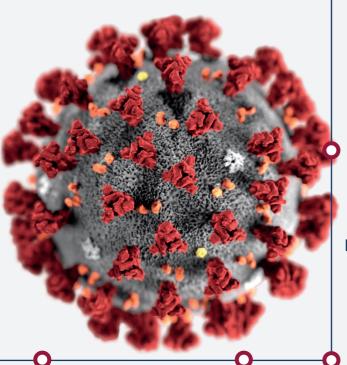




Ingeniería de Sistemas

Volumen XXXIV, Diciembre 2020



Volumen Especial Covid-19

Efecto de la correlación y los falsos negativos en la estrategia de testeo grupal para detección de COVID-19 Leonardo Basso / Vicente Salinas Denis Sauré / Charles Thraves Natalia Yankovic

Iniciativas matemático-computacionales desde la Universidad de Buenos Aires para contribuir a la toma de decisiones en el contexto del COVID-19 en Argentina Mehrnoosh Arrar/Laouen Belloli Ana María Bianco / Leonardo Boechi Rodrigo Castro / Guillermo Durán / et al.

Análisis de movilidad durante la pandemia de COVID-19 en las regiones del Biobío y Nuble utilizando datos de teléfonos móviles Sebastián Cancino / Helen De la Fuente Rodrigo Victoriano / Sebastián Astroza Julio Covarrubias / Matías Durán Juan Antonio Carrasco

> y brotes en tiempo real a través de llamadas telefónicas Ezequiel Álvarez / Daniela Obando

Ezequiel Alvarez / Daniela Obando Sebastián Crespo / Enio García Nicolás Kreplak / Franco Mársico

Impactos del COVID-19 en el consumo eléctrico chileno

Rodrigo Moreno / Miguel Sánchez Carlos Suazo / Matías Negrete Daniel Olivares / Diego Alvarado



R E V I S T A

INGENIERIA DE SISTEMAS

Volumen XXXIV Diciembre 2020

Efecto de la Correlación y los Falsos Negativos en la Estrategia de Testeo Grupal para detección de COVID-19 Leonardo Basso, Vicente Salinas, Denis Sauré, Charles Thraves, Natalia Yankovic.	5
Iniciativas matemático-computacionales desde la Universidad de Buenos Aires para contribuir a la toma de decisiones en el contexto del COVID-19 en Argentina Mehrnoosh Arrar, Laouen Belloli, Ana María Bianco, Leonardo Boechi, Rodrigo Castro, Guillermo Durán, Roberto Etchenique, Natalia Fernández, Luciana Ferrer, Diego Garbervetsky, Rodrigo Goldsmit, Carolina Grillo Vidal, Juan Kamienkowski, Pablo Laciana, Esteban Lanzarotti, Mario Lozano, Rodrigo Maidana, Mauricio Mendiluce, Sol Minoldo, Leonardo Pepino, Ezequiel Pecker-Marcosig, Ezequiel Puerta, Rodrigo Quiroga, Guillermo Solovey, Marina Valdora, Mariano Zapatero.	31
Análisis de movilidad durante la pandemia de COVID-19 en las regiones del Biobío y Ñuble utilizando datos de teléfonos móviles Sebastián Cancino, Helen De la Fuente, Rodrigo Victoriano, Sebastián Astroza, Julio Covarrubias, Matías Durán, Juan Antonio Carrasco.	79
Estimación de los casos de COVID-19 y brotes en tiempo real a través de llamadas telefónicas Ezequiel Álvarez, Daniela Obando, Sebastián Crespo, Enio García, Nicolás Kreplak, Franco Marsico.	99
Impactos del COVID-19 en el Consumo Eléctrico chileno Rodrigo Moreno, Miguel Sánchez, Carlos Suazo, Matías Negrete, Daniel Olivares, Diego Alvarado, Carlos Sepúlveda, Héctor Otárola, Leonardo Basso.	119
Publicada por el DEPARTAMENTO DE INGENIERÍA INDUSTRIAL FACILITAD DE CIENCIAS DÍSICAS Y MATEMÁTICAS	

FACULTAD DE CIENCIAS FISICAS Y MATEMATICAS UNIVERSIDAD DE CHILE

REVISTA **INGENIERIA DE SISTEMAS**

EDITOR

Guillermo Durán

Departamento de Matemática e Instituto de Cálculo FCEN-UBA y CONICET, Argentina Departamento de Ingeniería Industrial Universidad de Chile

EDITOR ASOCIADO

Richard Weber

Departamento de Ingeniería Industrial Universidad de Chile

AYUDANTE DE EDICIÓN

Cinthya Vergara

Departamento de Ingeniería Industrial Universidad de Chile

COMITÉ EDITORIAL

René Caldentey

University of Chicago, Estados Unidos

Universidad de la República, Uruguay

Rafael Epstein

Universidad de Chile, Chile

Luis Llanos

Héctor Cancela

Westgate Energía, Chile

Javier Marenco

Universidad Nacional de General Sarmiento, Argentina

Juan de Dios Ortúzar

Pontificia Universidad Católica de Chile, Chile

Víctor Parada

Universidad de Santiago, Chile

Oscar Porto

GAPSO, Brasil

Lorena Pradenas

Nicolás Stier

Universidad de Concepción, Chile

Facebook Core Data Science. Estados Unidos

La Revista Ingeniería de Sistemas (RIS) es financiada parcialmente por el Instituto Sistemas Complejos de Ingeniería (ISCI), como reconocimiento a la difusión de las materias abordadas y de sus participantes.

Las opiniones y afirmaciones expuestas representan los puntos de vista de sus autores y no necesariamente coinciden con las del Departamento de Ingeniería Industrial de la Universidad de Chile.

Los artículos sólo pueden ser reproducidos previa autorización del Editor y de los autores.

Representante legal: Richard Weber Correo electrónico: ris@dii.uchile.cl Diagramación y Diseño:

Imagen COVID-19 Portada:

Dirección: Av. Beauchef 851, Santiago, Chile.

Web URL: www.dii.uchile.cl/ris

Cinthya Vergara

Centers for Disease Control and Prevention's Public Health Image Library N° 23312

Carta Editorial Volumen XXXIV

El COVID-19 cambió al mundo a lo largo del año 2020. Una enfermedad prácticamente desconocida se expandía por todo el planeta generando una gran cantidad de contagios y fallecimientos, retraso económico y largos aislamientos sociales.

El mundo de la ciencia y la academia centralizó su tiempo, sus conocimientos y sus ideas para colaborar con los tomadores de decisiones en la lucha contra esta feroz pandemia.

En esta oportunidad nos es muy grato presentar este nuevo número de la Revista de Ingeniería de Sistemas (RIS) dedicado a temas de frontera en Investigación de Operaciones, Gestión y Tecnología vinculados a la lucha contra el COVID-19. Queremos agradecer al Instituto Sistemas Complejos de Ingeniería (ISCI) por su colaboración para hacer posible esta publicación.

Este número contiene artículos de grupos que trabajaron intensamente en temas relacionados con la pandemia tanto en Chile como en Argentina. Incluso en alguno de ellos son co-autores los propios decisores de las políticas públicas.

Nuestro objetivo a través de esta publicación es contribuir a la generación y difusión de las tecnologías modernas de gestión y administración, y al uso de ellas para mejorar las decisiones en los ámbitos público y privado.

La revista pretende destacar la importancia de generar conocimiento en estas áreas, orientado tanto a problemáticas nacionales como a la realidad de países de características similares de la región.

Estamos seguros de que los artículos publicados en esta oportunidad muestran formas de trabajo innovadoras que serán de gran utilidad e inspiración para todos los lectores, ya sean académicos o profesionales, por lo que esperamos que esta iniciativa tenga la recepción que creemos se merece.

Guillermo Durán Editor

Richard Weber Editor Asociado

Llamado a Presentar Trabajos

La Revista Ingeniería de Sistemas (RIS) busca constituir un canal de divulgación de los avances en las áreas de Gestión de Operaciones, Tecnologías de Información e Investigación Operativa, que incluya los mundos académico y empresarial. Son particularmente apropiados artículos orientados a la práctica de estas disciplinas, que estimulen su uso o den cuenta de aplicaciones innovadoras de ellas, especialmente en América Latina.

También son bienvenidos artículos con análisis del estado del arte en un campo particular y de la forma en que los avances en dicho campo se han utilizado en la práctica.

Se espera que los artículos estén escritos de manera que puedan ser leídos por personas no especialistas en el tema tratado. Se recomienda incluir una lista de lecturas sugeridas para que los lectores no especialistas puedan profundizar en el tema.

Formato del Manuscrito

Los autores deben enviar un archivo en formato PDF del manuscrito que desean someter a referato a:

Comité Editorial Revista Ingeniería de Sistemas, Departamento de Ingeniería Industrial, Universidad de Chile, Santiago, Chile. Email: ris@dii.uchile.cl

Los manuscritos deben estar formateados para hojas tamaño carta, a doble espacio, márgenes de 2,5 centímetros en todos los lados y su extensión no debe exceder las 20 hojas.

La primera hoja debe contener el título del trabajo, nombre y dirección de los autores (teléfono y correo electrónico del autor de contacto) y un resumen de no más de 150 palabras.

Referencias

Las referencias se deben citar en el cuerpo del texto usando el nombre del autor y el año de publicación, e.g., Morton (1998). Al final del artículo se debe incluir la lista en orden alfabético de las referencias citadas en el texto. Para referencias de revistas científicas el formato es el siguiente: Autor(es), Año de publicación. Título. Nombre completo de la revista , Volumen e.g.:

Kodialam, M. y H. Luss, 1998. Algorithms for Separable Nonlinear Resource Allocation Problems. Operations Research, 44(2), 272-284.

Para referencias de libros el formato es el siguiente: autor(es), año de publicación. Título. Editorial, Ciudad; e.g.:

Kleinrock, L., 1975. Queueing Systems . John Wiley, New York.

En caso de haber más de una referencia con el mismo autor y año de publicación, se debe usar "a", "b", etc. como sufijo del año de publicación para diferenciarlas.

Detalles en www.dii.uchile.cl/ris

Efecto de la Correlación y los Falsos Negativos en la Estrategia de Testeo Grupal para detección de COVID-19

Leonardo Basso * †

Vicente Salinas * ‡

Denis Sauré * §

Charles Thraves * §

NATALIA YANKOVIC ¶

Resumen

La pandemia de COVID-19 ha puesto a prueba los sistemas de salud en todo el mundo. En situaciones como la actual, donde el testeo activo forma parte de las estrategias de respuesta y la disponibilidad de tests de detección es limitada, el método de testeo grupal surge como una alternativa para aumentar la capacidad de detección. La ganancia en eficiencia obtenida mediante el testeo grupal, derivada de testear múltiples muestras combinadas simultáneamente, depende del estado de la epidemia en la población objetivo. La progresión de la infección en poblaciones cerradas, tales como Establecimientos de Larga Estadía para Adultos Mayores (ELEAM), anticipa una correlación significativa en el contagio. Frente a ello, se desarrolló un modelo probabilístico que incorpora dicha correlación en el cálculo del tamaño de grupo óptimo para esquemas de testeo grupal de dos etapas. A su vez, en base a una experiencia previa implementando pilotos de testeo grupal en el SENAMA (Servicio Nacional del Adulto Mayor), se desarrolló un modelo de simulación de la dinámica de contagios en un ELEAM y se estimó la correlación del contagio entre residentes y funcionarios. Los resultados muestran que es posible aplicar el testeo grupal de forma efectiva en el contexto local, considerando distintos escenarios prevalencia y correlación, y manteniendo bajo control el riesgo de falsos negativos, y que la correlación en contagio puede ser alta, dependiendo de la política de cuarentena implementada por el ELEAM, validando la utilización de modelos que incorporan correlación en el cálculo de tamaños de grupo.

Palabras Clave: Testeo grupal, COVID-19, Simulación, Beta-Binomial

^{*}Instituto Sistemas Complejos de Ingeniería (ISCI), Santiago, Chile.

[†]Departamento de Ingeniería Civil - Transporte, FCFM, Universidad de Chile, Chile.

[‡]Departamento de Física, FCFM, Universidad de Chile, Chile.

[§]Departamento de Ingeniería Industrial, FCFM, Universidad de Chile, Chile.

[¶]ESE Business School, Universidad de los Andes, Chile.

1. Motivación y antecedentes

El virus causante de la pandemia de COVID-19, SARS-CoV-2, ha puesto bajo estrés a prácticamente la totalidad de los sistemas de salud del mundo. Los recursos de todo tipo han mostrado ser escasos y la detección y trazabilidad de los posibles contagiados se ha presentado como una estrategia efectiva para la contención de la pandemia.

En este contexto, las políticas de contención propuestas a la fecha consideran típicamente el testeo de todos los contactos estrechos de pacientes confirmados o sospechosos de COVID-19 [2]. En Chile, los contactos estrechos se definen como todas las personas que han estado en contacto con un caso confirmado con COVID-19, entre dos días antes del inicio de síntomas y 14 días después del inicio de los síntomas del enfermo, o 14 días después de confirmar el diagnóstico¹ en caso de ser un paciente asintomático [4]. El contacto estrecho supone haber mantenido: (i) más de 15 minutos de contacto cara a cara, a menos de un metro, sin mascarilla; (ii) haber compartido un espacio cerrado por dos horas o más, en lugares como oficinas, trabajos, reuniones, colegios, entre otros, sin mascarilla; (iii) haberse trasladado en cualquier medio de transporte cerrado a una proximidad menor de un metro con otro ocupante del medio de transporte que esté contagiado, sin mascarilla; o (iv) vivir o pernoctar en el mismo hogar o lugares similares a un hogar, tales como, hostales, internados, instituciones cerradas, ELEAM, hoteles, residencias, entre otros.

Bajo esta definición de contacto estrecho, es necesario testear poblaciones completas viviendo juntas, poniendo bajo presión la capacidad de realizar las pruebas diagnósticas de manera oportuna. En este contexto, es razonable esperar un alto grado de **correlación** en el contagio de ciertos grupos de pacientes, como por ejemplo aquellos habitando Establecimientos de Larga Estadía para Adultos Mayores (ELEAM) gestionados por el Servicio Nacional del Adulto Mayor (SENAMA).

En situaciones donde la disponibilidad de tests de detección para el cribado de pacientes infectados de una población es limitado, una posible estrategia que permite ahorrar, potencialmente, un número significativo de tests es el **testeo grupal**. En esta estrategia, las muestras individuales se combinan y se analizan de manera agrupada en búsqueda de un patógeno: si el resultado del test es negativo, se considera que todos los pacientes del grupo son negativos;

¹La confirmación de los casos de COVID-19 está basado en la detección de secuencias únicas del ARN del virus mediante pruebas de amplificación del ácido nucleico (PCR).

por otro lado, si el resultado del test es positivo, se concluye que al menos uno de los pacientes es positivo. Cuando el objetivo del test es diagnosticar a cada paciente, un resultado positivo requiere pruebas adicionales para determinar a los pacientes infectados.

A modo de ejemplo, considere un grupo de 10 pacientes cuyas muestras se analizan en un único grupo: un resultado negativo implica un ahorro de un 90 % de tests (relativo al testeo individual), incluyendo ahorros en reactivos asociados y tiempos de proceso. En la medida que un resultado negativo es dominante (lo cual depende de los supuestos subyacentes, ver Sección 1.2), esta estrategia aparece como un método efectivo para ampliar la capacidad de testeo de un sistema de salud, lo que posibilitaría la implementación de estrategias de testeo masivas, como las que se han llevado a cabo en otros países [16], o las que están siendo impulsadas por el gobierno [10].

El problema de minimizar el número de tests requeridos para diagnosticar un conjunto de pacientes mediante la estrategia de testeo grupal ha sido estudiado extensamente en los últimos 80 años (ver la Sección 1.2). Sin embargo, la estrategia de testeo óptima es muy difícil de encontrar, su cálculo es sumamente caro, en término de recursos computacionales y sumamente compleja de implementar, debido a su carácter adaptativo.

En el caso de los pacientes con COVID-19 se presentan una serie de dificultades para la implementación del testeo grupal. Por una parte, los falsos negativos han sido una preocupación en la lucha contra la pandemia [15] y en la práctica, las estrategias basadas en testeo grupal pueden ser más arriesgadas en relación a los potenciales falsos negativos. Por otra parte, la existencia de correlación entre pacientes ha sido evidente (por ejemplo, los ELEAM han sido foco de gran cantidad de contagios y muertes por COVID-19 en particular en España e Italia), pero no ha sido incorporada en los modelos. Finalmente, la toma de muestra nasofaríngea necesaria para el PCR es un procedimiento molesto y desagradable, que requiere de un personal sanitario cualificado, por lo que repetir la toma de muestras debería evitarse, limitando el diseño de la estrategia de testeo grupal a la clásica de dos etapas.

1.1. Objetivo

Este trabajo busca presentar una estrategia de testeo grupal que se haga cargo de la realidad de los pacientes COVID-19 en entornos donde se realiza testeo activo por contagios². Para esto, En primer lugar, se plantea un modelo probabilístico de contagio en una población cerrada, que incluye como

²Para realizar el análisis, no se necesita la revisión de una Junta de Repaso Institucional (IRB, por sus siglas en inglés) al tratarse de un trabajo de modelamiento y optimización teórico y predictivo, donde no se han usado datos individuales de pacientes.

input los parámetros de operación de la técnica de testeo, así como también la correlación en la infección en la población objetivo, extendiendo así modelos tradicionales para el cálculo de tamaños de grupo óptimo que no incluyen correlación. El modelo toma como base el modelo tradicional utilizado en la literatura de testeo grupal, el que asume independencia en el contagio entre los individuos en la población y lleva a una distribución binomial del número total de contagios, e introduce correlación en el contagio incorporando aleatoriedad en la probabilidad de contagio de los individuos. Dicho modelo permite estudiar, entre otras cosas, el efecto de la correlación en los tamaños de grupo óptimos y en el riesgo de falsos negativos.

En segundo lugar, se desarrolla un modelo de simulación de la dinámica de contagios en una población cerrada. Dicho modelo es capaz de incorporar distintas dinámicas de interacción entre subgrupos de la población, por ejemplo, funcionarios y residentes de un ELEAM, así como replicar distintas políticas de cuidado, cuarentena y testeo dentro de dicha población. Utilizando este modelo es posible, entre otras cosas, estimar el grado de correlación en el contagio en la población, posibilitando la validación de la intuición al respecto, y estudiar el efecto de las distintas estrategias de contención del virus en dicha correlación, y la progresión del contagio en general.

