendo los reportes hasta el tiempo  $T$  es:

$$\widehat{NI}(t, T; \infty) = \frac{1}{\mathbb{P}(\text{retraso} \leq T - t)} NI(t, T) = \frac{N(T, T)}{N(T - t, T)} NI(t, T).$$

Estimamos de la misma manera las otras variables así:

- $\widehat{NI}_{sin}(t, T; \infty) = \frac{N(T, T)}{N(T - t, T)} NI_{sin}(t, T)$ : número estimado del total de nuevos infectados en el día  $t$  que no son asintomáticos a partir de la información del día  $T$ .
- $\widehat{NI}_{asin}(t, T; \infty) = \frac{N(T, T)}{N(T - t, T)} NI_{asin}(t, T)$ : número estimado del total de nuevos infectados asignados al día  $t$  que son asintomáticos a partir de la información del día  $T$ .
- $\widehat{I}_{tot}(t, T; \infty)$ : acumulado de infectados estimados al día  $t$  a partir de la información del día  $T$ .
- $\widehat{I}(t, T; \infty) = I_{tot}(t, T; \infty) - R(t) - M(t)$ : número estimado de infectados activos en el día  $t$  a partir de la información del día  $T$ .

Tengamos en cuenta dos cosas en esta estimación. Primero, si  $NI(t, T) = 0$  entonces la estimación también es cero. Es decir sólo se estiman casos con un retraso en la medida que haya casos con ese retraso. Si no hay nuevos infectados reportados en los últimos días (lo cual es buena noticia) no se estimará el valor de los nuevos infectados usando este procedimiento.

Y segundo, esta estimación funciona en la medida que la probabilidad de que los casos se reporten en menos que  $d$  días sea mayor que 0. Es decir, para esta estimación sólo se estimarán nuevos infectados en el tiempo  $t$  con la información recopilada hasta  $T$  si

$$\mathbb{P}(\text{retraso} \leq T - t) > 0.$$

La estimación presentada arriba se puede mejorar cuando se varía la ventana de tiempo en la cual se calcula la fracción de casos con retraso específico. Dicho análisis es relevante pero no se presentará en este trabajo.

## Cálculo de $\widehat{\mathfrak{R}}_t$ incorporando la estimación de los nuevos infectados (nowcast)

Denotamos por  $\widehat{\mathfrak{R}}_t$  a la estimación del número básico de reproducción temporal basada en la información de los infectados activos estimados hasta la fecha  $T$ :

$$\widehat{\mathfrak{R}}_t = \widehat{D}_t \cdot [\log(\widehat{I}(t, T; \infty)) - \log(\widehat{I}(t - 1, T; \infty))] + 1.$$

El cálculo de  $\widehat{\mathfrak{R}}_t^s$  suavizado es::

$$\widehat{\mathfrak{R}}_t^s = \frac{\widehat{\mathfrak{R}}_t + \widehat{\mathfrak{R}}_{t-1} + \widehat{\mathfrak{R}}_{t-2}}{3}.$$

## Relación con el modelo SIR

El modelo SIR es un sistema de ecuaciones diferenciales ordinarias con tres compartimientos: susceptibles, infectados y recuperados, expuesta hace más de noventa años para modelar epidemias<sup>8</sup>. Asume que la población total  $N$  es constante y que  $N = S(t) + I(t) + R(t)$  donde  $S(t)$  denota a las personas susceptibles en un tiempo  $t$ . Las ecuaciones son<sup>9</sup>:

$$\begin{aligned}\frac{d(S(t))}{dt} &= -\beta \frac{S(t)}{N} I(t) \\ \frac{d(I(t))}{dt} &= \beta \frac{S(t)}{N} I(t) - \gamma I(t) \\ \frac{d(R(t))}{dt} &= \gamma I(t)\end{aligned}$$

donde podemos ver que en nuestra notación anterior  $\gamma = \frac{1}{D}$  y  $\mathfrak{R}_t = \frac{\beta}{\gamma} \frac{S(t)}{N}$ .

