

INFORMACIÓN ESCONDIDA EN LOS DATOS INCIERTOS SOBRE EL COVID-19 EN CHILE

HÉCTOR PASTÉN, JORGE CASTILLO SEPÚLVEDA

1. RESULTADOS (PARA UNA AUDIENCIA GENERAL)

Gran parte de la discusión científica actual en torno a los datos oficiales sobre el COVID-19 en Chile se ha centrado en la baja calidad de los mismos. Sin embargo, mientras estas conversaciones (¡válidas y necesarias!) ocurren, la epidemia sigue avanzando. Se hace necesario extraer información útil de los datos existentes, a pesar de que la calidad de estos datos no es la óptima.

Como contribución a la comunidad, a continuación presentamos varias conclusiones *confiables* obtenidas de los datos existentes. Nos limitaremos a entregar solamente información que es válida a pesar de la incertidumbre de los datos. Los detalles técnicos del análisis están en la sección siguiente.

Importante. Suponemos que el total de casos *confirmados* acumulados y que el total de fallecidos *confirmados* que son entregados diariamente por el MINSAL son datos reales y no son estimaciones.

1.1. Tasa de mortalidad sobre casos confirmados. Hasta hoy, entre los casos confirmados de COVID-19 en Chile un 2,1 % de ellos terminan en muerte (cercano a 1 entre 50). Es decir, la tasa de mortalidad sobre casos confirmados (*case fatality rate*) observada a la fecha es de 2,1 %.

En la prensa se ha informado que la tasa de mortalidad sobre casos confirmados en Chile es mucho más baja, del orden de 1,3 % - 1,4 %, pero ese cálculo es completamente incorrecto y se basa en un error lamentablemente frecuente, ya documentado en la literatura [3, 9].

El 2,1 % indicado anteriormente se ha mantenido relativamente estable a lo largo de la epidemia. Sin embargo, es sobre casos *confirmados*, por lo que podría aumentar si los testeos se hacen insuficientes, o disminuir si se comienza a testear más.

1.2. Tiempo desde diagnóstico a fallecimiento. A la fecha, desde que un paciente es diagnosticado con COVID-19 en Chile hasta que fallece, pasan 9,2 días en promedio.

1.3. Tiempo desde inicio de síntomas a fallecimiento. La OMS (ver p. 14 de [4]) reporta que desde el inicio de síntomas hasta el fallecimiento por COVID-19 transcurren entre 2 y 8 semanas. ¿Entonces por qué el tiempo promedio de diagnóstico a fallecimiento es tan breve en Chile, de solo 9,2 días? Dos posibles razones:

- Hay un desfase desde el inicio de síntomas hasta el diagnóstico. Si este retraso en Chile fuese de una semana (lo cual ya sería bastante lento) significaría que el tiempo desde inicio de síntomas a fallecimiento sería de unos $7 + 9 = 16$ días. Esto nos deja muy abajo en el rango 14 - 56 reportado por la OMS, por lo que no puede ser la única explicación.
- Otra posible explicación es que una proporción significativa de los pacientes que fallecen por COVID-19 en Chile no estarían recibiendo tratamiento intensivo. Esta posibilidad es consistente con datos oficiales del MINSAL: por ejemplo, el 1 de Mayo el MINSAL informó el deceso de 7 pacientes, de los cuales 5 no recibieron tratamiento intensivo. El patrón se repite en muchos de los informes diarios.

Sería interesante analizar este fenómeno en más detalle.

Date: 3 de Mayo de 2020.

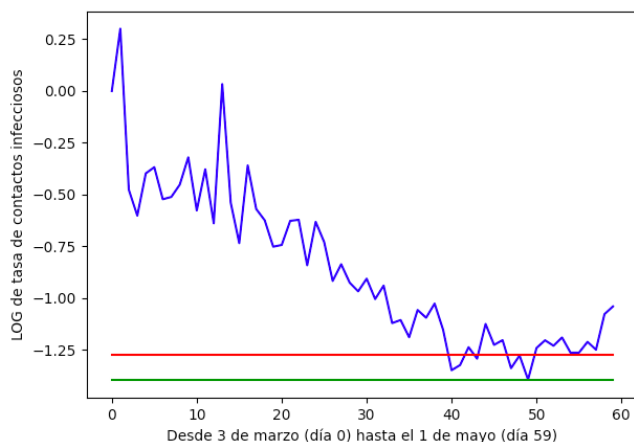
Este estudio finalizó el 1 de Mayo de 2020. No se incluyen datos de fechas posteriores.