1.2. Contribución

Como resultado del trabajo realizado, se espera extender los cálculos de tamaños de grupo óptimos a escenarios con correlación en el contagio en la población objetivo, contribuyendo a la literatura existente en el tema. Así mismo, el análisis incluye aristas de la estrategia no cubierta por la literatura en testeo grupal. Este es el caso del riesgo de falsos negativos, cuyo análisis es de interés independiente de la extensión a escenarios con correlación positiva. En relación a este punto, una segunda contribución consiste en mostrar la existencia de un trade-off entre el número esperados de tests a utilizar y la varianza del número de falsos negativos, que se acentúa en escenarios con correlación positiva. Al respecto, el análisis indica que focalizar la elección de tamaño de grupo solamente en minimizar el número esperado de tests a utilizar puede resultar en decisiones arriesgadas, que apuestan fuertemente en la generación de falsos negativos para efectuar dicha minimización. Este es sin duda un factor a considerar al momento de decidir una estrategia de testeo, y sin duda el presente análisis es novedoso y de interés para los tomadores de decisiones.

Por último, el trabajo contribuye en mostrar que es posible enfrentar escenarios con alta correlación en contagio cuando la población objetivo es relativamente cerrada, como lo es el caso de los ELEAM operados por SENAMA. En particular, se muestra cómo la prevalencia y correlación observada al momento de realizar campañas de testeo dependen de la política general de cuidado, cuarentena y testeo en poblaciones cerradas. De forma prescriptiva, el análisis puede ser utilizado para recomendar técnicas y tamaños de grupo para realizar dichas campañas, de forma de minimizar la utilización de recursos escasos.

El trabajo está organizado de la siguiente forma. En la Sección 2 se revisa la literatura relevante. Luego, en la Sección 3, se presenta el modelo probabilístico utilizado para calcular medidas de desempeño asociadas a un tamaño de grupo bajo la estrategia de testeo grupal en dos etapas, esto es, el número esperado de tests, y el número esperado y varianza de falsos negativos. A continuación, en la Sección 4, se analizan los resultados de utilizar la estrategia propuesta, calculando los potenciales ahorros en el número de tests y el riesgo en el número de falsos negativos. En la Sección 5 se incluye una aplicación de la estrategia propuesta aplicada en el contexto de los ELEAM gestionados por el SENAMA. Finalmente, en la Sección 6, se presentan las reflexiones finales y las conclusiones de este trabajo.

2. Revisión bibliográfica

Las estrategias de testeo grupal no son nuevas. En [5] se presenta la estrategia original de dos etapas para pacientes infectados de sífilis en el contexto del ejército de los Estados Unidos. El modelo considera sensibilidad perfecta del test, por lo que los falsos negativos no son un problema, y se muestra que el número de tests necesarias para testear una población usando la estrategia de muestreo grupal es mucho menor con respecto a la estrategia tradicional de testeo individual, especialmente en situaciones de baja prevalencia.

A partir de este estudio seminal, numerosos trabajos han abordado el problema de testeo grupal, realizando modificaciones relacionadas con qué hacer en caso de que el test agrupado resulte positivo. Por ejemplo, [7] y [13] proponen continuar testeando de manera agrupada, ya sea en un único test adicional o recursivamente en dos subgrupos, toda vez que el resultado de una muestra agrupada sea positivo.

Las políticas de testeo grupal pueden ser clasificadas en adaptativas y no adaptativas. El primer caso se caracteriza por permitir más de una secuencia de testeo, donde los agrupamientos sucesivos son dependientes de los resultados de los tests previamente realizados. En este sentido, el modelo clásico de testeo grupal propuesto en [5] corresponde a una política de testeo adaptativa con

dos etapas. Otros trabajos como [14] han aplicado programación dinámica para optimizar la selección de subgrupos con el objetivo de minimizar los costos de falsos-negativos, falsos-positivos, y número de tests.

En el segundo caso, el de políticas no adaptativas, los grupos a testear son definidos previo a conocerse cualquier resultado intermedio de testeo. En [1] se presenta una revisión bibliográfica sobre políticas no adaptativas.

Los problemas de testeo grupal también pueden clasificarse según los supuestos sobre la población contagiada, la que puede ser tratada de manera probabilística o combinatorial. En el caso probabilístico, se define una distribución de probabilidad para modelar la condición de salud de los pacientes (infectados o no), la que por lo general considera que las probabilidades de infección son independientes. En el caso combinatorial, se considera que existe un conjunto (no conocido) de individuos que está infectado y es ese conjunto el que se desea identificar.

El testeo grupal adaptativo ha sido estudiado para pacientes COVID-19. Casos con múltiples etapas de testeo grupal han sido analizadas teóricamente por [9]. La estrategia clásica de testeo grupal en dos etapas para casos de COVID-19 ha sido validada internacionalmente en estudios realizados en Israel [17]. En Chile, se ha validado el PCR para muestras agrupadas, donde las muestras individuales se dividen en dos: una para el análisis agrupado y la otra para realizar un PCR individual en caso que el resultado del test agrupado sea positivo [6]. En [3] se utiliza simulación para evaluar el impacto de las políticas de testeo agrupado versus el testeo individual. En este trabajo no se incluye la posible correlación de la población ni tampoco se realiza una estimación del riesgo de falsos negativos de manera explícita.

3. Modelo de prevalencia y correlación en infección

Considerando el problema de testear a una población de N pacientes mediante la técnica de testeo grupal, y siguiendo la literatura, se supondrá que cada paciente en la población presenta el patógeno (a niveles detectables por la prueba PCR) con probabilidad $p \in (0,1)$. Esto es, p representa la prevalencia de SARS-CoV-2 en la población. Sin embargo, se tiene como supuesto que existe correlación en la presencia del patógeno en dos individuos cualesquiera en la población. Se denota $\rho \in [0,1)$ a esta correlación, y se restringe su rango a valores no-negativos, bajo la lógica que, en las aplicaciones de interés en donde hay algún grado de contacto entre los integrantes de una población, la

presencia del patógeno en un individuo aumenta la posibilidad de su presencia en el resto.

Formalmente, se introduce la correlación ρ mediante el siguiente modelo. Para $i \leq N$, se define

$$X_i := \begin{cases} 1 & \text{si paciente } i \text{ presenta el patógeno,} \\ 0 & \sim, \end{cases} \quad i \le N,$$

y se asume que, dado un valor $q \in (0,1)$, X_i distribuye Bernoulli(q) (es decir, $\mathbb{P}\{X_i=1\}=q$) para $i \leq N$ y la secuencia $\{X_i, i \leq N\}$ es independiente e idénticamente distribuida (notar que esta independencia está condicionada en el valor de q).

Cuando $\rho = 0$, se tiene que q = p, y el número de pacientes con el patógeno en un grupo de tamaño n distribuye Binomial(n, p). El analísis de este caso es el presentado en [5]. Por otro lado, cuando $\rho > 0$, se asume que q es una variable aleatoria distribuida $Beta(\alpha, \beta)$, donde

$$\alpha = p(\rho^{-1} - 1), \quad \beta = (1 - p)(\rho^{-1} - 1),$$
 (1)

(En rigor, tanto α como β son funciones de la prevalencia p y la correlación ρ , sin embargo se omitirá esta dependencia, para mantener la claridad de la exposición.) Esto tiene dos importantes consecuencias: i) la aleatoriedad en q introduce correlación en la presencia del patógeno en la población; y ii) la distribución del número de pacientes con el patógeno en un conjunto de tamaño k sigue una distribución $BetaBinomial(k,\alpha,\beta)$. El siguiente lema formaliza estas propiedades.

Lema 1. Considerando un conjunto de pacientes $S \subseteq \{1, ..., N\}$, $y X(S) := \sum_{i \in S} X_i$, la cantidad de pacientes en S con el patógeno. Se tiene que $X(S) \sim BetaBinomial(|S|, \alpha, \beta)$, esto es

$$\mathbb{P}\left\{X(S)=k\right\} = \binom{|S|}{k} \frac{B(k+\alpha,|S|-k+\beta)}{B(\alpha,\beta)}, \quad k \leq |S|,$$

donde |S| denota la cardinalidad del conjunto S y $B(\cdot, \cdot)$ es la función Beta. Adicionalmente, para $i \neq j$ se tiene que

$$Corr(X_i, X_j) := \rho, \quad \mathbb{E}\{X_i\} = p, \quad y \quad Var(X_i) = p(1-p).$$

Prueba. Antes de comenzar, notar que si se toma el caso $\rho > 0$ y, utilizando la Ecuación (1), se puede mostrar que

$$p = \frac{\alpha}{(\alpha + \beta)}, \quad \rho = \frac{1}{(\alpha + \beta + 1)}.$$

Estas relaciones se utilizarán en reiteradas ocasiones en el resto del documento.

La primera parte del Lema es consecuencia directa de la definición de una distribución BetaBinomial. Se tiene que

$$\mathbb{P}\left\{X(S) = k\right\} = \int_{0}^{1} \mathbb{P}\left\{X(S) = k|q = x\right\} \frac{x^{\alpha - 1}(1 - x)^{\beta - 1}}{B(\alpha, \beta)} dx
= \int_{0}^{1} \binom{|S|}{k} x^{k} (1 - x)^{|S| - k} \frac{x^{\alpha - 1}(1 - x)^{\beta - 1}}{B(\alpha, \beta)} dx
= \binom{|S|}{k} \frac{B(\alpha + k, \beta + |S| - k)}{B(\alpha, \beta)} \int_{0}^{1} \frac{x^{\alpha + k - 1}(1 - x)^{\beta + |S| - k - 1}}{B(\alpha + k, \beta + |S| - k)} dx
= \binom{|S|}{k} \frac{B(\alpha + k, \beta + |S| - k)}{B(\alpha, \beta)}.$$

En este desarrollo, primero se condiciona sobre el valor de q (utilizando la densidad de una variable aleatoria $Beta(\alpha,\beta)$), y después se usa el hecho que, condicional en el valor de q, X(S) distribuye binomial. Notar que, la última igualdad de la ecuación anterior, sigue de reconocer la integral de la densidad de una variable aleatoria $Beta(\alpha + k, \beta + |S| - k)$ sobre su dominio.

Respecto a la segunda parte del Lema 1, se tiene que

$$\mathbb{E}\left\{X_{i}\right\} = \int_{0}^{1} \mathbb{E}\left\{X_{i}|q=x\right\} \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} dx$$
$$= \int_{0}^{1} x \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} dx = \frac{\alpha}{\alpha+\beta} = p,$$

donde en la última igualdad se encuentra la esperanza de una variable aleatoria de distribución $Beta(\alpha,\beta)$ y se toma la definición de α y β en términos de p y ρ . Por otro lado, se tiene que

$$\mathbb{E}\left\{X_{j} X_{i}\right\} = \int_{0}^{1} x^{2} x^{\alpha-1} (1-x)^{\beta-1} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} dx$$

$$= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} \int_{0}^{1} x^{\alpha+1} (1-x)^{\beta-1} \frac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+2)\Gamma(\beta)} dx$$

$$= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)},$$

donde se establece $Beta(\cdot,\cdot)$ en términos de la función $\Gamma(\cdot)$ (gamma). Con esto, se tiene que

$$\operatorname{Cov} \{X_i, X_j\} = \mathbb{E} \{X_i X_j\} - \mathbb{E} \{X_i\} \mathbb{E} \{X_j\}$$
$$= \frac{\alpha}{\alpha + \beta} \left(\frac{\alpha + 1}{\alpha + \beta + 1} - \frac{\alpha}{\alpha + \beta} \right) = \frac{\alpha \beta}{(\alpha + \beta)^2 (\alpha + \beta + 1)} = p(1 - p)\rho.$$

Debido al carácter binario de X_i , se tiene que $\mathbb{E}\left\{X_i^2\right\} = \mathbb{E}\left\{X_i\right\} = \alpha/(\alpha+\beta)$. Esto implica que

$$\operatorname{Var}(X_i) = \frac{\alpha}{\alpha + \beta} \left(1 - \frac{\alpha}{\alpha + \beta} \right) = \frac{\alpha \beta}{(\alpha + \beta)^2} = p(1 - p).$$

Finalmente, se llega a que, para $i \neq j$,

$$Corr(X_i, X_j) = \frac{Cov(X_i, X_j)}{\sqrt{Var(X_i) Var(X_j)}} = \frac{Cov(X_i, X_j)}{Var(X_j)} = \rho.$$

Lo que concluye la prueba del Lema 1.

Considerando el caso $\rho > 0$ y n el tamaño de grupo utilizado en la estrategia de testeo grupal de dos etapas. Para calcular el número esperados de teste a utilizar bajo esta estrategia, se considera la especificidad y la sensibilidad de la técnica de testeo. Esto es, se denota por $S_e \in [0,1]$ a la probabilidad que un test entregue un resultado positivo cuando la muestra efectivamente contiene el patógeno (la sensibilidad del test), y por $S_p \in [0,1]$ a la probabilidad que un test entregue un resultado negativo cuando la muestra efectivamente no contiene el patógeno (la especificidad del test). Al respecto, se toma como supuesto que, estos parámetros de operación, no son afectados por el tamaño del agrupamiento, y que cada test falla en el diagnóstico de forma independiente, incluso si se utiliza la misma muestra (en un testeo sucesivo).

Prueba. Estos supuestos son razonables en el marco de la especificidad del test PCR, puesto que los falsos positivos ocurren mayoritariamente debido a problemas en el manejo de las muestras. Los supuestos son más fuertes en el caso de la sensibilidad, puesto que los falsos negativos ocurren cuando, por ejemplo, una de las muestras incluidas en el agrupamiento se encuentra muy cerca pero bajo del umbral de detección³ (por lo que la muestra, testeada de forma individual, entrega un resultado positivo); en este caso es posible que se produzca dilución en la muestra, lo que ubique a la muestra combinada ligeramente por sobre el umbral de detección, con lo que la muestra es incorrectamente etiquetada como libre del patógeno. En la práctica, sin embargo, es muy díficil encontrar muestras cercanas al umbral de detección [6].

Sea T el número de tests utilizados para diagnosticar a toda la población. El Lema 2 entrega un expresión para la esperanza de T.

³La técnica de testeo PCR mide cuantos ciclos (de calentamiento/enfriamiento de la muestra) se requieren para amplificar la presencia del patógeno de forma de hacerla detectable. Por esto, si dicho tiempo (en ciclos) es menor que cierto umbral, se concluye que el resultado es positivo.

Lema 2. Suponiendo que n es múltiplo de N, entonces,

$$\mathbb{E}\left\{T\right\} = N\left(\frac{1}{n} + S_e + (1 - S_e - S_p)\frac{B(\alpha, n + \beta)}{B(\alpha, \beta)}\right).$$

Prueba. Sea S_k el conjunto de pacientes incluidos en el grupo k a testear, y se denota por T_k al número de tests necesarios para diagnosticar a los pacientes en el grupo k. (Notar que $\{S_k, k = 1...N/n\}$ forma una partición de la población). Se tiene que

$$\mathbb{E}\left\{T\right\} = \sum_{k=1}^{N/n} \mathbb{E}\left\{T_{k}\right\}$$

$$= \sum_{k=1}^{N/n} \left(\left(1 + n(1 - S_{p})\right) \mathbb{P}\left\{X(S_{k}) = 0\right\} + \left(1 + n S_{e}\right) \left(1 - \mathbb{P}\left\{X(S_{k}) = 0\right\}\right)\right)$$

$$= \sum_{k=1}^{N/n} 1 + nS_{e} + n\left(1 - S_{e} - S_{p}\right) \mathbb{P}\left\{X(S_{k}) = 0\right\}$$

$$= N\left(\frac{1}{n} + S_{e} + \left(1 - S_{p} - S_{e}\right) \frac{B(\alpha, n + \beta)}{B(\alpha, \beta)}\right).$$

Donde la primera igualdad viene de la linealidad de la esperanza, la segunda de condicionar sobre el número de pacientes con el patógeno en el grupo k y la última del Lema 1, y hecho que el número de infecciones en un grupo se distribuye de igual forma en cada grupo.

Cuando no existe correlación, se obtiene el resultado presentado en [3], que extiende el desarrollo en [5] considerando la especificidad y sensitividad del test.

Notar que la expresión arriba es muy fácil de evaluar (la función Beta viene incorporada en la mayoría de los software estadísticos). Considerando que existen limites naturales al tamaño de grupo a utilizar en la práctica (por ejemplo, a la fecha, el método de testeo grupal ha sido válido en laboratorio para grupos de hasta 32 muestras sin modificar el protocolo del test: adicionalmente, en ausencia de tecnologías de procesamiento automatizado, el personal de laboratorio puede manejar tamaños de grupo relativamente bajos), esta expresión puede ser usada directamente para el cálculo del tamaño óptimo, vía enumeración.

Podemos obtener del Lema 2 que, para un tamaño de grupo dado, los parámetros de operación de la técnica de testeo afectan directamente el número esperado de tests requeridos. En particular, mientras mayor sea la especificidad del test, menor es el número esperado de tests requeridos (en la práctica, la especificidad del test PCR es cercana al 100 %). Por otro lado, el efecto de la

sensibilidad es el inverso, y a mayor sensibilidad, mayor es el número esperado de tests a utilizar (debido al aumento del riesgo de obtener falsos negativos). Ciertamente, el riesgo de falsos negativos es algo que debiese considerarse al momento de determinar tamaños de grupo. A continuación se realiza un análisis de este riesgo. Se denota por F el número de falsos negativos asociados al diagnóstico de la población. El Lema 3 caracteriza el valor esperado y la varianza de F, en función de los parámetros de operación, y el tamaño de grupo.