El parámetro  $\mathfrak{R}_t$  incorpora la probabilidad  $S(t)/N$  de que el contacto de una persona infectada la haga con una persona susceptible. Al principio de la epidemia la mayoría de las personas son susceptibles y por lo tanto se puede asumir que  $S(t)/N \approx 1$ . Sin embargo este factor hará que el parámetro  $\mathfrak{R}_t$  tienda a cero en la medida que el tiempo transcurra.

Este modelo se ha usado (y también mal usado) para proyectar la epidemia de COVID-19 asumiendo que las variables  $\beta$ ,  $\gamma$  y  $N$  son constantes. Esto en particular implicaría que el valor de  $\mathfrak{R}_t$  solo dependería de  $\mathfrak{R}_0$  y del número de personas susceptibles en el tiempo  $t$ . Este no es el caso en esta epidemia. Hay que tener mucho cuidado al interpretar las proyecciones de la evolución de la epidemia por medio de un modelo SIR que no tiene en cuenta que sus parámetros estructurales pueden cambiar con el tiempo.

## Tablero dinámico

Hemos dispuesto de un tablero dinámico para esta segunda versión que es accesible por medio del enlace siguiente:

<https://rtcolombia.herokuapp.com/>

La versión anterior se puede visualizar en el enlace:

<https://rtcolombiaalpha.herokuapp.com/>

Estos tableros dinámicos fueron elaborados por el primer autor Jairo A. Diaz y el tercer autor Héctor López.

Allí aparecen la gráfica de  $\mathfrak{R}_t$ , las gráficas de barras del número de infectados y fallecidos por día y las de infectados y fallecidos acumulados, la gráfica del logaritmo de los infectados activos y un los datos de los últimos días.

<sup>8</sup>Kermack, W. O.; McKendrick, A. G. (1927). "A Contribution to the Mathematical Theory of Epidemics". *Proceedings of the Royal Society A*. 115 (772): 700–721.

<sup>9</sup>Wikipedia: SIR model.

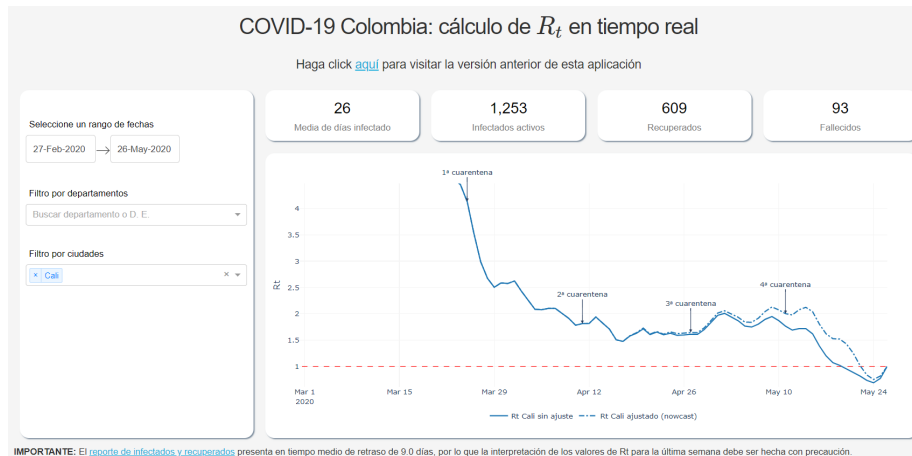


Figura 1: Evolución del número básico de reproducción temporal  $\mathfrak{R}_t$  y del estimado  $\hat{\mathfrak{R}}_t$  en la ciudad de Cali hasta el día 26 de mayo de 2020. La línea roja punteada representa el valor constantes 1. Si el valor de  $\mathfrak{R}_t$  está por encima de la línea roja, la epidemia crece en términos de infectados. Si está por debajo, la epidemia decrece. La duración media del periodo en que las personas han permanecido infecciosas, el número de personas infectadas activas, de recuperadas y de fallecidas en la ciudad de Cali aparece arriba de la gráfica.