1.4. Tiempo desde contagio hasta diagnóstico. El tiempo promedio que pasa desde que un paciente es contagiado hasta que es diagnosticado con COVID-19 en Chile es de *al menos* 10 días.

Por lo tanto, si la cuarentena solo comienza al confirmarse el diagnóstico, el contagiado tiene un rango no menor a una semana donde puede contagiar a otras personas.

1.5. Número real de contagiados. Para hacernos una idea de cuántos casos reales hay en comparación a los reportados en una fecha dada, calculamos que el número *real* de contagiados por COVID-19 en Chile el 19 de Abril¹ era mayor a 40 mil, mientras que la cifra confirmada ese día era de poco más de 10 mil.

1.6. Tasa de contactos infecciosos. La tasa de contactos infecciosos gobierna la velocidad de crecimiento de una epidemia y por ende, su evolución. Hay tres zonas para este indicador: verde (menos de 6%, evolución auspiciosa), gris (entre 6% y 8%, evolución compleja pero no catastrófica) y roja (más de 8%, evolución catastrófica). Los detalles del cálculo de estas tres zonas para el COVID-19 en Chile están en [6]. Este es el gráfico de la tasa de contactos infecciosos para el COVID-19 en Chile, al 1 de Mayo de 2020:



Comenzamos en zona roja (sobre la línea roja), bajando en dirección a la verde (bajo la línea verde) pero luego de Semana Santa ($n = 40$) se estancó por un tiempo en la gris (entre la línea verde y la roja). Desde el 22 de Abril aproximadamente ($n = 50$), esta tasa comenzó una tendencia en la zona roja.

1.7. Pronóstico para final de Mayo. De permanecer en la zona roja, a fin de Mayo acumularemos del orden de 40 mil casos confirmados y un total del orden de 650 fallecidos. Estos números serán más grandes si la tasa de contactos infecciosos se adentra más en la zona roja (por ejemplo, si se congregan grandes cantidades de personas en lugares cerrados).

Ese escenario se podría evitar bajando la tasa de contactos infecciosos lo antes posible, lo cual requeriría medidas de aislamiento y distanciamiento más efectivas que las vigentes hoy.

2. DETALLES TÉCNICOS (PARA UNA AUDIENCIA MÁS ESPECIALIZADA)

Los datos nacionales que utilizamos como insumo de este análisis solamente serán números oficiales confirmados por el MINSAL. Aclaramos que el polémico número estimado de “recuperados” no será utilizado.

¹La comparación en días posteriores es similar, pero el cálculo para una fecha dada necesita datos de varios días siguientes, por eso solo calculamos el número de casos reales para una fecha pasada reciente. También podemos hacer proyecciones a fechas futuras usando nuestros modelos predictivos, pero en este punto preferimos reportar sobre datos observados, no predicciones.

2.1. Fatalidad relativa a casos reportados. La cantidad de casos reales acumulados hasta el día t se anota $T_r(t)$. La cantidad de casos activos el día t es $I_r(t)$ y la cantidad de casos nuevos es $I_r^*(t)$. El total real de muertes acumuladas al día t es $M_r(t)$, y el de muertes nuevas reales el día t es $M_r^*(t)$. Las cantidades anteriores *reportadas* al día t en vez de reales se anotan con los mismos símbolos pero sin el subíndice r .

Sea d_0 el tiempo promedio desde contagio a diagnóstico en los casos diagnosticados. Entonces $I^*(t + d_0) = x \cdot I_r^*(t)$ para un cierto porcentaje x que depende de la capacidad de testeo. Si este porcentaje x fuera aproximadamente constante obtendríamos

$$(2.1) \quad T(t + d_0) = \int_0^t I^*(s + d_0) ds \approx x \int_0^t I_r^*(s) ds = x \cdot T_r(t).$$

Similarmente $M^*(t) = y \cdot M_r^*(t)$ para un cierto porcentaje y que depende de la capacidad de atribuir las muertes correctamente al COVID-19. Si y fuera aproximadamente constante obtendríamos

$$M(t) = \int_0^t M^*(s) ds \approx y \int_0^t M_r^*(s) ds = y \cdot M_r(t).$$

Sea D la duración promedio desde diagnóstico a fallecimiento, de modo que $d_0 + D$ es el tiempo promedio desde contagio a fallecimiento. Entonces la tasa de mortalidad intrínseca de la enfermedad durante el desarrollo de la epidemia debería ser aproximadamente constante igual a

$$\tau_M = \frac{M_r^*(t + d_0 + D)}{I_r^*(t)} \approx \frac{y^{-1} \cdot M^*(t + d_0 + D)}{x^{-1} \cdot I^*(t + d_0)}$$