Lema 3. Suponiendo que n es múltiplo de N, entonces se tiene que $\mathbb{E}\left\{F\right\} = N(1 - S_e^2)$ y

$$Var(F) = N(1 - S_e^2)p - N(1 - S_e)(1 + S_e - S_e^2 - nS_e^3)(p^2 + p(1 - p)\rho) + N^2(1 - S_e^2)^2p(1 - p)\rho.$$

Prueba. Siguiendo la prueba del Lema 2, sea F_k el número de falsos negativos obtenidos al testear al grupo k. Se tiene que

$$\mathbb{E} \{F\} = \sum_{k=1}^{N/n} \mathbb{E} \{F_k\}$$

$$= \sum_{k=1}^{N/n} \sum_{i=1}^{n} \mathbb{E} \{F_k | X(S_k) = i\} \mathbb{P} \{X(S_k) = i\}$$

$$\stackrel{(a)}{=} \sum_{k=1}^{N/n} \sum_{i=1}^{n} (i (1 - S_e) + i (1 - S_e) S_e) \mathbb{P} \{X(S_k) = i\}$$

$$= \sum_{k=1}^{N/n} (1 - S_e^2) \mathbb{E} \{X(S_k)\} = \frac{N}{n} (1 - S_e^2) n \frac{\alpha}{\alpha + \beta} = N(1 - S_e^2) p.$$

Se observa que en (a) se utiliza el hecho que, cuando $X(S_k) = i$, si el test grupal resulta en un falso negativo (lo que ocurre con probabilidad $(1 - S_e)$), esto resulta en i falsos negativos, y cuando el test grupal entrega el resultado correcto (lo que ocurre con probabilidad S_e), esto resulta, en promedio, en $(1 - S_e)i$ falsos negativos, provenientes de los tests individuales. La última igualdad utiliza el hecho que la esperanza de una variable aleatoria distribuida $BetaBinomial(k, \alpha, \beta)$ es $k(\alpha/(\alpha + \beta))$.

Considerar ahora el cálculo de la varianza de F. Primero, se debe notar que condicional en q, los falsos negativos en cada grupo son variables aleatorias independientes, por lo que se tiene que

$$\mathbb{E}\left\{F^{2}|q\right\} = \sum_{k=1}^{N/n} E\left\{F_{k}^{2}|q\right\} + \frac{N}{n} \left(\frac{N}{n} - 1\right) \left(n \, q \, (1 - S_{e}^{2})\right)^{2}.$$

Para desarrollar el término asociado a cada grupo, recordar que si $X \sim Binomial(n,q)$, entonces

$$E\{X^2\} = Var(X) + \mathbb{E}\{X\}^2 = nq(1-q) + (nq)^2.$$

Luego, tomando que, condicional en $X(S_k) = i$ y el test grupal no falló, el número de falsos negativos obtenidos en el grupo k distribuye $Binomial(i, (1 - S_e))$. Dejando a G denotar al evento en el que test grupal no falla, se tiene que

$$E\left\{F_{k}^{2}|q\right\} = \sum_{i=1}^{n} \left((1-S_{e})\mathbb{E}\left\{F_{k}^{2}|X(S_{k})=i,\bar{G}\right\} + S_{e}\mathbb{E}\left\{F_{k}^{2}|X(S_{k})=i,G\right\}\right)$$

$$\mathbb{P}\left\{X(S_{k})=i|q\right\}$$

$$= \sum_{i=1}^{n} \left((1-S_{e})i^{2} + S_{e}\left(iS_{e}(1-S_{e}) + i^{2}(1-S_{e})^{2}\right)\right)\mathbb{P}\left\{X(S_{k})=i|q\right\}$$

$$= S_{e}^{2}(1-S_{e})nq + (1-S_{e})(1+S_{e}(1-S_{e}))\left(q(1-q)n + n^{2}q^{2}\right)$$

$$= \left(S_{e}^{2}(1-S_{e}) + (1-S_{e})(1+S_{e}(1-S_{e}))nq\right)$$

$$- \left((1-S_{e})(1+S_{e}(1-S_{e}))q^{2}n(1-n)\right)$$

$$= \left(1-S_{e}^{2}\right)nq - \left((1-S_{e})(1+(1-S_{e})S_{e})q^{2}n(1-n)\right).$$

Con esto, tenemos que

$$\mathbb{E}\left\{F^{2}|q\right\} = N(1 - S_{e}^{2})q - ((1 - S_{e})(1 + (1 - S_{e})S_{e})q^{2}N(1 - n) + N^{2}q^{2}((1 - S_{e}^{2}))^{2} - Nnq^{2}((1 - S_{e}^{2}))^{2}.$$

Ahora, notamos que

$$\mathbb{E}\left\{q^{2}\right\} = \operatorname{Var}(q) + \mathbb{E}\left\{q\right\}^{2} = \frac{\alpha\beta}{(\alpha+\beta)^{2}(\alpha+\beta+1)} + \frac{\alpha^{2}}{(\alpha+\beta)^{2}} = p(1-p)\rho + p^{2}.$$

Finalmente, tomando la esperanza (respecto a q) sobre $\mathbb{E}\left\{F^2|q\right\}$, y substrayendo $\mathbb{E}\left\{F\right\}^2$, se tiene

$$\operatorname{Var}\left\{F\right\} = N\left(1 - S_{e}^{2}\right)p - \left(\left((1 - S_{e})(1 + (1 - S_{e})S_{e})(1 - n) + N^{2}\left((1 - S_{e}^{2})\right)^{2} - Nn\left((1 - S_{e}^{2})\right)^{2}\right) \left(p(1 - p)\rho + p^{2}\right) - N^{2}(1 - S_{e}^{2})^{2}p^{2}$$

$$= N(1 - S_{e}^{2})p - N(1 - S_{e})(1 + S_{e} - S_{e}^{2} - nS_{e}^{3})(p^{2} + p(1 - p)\rho) + N^{2}(1 - S_{e}^{2})^{2}p(1 - p)\rho.$$

donde la última igualdad viene de agrupar términos de acuerdo a sus dependencias, después de algo de álgebra. \Box

Notar que el número esperado de falsos negativos es independiente del tamaño de grupo utilizado. Erróneamente se podría concluir que el tamaño de grupo no afecta a F. La segunda parte del Lema 3, sin embargo, nos dice que a medida que el tamaño de grupo aumenta, también lo hace la varianza de F. La consecuencia para la selección del tamaño del agrupamiento es directa: a mayor tamaño mayor es el riesgo asociado de tener falsos negativos (tomar como referencia el caso extremo donde n=N). Estos aspectos se revisarán numéricamente en la Sección 4.

4. Resultados

En la Tabla 1 se presentan los resultados del modelo obtenidos con una población de 100 pacientes (N=100) para prevalencias que varían desde 0.01% a 40%, considerando 4 niveles de correlación. Se incluye el agrupamiento óptimo, el número esperado de tests y los ahorros en el número de tests comparado con la estrategia de testeo individual y la estrategia de testeo grupal, pero utilizando el agrupamiento que se obtiene si se ignora la correlación. Adicionalmente, en el caso que el test no sea perfecto y pueda arrojar falsos negativos, se incluye el valor esperado de los falsos negativos y su desviación estándar. Para evaluar el impacto del testeo grupal en el riesgo de los falsos negativos se consideró $S_e=0.7$. En la práctica, la sensibilidad del test PCR varía dependiendo del método de recolección de la muestra, e.g. raspado nasofaríngeo, saliva, etc. En este modelo se consideró el escenario $S_e=0.7$ como un extremo y $S_e=1.0$ el otro.

Los ahorros en el número de tests pueden ser tan grandes como del $99\,\%$ en comparación a realizar pruebas individuales y de hasta el $60\,\%$ en comparación con realizar testeo grupal pero utilizando el agrupamiento calculado a partir de un modelo que no considere la correlación.

Adicionalmente, se observa que para prevalencias bajas los grupos de testeo son grandes, iguales al tamaño de la población. El tamaño del agrupamiento óptimo disminuye al aumentar la prevalencia para $S_e = 1$. En el caso de tener tests imperfectos ($S_e < 1$) se produce una discontinuidad y vuelve a ser óptimo testear a toda la población en un único grupo al aumentar la prevalencia.

La Figura 1 muestra los ahorros en el número de tests al utilizar el agrupamiento óptimo del modelo (que incluye explícitamente la correlación) versus el caso donde la correlación no se considera, en función de la correlación para distintos niveles de prevalencia ($S_e = 0, 7, N = 100$).

Si bien el número esperado de falsos negativos no depende del tamaño del grupo, su desviación estándar es creciente en n. En la Figura 2 se presenta el agrupamiento óptimo en función de la prevalencia, para distintos niveles de correlación. Se ve claramente que el testeo grupal en escenarios de alta prevalencia y de baja correlación se produce una apuesta por tener un resultado negativo en un único test agrupado (i.e., n=100), el que con alta probabilidad será un falso negativo producto de la sensibilidad del test ($S_e=0,7,$ N=100).

La Figura 3 muestra como la desviación estándar de falsos negativos disminuye al reducir el tamaño del grupo, a la vez que el número esperado de tests crece. Podemos ver que usando n=6 en vez del agrupamiento óptimo $(n^*=100)$ aumentamos en 8.5% el número esperado de tests (de 71 a 77) pero disminuimos la desviación estándar de los falsos negativos en un 39% (de 15 a 9.2).

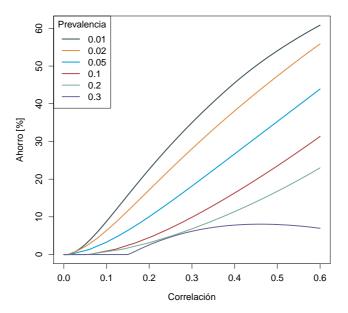


Figura 1: Ahorros en el número esperado de tests considerando la correlación de manera explícita vs testeo grupal sin incluir correlación ($S_e = 0, 7, N = 100$)

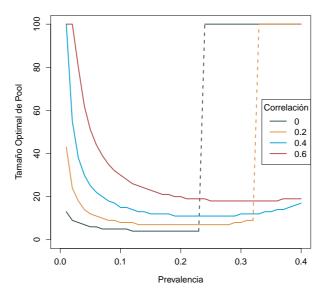


Figura 2: Tamaño del agrupamiento óptimo en función de la prevalencia ($S_e=0,7,\,N=100$)

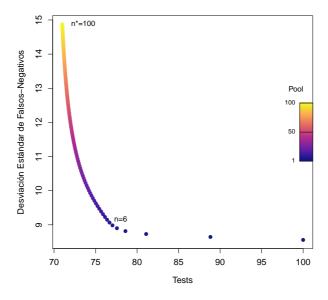


Figura 3: Desviación estándar y número esperado de tests para distintos tamaños de grupos ($S_e=0,7,\,N=100$)

Tabla 1: Agrupamiento óptimo, número esperado de tests y porcentaje de ahorro en relación a estrategia de testeo individual y de la estrategia de testeo grupal ignorando la correlación. A la izquierda $S_e=1$ y a la derecha $S_e=0,7$ donde se reporta además el número esperado de falsos negativos y su desviación estándar (N=100).

	`	/	Ahorros	s [%]			Ahorros	Falsos Negativos				
Prevalencia	n*	Exp.	Testeo	$\rho = 0$	n*	Exp.	Testeo	$\rho = 0$	Exp. (Std.)			
Trevalencia	11	tests	ind.	$\rho = 0$	11	tests	ind.	$\rho = 0$	Exp. (Std.)			
Correlación = 0												
0,0001	100	2	98	0	100	2	98	0	0 (0)			
0,001	32	6	94	0	39	5	95	0	0 (0)			
0,01	11	20	80	0	13	16	84	0	1 (1)			
0,02	8	27	73	0	9	23	77	0	1 (1)			
0,05	5	43	57	0	6	35	65	0	3(2)			
0,1	4	59	41	0	5	49	51	0	5 (2)			
0,2	3	82	18	0	4	66	34	0	10 (3)			
0,3	3	99	1	0	100	71	29	0	15 (10)			
0,4	1	100	0	0	100	71	29	0	20 (13)			
				Correla	ción = 0	.2						
0,0001	100	1	99	0	100	1	99	0	0 (0)			
0,001	100	2	98	42	100	2	98	40	0 (1)			
0,01	31	12	88	18	43	9	91	23	1 (3)			
0,02	17	19	81	12	24	15	85	17	1 (3)			
0,05	9	34	66	7	12	27	73	10	3(5)			
0,1	6	51	49	3	8	40	60	5	5 (7)			
0,2	4	73	27	2	7	57	43	3	10 (10)			
0,3	4	90	10	1	8	68	32	3	15 (11)			
0,4	1	100	0	0	100	71	29	0	20 (19)			
				Correla	ción = 0	.4						
0,0001	100	1	99	0	100	1	99	0	0 (0)			
0,001	100	2	98	54	100	1	99	50	0 (1)			
0,01	72	8	92	40	100	6	94	46	1 (4)			
0,02	38	13	87	31	55	10	90	38	1 (5)			
0,05	18	25	75	22	25	19	81	27	3 (8)			
0,1	11	41	59	13	15	31	69	16	5 (10)			
0,2	7	62	38	10	11	47	53	11	10 (14)			
0,3	7	78	22	6	12	58	42	8	15 (16)			
0,4	8	90	10	10	17	66	34	3	20 (18)			
				Correla	ción = 0	.6						
0,0001	100	1	99	0	100	1	99	0	0 (0)			
0,001	100	1	99	60	100	1	99	54	0 (1)			
0,01	100	5	95	58	100	4	96	61	1 (5)			
0,02	81	9	91	49	100	6	94	56	1 (7)			
0,05	36	18	82	39	51	13	87	44	3 (10)			
0,1	20	31	69	28	30	23	77	31	5 (13)			
0,2	13	50	50	21	20	37	63	23	10 (17)			
0,3	12	65	35	15	18	48	52	7	15 (19)			
0,4	12	77	23	23	19	56	44	5	20 (21)			

5. Caso de estudio: Correlación en un ELEAM

El modelo de prevalencia y correlación en los contagios presentado en la Sección 3 está motivado por la realidad de los ELEAM gestionados por el SENAMA. En estos centros, un grupo de adultos mayores vive bajo los cuidados de un equipo de profesionales de la salud.

Para utilizar la estrategia de testeo grupal analizado, es necesario conocer la prevalencia y la correlación de la población para poder escoger adecuadamente el tamaño del grupo. Con este propósito, se desarrolló una herramienta que permite simular la evolución de infectados y del número de pacientes sintomáticos entre la población que compone un ELEAM. A continuación se detalla dicha herramienta, y los resultados obtenidos.

5.1. Detalles de la rutina de simulación

En la simulación se consideraron dos grupos de individuos: residentes y funcionarios. En ella, se generó una matriz de interacciones entre la población, donde se establece la probabilidad que dos individuos (funcionarios o residentes) entren en contacto durante un turno: mientras mayor sea la probabilidad de las interacciones, se espera una difusión de los contagios más rápida. Con ello, se supuso que la probabilidad de interacción diaria entre dos miembros cualesquiera de la población es fija (se realizan simulaciones con diferentes valores para este valor).

En términos del modelo epidemiológico, el tiempo de incubación se modeló con una distribución Lognormal [8], mientras que la capacidad de contagio sigue la evolución de la infección del paciente de manera escalada [11]. A su vez, todos los pacientes contagiados tienen la capacidad de contagiar, aunque existe una proporción que nunca presenta síntomas. La simulación consideró que un 30 % de los pacientes asintomáticos, independiente del grupo al que pertenezcan (en línea con la evidencia internacional [12]). Luego, los principales supuestos y parámetros utilizados fueron⁴:

Supuestos.

 Las personas se infectan sólo una vez. Al recuperarse, quedan inmunizados.

⁴El detalle de esta simulación se encuentra disponible en https://covidanalytics.isci.cl/simulacion.

- Las tareas de los funcionarios contagiados y detectados son cubiertas por el resto de los funcionarios.
- La infecciosidad, largo del período de incubación y presencia/ausencia de síntomas son independientes entre individuos.
- El tiempo de incubación sigue una distribución Lognormal [8] y el tiempo de recuperación distribuye uniforme entre 2 y 4 semanas.
- La probabilidad de testear positivo está basada en la curva de positividad presentada en [11], considerando la evolución de la enfermedad en un paciente.
- La cuerva de infecciosidad es proporcional a la curva de positividad.

Parámetros.

- Matriz de turnos: indica las horas de interacción dentro de la población cerrada. Por ejemplo, para residentes, es una matriz de unos, y para funcionarios contiene unos (ceros) por las noches y ceros (unos) para el resto del día, para el turno nocturno (diurno).
- Matriz de interacción: indica la probabilidad de interacción entre individuos de distintos subgrupos de la población.
- Grupos: grupos considerados, en este caso, residentes (28) y funcionarios (22).
- Pacientes asintomáticos: Cada paciente es asintomático de forma independiente, con probabilidad 30 % [12].
- Probabilidad de contagio externa funcionarios: corresponde a la probabilidad diaria de contagio externa igual a 0.1 %.
- Cuarentena preventiva de pacientes sintomáticos: corresponde a todos los pacientes dos días tras desarrollar síntomas comienzan una cuarentena preventiva durante 14 días.

El proceso de simulación, se realizó mediante una simulación de Montecarlo para estudiar la evolución díaria de la infección. A su vez, se registró el número de personas infectadas (independiente de la presencia de síntomas), y de su estado de cuarentena. Cada ejecución de la simulación se comienza con toda la población no infectada.

 En t = 0 se asignó al azar la condición de sintomático/asintomático a cada individuo, que se utiliza en caso el individuo se contagie en algún momento durante la simulación.

- Para cada día, se calculó la probabilidad que cada paciente se contagie durante el transcurso de la jornada.
- Los contagios se generaron de acuerdo a las probabilidades calculadas arriba. Para cada persona infectada, se generaron tiempos de incubación, infecciosidad y recuperación.
- Tras esto, la simulación se mueve al siguiente día.

5.2. Caso de estudio

Como caso de estudio, se consideró un ELEAM con 28 residentes y 22 funcionarios en dos turnos de 11 personas cada uno. La simulación se inicia con toda la población del centro libre del virus, donde los funcionarios son quienes potencialmente introducen la infección con una probabilidad que depende de la prevalencia global.

Los residentes sólo pueden contagiarse por interacciones con funcionarios o residentes del centro, mientras que los funcionarios pueden contagiarse en sus interacciones en el centro o de manera exógena fuera del trabajo. En este escenario, las interacciones entre dos residentes tienen un mayor riesgo en comparación con las existentes entre un funcionario y un residente o entre dos funcionarios con igual turno, pues los contactos entre residentes suelen ser más largos. En ese sentido, un individuo sano puede contagiarse si entra en contacto con un paciente con potencial infeccioso, donde la probabilidad de contagio dependerá de la capacidad de contagio y del riesgo de la interacción.