## Gráfica de $\mathfrak{R}_t$

La primera gráfica muestra la evolución del número básico de reproducción temporal  $\mathfrak{R}_t$  y su versión estimada  $\hat{\mathfrak{R}}_t$  desde el inicio de la epidemia como se puede ver en la Figura 1. También muestra el valor medio de la duración del periodo en que las personas han permanecido infecciosas, el número de personas infectadas activas, de recuperadas y de fallecidas.

Hemos marcado las fechas de los inicios de las políticas de aislamiento social. Estas medidas se han tomado por lapsos de dos semanas comenzando el 24 de marzo y prolongándose el 13 de abril, el 27 de abril y el 11 de mayo.

La línea roja representa el valor constante 1. De especial importancia es el hecho de que  $\mathfrak{R}_t$  sea mayor o menor que 1. Si  $\mathfrak{R}_t$  es mayor que 1 la epidemia crece en términos de infectados. Si  $\mathfrak{R}_t$  es menor que 1 la epidemia decrece. La meta es mantener el valor de  $\mathfrak{R}_t$  menor que 1.

## Gráficas de conteo de casos

En las gráficas de conteo de casos que se puede ver en la Figura 2 se presenta la información a la fecha relativa al conteo de casos del departamento o municipio respectivo. En particular se presentan las siguientes gráficas:

- *Infectados diarios*: Número de infectados nuevos por día.
- *Fallecidos diarios*: Número de fallecidos por día.



Figura 2: Gráficas con la información de la ciudad de Cali al día 26 de mayo de 2020. Arriba a la izquierda está el conteo diario de nuevos infectados junto con su proyección del nowcast, arriba a la derecha el conteo diario de fallecidos, abajo a la izquierda el acumulado diario de infectados junto con los proyectados por el nowcast y abajo a la derecha el acumulado diario de fallecidos.

- *Infectados acumulados*: Número total acumulado de infectados al día específico.
- *Fallecidos acumulados*: Número total de fallecidos acumulados al día específico.

Vale la pena resaltar que estas gráficas están relacionadas. Las gráficas de los infectados nuevos por día y de fallecidos por día tienden a tener la misma forma, así como la de infectados acumulados y los fallecidos totales. Los fallecidos tienden a ser un porcentaje de los infectados.

## Gráfica de $\log(I(t))$ y tablas de información de casos

La primera gráfica corresponde a la evolución del logaritmo natural del número de infectados activos y de los infectados activos estimados (nowcast) y se puede ver en la Figura 3. Se grafica el logaritmo natural porque la evolución del número de infectados es exponencial. Notemos que la pendiente del logaritmo natural de infectados activos es:

$$\frac{d \log(I(t))}{dt} = \frac{1}{I(t)} \frac{dI(t)}{dt} = \frac{\mathfrak{R}_t - 1}{D}.$$

El logaritmo natural de los infectados activos cambia a una razón de cambio dada por el valor  $(\mathfrak{R}_t - 1)/D$ . Debido a esto las gráficas de  $\mathfrak{R}_t$  y de  $\log(I(t))$  están relacionadas. El logaritmo natural de los infectados crece cuando  $\mathfrak{R}_t > 1$  y decrece cuando  $\mathfrak{R}_t < 1$ .