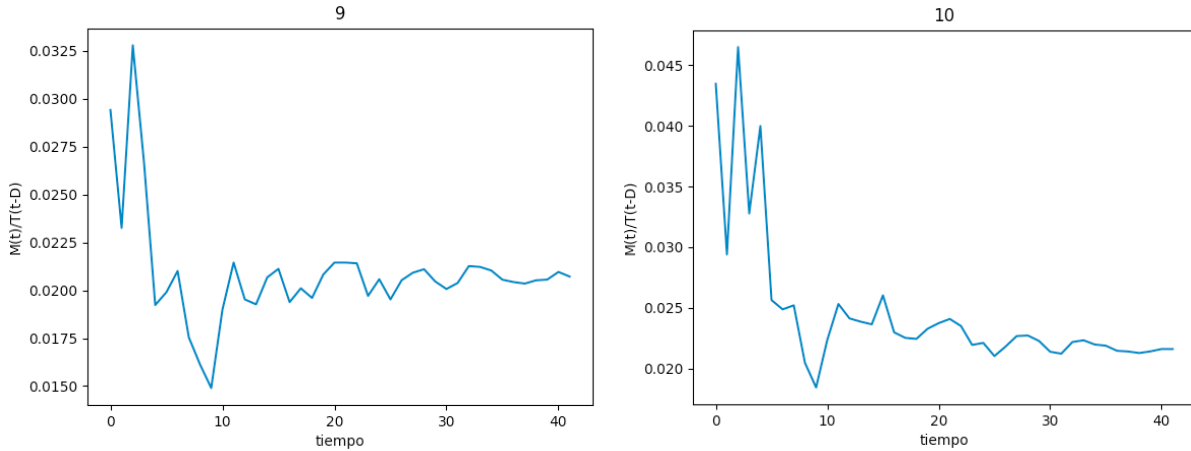
Sea $z = z(t)$ el porcentaje de los casos reportados en un cierto día t que terminan en muerte. Entonces de la ecuación anterior (con el cambio de variable $t \mapsto t + d_0$) obtenemos

$$(2.2) \quad z = M^*(t + D)/I^*(t) \approx \tau_M \cdot y/x.$$

En principio z puede variar con el día pero esta aproximación nos dice que debería ser aproximadamente constante. Integrando la ecuación $M^*(t + D) = z \cdot I^*(t)$ obtenemos

$$M(t + D) = \int_0^t M^*(s + D) ds = z \int_0^t I^*(s) ds = z \cdot T(t).$$

Así, $z = M(t)/T(t - D)$ debería ser aproximadamente constante. Con los datos reales, intentamos varios D y a continuación presentamos los casos más horizontales a la vista, $D = 9$ y $D = 10$. Dicha horizontalidad confirma que los supuestos anteriores se ajustan aproximadamente a la realidad.



Usando combinaciones convexas vemos que $D = 9,2$ es el más adecuado, correspondiente a $0,8 \cdot T(t - 9) + 0,2 \cdot T(t - 10)$, donde evaluamos la “horizontalidad” del gráfico haciendo una regresión lineal por mínimos cuadrados de forma que la pendiente sea cercana a 0.

Esta regresión lineal sobre combinaciones convexas no solo entrega el valor $D = 9,2$ sino que además da la aproximación $z = 0,02095 \approx 2,1\%$. El ajuste de $M(t) \approx z \cdot T(t - D)$ con estos valores, calculado con datos al 1 de Mayo está graficado (en escala logarítmica) a continuación:

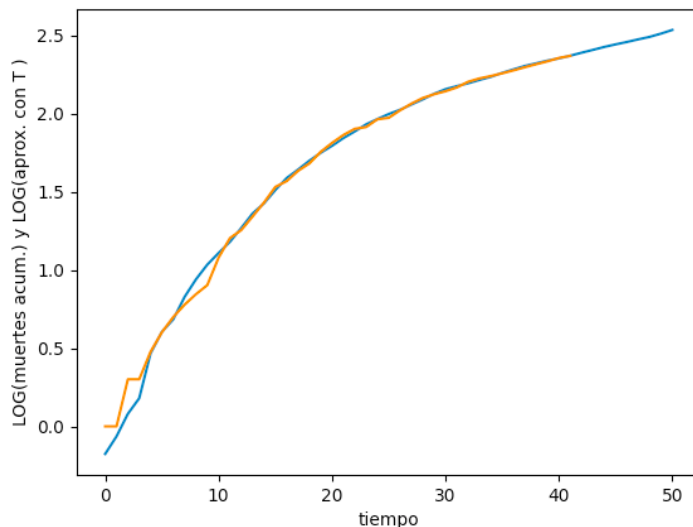


FIGURA 1. Ajuste $M(t) \approx z \cdot T(t - D)$ con $z = 2,1\%$ y $D = 9,2$ días, escala \log