Se toma como supuesto que todo paciente sintomático al segundo día de presentar síntomas se aísla en cuarentena preventiva y no tiene la posibilidad de contagiar. Esto último, busca replicar la estrategia de testeo reactivo que se ha aplicado en general por SENAMA, donde se testea de manera selectiva a quienes presentan síntomas de la enfermedad.

La Figura 4 presenta la evolución del número de casos activos y de los casos sintomáticos para tres niveles de probabilidad de interacción, desde el primer día que un paciente presenta sintómas. Es posible observar que, en caso de un alto nivel de interacciones, si no se realiza ninguna intervención al cabo de 22 días prácticamente la totalidad de la población estaría contagiada.

La Figura 5 presenta la evolución de la prevalencia y de la correlación para la población total del centro (N=28+11+11=50) en función de los días de simulación, mientras que la Figura 6 considera como primer día de referencia aquel cuando se presenta el primer caso sintomático (el que es sospechoso de COVID-19, pero no ha sido confirmado con una prueba PCR). Ambas gráficas fueron construidas para tres niveles en la probabilidad de interacción

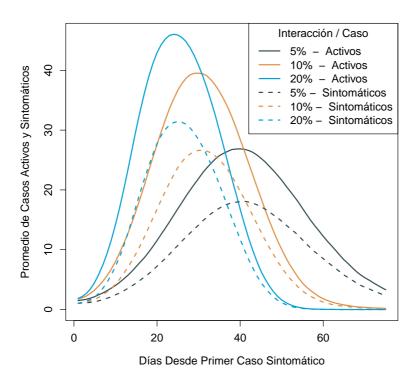


Figura 4: Número esperado de casos activos y de casos sintomáticos, para probabilidades de interacciones diarias del 5, 10 y 20 % (N = 50).

entre individuos, $5\,\%,\ 10\,\%$ y $20\,\%,$ entre cualquier par de miembros de la comunidad.

Es posible ver que, tanto la prevalencia como la correlación, aumentan a medida que avanzan los días desde el primer paciente con síntomas, hasta llegar a un peak y luego decaer. Cabe destacar además que, dicho peak, tanto para la prevalencia como para correlación, ocurren casi en el mismo día (para un mismo grado de interacción). Además, se aprecia que el valor máximo de la prevalencia está estrechamente relacionado de manera creciente con el grado de interacción (ver panel izquierdo de las Figuras 5 y 6). A su vez, se aprecia que el peak de la correlación (con respecto del día de inicio de la simulación), aumenta con el grado de interacción de los individuos, ver panel derecho de la Figura 5, pero no ocurre lo mismo al considerar el día con respecto al primer caso sintomático. Otro elemento a destacar, es la tasa a la que crece la prevalencia de la enfermedad y la correlación en ambas figuras, donde se observa que la prevalencia crece más rápido en el caso de considerar los días con res-

pecto al primer caso sintomático, mientras que, en el caso de la correlación, se tiene un crecimiento más rápido en el caso contrario. Es importante destacar que considerar el primer día de simulación o el primer día en que existe un caso sintomático sólo cambia el horizonte temporal de las Figuras 5 y 6, por lo que se esperaría similar comportamiento si se dejara evolucionar la simulación más allá del día 50 desde su inicio.

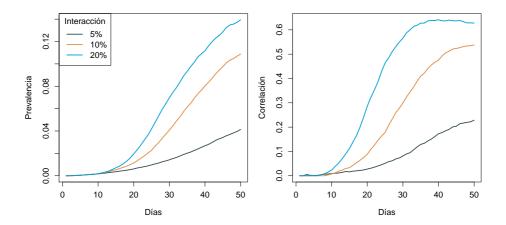


Figura 5: Prevalencia y correlación en la población, para probabilidades de interacciones diarias del 5, 10 y 20 % considerando el primer día de simulación (N=50).

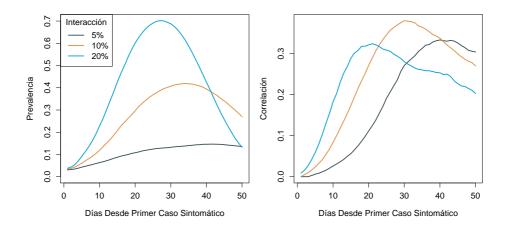


Figura 6: Prevalencia y correlación en la población, para probabilidades de interacciones diarias del 5, 10 y 20 % considerando el primer día en que existen pacientes sintomáticos (N=50).

Dada la evolución temporal de la prevalencia y de la correlación, la situa-

ción epidemiológica de la población bajo estudio puede ser muy diferente entre un día y otro. Estos resultados confirman la intuición detrás de la definición de contacto estrecho y las recomendaciones de testeo establecidas por el Ministerio de Salud [10], puesto que se observa una correlación significativa en contagios. Esto implica que, los agrupamientos recomendados, si se utiliza una estrategia de testeo grupal, serán diferentes dependiendo de la etapa en la que se encuentre la problación.

Las Tablas 2 y 3 presentan la prevalencia, correlación y los agrupamientos recomendados por el modelo presentado en la Sección 3 y el modelo de agrupamiento que no incorpora explícitamente la correlación. En la última columna se incluye el ahorro en el número esperado de tests de seguir el agrupamiento óptimo que considera la correlación versus el caso de agrupamiento que omite dicha correlación (es decir, considera $\rho=0$). Este análisis supone que $S_e=1$, para prevenir el efecto indeseable de "apostar" a un resultado negativo de todo el grupo producto de la sensibilidad de la prueba, como se discutió en la Sección 4. En el caso de la Tabla 2, los días hacen referencia al inicio de la simulación, mientras que, en la Tabla 3, se considera el primer día como aquel en que se tiene un paciente sintomático en la población.

Tabla 2: Para cada día de simulación: prevalencia, correlación, número óptimo de grupo (n^*) , número óptimo de grupo omitiendo correlación (n^{\diamond}) , y ahorro en número de tests a usar. Probabilidades de interacciones diarias del 5, 10 y 20 % de izquierda a derecha.

Día	Prev.	ρ	n^*	n^{\diamond}	Ahorro [%]	Prev.	ho	n^*	n^{\diamondsuit}	Ahorro [%]	Prev.	ho	n^*	n^{\diamondsuit}	Ahorro $[\%]$
5	0,00	0,00	46	46	0,0	0,00	0,00	47	47	0,0	0,00	0,00	50	50	0
10	0,00	0,01	25	25	0,0	0,00	0,01	24	24	0,0	0,00	0,02	31	24	2,3
15	0,00	0,02	20	17	0,8	0,01	0,03	18	14	2,1	0,01	0,11	25	12	11,8
20	0,01	0,03	16	13	1,5	0,01	0,09	15	10	5,1	0,02	$0,\!29$	25	8	20,5
25	0,01	0,05	14	11	1,9	0,02	0,18	14	7	10,8	0,04	$0,\!46$	25	5	32,3
30	0,01	0,08	13	9	3,7	0,04	0,30	15	6	13,0	0,07	$0,\!57$	24	4	34,6
35	0,02	0,13	13	8	5,2	0,06	0,41	15	5	17,7	0,09	0,63	24	4	31,9
40	0,03	$0,\!17$	12	7	7,2	0,08	0,48	16	4	23,5	0,11	0,64	22	4	28,3
45	0,03	0,20	12	6	9,3	0,10	$0,\!52$	16	4	22,4	0,13	0,64	19	3	34,1
50	0,04	$0,\!23$	11	5	12,9	0,11	$0,\!54$	15	4	20,9	0,14	0,63	18	3	31,3

Se puede apreciar, de la Tabla 2, cómo el tamaño de grupo recomendado, tanto considerando como no considerando correlación, baja a medida que avanzan los días. Esto se debe al aumento sostenido de prevalencia y correlación. Además, se aprecia que los casos con mayor correlación ocurren al asumir un mayor grado de interacción (probabilidad de interacción del 20%). Esto trae como consecuencia que omitir tales niveles de correlación implican una pérdida de eficiencia en el número esperado de tests a utilizar. Por ejemplo, se aprecia que luego de un mes de la simulación, usar una estrategia de tes-

Tabla 3: Para cada día de simulación con respecto al primer caso sintomático: prevalencia, correlación, número óptimo de grupo (n^*) , número óptimo de grupo omitiendo correlación (n^{\diamond}) , y ahorro en número de tests a usar. Probabilidades de interacciones diarias del 5, 10 y 20 % de izquierda a derecha.

Día	Prev.	0	n^*	n^{\diamond}	Ahorro	Prev.	ρ	n^*	n^{\diamond}	Ahorro	Prev.	ρ	n^*	n^{\diamond}	Ahorro
Dia Tiev. p	Ρ	10	10	[%]	1100.	Ρ	10	10	[%]	l icv.	Ρ	16	16	[%]	
5	0,04	0,01	5	5	0,0	0,06	0,02	5	5	0,0	0,09	0,06	4	4	0
10	0,06	0,03	5	5	0,0	0,12	0,08	4	4	0,0	0,23	0,18	4	3	1,4
15	0,09	0,06	4	4	0,0	0,21	0,17	4	3	1,6	0,43	0,29	9	1	1,3
20	0,11	0,11	5	4	0,2	0,30	$0,\!27$	5	3	2,5	0,60	0,32	1	1	0
25	0,12	0,19	5	3	6,2	0,37	$0,\!35$	7	1	10,1	0,69	0,31	1	1	0
30	0,13	$0,\!27$	6	3	9,1	0,41	$0,\!38$	8	1	8,2	0,69	$0,\!28$	1	1	0
35	0,14	0,31	7	3	10,3	0,42	$0,\!36$	8	1	6,3	0,59	$0,\!26$	1	1	0
40	0,15	$0,\!33$	7	3	10,8	0,39	0,34	7	1	6,9	0,43	$0,\!25$	1	1	0
45	0,14	0,33	7	3	10,6	0,34	0,30	6	1	10,5	0,26	0,23	4	3	1,8
50	0,14	0,30	7	3	10,5	0,27	$0,\!27$	5	3	2,7	0,13	0,20	5	3	5,9

teo que considere la correlación puede traer un ahorro en el número de tests de 34,6 % (versus testeo grupal miope a la correlación). Al realizar el mismo ejercicio sobre los resultados de la Tabla 3, se observa que el ahorro en número esperado de tests es de menor magnitud que en el caso anterior. Esto ocurre porque los niveles de correlación son menores al mirar los escenarios desde el día del primer caso sintomático versus desde el día de inicio de la simulación.

6. Discusión y conclusiones

Frente a la preocupante progresión de la infección por COVID-19 en poblaciones cerradas, tales como establecimientos de larga estadía para adultos mayores (ELEAM), este trabajo presenta un modelo de testeo grupal de dos etapas que incorpora de manera explícita la correlación en la probabilidad de estar infectado, y que puede ser usado para minimizar el número esperado de tests. En caso de tests con sensibilidades menores a uno se presenta una fórmula explícita para evaluar el riesgo en los falsos negativos, que es creciente con el tamaño del agrupamiento. Lo anterior hace evidente el trade-off que puede existir entre minimizar el número esperado de tests versus el número de falsos negativos.

El contexto de aplicación de este modelo son las poblaciones (parcialmente) cerradas, que en nuestro caso de estudio representa la población de un ELEAM. Para estimar la prevalencia y la correlación presente en un ELEAM se construyó un modelo de simulación que permite seguir la evolución de los pacientes contagiados, utilizando parámetros de la literatura y de las políticas

implementadas en la realidad. Ajustando una distribución Beta-Binomial, se estimó la prevalencia y correlación para cada día y se obtuvo como resultado el tamaño del agrupamiento óptimo en el contexto de testeo grupal de dos etapas.

El análisis desarrollado caracteriza el ahorro en el número de tests a utilizar al usar la política óptima de testeo grupal versus su versión tradicional que omite correlación. Lo anterior es significativamente más pronunciado al mirar los datos de las simulaciones desde el inicio de ésta. Además, los resultados resaltan la importancia de acudir prontamente a los ELEAM una vez que se han detectado casos sintomáticos, debido al rápido crecimiento de la prevalencia en los días inmediatamente posteriores como se ilustra en la Figura 6.

En este mismo sentido, los resultados oportunos son los que realmente permiten gestionar los aislamientos preventivos, puesto que de nada sirve testear a una población si no se es capaz de minimizar los riesgos de contagio mientras se está a la espera de los resultados.

Estos resultados resaltan la importancia de disponer de un mecanismo que permita anticiparse a un elevado número de contagios, por ejemplo, realizando con frecuencia tests a todos los miembros de la población o recomendando testeos grupales periódicos en grupos cerrados, cada dos o tres semanas en nuestro caso de estudio. De esta manera, sería posible identificar a tiempo cualquier potencial brote utilizando un número muy limitado de tests, puesto que esto "reiniciaría" la dinámica de la evolución de la infección (tal como se ilustra en la Figura 5).

En términos de direcciones futuras de investigación, el siguiente paso natural consiste en la validación de la dinámica de los contagios en el modelo de simulación presentado en este trabajo. Una vez validado, el modelo puede servir de base para la evaluación de estrategias de testeo y de cuarentenas preventivas, que podrían incluir el testeo agrupado de toda la población bajo estudio. Por otro lado, el modelo toma un número de supuestos respecto a la evolución temporal de la infección, y a la dinámica de contagio, basado en evidencia parcial, recolectada hasta el momento sobre la pandemia. En la medida que el conocimiento acerca del virus se perfeccionen, nuevos y mejores modelos podrán ser planteados y utilizados.

Agradecimientos: Esta investigación ha sido financiada en parte por el Instituto Sistemas Complejos de Ingeniería ISCI (ICM-FIC: P05-004-F, CO-NICYT: FB0816). Los autores agradecen el apoyo de ANID PIA/BASAL AFB180003.

Referencias

- [1] D. Balding, W. Bruno, D. Torney, y E. Knill. A comparative survey of non-adaptive pooling designs. En *Genetic mapping and DNA sequencing*, páginas 133–154. Springer, 1996.
- [2] CDC. Contract tracing plan, covid-19. Accessed July 22, 2020, 2020.
- [3] A. Cherif, N. Grobe, X. Wang, y P. Kotanko. Simulation of pool testing to identify patients with coronavirus disease 2019 under conditions of limited test availability. *JAMA Network Open*, 6(3), 2020.
- [4] Diario Oficial. Resolución 424 del 9 de junio de 2020, ministerio de salud, gobierno de chile, 2020.
- [5] R. Dorfman. The detection of defective numbers of large populations. *Annals of Mathematical Statistics*, 1(14):436–440, 1943.
- [6] M. Farfan, J. Torres, M. O'Ryan, M. Olivares, P. Gallardo, y S. C. Optimizing rt-pcr detection of sars-cov-2 for developing countries using pool testing. medRxiv 2020.04.15.20067199, 2020.
- [7] A. Gill y D. Gottlieb. The identification of a set by successive intersections. *Information and Control*, 24(1):20–35, 1974.
- [8] X. He, E. H. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y. C. Lau, J. Y. Wong, Y. Guan, X. Tan, et al. Temporal dynamics in viral shedding and transmissibility of covid-19. *Nature medicine*, 26(5):672–675, 2020.
- [9] C. Mentus, M. Romeo, y C. DiPaola. Analysis and applications of adaptive group testing methods for covid-19. *medRxiv*, 2020.
- [10] Ministerio de Salud. Protocolo de coordinación para acciones de vigilancia epidemiológico durante la pandemia covid-10 en chile: Estrategia nacional de testeo, trazabilidad y aislamiento, 2020.
- [11] N. Sethuraman, S. S. Jeremiah, y A. Ryo. Interpreting Diagnostic Tests for SARS-CoV-2. *JAMA*, 323(22):2249–2251, 2020.
- [12] Q. Stephen A. Lauer, Kyra H. Grantz, Bi, et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Annals of Internal Medicine*, 172(9):577–582, 2020.
- [13] A. Sterrett. On the detection of defective members of large populations. *The Annals of Mathematical Statistics*, 28(4):1033–1036, 1957.

- [14] L. M. Wein y S. A. Zenios. Pooled testing for hiv screening: capturing the dilution effect. *Operations Research*, 44(4):543–569, 1996.
- [15] S. Woloshin, N. Patel, y A. Kesselheim. False negative tests for sars-cov-2 infection - challenges and implications. New England Journal of Medicine, 38(383), 2020.
- [16] WSJ. Wuhan tests nine million people for coronavirus in 10 days. Accessed September 11, 2020, 2020.
- [17] I. Yelin, N. Aharony, E. Shaer-Tamar, A. Argoetti, E. Messer, D. Berenbaum, E. Shafran, A. Kuzli, N. Gandali, T. Hashimshony, et al. Evaluation of covid-19 rt-qpcr test in multi-sample pools. *MedRxiv*, 2020.

Iniciativas matemático computacionales desde la Universidad de Buenos Aires para contribuir a la toma de decisiones en el contexto del COVID-19

Mehrnoosh Arrar *
Rodrigo Castro †
Luciana Ferrer †
Juan Kamienkowski †
Rodrigo Maidana ††
E. Pecker-Marcosig †
Marina Valdora *

Laouen Belloli †
Guillermo Durán * †
Diego Garbervetsky †
Pablo Laciana †
Mauricio Mendiluce |
Ezequiel Puerta †
Mariano Zapatero †

Ana María Bianco *
Roberto Etcheñique §
Rodrigo Goldsmit |
Esteban Lanzarotti †
Sol Minoldo ‡‡
Rodrigo Quiroga §§

Leonardo Boechi *
Natalia Fernández ¶
Carolina Grillo |
Mario Lozano **
Leonardo Pepino †
Guillermo Solovey *

Resumen

Con la llegada de la pandemia a la Argentina, en marzo de 2020, se creó un grupo multidisciplinario con base en la Universidad de Buenos Aires y amplia trayectoria y experiencia en el desarrollo e investigación de herramientas matemático-computacionales, para colaborar en la toma de decisiones en el contexto del COVID-19. Análisis de datos en el país y en el mundo, simulación de escenarios, y proyectos en territorio fueron parte del desafío encarado. En este artículo se reseñan algunas de las actividades realizadas por el grupo y se analiza el impacto de ellas.

Palabras Clave: Pandemia, COVID-19, Políticas Públicas, Análisis de datos, Simulación.