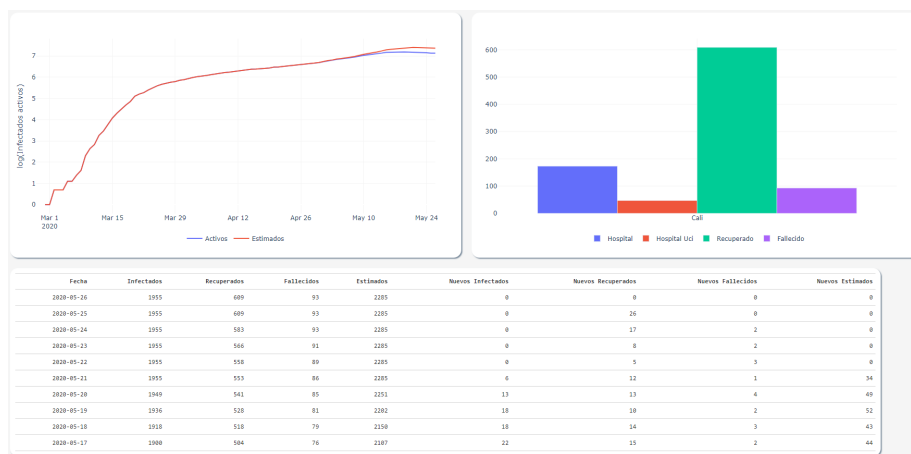


Figura 3: Evolución del logaritmo natural del número de infectados activos y de activos estimados (nowcast) en la ciudad de Cali hasta el día 26 de mayo de 2020. La curva es ascendente cuando aumentan los infectados, es decir cuando  $\mathfrak{R}_t > 1$ , y descendente cuando disminuyen, es decir  $\mathfrak{R}_t < 1$ . Se grafica el logaritmo natural de los infectados porque el comportamiento de la evolución de los infectados es exponencial. Además el cambio de esta variable con respecto al tiempo es  $(\mathfrak{R}_t - 1)/D$ . Se incluyen el número de personas hospitalizadas, en unidades de cuidados intensivos, recuperadas y fallecidas. Al final se muestra la información de los últimos días.

Para obtener el número de infectados activos de la gráfica del logaritmo natural de los infectados activos se evalúa el valor dado en la gráfica en la función exponencial  $I(t) = e^{\log(I(t))}$ .

## Interpretación de la gráficas

La información que se puede inferir de las gráficas antes descritas es variada. En esta sección explicaremos cómo obtener cierta información de las gráficas que consideramos relevante e importante.

### Interpretación de $\mathfrak{R}_t$

La estimación que se hace del número básico de reproducción temporal depende del tiempo de infecciosidad de las personas contagiadas. En este trabajo hemos decidido tomar dicho tiempo como el periodo medio en que las personas son reportadas como infecciosas. Al analizar la información por departamentos y municipios hemos visto que este periodo oscila entre 20 y 30 días.

Se podría asumir que el periodo de infecciosidad es un parámetro externo a los datos y por lo tanto la estimación del  $\mathfrak{R}_t$  con ese periodo diferiría de la presentada en este texto. Hemos visto como este periodo se ha asumido como un valor entre 5 y 25 días y que su determinación ha dependido del estudio médico realizado de los casos.

Ahora, sea cual fuere el valor del periodo de infecciosidad que se asuma, ya sea el que se estima por medio de los datos o el que se determina con estudios médicos, podemos relacionar los valores de  $\mathfrak{R}_t$  respectivos mediante el siguiente argumento.

Denotemos por  $\mathfrak{R}_t(\text{periodo } D)$  y  $\mathfrak{R}_t(\text{periodo } P)$  al valor de  $\mathfrak{R}_t$  asociado al periodo de infecciosidad  $D$  y  $P$  respectivamente. Denotemos por  $I_D(t)$  y  $I_P(t)$  al número de los infectados activos en el tiempo  $t$  asumiendo un periodo de infecciosidad de  $D$  y  $P$  días respectivamente. Notemos que en  $I_D(t)$  una persona se tiene en cuenta en el conteo por  $D$  días mientras que en  $I_P(t)$  se tiene en cuenta por  $P$  días. Esto quiere decir que se tiene la siguiente relación

$$I_P(t) = \frac{P}{D} \cdot I_D(t),$$

que implica  $\log(I_D(t)) + \log(P) = \log(I_P(t)) + \log(D)$  y cuya derivada nos permite concluir

$$\frac{\mathfrak{R}_t(\text{periodo } D) - 1}{D} = \frac{d \log(I_D(t))}{dt} = \frac{d \log(I_P(t))}{dt} = \frac{\mathfrak{R}_t(\text{periodo } P) - 1}{P}.$$