Observación. La tasa de mortalidad sobre casos diagnosticados (CFR: *case fatality rate*) de una epidemia se calcula más acertadamente al final de la misma, como el número de defunciones dividido por casos reportados totales. Durante el desarrollo de la epidemia, sin embargo, es más complicado porque las muertes al día N no reflejan correctamente los casos al día N que eventualmente pueden terminar en muerte. Esta no es solamente la opinión de los autores, sino que es un hecho importante bien sabido por los expertos; ver por ejemplo la discusión detallada en el párrafo *How to calculate the mortality rate during an outbreak* de [9]. El hecho particular que el estimador $M(t)/T(t)$ no es apropiado para estimar el CFR mientras una epidemia todavía está en desarrollo es justamente el punto central del artículo [3].

Lamentablemente, este error se comete frecuentemente en los medios chilenos (e internacionales) en las estimaciones de la tasa de mortalidad sobre casos reportados, lo cual lleva a números errados mucho menores al real, del orden del 1,3% - 1,4% para Chile. Estos números son mucho menos informativos y pueden llevar a conclusiones erradas.

Nota agregada en edición. Mientras este reporte se terminaba de editar el 2 de Mayo, fuimos informados que R. Baeza-Yates publicó el mismo día en Twitter [1] una imagen similar a la Figura 1 con un ajuste correspondiente a $D = 10$ y $z = 2,22\%$. Al consultar con Baeza-Yates, el 3 de Mayo nos informó que la estimación $D = 10$ la había hecho con anterioridad y que apareció mencionada el 8 de Abril en [2] (sin dar detalles técnicos, pues no era un artículo especializado). Aunque sus métodos no están publicados, nos confirmó que son distintos a los nuestros y usa variable discreta, por lo cual su resultado daba un entero. Visualmente, en el gráfico de [1] no parece haber mayores diferencias con nuestro gráfico en la Figura 1, pero nuestro método de regresión lineal sobre combinaciones convexas nos permite estimar los parámetros minimizando errores no perceptibles a la vista sin restringirnos al caso cuando D es entero.

En cualquier caso, el método independiente de Baeza-Yates lleva a números bastante similares, lo que confirma mutuamente los resultados a pesar de las diferencias en los métodos.

2.2. Tiempos desde contagio a diagnóstico y fallecimiento. El tiempo medio de incubación (desde contagio a inicio de síntomas) es de 5,1 días [7]. En los casos que terminan en muerte en Chile el tiempo medio desde diagnóstico a fallecimiento es $D = 9,2$ días (sección anterior). Recordemos que d_0 es el tiempo promedio desde el contagio hasta el diagnóstico en Chile (no sabemos su valor exacto). Entonces el tiempo medio desde inicio de síntomas a fallecimiento es $d_0 + D - 5,1 = d_0 + 4,1$.

La OMS [4] reporta que el tiempo usual desde inicio de síntomas hasta fallecimiento es entre 2 y 8 semanas. Si es que Chile está en ese rango, entonces $d_0 + 4,1 \geq 14$ lo que significa que el tiempo desde el contagio hasta el diagnóstico es $d_0 \geq 9,9$. Es decir, al menos unos 10 días.

Observación. Por supuesto que d_0 calculado solo en los casos graves (como lo hicimos) podría ser menor al d_0 calculado sobre todos los casos, pero eso solo aumentaría nuestra cota inferior.

2.3. El número de casos reales. Estudios internacionales [8] estiman que la tasa de mortalidad sobre casos reales para el COVID-19 es del orden de $\tau_M = 0,66\%$. Para Chile, usaremos la cota superior $\tau_M \leq 0,7\%$ (la población en Chile es relativamente joven, así que esta cota superior es bastante confiable porque de hecho $0,66\%$ ya debería ser una cota superior luego de reajuste etario). Usando las ecuaciones (2.1) y (2.2) obtenemos

$$T_r(t) = x^{-1}T(t + d_0) \approx \frac{z}{y \cdot \tau_M}T(t + d_0)$$

Usando nuestra estimación $z \approx 2,1\%$ y las cotas $d_0 \geq 9$ (sección anterior), $y \leq 1 (= 100\%)$, y $\tau_M \leq 0,7\%$ obtenemos

$$T_r(t) \geq \frac{2,1\%}{0,7\%}T(t + 9) = 3 \cdot T(t + 9).$$

El día 28 de Abril (último día antes de la nueva metodología de contar casos asintomáticos) hubo 14365 casos reportados acumulados. Entonces el 19 de Abril hubo al menos $3 \cdot 14365 = 43095$ casos reales. Ese mismo día se reportaban 10088 casos acumulados.