^{*}CONICET - Universidad de Buenos Aires. Instituto de Cálculo (IC), Argentina.

[†]CONICET - Universidad de Buenos Aires. Instituto de Ciencias de la Computación (ICC), Argentina.

[‡]Departamento Ingeniería Industrial, FCFM, Universidad de Chile e Instituto Sistemas Complejos de Ingeniería (ISCI), Santiago, Chile.

[§]CONICET - Universidad de Buenos Aires. DQIAyQF - Instituto de Química-Física de los Materiales, Medio Ambiente y Energía (INQUIMAE), Argentina.

[¶]CONICET - Universidad de Buenos Aires. Instituto de Biociencias, Biotecnología y Biología Traslacional (iB3), Argentina.

FCEyN - Universidad de Buenos Aires, Argentina.

^{**}Departamento de Ciencia y Tecnología, Universidad Nacional de Quilmes, Argentina.

^{††}FCE, Universidad Nacional de La Plata, Argentina.

^{‡‡}CONICET - Universidad de Córdoba. Centro de Investigaciones y Estudios sobre Cultura y Sociedad (CIECS), Argentina.

^{§§}CONICET - Universidad Nacional de Córdoba, Instituto De Investigaciones en Físico-Química de Córdoba (INFIQC), Argentina.

1. Introducción

La pandemia producida por el COVID-19, causada por la infección del virus SARS-CoV-2, ha provocado durante 2020 más de un millón de muertes en todo el mundo [26]. Desde principios de ese año, miles de académicos en todo el mundo se pusieron a trabajar en temas vinculados a la pandemia siendo, probablemente, los más citado el grupo de Neil Ferguson perteneciente al Imperial College de Londres, que publicó su primer reporte sobre el tema en enero de 2020 [9] y el ingeniero franco-español Tomás Pueyo, con su renombrada anécdota del martillo y la danza como herramientas claves para enfrentar al Covid-19 [22].

Debido a la gran dificultad que han enfrentado los organismos de salud para encontrar tratamientos y/o vacunas efectivas en el corto plazo, y así contener el avance de la pandemia, han surgido nuevos enfoques que buscan utilizar la información, existente previamente o generada durante la pandemia, para ayudar a la toma de decisiones de quienes lideran las instituciones. Se trata de enfoques multidisciplinarios que consideran aspectos logísticos, económicos, políticos y sociales.

Desde el Instituto de Cálculo (IC) y el Instituto de Ciencias de la Computación (ICC) de la Facultad de Ciencias Exactas y Naturales de la Universidad de Buenos Aires (FCEyN-UBA) y del CONICET se creó un grupo multidisciplinario, con amplia trayectoria y experiencia en el desarrollo e investigación de herramientas matemático-computacionales, asociado con científicos y científicas de otras instituciones académicas nacionales, que buscó colaborar con diferentes jurisdicciones, la gran mayoría de ellas ubicadas en la Provincia de Buenos Aires (PBA), en la toma de decisiones para luchar contra la pandemia.

En este artículo se resumen las principales acciones llevadas a cabo por este núcleo científico. Estas acciones corresponden a un conjunto amplio de contribuciones, de naturaleza variada, que van desde análisis de datos y asesoramiento a diferentes distritos, hasta actividades directas en los territorios. Estos desarrollos permiten o facilitan la adquisición de datos, la organización de recursos disponibles, cuantificar y predecir el avance de la pandemia, el análisis cuantitativo de escenarios para predecir efectos de determinadas políticas públicas, entre otras actividades de suma utilidad. Varios de los trabajos responden a necesidades y pedidos específicos realizados por los principales actores de esta pandemia: directores de hospitales, autoridades de los dife-

rentes Ministerios de Salud, intendentes, gobernadores, entre otros, y fueron diseñados y desarrollados conjuntamente con ellos. Otros aportes surgieron de ideas de los propios investigadores, quienes se pusieron a disposición de los tomadores de decisiones o de la población.

El presente trabajo presenta las diferentes actividades científico/tecnológicas realizadas, divididas en dos secciones. La primera de ellas está orientada al análisis de datos, y la segunda, a los trabajos en territorio, de acuerdo con la siguiente estructura:

Análisis de Datos:

- Evaluación de diferentes modelos para la estimación de métricas ampliamente utilizadas por parte de las autoridades, tales como el número reproductivo básico (R_0) y efectivo $(R_e f)$ (Sección 2.1), el tiempo de duplicación (TD) de cantidad de casos (Sección 2.2), entre otros.
- Análisis de datos de pacientes recolectados por el Ministerio de Salud de la Provincia de Buenos Aires para el desarrollo de modelos para Prognosis Automática (Sección 2.3.1), y poder estimar probabilidad de que un individuo infectado necesite internación en Unidades de Cuidados Intensivos (UCI) (Sección 2.3.2).
- Desarrollo de un portal donde se presentan varios indicadores de la evolución de la pandemia, a diferentes niveles de resolución: nivel nacional, provincial, municipal y otras agrupaciones geográficas o administrativas (Sección 2.4).
- Estimaciones de prevalencia a nivel nacional, provincial, municipal en base a estimaciones del Infection Fatality Rate (IFR) e información de fallecimientos, por rango etario y sexo (Sección 2.5).
- Análisis de diferentes tipos de inmunidad y sus consecuencias, considerando distintas métricas (Sección 2.6).

Trabajos en territorio:

- Desarrollo de una aplicación que permite organizar las camas en una red de hospitales, y su aplicación en la región sudeste del Gran Buenos Aires (Sección 3.1).
- La puesta en marcha de un Centro de Prevención Covid en la Facultad de Ciencias Exactas y Naturales de la Universidad de Buenos Aires (Sección 3.2).
- Simulaciones basadas en agentes en el distrito de Marcos Paz, en la Provincia de Buenos Aires (Sección 3.3).

En todos los proyectos se consideraron variables fundamentales, tales como la privacidad de los datos, el uso responsable de los mismos, así como la urgencia de las soluciones y el acotado presupuesto disponible para la desafiante y necesaria tarea de entregar apoyo a la lucha contra la pandemia en la Argentina.

2. Análisis de Datos

2.1. Estimación del número reproductivo R de la pandemia a lo largo del tiempo

El manejo de una epidemia por parte de autoridades públicas se nutre necesariamente de múltiples métricas e indicadores que dan cuenta de distintos aspectos sociales y biológicos que definen la dinámica de la enfermedad en una sociedad.

Una métrica vastamente aceptada para caracterizar la incidencia es el número reproductivo, también conocida como R, que se define como el número promedio de encuentros con transmisiones exitosas (contagios) generados por cada individuo infectivo.

Al inicio de una epidemia toda la población puede ser considerada susceptible al contagio y la dinámica de encuentros sociales no se ve afectada aún por medidas sanitarias. En esta situación, se asume R en su valor conocido como número reproductivo básico y se denota como R_0 . Luego de esa situación, cuando puede haber personas inmunes al contagio o la población adopta medidas especiales de higiene y distanciamiento, R se denomina número reproductivo efectivo, o R_{ef} , verificando usualmente $R_0 \geq R_{ef} \geq 0$.

Luego, R_{ef} sirve como línea guía para sintetizar el estado de la epidemia en cuanto a la dinámica de cambio del número de infectados activos: valores $R_{ef} > 1$ representan una fase de crecimiento exponencial (o quizá sub-exponencial) mientras que valores $R_{ef} < 1$ dan cuenta de una fase de decrecimiento (o supresión) en la cual la epidemia tiende a su extinción.

Con ello, se propone el desarrollo de una metodología que permita estimar el valor de R_{ef} para diversas zonas relevantes del país.

Metodología para estimar el valor de R_{ef}

Existen diversas técnicas para determinar el valor de R_{ef} a partir de la evidencia que se acumula a medida avanza el desarrollo de una epidemia. Una familia de métodos, se basa en procedimientos estadísticos, implementados

en bibliotecas de software como Epicontacts [20] o EpiEstim [1], requiriendo usualmente información sobre cadenas de causa-efecto con detalle a nivel de individuos. Estos datos son, en general, difíciles de obtener, por lo cual se aplica una serie de hipótesis sobre distribuciones estadísticas para completar la información faltante.

Otra familia de métodos, parte de modelos que definen explícitamente la mecánica de contagio en una población, como los modelos compartimentales [24]. Allí la población se subdivide en categorías (compartimentos) como, por ejemplo, en Susceptibles (S), Infectados (I) y Recuperados (R) (en su versión más compacta, conocida como modelo SIR), mientras que un sistema de ecuaciones diferenciales (tiempo continuo) o en diferencia (tiempo discreto) definen la dinámica de la evolución de un compartimiento a otro. En este marco, $R_{ef}(t)$ es una variable que depende de parámetros del sistema, como por ejemplo en el siguiente sistema de ecuaciones diferenciales:

$$dS/dt = -\beta(t) \times I(t) \times S(t)$$
 (tasa promedio de infección de susceptibles)

 $dI/dt = \beta(t) \times I(t) \times S(t) - \gamma \times I(t)$ (tasa promedio de cambio de contagiados infectivos)

$$dR/dt = \gamma \times I(t)$$
 (tasa promedio de recuperación y/o muerte)

donde

$$N=S(t)+I(t)+R(t)$$
 (población total constante)
 $R_{ef}(t)=\beta(t)/\gamma\times S(t)$ (número reproductivo efectivo, adimensional)

 $\beta(t)$ corresponde a la tasa promedio de infección per cápita, que es dependiente del comportamiento social, mientras que γ es la tasa promedio de remoción del sistema (recuperación o muerte). Dado que S(t) es la cantidad de individuos susceptibles y que la misma decrece monótonamente a medida que avanza la epidemia, este valor es uno de los factores determinantes para que, al cabo de cierto tiempo, $R_{ef}(t)$ cruce (hacia abajo) el umbral de valor 1, indicando el comienzo de la extinción del proceso.

Este tipo de modelos permite comprender mejor la evolución del sistema y realizar proyecciones dentro de ventanas temporales en el futuro, asumiendo escenarios hipotéticos de cambio típicamente mediante intervenciones no farmacológicas (políticas públicas) reflejadas como cambios en $\beta(t)$.

Es importante notar que, todo modelo matemático, es una simplificación de la realidad cuya utilidad y validez están limitadas a cierto tipo de preguntas. Por ejemplo, este modelo asume mezcla homogénea — equiprobabilidad de

encuentros entre cualquier par de individuos — requiriendo aplicar criterios al elegir la escala espacio-temporal de validez del modelo.

Asimismo, se aplicó una extensión a este modelo, conocida como SEIR, para la estimación de $R_{ef}(t)$ de acuerdo a etapas de la evolución de la epidemia y los cambios del comportamiento social. Un modelo SEIR agrega el compartimiento de Expuestos (E) para representar a la población infectada que aún no contagia, lo cual es relevante en COVID-19. Esto se refleja reemplazando dI/dt y agregando dE/dt en el sistema de ecuaciones de SIR del siguiente modo:

 $dE/dt = \beta(t) \times I(t) \times S(t) - \alpha \times E(t)$ (tasa promedio de cambio de expuestos)

 $dI/dt = \alpha \times E(t) - \gamma \times I(t)$ (tasa promedio de cambio de contagiados infectivos)

Donde α representa la inversa del tiempo promedio de incubación de la enfermedad.

Resultados

Asumiendo γ conocido (en su valor promedio) y disponiendo de los datos para los infectados reportados $I_r(t)$, se realizó un ajuste de parámetros constante a tramos, de modo tal que, se minimice la norma del error cuadrático medio entre I(t) (modelo) e $I_r(t)$ (datos). Este procedimiento implica identificar la mínima cantidad de instantes t_k , con k=0,...,M en los cuales es necesario subdividir los tramos. Otra opción, es elegir a priori un período fijo para los tramos. La primera estrategia apunta a identificar cuándo hubo cambios significativos en la dinámica social mientras que, la segunda, es más útil para comparar tramos idénticos entre diferentes jurisdicciones.

La Figura 1 muestra un ejemplo de ajuste para Argentina eligiendo tramos de 14 días (un lapso suficiente para capturar efectos de cambio sobre los datos de cantidad de nuevos infectados, recuperados y muertes). Utilizando escala semilogarítmica y datos oficiales de casos reportados por fecha de inicio de síntomas, se observa el ajuste del modelo (línea llena) con los datos (puntos). Se muestran los nuevos casos confirmados diarios y los casos confirmados acumulados, junto a los sucesivos valores ajustados para $R_{ef}(t_k)$.

Aplicando el mismo criterio para varias jurisdicciones, es posible comparar sus evoluciones temporales, como se muestra en la Figura 2.

La aplicación de los métodos de tramos variables y tramos fijos mostró resultados consistentes. A su vez, la presentación visual online, actualizada

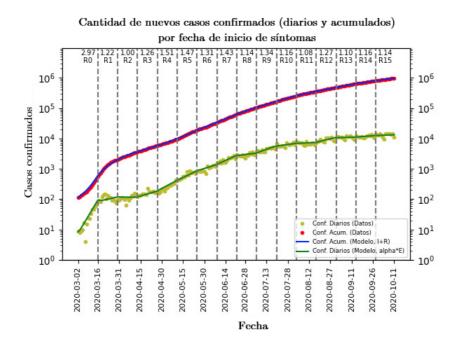


Figura 1: Evolución de $R_{ef}(t)$ para Argentina ajustando un modelo tipo SEIR.

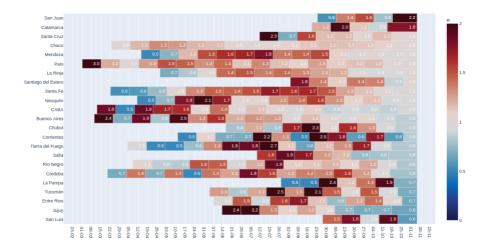


Figura 2: Mapa de calor temporal mostrando evoluciones del número reproductivo efectivo $R_{ef}(t)$ para las provincias argentinas (foto al 15/11/2020). Nota: Esta visualización es parte de los análisis automatizados disponibles en el sitio web descrito en la Sección 2.4.

diariamente, mostró ser informativa para hacer análisis integrales de la progresión temporal de las etapas de la epidemia en distintas jurisdicciones del país. Se destaca la flexibilidad para definir de manera sencilla nuevas regiones geográficas de interés para ser monitoreadas mediante el número efectivo de reproducción, lo cual permitió reaccionar rápidamente a demandas cambiantes de información.

2.2. Tiempos de duplicación

Un parámetro muy utilizado por gobiernos de todo el mundo para conocer la tasa a la que crecen los contagios es el tiempo de duplicación de casos (TD). En Argentina, el gobierno nacional ha definido fases de la pandemia también de acuerdo al TD [4]. Según este criterio, si en un distrito el tiempo de duplicación supera los 25 días, ese distrito podría cambiar de fase en el proceso de aislamiento y, consecuentemente, reabrir actividades que habían sido prohibidas. Sin embargo, es llamativamente poco frecuente que se reporte la incerteza asociada al TD inferido. Esto es importante dado que, por ejemplo, un tiempo de duplicación de 15,2 \pm 0,1 días debería interpretarse diferente a uno de 15,2 \pm 2 días. Por lo tanto, para un diagnóstico confiable del avance de la pandemia es necesario conocer también la precisión de la estimación.

Luego se propone estimar el tiempo de duplicación de casos confirmados y el intervalo de confianza de esa estimación utilizando dos variantes del modelo de crecimiento exponencial.

Metodología para estimar el tiempo de duplicación de casos

Sea y_n el número de casos confirmados hasta el día n. Para estimar el tiempo de duplicación en el día n, se toman datos de los últimos 7 días, $y_{(n-6)}, \ldots, y_n$, y se ajustan los siguientes modelos:

Modelo 1:
$$log(y_k) = \alpha + \beta_k + \epsilon_k$$
, con $k = n - 6, \dots, n$.

Modelo 2:
$$y_k = \gamma \exp^{(\beta_k)} + \epsilon_k$$
, con $k = n - 6, \dots, n$.

El tiempo de duplicación esperado es $log(2)/\beta$. Por el método mínimos cuadrados calculamos un estimador de β , lo llamamos β^* , y deducimos que un estimador para el tiempo de duplicación es $log(2)/\beta^*$.

En el Modelo 1, la distribución de β^* es conocida, bajo el supuesto de que los errores tienen distribución normal. En el Modelo 2, sin embargo, sólo es posible hallar la distribución asintótica o aproximada de β^* y, dado que se espera generar una estimación con sólo 7 observaciones, esta aproximación puede no ser muy buena. Sin embargo, se calcularon los intervalos de confianza bajo

ambos modelos y se realizó una simulación de Monte Carlo para estudiar su nivel de cobertura empírico. Se ejecutaron experimentos con errores normales y con errores con colas pesadas (distribución t_3).

Resultados

Los resultados del estudio de Monte Carlo se encuentran resumidos en la Tabla 1. En base a estos resultados, se decidió utilizar sólo los intervalos calculados con el Modelo 1. Un intervalo de confianza de nivel $1-\alpha$ para el tiempo de duplicación bajo este modelo es [$log(2)/(\beta^*+t_{5,\alpha/2}~SE(\beta^*))$, $log(2)/(\beta^*-t_{5,\alpha/2}~SE(\beta^*))$].

Se observa que, algunos días, se retrasa la carga de datos y se compensa en los días siguientes, ocasionando una variabilidad en los datos que no refleja la realidad de la epidemia. Es por eso que se decidió reemplazar las variables aleatorias y_n por variables $z_n = (y_n + y_{n-1} + y_{n-2})/3$.

Las simulaciones preliminares mostraron que los intervalos de confianza basados en 7 días dan buenos resultados de cobertura para el Modelo 1, como se observa en la Tabla 1, para los dos tipos de errores considerados. Sin embargo, para el Modelo 2, la Tabla 1 muestra que el nivel de cobertura empírico está levemente por debajo del valor nominal. Por esta razón, el análisis posterior con datos realizó con el Modelo 1.

Tabla 1: Errores de estimación obtenidos por simulaciones de Monte Carlo para cada uno de los 2 modelos

	Modelo 1	Modelo 2
Errores normales	0.9499	0.8916
Errores t_3	0.9497	0.8890

Este método se utilizó para estimar el tiempo de duplicación de distintos distritos de la Argentina (ver Figura 3) y se creó una aplicación para visualizar los resultados públicamente [2]:

Finalmente, la estimación del tiempo de duplicación resultó similar para los dos métodos de ajuste. Es importante resaltar que el tiempo de duplicación es dependiente del modelo y debe ser interpretado como una medida de la "velocidad instantánea" de crecimiento de los contagios, no como una predicción.