Deducimos entonces la relación entre los diferentes valores de  $\mathfrak{R}_t$  para diferentes periodos de infecciosidad:

$$\mathfrak{R}_t(\text{periodo } P) = (\mathfrak{R}_t(\text{periodo } D) - 1) \cdot \frac{P}{D} + 1.$$

Número de días en que se duplican los casos

<b>3,5</b>	1	2	2	2	2	3	3	3	4	4	4	4	5	5	5	6	6	6	7	7	7	7	8	8	8			
<b>3,4</b>	1	2	2	2	3	3	3	3	4	4	4	5	5	5	5	6	6	6	7	7	7	8	8	8	8	9		
<b>3,3</b>	2	2	2	2	3	3	3	4	4	4	5	5	5	5	6	6	6	7	7	7	8	8	8	8	9	9		
<b>3,2</b>	2	2	2	3	3	3	3	4	4	4	5	5	5	6	6	6	7	7	7	8	8	8	8	9	9	9		
<b>3,1</b>	2	2	2	3	3	3	4	4	4	5	5	5	6	6	6	7	7	7	8	8	8	8	9	9	9	10	10	
<b>3</b>	2	2	2	3	3	3	4	4	4	5	5	5	6	6	6	7	7	7	8	8	8	8	9	9	9	10	10	
<b>2,9</b>	2	2	3	3	3	4	4	4	5	5	5	6	6	7	7	7	8	8	8	9	9	9	9	10	10	11	11	
<b>2,8</b>	2	2	3	3	3	4	4	5	5	5	6	6	7	7	7	8	8	8	9	9	9	10	10	10	11	11	12	
<b>2,7</b>	2	2	3	3	4	4	4	5	5	6	6	7	7	7	8	8	9	9	9	10	10	10	10	11	11	11	12	12
<b>2,6</b>	2	3	3	3	4	4	5	5	6	6	6	7	7	8	8	9	9	10	10	10	10	11	11	12	12	13	13	13
<b>2,5</b>	2	3	3	4	4	5	5	6	6	6	7	7	8	8	9	9	10	10	10	11	11	12	12	12	13	13	14	14
<b>2,4</b>	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9	10	10	11	11	12	12	13	13	14	14	15	15	15
<b>2,3</b>	3	3	4	4	5	5	6	6	7	7	8	9	9	10	10	10	11	11	12	12	13	13	14	14	15	15	16	16
<b>2,2</b>	3	3	4	5	5	6	6	7	8	8	9	9	10	10	10	11	12	12	13	13	14	14	15	16	17	17	17	17
<b>2,1</b>	3	4	4	5	6	6	7	8	8	9	9	10	11	11	12	13	13	14	14	15	16	16	17	18	18	19	19	19
<b>2</b>	3	4	5	6	6	7	8	8	9	10	10	11	12	12	13	14	15	15	16	17	17	18	19	19	20	21	21	21
<b>1,9</b>	4	5	5	6	7	8	8	9	10	11	12	12	13	14	15	15	16	17	18	18	19	20	21	22	22	23	23	23
<b>1,8</b>	4	5	6	7	8	9	10	10	11	12	13	14	15	16	16	17	18	19	20	21	22	23	23	24	25	26	26	26
<b>1,7</b>	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	30	30
<b>1,6</b>	6	7	8	9	10	12	13	14	15	16	17	18	20	21	22	23	24	25	27	28	29	30	31	32	34	35	35	35
<b>1,5</b>	7	8	10	11	12	14	15	17	18	19	21	22	24	25	26	28	29	30	32	33	35	36	37	39	40	42	42	42
<b>1,4</b>	9	10	12	14	16	17	19	21	23	24	26	28	29	31	33	35	36	38	40	42	43	45	47	49	50	52	52	52
<b>1,3</b>	12	14	16	18	21	23	25	28	30	32	35	37	39	42	44	46	49	51	53	55	58	60	62	65	67	69	69	69
<b>1,2</b>	17	21	24	28	31	35	38	42	45	49	52	55	59	62	66	69	73	76	80	83	87	90	94	97	101	104	104	104
<b>1,1</b>	35	42	49	55	62	69	76	83	90	97	104	111	118	125	132	139	146	152	159	166	173	180	187	194	201	208	208	208
<b>↖</b>	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	30	30