Nota agregada en edición. El 3 de Mayo, R. Baeza-Yates nos informó que en [2] él presentó un método distinto para estimar el número de casos reales, también usando como insumo los datos del total de defunciones. Aunque ambos métodos son distintos, llevan a resultados del mismo orden de magnitud, lo que confirma mutuamente la validez de los resultados. La conclusión a la que llevan ambos estudios es que los casos confirmados difieren de los reales al menos por un factor entre 3 y 4, con la metodología usada hasta ahora por el MINSAL para confirmar y reportar casos.

2.4. Pronóstico para Mayo. El pronóstico de casos totales confirmados fue calculado con los métodos ya expuestos en [6]. Se usa una modificación del modelo SIR desarrollada por los autores, con un parámetro de delay en lugar de tasa de remoción, y con tasa de contactos infecciosos variable con componente aleatoria. La predicción se calcula con distintos valores del parámetro de delay ($d = 14, \dots, 21$) y resulta ser robusta. Una introducción a los modelos tipo SIR es presentada en el artículo de divulgación [5].

A partir de nuestra proyección de casos totales confirmados, obtenemos una proyección del total acumulativo de fallecidos usando nuestros resultados de la sección 2.1, especialmente las estimaciones $D = 9,2$ y $z = 2,1\%$. Así llegamos a las proyecciones reportadas en la sección 1.7.

3. AGRADECIMIENTOS

Agradecemos a Natalia García-Fritz por algunos comentarios de carácter técnico.

Agradecemos también a Ricardo Baeza-Yates por informarnos sobre su trabajo en temas relacionados a los estudiados en este artículo. Fue una grata sorpresa que sus estimaciones del periodo promedio diagnóstico-muerte y su estimación de casos reales, a pesar de ser obtenidas con métodos distintos, llevan a resultados similares a los nuestros. Esto confirma mutuamente los resultados,

al menos en esos dos temas. Además, en comunicación personal, R. B.-Y. no solo se refirió a su trabajo, sino que además tuvo la gentileza de leer un primer borrador de nuestro artículo y darnos una serie de comentarios constructivos. Le agradecemos profundamente.

REFERENCIAS

- [1] R. Baeza-Yates, <https://twitter.com/PolarBearby/status/1256673286022295553>
- [2] R. Baeza-Yates, *Datos de calidad y el corona virus*. https://medium.com/@rbaeza_yates/datos-de-calidad-y-el-corona-virus-98893b7600e3
- [3] Ghani, A. C., Donnelly, et al. *Methods for estimating the case fatality ratio for a novel, emerging infectious disease*. (2005) American journal of epidemiology, 162(5), 479-486.
- [4] OMS, *Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19)*. (2020) Disponible en <https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf>
- [5] H. Pastén, J. Castillo Sepúlveda, *Evolución de las epidemias: la matemática de aislarse*. (2020) Disponible en <http://www.mat.uc.cl/~hector.pasten/preprints/Epidemia2020.pdf>
- [6] H. Pastén, J. Castillo Sepúlveda, *COVID-19 y la tasa de contactos infecciosos: una historia narrada a color*. (2020) Disponible en <http://www.mat.uc.cl/~hector.pasten/preprints/TasaColor.pdf>
- [7] Stephen A. Lauer, et al. *The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application*. Annals of Internal Medicine (2020) DOI: 10.7326/M20-0504
- [8] Verity, Robert, et al. *Estimates of the severity of coronavirus disease 2019: a model-based analysis*. The Lancet Infectious Diseases (2020). DOI:[https://doi.org/10.1016/S1473-3099\(20\)30243-7](https://doi.org/10.1016/S1473-3099(20)30243-7)
- [9] <https://www.worldometers.info/coronavirus/coronavirus-death-rate/>

(H. Pasten) FACULTAD DE MATEMÁTICAS, UC CHILE
Email address: hector.pasten@mat.uc.cl

(J.Castillo-Sepúlveda) Z DATA LAB
Email address: jorgecastillo@zdatalab.com