Duplicación de casos de COVID-19 en Argentina

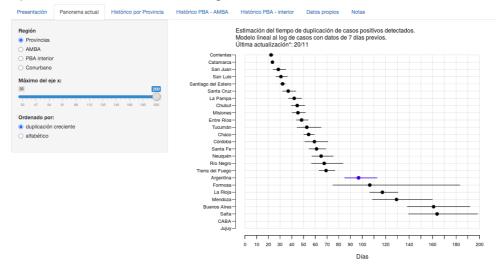


Figura 3: Captura de pantalla de la aplicación.

Nota: Disponible en https://institutodecalculo.shinyapps.io/tdcovid/

2.3. Análisis y predicción de la evolución de los casos

Los sistemas de pronóstico (o modelos de prognosis) pueden asistir al personal de salud en la toma de decisiones, identificando casos sobre los cuales se requiere una atención inmediata o un seguimiento frecuente. A su vez, la implementación de dichos modelos permite explorar cuáles son los atributos más importantes en la decisión automática y, de esta manera, descubrir potenciales nuevos marcadores tempranos de la enfermedad.

En la gran mayoría de los seres humanos, la infección por el nuevo coronavirus SARS-CoV-2, desarrolla formas leves o moderadas de la enfermedad por COVID-19. Sin embargo, algunos infectados pueden presentar síntomas severos que requieren internación en Unidad de Cuidados Intensivos (UCI). En este contexto, es necesario identificar quiénes son aquellos pacientes que podrían requerir una cama de UCI y, a tal efecto, es de interés estudiar qué comorbilidades tienen mayor asociación con el hecho de necesitar internación en UCI.

Utilizando datos provistos por el Ministerio de Salud de la Provincia de Buenos Aires, en el marco de un convenio de cooperación con la Facultad de Ciencias Exactas y Naturales de la UBA, se realizaron dos trabajos. El primero, tiene por objetivo predecir la probabilidad de que un individuo desarrolle un caso grave de la enfermedad, usando como covariables la edad, el sexo y las comorbilidades (Sección 2.3.1). El segundo, tiene como objetivo predecir si un

individuo deberá ser internado en UCI o no en el transcurso de la enfermedad (Sección 2.3.2). Si bien, la mayoría de los casos graves son internados en UCI, no todos lo son, lo cual se evidencia por el hecho de que hay un porcentaje de individuos fallecidos que no pasaron por una UCI, lo que podría explicar las pequeñas diferencias en las conclusiones en ambas secciones respecto a los factores de riesgo.

2.3.1. Prognosis automática

Los modelos de prognosis, o sistemas de pronóstico de la evolución de una enfermedad, pueden asistir al personal de salud en la toma de decisiones identificando casos sobre los cuáles se requiere una atención inmediata o un seguimiento frecuente. A su vez, la implementación de dichos modelos, permiten explorar cuáles son los atributos más importantes en la decisión automática y, de esta manera, descubrir potenciales nuevos marcadores tempranos de la enfermedad.

Dada la diversidad de datos y experiencias los modelos de prognosis existentes, en el marco del Covid-19, difieren tanto en la variable a predecir como en los atributos utilizados. Respecto a las variables a predecir, la mayoría de los modelos buscan predecir la mortalidad, sin embargo, también se han desarrollado modelos para predecir la gravedad, determinada, por ejemplo, como estadías mayores a 10 días o la necesidad de UCI [6, 27]. Esto es particularmente importante en un contexto de escasez de recursos.

En [6], se presenta una revisión sistemática de modelos comparando 22 de ellos de acuerdo con su capacidad para predecir deterioro, a lo largo de la estadía en el hospital, y mortalidad. Esta revisión utiliza una muestra para validación de 411 casos y concluye que ningún modelo supera el análisis univariado de los atributos más explicativo, que corresponden a la saturación de oxígeno al entrar al hospital y la edad para deterioro y mortalidad. Yan, Zhang, Goncalves et al. [29] utilizan un modelo basado en árboles de decisión, aplicados sobre una muestra de alrededor de 500 casos, donde muestra que, utilizando estudios de sangre, se obtienen valores de F1 cercanos a 1 en la predicción de mortalidad.

Los estudios reportados de modelos automáticos de prognosis utilizan, generalmente, muestras pequeñas para entrenar y validar, entre 100 y 500 casos. En [25] se explora la asociación de atributos de los pacientes a la mortalidad en una muestra mucho más grande ($N\sim 107$ personas). Este estudio muestra que la edad es la variable que mejor explica el riesgo, aumentando con la edad. Dentro de las comorbilidades, las de mayor impacto son diabetes, cáncer, trasplantes de órganos, deficiencias en el funcionamientos de riñón o hígado,

enfermedades neurológicas, accidentes cerebrovasculares previos, o déficit en el sistema inmune. Finalmente, otras variables demográficas o socioeconómicas como el sexo, la etnia o el contexto socioeconómico tienen una incidencia moderada en el riesgo. Los hallazgos en este estudio son consistentes con los estudios de pronóstico automático en poblaciones más acotadas [6, 27].

Metodología modelo automático de pronóstico de la gravedad de un cuadro causado por COVID-19

Con el fin de desarrollar un modelo modelo automático de pronóstico de la gravedad del cuadro que tendrá una cierta persona al contraer Covid, se desarrolló un modelo de clasificación, entrenado sobre una muestra de ~ 105 casos, en base a atributos como su edad, sexo, comorbilidades y síntomas.

Para su costrucción, se utilizaron datos provistos por el Ministerio de Salud de la Provincia de Buenos Aires al día 28 de agosto de 2020. En particular:

- Edad y sexo de la persona;
- Síntomas (19 valores): Anosmia de reciente aparición, disgeusia de reciente aparición, disnea, odinofagia, fiebre, tos, artralgia, cefalea, confusión mental, convulsiones, diarrea, dolor abdominal, dolor torácico, inyección conjuntival, irritabilidad, malestar general, mialgias, rechazo del alimento y vómitos;
- Comorbilidades (22 valores): Diabetes, embarazo, enfermedad hepática crónica, enfermedad neurológica previa, enfermedad oncológica previa, asma, enfermedad pulmonar obstructiva crónica (EPOC), bajo peso al nacer, bronquiolitis previa, diálisis aguda, diálisis crónica, ex fumador, fumador, hipertensión arterial, inmunosupresión congénita o adquirida, insuficiencia cardíaca, insuficiencia renal, neumonía adquirida en la comunidad (NAC) previa, obesidad, prematuro, tuberculosis y una característica que indica la falta de comorbilidades;
- Datos de internación y fallecimiento: Un indicador de si el paciente fue internado y en caso que sí, si fue en terapia intensiva, y si el paciente falleció.

Todos los atributos, salvo la edad, que está medida en años, y la fiebre que toma tres valores (sin fiebre, fiebre menor a $38^{\circ}C$ y fiebre mayor a $38^{\circ}C$), toman valores binarios, donde 1 codifica si el síntoma o la comorbilidad están presentes en el paciente. Se consideraron únicamente los casos confirmados como Covid positivos.

Para el modelado se utilizó Random Forest, un método estándar de aprendizaje automático, elegido por su robustez, flexibilidad e interpretabilidad. El modelo se entrenó con los atributos, listados anteriormente, como entrada para predecir la gravedad de cada caso. La gravedad se definió como leve, si el paciente se recupera sin o con internación pero sin cuidados intensivos, y como grave, si requiere cuidados intensivos y/o fallece. Es importante notar que, esta definición, es una aproximación a la gravedad del caso pero no es perfecta, dado que es posible que haya casos que fueron internados en UCI sólo por prevención sin que el cuadro fuera tan grave (por ejemplo, en el caso de embarazadas o niñas y niños pequeños). Durante el entrenamiento, el costo correspondiente a cada muestra fue pesado por la inversa de la frecuencia correspondiente a su clase, para compensar el desbalance entre las dos clases.

Para el desarrollo de los modelos se utilizaron los casos con fecha de inicio de síntomas (FIS) anterior al día 31 de Julio de 2020, mientras que, para la evaluación final, se utilizaron los datos con FIS del 1 al 15 de Agosto de 2020, descartando los casos más recientes, los cuales no se sabe si se convertirán en casos graves. Sólo fueron incluidos los casos con al menos un síntomas. Los datos utilizados incluyeron 108.249 casos, de los cuales el $4.7\,\%$ fueron casos graves. La selección de parámetros se realizó con el método de validación cruzada en los datos de desarrollo. La evaluación se realizó entrenando un nuevo modelo en los datos de desarrollo y calculando el rendimiento en los datos de evaluación, los cuales incluyeron 47.014 casos, de los cuales 2.1% fueron graves. El porcentaje de casos graves es distinto en los datos de desarrollo y los de evaluación debido a que los datos fueron divididos temporalmente y, el porcentaje de casos graves, fue disminuyendo con el tiempo. A su vez, para entender qué atributos fueron más relevantes para la tarea, se utilizó un algoritmo de selección de atributos secuencial [19]. Este algoritmo, comienza ajustando el modelo considerando todas las atributos de interés, descartando en cada etapa, uno por vez, el que menos degrada o produce mayor mejora el rendimiento del sistema en los datos de desarrollo. El algoritmo genera así una lista ordenada de atributos según su importancia para la clasificación en los datos de desarrollo.

Resultados

El análisis resultante del algoritmo de selección de atributos sugiere que las comorbilidades más importantes para la clasificación son la diabetes, obesidad, insuficiencia renal y enfermedad oncológica previa. Las cuales, también, fueron encontradas relevantes para la prognosis en la literatura. Al agregar los síntomas al modelo, se obtiene una ganancia significativa. En este caso, el al-

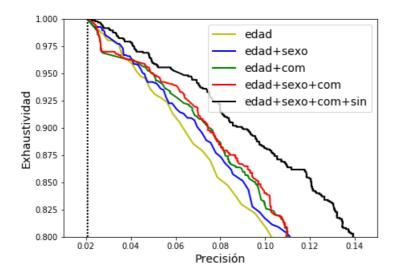


Figura 4: Curvas de exhaustividad y precisión en los datos de evaluación usando distintas listas de atributos de entrada.

Nota: La línea punteada vertical corresponde al porcentaje de casos graves. Las referencias 'com' y 'sin' indican comorbilidades y síntomas, respectivamente.

goritmo de selección indica que disnea, odinofagia y anosmia son los síntomas más relevantes para la clasificación. En el caso de la odinofagia y la anosmia, la existencia del síntoma ocurre más frecuentemente en los casos leves que en los graves lo cual ya había sido observado por [13].

La Figura 4 muestra las curvas de exhaustividad (porcentaje de los casos graves que son detectados como tales), versus precisión (porcentaje de los casos detectados como graves que efectivamente fueron graves), para modelos entrenados con distintas listas de atributos de entrada al variar el umbral de decisión sobre la salida del sistema. A su vez, muestra que la edad es un excelente predictor de la gravedad con que se desarrollará un caso y que el sexo agrega información relevante para predecir la gravedad del caso.

Los resultados coinciden con lo encontrado en la literatura [25]. Adicionalmente, se identifica que, agregar las comorbilidades a la edad, mejora más el rendimiento que agregar el sexo, lo cual sugiere que la información que extrae el sistema del atributo sexo está al menos parcialmente correlacionada con las comorbilidades.

También se identifica que, otros síntomas y comorbilidades, además de los seleccionados como importantes por el algoritmo, correlacionan fuertemente con la gravedad del caso, como la diálisis o insuficiencia cardíaca al analizarlas

independientemente de las demás. Sin embargo, es posible que estos atributos no sean elegidos por estar muy correlacionados con otros que sí selecciona el modelo, como la edad. En la Sección 2.3.2 se describe un trabajo enfocado más específicamente en el análisis de los factores de riesgo.

Finalmente, el sistema de prognosis automático descrito, podría ser usado para hacer triage en caso de saturación del sistema de salud y que requiera posponer la atención de algunos pacientes en beneficio de otros con peor prognosis. Es importante aclarar que, dadas ciertas comorbilidades que afectan significativamente la prognosis suceden muy infrecuentemente en los datos, el modelo probablemente no está aprendiendo a predecir correctamente la gravedad para esos casos. Por esta razón, la salida del sistema, sólo debería ser tomada como una recomendación que debe ser confirmada por un profesional de la salud.

2.3.2. Comorbilidades y riesgo de internación en UCI

Durante la pandemia por COVID-19, se produzco un colapso generalizado de los centros hospitalarios frente a una demanda, que a requerido internación de pacientes en Unidades de Cuidados Intensivos (UCI), por sobre la capacidad del sistema. Se consideró de interés el análisis de factores de riesgo asociados al contagio por COVID-19 y la estimación de probabilidad que un individuo infectado necesite internación en UCI, en base a variables como la edad, el sexo y las 22 comorbilidades, detalladas en la Sección 2.3.1.

El presente estudio, se basa en los datos de casos confirmados con COVID-19, recolectados por el Ministerio de Salud de la Provincia de Buenos Aires, reportados el 7 de julio de 2020. En particular, los datos que se analizan, corresponden a los casos confirmados de individuos con edad menor a 90 años en la Provincia de Buenos Aires hasta el 22 de junio de 2020, lo que comprende un total de 22306 casos. Se descartaron en este análisis los casos confirmados entre el 23 de junio y el 7 de julio debido a que, a lo largo del desarrollo de la pandemia en Argentina, se ha observado que desde la confirmación de un caso, es necesario dejar pasar un lapso de tiempo para que los casos evolucionen. De acuerdo a los datos, más del 99 % de los individuos que fueron internados en UCI, la internación ocurrió en 15 días o menos desde la confirmación del diagnóstico¹. La proporción de pacientes en esta muestra que requirieron cuidados intensivos fue de 0.025 (2.5 %).

¹Un análisis de los tiempos obtenidos a partir de los datos en todo el país puede verse en http://covid.ic.fcen.uba.ar:3838/apps_ic/tiempoeventos/

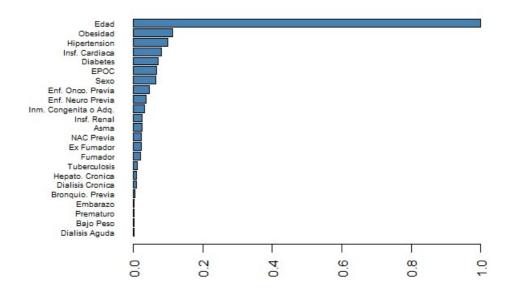


Figura 5: Importancia de los factores de riesgo a partir de 1000 árboles construidos usando remuestreo de los datos originales

Metología para la estimación de probabilidad del riesgo de internación

Con el fin de explorar la asociación entre un paciente de COVID-19 requiera UCI o no y las comorbilidades. Para ello, se utilizaron árboles de clasificación, que permiten realizar análisis descriptivos, con apoyo gráfico, y que utilizan el método de ramificación para representar los posibles resultados de una decisión binaria [13]. Una medida de resumen de la importancia de las variables explicativas en un árbol, puede computarse mediante el decrecimiento del índice de Gini asociado a cada variable. La Figura 5 muestra una representación gráfica de la importancia de las comorbilidades, la edad y el sexo utilizando un árbol de clasificación. Estas medidas se calcularon a partir de 1000 árboles construidos usando remuestreo de los datos originales, y se computaron como el promedio de la importancia de cada variable relativa a la de mayor importancia. Este estudio descriptivo sugiere a edad, obesidad e hipertensión como las variables de mayor importancia.

Adicionalmente, se realizó un ejercicio de inferencia estadística mediante el ajuste de un modelo de regresión logística [7] categorizando los datos en 2 clases. La primera representa a quienes requirieron UCI y, la segunda, a quienes no lo hicieron, empleando las comorbilidades como variables explicativas. A tal fin, si $y \in \{0,1\}$ representa a la respuesta dicotómica y $x\epsilon$ representa al vector

de covariables. El modelo de regresión logística asume que $P(y=1 \mid x) = expit(\beta Tx)$, donde la función expit(t) = 1/(1 + exp(-t)). Su inversa es la función logit(p) = log(p/(1-p)) que involucra a la chance (odds en inglés). En efecto, si p es la probabilidad de un evento, la cantidad p/(1-p) se llama chance del evento en cuestión. Por ejemplo, la probabilidad de obtener 6 en un dado es 1/6, y la chance es 1/5. Esto significa que por cada vez que sale un 6, se espera que aproximadamente habrá 5 veces en que no sale 6. La Figura 6 muestra las proporciones observadas de casos confirmados que necesitaron internación en UCI por edad y el logit de esas mismas proporciones. Se debe tener en cuenta que, la proporción de infectados que requieren cuidados intensivos, disminuye hasta aproximadamente los 30 años de edad, luego aumenta hasta alrededor de los 70 años y, finalmente, parece disminuir para las mujeres, mientras que, para los hombres, parece seguir aumentando, aunque quizás más lentamente.

Estas tendencias se aprecian más claramente en la escala logit, donde las observaciones nos llevan a incluir en el modelo logístico dos splines para explicar el efecto asociado a tener más de 30 años, el de tener más de 70 años y las posibles interacciones entre sexo y edad. La inclusión de una spline para obtener una relación log-lineal entre edad y probabilidad de desarrollar una forma grave de la enfermedad, ya fue considerada por Williamson et al. [25]. Asimismo, como los infectados mayores de 90 años son muy pocos, se restringió el análisis a los casos confirmados de menos de 90 años. Utilizando varias técnicas de selección de variables (best subset selection, forward, backward, stepwise y lasso) y validación cruzada con diversas medidas de evaluación del ajuste (AIC, ECM, AUC, deviance), se eligió el procedimiento para ajustar el modelo e identificar las variables a incluir. En la Tabla 2 se muestran los coeficientes significativos y sus intervalos de confianza con un nivel global de 0,05.