Días de infecciosidad

Figura 4: Tiempo en días de duplicación de casos de infectados activos, y por lo tanto de fallecidos y recuperados por día, teniendo en cuenta el valor de  $\mathfrak{R}_t$  y el tiempo promedio de infecciosidad  $D$ .

Esta relación entre los diferentes valores de  $\mathfrak{R}_t$  permite comparar las diferentes estimaciones que se hacen de este parámetro.

El valor de  $\mathfrak{R}_t$  junto con el periodo de infecciosidad  $D$  nos permite determinar el número de días en que los casos de infectados activos se duplicarían. Asumiendo que  $\mathfrak{R}_t$  permaneciera constante podríamos integrar la ecuación diferencial de los infectados activos (1) obteniendo la siguiente relación para  $t \leq T$ :

$$I(T) = I(t)e^{\frac{\mathfrak{R}_t-1}{D}(T-t)}.$$

Los casos se duplican cuando  $\frac{\mathfrak{R}_t-1}{D}(T-t) = \log(2)$  y por lo tanto el periodo de tiempo en que los casos se duplican es:

$$\text{Tiempo en se duplican los casos} = \frac{\log(2) \cdot D}{\mathfrak{R}_t - 1}.$$

Notemos que el tiempo de duplicación sólo depende de la fracción  $\frac{\mathfrak{R}_t-1}{D}$ , y que como vimos antes, es independiente del periodo de infecciosidad que asumamos.

En la tabla de la Figura 4 podemos ver el tiempo de duplicación de casos de infectados activos para diferente valores de  $\mathfrak{R}_t$  y del tiempo promedio de infecciosidad  $D$ . Enfatizamos que el tiempo de duplicación de casos es el mismo para el  $\mathfrak{R}_t$ (periodo  $D$ ) con periodo  $D$  que para el  $\mathfrak{R}_t$ (periodo  $P$ ) con periodo  $P$ .

## Interpretación del conteo de casos

Las gráficas de barras del número de infectados nuevos por día y del número de fallecidos por día están relacionadas de la siguiente manera. El cambio de los fallecidos por día es proporcional a los infectados activos debido a que se ha observado que un porcentaje relativamente constante de infectados fallece debido al virus. Este porcentaje se puede determinar de manera local en cada departamento o municipio y puede variar de lugar en lugar. El tiempo de duplicación de casos calculado para los infectados activos se aplica también para los fallecidos por día debido al mismo argumento anterior. Tomando el tiempo de duplicación de casos que nos define el  $\mathfrak{R}_t$  y el tiempo promedio de infecciosidad  $D$  podemos proyectar que en ese tiempo se doblarían el número de fallecidos por día.

Veamos un ejemplo. El 26 de abril el  $\mathfrak{R}_t$  estimado de Colombia fue aproximadamente de 2.2 y el periodo de recuperación fue de 22 días. Según la tabla de la Figura 4 vemos que esto implicaría una duplicación de casos en 13 días. Podemos ver que alrededor del 20 de abril (seis días antes) había aproximadamente 340 nuevos infectados y alrededor del 3 de mayo (siete días después) había aproximadamente 610 infectados. Así mismo, alrededor del 20 de abril hubo 9 fallecidos por día y alrededor del 3 de mayo hubo 20. En ambos casos se puede apreciar el efecto aproximado de la duplicación de los casos en el periodo de 13 días