Resultados

El análisis estadístico presentado permite identificar factores de riesgo que resultan importantes para desarrollar formas severas de COVID-19 que requieran internación en UCI. La chance de necesitar internación en UCI, entre los individuos con un factor de riesgo, como la obesidad, es el cociente entre la probabilidad de que un infectado con el factor de riesgo necesite internación en UCI y la probabilidad de que no la necesite. A su vez, el odds ratio (OR) se define como el cociente del odds de necesitar UCI para infectados que tienen el factor de riesgo de interés y el odds de los que no lo tienen, manteniendo todas las demás variables explicativas constantes. Por ejemplo, de acuerdo con la Tabla 2, la chance de que un infectado obeso necesite internación en UCI se

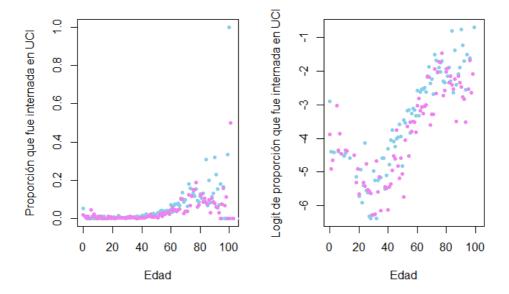


Figura 6: Proporción (panel izquierdo) y proporción en escala logit (panel derecho) de casos confirmados internados en UCI por edad y sexo (en celeste sexo masculino y en rosa femenino)

Tabla 2: Odds ratio estimados y sus intervalos de confianza de nivel 0.95.

Factor de Riesgo	OR Estimado	2.5%	97.5%
Diabetes	1.67	1.12	2.48
Embarazo	7.31	1.70	31.38
Enfermedad Oncológica Previa	2.65	1.47	4.78
EPOC	1.99	1.15	3.45
Hipertensión	1.72	1.23	2.41
Inm. Congénita o Adquirida	3.32	1.43	7.71
Insuficiencia Cardíaca	1.75	1.06	2.91
Obesidad	3.48	2.22	5.44
Más de 30	1.14	1.07	1.21
Más de 70	0.95	0.91	0.99
Edad	0.95	0.90	0.99

estima como aproximadamente el triple de la chance para un no obeso con la misma edad, del mismo sexo y con todos los otros factores de riesgo iguales.

Por su parte, los intervalos de confianza indican cuál es la precisión de las chances que se han estimado. Por ejemplo, la chance de necesitar terapia intensiva de los obesos, se estima que es entre 2.22 y 5.44 veces la de los no obesos, mientras que, la chance de necesitar terapia intensiva de los diabéticos, se estima que es entre 1.12 y 2.48 veces la de los no diabéticos. Con el modelo ajustado es posible estimar la probabilidad de que una persona diagnosticada con COVID 19 necesite internación en UCI, conociendo su edad, sexo y comorbilidades.

De este estudio surgen como factores de riesgo relevantes: diabetes, embarazo, enfermedad oncológica previa, EPOC, hipertensión, inmunosupresión congénita o adquirida, insuficiencia cardíaca, obesidad, edad y sus dos splines asociados que indican tener más de 30 años y más de 70 años. En este conjunto están incluidas la mayoría de las comorbilidades detectadas como importantes para predicción de la gravedad presentada en la Sección 2.3.1.

2.4. Panel de Análisis y Control basado en Datos

La gestión de la pandemia se nutre de múltiples fuentes de información cuantitativa de actualización permanente. Al tratarse de un contexto de emergencia, las estructuras de las fuentes de datos están sujetas a cambios, errores, demoras, entre otros. A su vez, la cantidad de fuentes de información disponible suele crecer en el tiempo, pero la calidad de los datos no suele ser homogénea. Luego, surge la necesidad de contar con un recurso centralizado que aplique métodos sistemáticos, repetibles y comparables de ingesta, depurado, curado, persistencia y visualización de datos.

Cada día los datos, dispuestos por el Ministerio de Salud de la Nación, contienen información nueva y también actualizaciones o correcciones sobre información del pasado, debido a múltiples razones, como por ejemplo, demoras de carga o actualización del estado de salud de pacientes que fueron registrados por primera vez en el sistema un tiempo atrás. Asimismo, la información previa que no tiene actualizaciones es repetida cada día, generando una gran cantidad de redundancia.

Se identifica como necesidad construir un repositorio confiable y unificado de datos que ofrezca información sobre la evolución temporal de métricas relacionadas con el COVID-19 en Argentina. Sobre esta plataforma, adicionalmente, se pueden producir diversos análisis de datos con diferentes objetivos y la posibilidad de generar nuevas métricas más complejas.

Metodología para diseño de un repositorio de datos y métricas de COVID-19

El diseño del repositorio, se inicia con la implementación de una copia completa de cada actualización, con fines estadísticos sobre la evolución del proceso mismo de carga de datos, logrando obtener más de 170 millones de registros. Sin embargo, al momento de consolidar una base de datos para monitoreo de la epidemia, se mantienen operativos solamente los cambios en la condición epidemiológica de cada ciudadano.

Por su parte, se identifica que, para el análisis de indicadores relevantes de la epidemia, se requiere de diferentes agrupamientos territoriales dependiendo del funcionario, funcionaria o analista de datos que consume la información. Por ejemplo, desde el punto de vista sanitario, en la Provincia de Buenos Aires tiene relevancia analizar casos según los cordones del conurbano bonaerense, y también por regiones sanitarias [17]. Además, la información publicada en el sitio web, se construye a partir de múltiples fuentes de datos, entre las que se encuentran: Datos abiertos del Ministerio de Salud de la Nación [14], Sala de Situación COVID-19 (Ministerio de Salud de la Nación) [16] y Datos abiertos de situación epidemiológica de COVID-19 del Gobierno de la Ciudad de Buenos Aires [15], entre otros.

Durante el desarrollo de este estudio, el sistema tuvo a disposición conjuntos de datos completos actualizados diariamente desde el 15 de julio, conteniendo información epidemiológica dando cuenta de inicios de síntomas originados desde el 1 de enero de 2020. La presentación consolidada de los datos y principales indicadores, fueron desplegados a través de un sitio web, accesible en [12]. La arquitectura del sistema fue diseñada para que cualquier grupo de analistas de datos pueda crear nuevos contenidos visualizables en este sitio, aprovechando la infraestructura de datos disponible.

Dentro de el sitio, se proporciona la capacidad de definir unidades territoriales denominadas "Conglomerados", que consisten en agrupamientos relevantes de localidades. Esto puede verse en la pestaña: Provincia de Buenos Aires y AMBA (ciudad de Buenos Aires más su conurbano) — > Regiones Sanitarias; y también en — > Cordones Conurbanos. Al momento, el sistema presenta información sobre 550 departamentos a nivel nacional y tiene definidos 26 conglomerados de interés.

En líneas generales, se muestran evoluciones temporales de la cantidad de casos diarios y la cantidad de casos acumulados (ambos por fecha de inicio de síntomas) y de fallecimientos diarios y acumulados (por fecha de fallecimiento). También se muestran proporciones de concentración geográfica de casos confirmados. La información puede filtrarse por zona geográfica y consultarse

en escala lineal o logarítmica, en cifras totales o cada 100.000 habitantes y en valores diarios o promedios móviles cada 7 días. Los gráficos son interactivos, de manera tal que, el usuario, pueda modificar aspectos relevantes de cada visualización enfocándose en aquello que más le interesa.

A noviembre de 2020, el sistema ha sido consultado por más de 850 usuarios diferentes con más de 3800 visitas, provenientes en un 95 % de Argentina, e incluyendo también visitas desde países de América del Sur, América del Norte y Europa. Asimismo, las visualizaciones actualizadas diariamente sirven como insumo para compilar reportes que se proveen a pedido y también periódicamente a diferentes administraciones públicas, como por ejemplo, al Ministerio de Salud de la Provincia de Buenos Aires.

Las Figuras 7 y 8 muestran vistas referenciales de las páginas del sitio. La estructura del sitio, se organiza en divisiones políticas y geográficas: Provincia de Buenos Aires y AMBA (que incluye la Ciudad de Buenos Aires y su Conurbano), Argentina y Provincias, sumado a una sección de Análisis donde, al momento, se presenta la evolución del número reproductivo efectivo $R_{ef}(t)$ para cada provincia (ver Sección 2.1, Figura 1) y un análisis de demora de carga de fallecimientos (ver Figura 9).

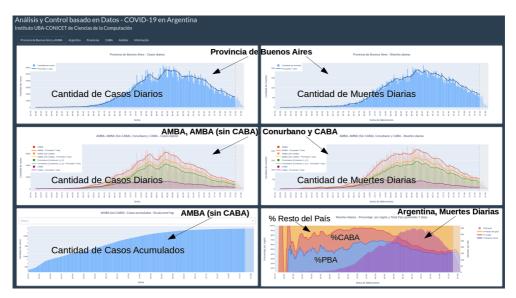


Figura 7: Vista del sitio mostrando casos confirmados diarios y fallecimientos diarios para la Provincia de Buenos Aires, AMBA, AMBA sin incluir la Ciudad Autónoma de Buenos Aires (CABA), Cordones 1 y 2 del Conurbano Bonaerense y CABA.

Nota: Incluye un análisis de la proporción de fallecimientos producidos en CABA, en la Provincia de Buenos Aires y en el resto del país.

Puede notarse que, hacia el final de las gráficas, hay una línea punteada vertical que separa los últimos 12 días de datos del resto de la información, y que los promedios móviles finalizan allí. Esto es consecuencia de que el sitio usa como base de fechas a la Fecha de Inicio de Síntomas (FIS), en vez de la Fecha de Reporte para los casos. Esto es así ya que la FIS da cuenta de un indicador más estable y fidedigno de la dinámica real de la epidemia, por estar menos sujeta a la variabilidad del proceso burocrático de carga de datos. Sin embargo, utilizar la FIS, tiene como aspecto negativo una tardanza, estimada en 12 días, hasta que se reportan todos los casos que comparten una misma FIS.



Figura 8: Vista del sitio mostrando el detalle de evolución de casos confirmados diarios para Municipios de la Provincia de Buenos Aires

Finalmente, se presenta un ejemplo de análisis de datos que es posible realizar, exclusivamente, si se mantiene un historial de todas las actualizaciones diarias de datos y se aplica un análisis de cambios de estados para cada individuo a lo largo del tiempo (es decir, no es suficiente acceder a "la última foto" de las bases de datos de acceso público).

En la Figura 9, se muestra la evolución del proceso de reporte de fallecimientos en CABA y la PBA, haciendo foco en la demora producida en la carga de fallecimientos que suceden cada día. Los colores denotan la diferencia entre la fecha de fallecimiento y la fecha de carga, para todos los casos que fueron ingresados al sistema, día por día. Esta información se encuentra en la sección Análisis del sitio. Esta comparación, tiene relevancia para mostrar las distintas estrategias que usan diferentes jurisdicciones para lidiar con un problema conocido y presente en todo sistema de carga de datos en el mundo, que es el retraso en la carga de fallecimientos cuando los sistemas administrativos sanitarios se ven desbordados.

Se deduce con este análisis que la Provincia de Buenos Aires optó por una transparentación y remediación inmediata, de alta intensidad, mientras que la Ciudad Autónoma de Buenos Aires optó por una estrategia de remediación lenta y dosificada.

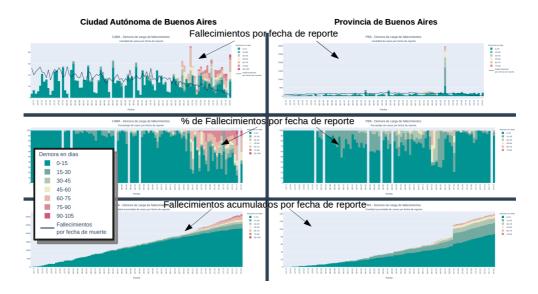


Figura 9: Vista del sitio mostrando la carga de fallecimientos y sus demoras. Nota: Izquierda: Ciudad Autónoma de Buenos Aires; Derecha: Provincia de Buenos Aires. Arriba: Cantidad de fallecimientos cargados por día; Centro: Porcentajes (normalizado a 100%) de demoras sufridas por los fallecimientos cargados para cada día; Abajo: Evolución acumulada de carga de fallecimientos

Este proyecto ha mostrado resultados muy satisfactorios, dado que ha probado ser de utilidad para sumar información relevante y de manera instantánea a numerosos debates y discusiones sobre la evolución de COVID-19 en Argentina, tanto en ambientes académicos como con funcionarios políticos y con la opinión pública. A su vez, las series temporales del sitio son utilizadas como insumo para ajustar modelos de simulación usados para proyectar posibles escenarios futuros y sugerir estrategias de manejo de la epidemia. En este sentido, contar con un panel unificado y confiable de datos, como el presentado en esta sección, viene cubriendo de manera exitosa una necesidad clave para el asesoramiento de tomadores de decisiones en políticas públicas mediante métodos matemático-computacionales.

2.5. Estimación de prevalencia a nivel nacional y distrital

Conocer el porcentaje de población infectada en cada distrito es de interés a la hora de tomar decisiones de políticas públicas, como ser políticas de testeo, rastreo, restricciones a la movilidad, etc. Para ello, se propone estimar el porcentaje real de población infectada en cada distrito utilizando como datos de entrada los fallecimientos discriminados por rango etario y por sexo, y la letalidad para cada uno de esos rangos, permitiendo, además, inferir la relación

entre infectados y casos confirmados.

La estimación de estos valores se realizará para el país en su totalidad, para las 24 provincias de la Argentina y para los 135 municipios de la Provincia de Buenos Aires (PBA). Calcularlos, con detalle para cada municipio de la PBA, es de importancia dada la estrecha colaboración entre el grupo de trabajo, el gobierno provincial y los diferentes municipios de la Provincia. Para ello, se utilizarán las siguientes fuentes de datos:

- Del Instituto Nacional de Estadística y Censos (INDEC) se obtendrán los porcentajes de población de cada rango etario y la población proyectada al 2020 para cada distrito.
- De los Datos Abiertos del Ministerio de Salud de la Nación se obtendrán los datos de casos y fallecidos. Los casos confirmados fueron clasificados por fecha de apertura. Los fallecidos por fecha de fallecimiento.

Metodología para estimación de prevalencia

El análisis para la estimación de prevalencia, se diseñó a partir de un estudio de seroprevalencia realizado en España, donde se obtuvo el IFR (Infection Fatality Rate) por decil etario y por sexo [21]. Se tomó con base la prevalencia española debido a que es el estudio con mayor nivel de detalle que se ha reportado en la literatura. De todos modos, otros estudios, en diferentes lugares del mundo, muestran valores similares. El resumen de dicha letalidad, para cada decil etario y cada sexo, puede verse en la Tabla 3.

Tabla	3:	IFR]	por rango		go etario y		de	España	a.
Estimado en [21]									
0	0-9 10-19		20	-29	30-	39	40-49		
${ m M}$	\mathbf{F}	${\bf M}$	\mathbf{F}	\mathbf{M}	\mathbf{F}	M	\mathbf{F}	\mathbf{M}	\mathbf{F}
$0{,}002\%$	$0{,}002\%$	$0{,}005\%$	$0{,}004\%$	$0{,}013\%$	$0{,}012\%$	$0{,}03\%$	$0{,}02\%$	$0{,}09\%$	$0,\!05\%$
50-59 60-69		70	-79	80-	89	90+			
${ m M}$	\mathbf{F}	${ m M}$	\mathbf{F}	\mathbf{M}	\mathbf{F}	M	\mathbf{F}	\mathbf{M}	\mathbf{F}
0,38%	$0{,}20\%$	$1{,}62\%$	$0,\!62\%$	$6{,}11\%$	$2,\!68\%$	$14{,}21\%$	$5{,}47\%$	$27{,}59\%$	$10{,}51\%$

En primer lugar, se estimó el llamado "IFR teórico", que proviene de considerar la letalidad de la Tabla 3 y la distribución etaria y por sexo de la población de la Argentina y de sus diferentes jurisdicciones. Con estos datos, podemos obtener este IFR teórico, que presupone que los contagios en todos los rangos etarios y en ambos sexos se distribuyen de manera homogénea. Una vez obtenido este IFR teórico, más la población del distrito en consideración y el número de fallecidos a la fecha, se puede estimar el porcentaje de población

realmente contagiada 2 semanas hacia atrás (asumiendo una media de 2 semanas entre el reporte de un caso y la fecha de su posible fallecimiento). Como el número reportado de casos confirmados es conocido, entonces es posible estimar el coeficiente de relación entre los casos reales y los confirmados, lo que se denomina como Factor Quiros $(FQ)^2$ [3]. Dado este FQ y los casos confirmados en las últimas 2 semanas, se logra estimar el porcentaje de contagiados al día de la fecha.

A continuación, para estimar el denominado IFR actual, se realizó un ajuste sobre estos cálculos, utilizando como datos de entrada los fallecimientos reales, con sus rangos etarios y sexo, para cada jurisdicción. Usando estos datos de fallecidos como punto de partida, más la prevalencia por rango etario y sexo española, fue posible estimar el número de contagiados reales en cada rango etario y sexo, en cada distrito, dos semanas hacia atrás desde el momento en que se realiza la estimación, lo que entreaga el porcentaje de población infectada. Con este porcentaje es posible deducir el IFR actual y, si además se utiliza el número de casos confirmados, es posible estimar también el FQ ajustado (por edades y sexo de los fallecidos). Asumiendo que el FQ se mantiene para las últimas 2 semanas, se puede estimar el porcentaje real de contagiados al día de la fecha, utilizando nuevamente el dato de los casos confirmados en las últimas 2 semanas.

Adicionalmente, es importante considerar que, dado que el número de fallecimientos en la franja etaria 0-19 es muy bajo y que la prevalencia tomada en España para esas franjas también fue calculada con muy pocos casos, no se utilizaron las prevalencias de las franjas 0-9 y 10-19 en la estimación del IFR actual y que, para estimar el número de casos reales en ambas franjas, se tomaron los porcentajes de infectados de la franja 60-69 para la franja 10-19 y de 70-79 para la franja 0-9 (para cada distrito), asumiendo que se puede establecer un paralelo entre el nivel de cuidados de esas respectivas franjas y por consiguiente que el porcentaje de contagios ha sido relativamente pareja.

Resultados

La Tabla 4 muestra la situación de la Argentina al 30 de octubre de 2020. Se exhiben los datos para el país en su conjunto, la ciudad de Buenos Aires, el conurbano bonaerense, el interior de la Provincia de Buenos Aires y el resto del país. Se puede ver que, para esa fecha, tomando la estimación derivada del cálculo que usa edades y sexos de los fallecimientos reales, la Argentina ya tenía alrededor de un 24 % de su población contagiada en el plano nacional,

²Notar que el FQ es igual a 1/IDR, donde el IDR es el Infection to Detection Ratio, es decir la proporción de la unidad de los casos detectados, definido en la literatura

Tabla 4: Situación de la Argentina en cuanto a prevalencia e IFR en los principales distritos del país para fines de octubre de 2020.