De la misma manera el número de infectados acumulados es proporcional al número de fallecidos acumulados. Como dijimos antes el porcentaje de fallecidos de los infectados se asume que se estabiliza. Por lo tanto las gráficas de los infectados acumulados y de los fallecidos acumulados tienen la misma forma. La gráfica de los fallecidos por día relativa a la de infectados nuevos, así como de la fallecidos acumulados relativa a la de infectados acumulados, están desplazadas  $D$  días a la derecha en el tiempo. Esto debido a que cuando un nuevo infectado se reporta, su fallecimiento se reportaría  $D$  días después. Las gráficas de los infectados nuevos y de los infectados acumulados nos van a permitir proyectar las gráficas de los fallecidos por día y los fallecidos acumulados  $D$  días después.

## Interpretación del logaritmo de los infectados activos

Como vimos anteriormente, los logaritmos de los infectados asumiendo periodo  $D$  y periodo  $P$  sólo difieren por una constante. Por lo tanto la gráfica del logaritmo de los infectados activos es un buen indicador de la evolución de la epidemia. La pendiente del logaritmo de los infectados es independiente del periodo de infecciosidad que se asuma. Dicha pendiente es igual a  $\frac{\mathfrak{R}_t - 1}{D}$ . El logaritmo de los infectados activos es un buen indicador de la evolución de la epidemia. Dicho indicador es independiente del  $\mathfrak{R}_t$  y del  $D$  que se estimen.



## Código del modelo

El código fue desarrollado en conjunto por todos los autores. Se puede descargar en:

<https://github.com/jairoadiazr/rtcolombia/>

Los autores autorizan el uso del código. Sin embargo se exhorta a los usuarios del código que otorguen el reconocimiento apropiado a los que diseñaron e implementaron el código.

## Página web

Toda la información relativa a este proyecto estará disponible en la página web:

<http://www.rtcolombia.com>

Allí se podrán encontrar las actualizaciones y nuevas versiones de este documento, así como las actualizaciones del tablero dinámico y del código del modelo.

La información sobre este proyecto también se podrá acceder por medio de la página web que la Sociedad Colombiana de Matemáticas ha dispuesto para listar los modelos de  $\mathfrak{R}_t$ :

<http://scm.org.co/r0-covid-19/>

## Consideraciones finales

- El número básico de reproducción temporal  $\mathfrak{R}_t$  es un indicador que nos permite saber si en un tiempo determinado se están contagiando más o menos personas que las que dejan de ser contagiosas. Tener una estimación de este valor es por lo tanto de suma importancia para tener una idea de cómo evoluciona una epidemia.
- Esta estimación se basa en los datos reportados y por lo tanto es una estimación del  $\mathfrak{R}_t$  de los datos reportados. Hay que tener en cuenta que puede haber muchos casos no reportados y el valor real de  $\mathfrak{R}_t$  puede ser diferente.
- Hay ocasiones en que el valor de  $\mathfrak{R}_t$  estimado es negativo. Estas son buenas noticias. El número de infectados disminuye rápidamente. Sin embargo no hay que interpretarlo como que cada infectado esté curando a otros infectados. Sencillamente hay que interpretarlo como que se están recuperando más personas que las que el modelo espera se recuperen. Recordemos que el tiempo medio de recuperación se calcula entre todos los casos de una región y se usa para estimar  $\mathfrak{R}_t$  en días consecutivos. El hecho de que se

recuperen los infectados más rápido que el tiempo promedio puede llevar a que  $\mathfrak{R}_t$  sea negativo.

- La estimación del periodo en que se estima que una persona es infecciosa  $D$  puede variar con el tiempo. Podría disminuir debido a la aplicación de medidas farmacológicas. Podría aumentar si no se reporta a tiempo la fecha de recuperación de las personas.
- El modelo presentado en este trabajo es un sólo un modelo. Esperamos que el lector analice con cuidado lo que el modelo estima, que lo contraste y compare con otras fuentes.

□