Población Población Población Infectada Población Infectada

Distritos	Población (2020)	Casos Confirmados	Fallecidos	Incidencia de Casos	Población Infectada (según IFR teórico)	IFR homogéneo	Factor Quiros Homogéneo	Población Infectada Proyectada (según IFR teórico)	Población Infectada Estimada (c/ edad fall. y sexo)	IFR Actual	Factor Quiros Actual	Población Infectada Proyectada (c/ edad fall. y sexo)
				16/10	16/10			30/10	16/10			30/10
Argentina	45.376.763	1.157.174	30.792	2,19%	11,16%	0,61%	5,10	13,02%	20,51%	0,33%	9,38	23,92%
Ciudad de Buenos Aires	3.075.646	146.758	4.907	4.57%	17.68%	0.90%	3.87	18.47%	34.76%	0.46%	7.61	36.31%
Conurbano Bonaerense	13.630.369	458.346	14.993	3.16%	19.55%	0.56%	6.18	20.79%	35.11%	0.31%	11.11	37.34%
Interior de la Prov. de Buenos Aires	3.910.772	86.033	2.324	1.75%	7.28%	0.82%	4.16	9.16%	12.03%	0.49%	6.88	15.13%
Resto de las Provincias	24.759.976	463.408	8.493	1.41%	6.14%	0.56%	4.34	8.13%	11.81%	0.29%	8.35	15.62%

valor que supera ampliamente a los porcentajes de contagiados en los países de Europa en la primera ola de contagio ocurrida en marzo/mayo. En la ciudad de Buenos Aires y el conurbano bonaerense, ese valor alcanzó cifras cercanas al 37 % (superando también fuertemente al nivel de contagios en otros grandes conglomerados urbanos en el mundo), mientras que, en el interior de la Provincia de Buenos Aires y en el resto del país, ese valor bajaba a cifras cercanas al 15 %.

A su vez, los valores de los IFR actuales son más altos que los de los IFR teóricos, producto que los contagios no han sido homogéneos por rango etario y por sexo. Lo cual es esperable frente a distintos comportamientos entre grupos etarios: la gente mayor se ha cuidado más y, por consiguiente, en términos proporcionales, se ha contagiado menos, bajando los valores del IFR en prácticamente todos los distritos (por ejemplo, a nivel nacional el IFR teórico es de 0,61, mientras que el actual es de 0,33).

En cuanto al FQ real a nivel nacional, este número es de alrededor de 9, mostrándose acorde con lo visto en otros lugares del mundo. Esto quiere decir que se estaría confirmando un caso por cada 9 contagios.

Se concluye que, el número de infectados reales (deducido a partir de la letalidad por rango etario y sexo, más los datos de fallecimientos), a fines de octubre, tanto en la ciudad de Buenos Aires como en el conurbano bonaerense, toma valores muy altos, que bordean el 40 % del total de la población (lo que ha implicado un alto número de fallecimientos, ubicando al país en su conjunto en el top-10 en el mundo en este rubro cada 100 mil habitantes a fines de octubre de 2020). A pesar de la alta tasa de movilidad, en Buenos Aires y sus alrededores, el número semanal de contagios disminuyó para ese momento.

Esto se debió probablemente a los cuidados personales puestos en práctica y a la fuerte disminución de individuos susceptibles de contagio, producto de la alta proporción de ya contagiados. En el resto del país, la situación aún presentaba valores no tan altos de infectados (rondando el 15 % si miramos de manera global tanto el interior de la Provincia de Buenos Aires como el interior del país), lo que obligaba a extremar las medidas de cuidados personales, restricción de la movilidad y rastreo permanente de contactos estrechos, para mantener la curva de contagios, y en consecuencia de fallecimientos, en niveles bajos. La actualización semanal de estos datos se encuentran en [10].

2.6. Inmunidad: de las causas biológicas a las sociales

El término inmunidad del "cagazo", del lunfardo argentino ("fright immunity" en inglés y con una terminología más formal), da cuenta de la toma de conciencia de los individuos después de conocer de cerca las consecuencias de COVID-19, incluyendo muertes y secuelas graves. Al sobrevenir el "cagazo", la mayoría se aísla y empieza a comportarse de una manera mucho más cautelosa. El número de reproducción efectiva $R_{ef}(t)$ para todo el conjunto de individuos "asustados" es muy inferior a 1. Sin embargo, una pequeña fracción de individuos persiste en un comportamiento peligroso (a quienes llamamos "temerarios", personas mayoritariamente jóvenes, que no tienen miedo y se comportan lo más similar posible a como actuaban anteriormente), que, aunque algo más suave que antes de la pandemia, sigue implicando un número de reproducción $R_{ef}(t)$ superior a la unidad. Muchos de estos individuos estaban conectados por la cercanía social y física, y el virus encontró una forma bastante sencilla de propagarse. Es importante señalar que estas fracciones no se definen en forma institucional, ni están necesariamente determinadas por su situación laboral. Muchos trabajadores esenciales exceptuados se comportaron siguiendo normas muy estrictas y protocolos de cuidado, y viceversa. Siendo los temerarios una fracción menor del conjunto, y siendo mucho más móviles, alcanzan (como subgrupo) una verdadera inmunidad de rebaño en los meses posteriores al alto pico pandémico. Sin embargo, la mayoría asustada nunca llega a alcanzar altos números de infección, aunque una pequeña fracción de ellos es infectada por la población móvil. Esta fracción pequeña de un total grande implica que el número total de cuarentenados infectados es similar al de los temerarios, mostrando una de las características notables de la epidemia de CoVid: unos transmiten, otros padecen. Esto se acentúa debido a la mayor letalidad entre los adultos mayores, generalmente menos arriesgados.

Se han planteado algunas especulaciones sobre la posibilidad que la inmunidad de rebaño, en un cierto grado, sea la principal causa de la disminución de

las infecciones en Europa después del brote inicial. Algunas estudios indicaron que, entre el $10\,\%$ y el $20\,\%$ de las infecciones son suficientes para detener la dinámica de crecimiento, en lugar del $60\text{-}70\,\%$ que indican los cálculos simples (se estima este valor en $100\times(R_0-1)/R_0$)). Entre las especulaciones sobre la inmunidad de rebaño, se mencionan varias causas posibles: cierta inmunidad previa en toda la población debido a la reacción cruzada con otros coronavirus, patrones de contagio altamente inhomogéneos, entre otras. Sin embargo, la mayoría de estas especulaciones cayeron cuando se conocieron los números reales de la inmunidad adquirida: menos del $10\,\%$ en la mayoría de las ciudades. El parámetro conocido como IFR (Infection-Fatality Ratio), que puede estimarse con suficiente precisión a partir del recuento de muertes, tal cual se vio en la subsección anterior, incluso en ausencia de amplios estudios de sero-prevalencia, muestra también valores de prevalencia bajos en todos los países europeos. Teniendo en cuenta estos hechos, se plantea la cuestión de la causa principal del rápido declive de la tasa de infección en los países europeos.

Para ello, se propuso una investigación para indagar y encontrar explicaciones alternativas relacionadas con la disminución de los casos producto de causas sociales y, no necesariamente, causas biológicas, en términos simples, el desplazamiento de la inmunidad de rebaño a la inmunidad del "cagazo".

Resultados

Es posible establecer un modelo simple basado en un SEIR de dos poblaciones, donde el parámetro de transmisión beta se desglosa en cuatro valores $\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}$, correspondientes a la transmisión entre las poblaciones 1 (asustados y en cuarentena) y 2 (temerarios y móviles). En la Figura 10 se muestra la ejecución de un modelo esquemático para España en el que ambas poblaciones se describen como homogéneas dentro de sí mismas. A partir del IFR, las muertes y la seroprevalencia se pueden derivar las cifras reales de infección. La fracción de individuos temerarios (f2) es un parámetro ajustado a los datos reales, así como los números de reproducción $R_1 = \beta_{11}/\gamma, R_2 = \beta_{22}/\gamma$ y el parámetro de transmisión cruzada $beta_{21}$.

Este tipo de comportamiento, debe considerarse como un cartoon, en la medida en que es indudable que existe todo un degradado de comportamientos asociados, en mayor o menor medida, con las posiciones extremas del modelo. Cuando se aplica el procedimiento de ajuste a diversos países o ciudades de Europa (Reino Unido, Alemania, Bélgica, Francia, Madrid, Londres), se observa que, en todos los casos, el ajuste arroja 2 valores bien diferenciados de R correspondientes a una situación de comportamientos diferentes. Distinto es cuando se trata con este mismo modelo los datos de la Ciudad Autónoma

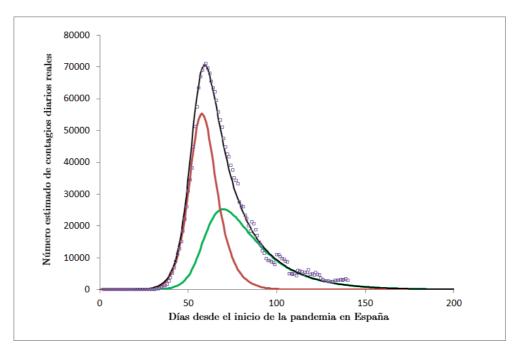


Figura 10: Patrón de inmunidad de "cagazo" para España. Nota: Línea roja: Infecciones diarias de la población móvil. Línea verde: Infecciones diarias de la población en cuarentena. Línea negra: suma de infecciones. Cuadrados: datos reales corregidos por factores de seroprevalencia e IFR. Población = 47×10^6 , $R_1 = 0.67$, $R_2 = 3.57$, $f_2 = 0.0253$, $\beta_{21} = 8.3 \times 10^{-3} \beta_{12} = 0$

de Buenos Aires. Los R ajustados son de 1.6 y 1.13 respectivamente, ambos mayores que 1 y cercanos entre sí, lo cual muestra que no hay dos poblaciones claramente diferenciadas como en los países europeos. Este bajo nivel de "cagazo", es muy compatible con el éxito en el objetivo buscado en "aplanar la curva" de contagios. La llegada paulatina de casos y fallecidos, y la no saturación del sistema de salud disminuyen en forma importante la percepción de peligro, y alientan un comportamiento más cercano a la normalidad, sólo detenido en parte por las restricciones a la circulación y a ciertas actividades. El análisis debe diferenciarse respecto de los barrios populares, donde las barreras a la transmisión son mínimas y los valores de R son mucho más altos, del nivel esperado para la curva epidémica natural de COVID. Las estrategias con que se pretendió morigerar esta situación fueron inefectivas, llegando a una prevalencia acumulada de más del 50 % en pocas semanas (por ejemplo, el Barrio Carlos Mugica) [3]. Una situación muy diferente ocurre en los barrios tradicionales de la Ciudad. Como se vio en el capítulo anterior, a fines de octubre, la prevalencia acumulada de CABA ya puede estimarse por encima del 35 %. Sin embargo, la fracción de infectados es muy diferente para los distintos rangos etarios.

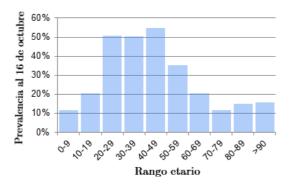


Figura 11: Prevalencia estimada para los diferentes rangos etarios en la Ciudad de Buenos Aires.

La Figura 11 muestra que, mientras los individuos más jóvenes se han comportado más temerariamente y alcanzado niveles de prevalencia importantes, aunque mucho más lentamente que la fracción temeraria en España y otros países de Europa, los adultos mayores han tenido mucho menos contacto con el virus, habiendo comprendido el riesgo asociado, particularmente alto a estas edades. Esta baja prevalencia de los mayores indica, adicionalmente, que para el alto número de muertes en la ciudad de Buenos Aires y su conurbano, el número de casos reales entre los individuos jóvenes ha sido varias veces superior a los ocurridos en España y demás países de Europa.

El modelo matemático propuesto muestra que tendría capacidad para explicar el porqué del descenso de la curva de contagios en los países de Europa a partir de mayo, a pesar del bajo número de prevalencia. Las causas de este comportamiento se mueven de las razones biológicas a las sociales. Mientras en Europa se observa un pico agudo con saturación del sistema de salud, seguido de una baja abrupta de casos producto de las medidas tomadas y la toma de conciencia por parte de la población, la curva epidémica en la ciudad de Buenos Aires y su conurbano es completamente diferente, no existiendo "cagazo" generalizado a pesar del alto número de casos y de muertos, por cuanto éstos aparecen en forma paulatina. Mientras los jóvenes presentan una prevalencia cercana a condiciones de "rebaño", señal de un comportamiento relativamente más móvil durante el aislamiento impuesto por el gobierno, lo cual ayuda a bajar el número de casos diarios (sumado seguramente al factor estacional que puede haber acentuado el descenso), los mayores han permanecido con

poco contacto. La baja prevalencia entre los rangos etarios más proclives a cuadros graves de COVID-19 representan un riesgo adicional en caso de que no se detenga la transmisión a niveles muy bajos antes de que las condiciones climáticas vuelvan a ser más favorables para la propagación del virus.

3. Desarrollos en territorio

3.1. Hospitales en red: el caso del Hospital del Cruce en Florencio Varela, en la Provincia de Buenos Aires

Debido al avance de la pandemia, los hospitales de la región sudeste del Gran Buenos Aires, que incluye los municipios de Almirante Brown, Quilmes, Florencio Varela y Berazategui, comenzaron a reestructurarse para trabajar como una única unidad funcional (ver Figura 12). Uno de los objetivos de formar esta red fue organizar, y así optimizar, la ocupación de camas frente a la pandemia que se avecinaba. La red cuenta con 7 hospitales, cuyo hospital central es el Hospital del Cruce de Florencio Varela. Durante la pandemia se construyeron además 3 hospitales Modulares y 4 Unidades de Pronta Atención (UPA), algunas de las cuales funcionan dentro de los mismos hospitales.

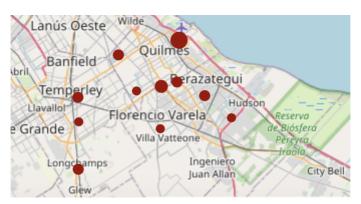


Figura 12: Localización de los hospitales de la región sudeste del Gran Buenos Aires, cuyo hospital central es el Hospital del Cruce de Florencio Varela. Nota: Se incluyen los Modulares y UPA que fueron construidos durante el desarrollo del presente proyecto.

Luego, se propone el desarrollo de una herramienta computacional para organizar a los pacientes y camas de esta red de hospitales, que pueda ser utilizada por cualquier otra red de hospitales del país, y que permita realizar análisis epidemiológicos posteriores con la información generada. Dentro del

proyecto participaron alrededor de 50 personas entre investigadores del CO-NICET, desarrolladores independientes, médicos y personal administrativo y jerárquico de los hospitales.

Metodología para herramienta computacional para organizar red de hospitales

Debido a la urgencia para comenzar a utilizar la herramienta, se decidió hacer una primera versión (producto mínimo viable) en pocos días, y luego se fue ampliando el desarrollo, a medida que se fueron definiendo los requisitos. La Figura 13 muestra la pantalla principal del sistema en uso para el mes de octubre.

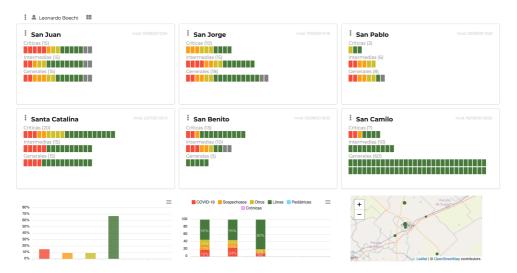


Figura 13: Panel Principal de Camas.

Nota: Se observan los hospitales de la red (7 hospitales más 7 Modulares/UPA) y las camas organizadas en Tipos: Generales, intermedias o Críticas. El color de cada cama corresponde con su Estado: Covid-19, Sospechoso, Otros, Libre, Inhabilitada. Cada bolita del mapa corresponde a un hospital/Modular/UPA de la red, el área y el color de las bolitas corresponde a la cantidad de camas que hay para ese Estado. El panel permite descargar datos. La imagen corresponde a un panel con datos ficticios.

Si bien el sistema respondió a los requerimientos de esta red de hospitales, se desarrolló para que cualquier red de hospitales pueda utilizarlo ("Software as a Service"). El sistema está localizado en un sitio accesible desde cualquier computadora. Todo es personalizable y autogestionable. A su vez, permite ingresar datos de los pacientes, los cuales pueden observarse en el Panel de

Pacientes (Figura 14), que es utilizado para organizar a los pacientes y a las camas de cada dependencia. Cabe aclarar que el sistema no releva información sensible de los pacientes como nombre, DNI, dirección, teléfono, etc.



Figura 14: Panel de Pacientes.

Nota: El tacómetro muestra el nivel de riesgo social, el hombrecito posee diferente color según el riesgo clínico, también se incluye la cantidad de días de internación, la edad del paciente, y si ha sido derivado de otro hospital (ambulancia).

Finalmente, la herramienta mostró gran aceptación, se encuentra ampliamente utilizada dentro de la red de hospitales [28] y está permitiendo organizar las camas y los pacientes dentro de los hospitales, así como los traslados entre los mismos. El proyecto continúa en etapa de desarrollo de nuevas funcionalidades y requerimientos. El link al sistema está disponible en [8].

3.2. Centro de Prevención COVID-19 en la Facultad de Ciencias Exactas y Naturales de la UBA

La mejor manera de prevenir una enfermedad viral que se propaga, principalmente, de persona a persona y, para la cual, no existe aún una vacuna, es evitar la exposición al virus. Implementar medidas de prevención ha sido el mejor accionar en el manejo y control sanitario del COVID-19; tanto por su impacto en la contención como en la mitigación de la enfermedad. Uno de los programas de prevención desarrollados por la Provincia de Buenos Aires es el

de **identificación** y **acompañamiento** de los contactos estrechos (CE³) de casos confirmados o sospechosos de COVID-19 para asegurarse que estos no sigan contagiando, y al mismo tiempo que éstos puedan recibir la asistencia que necesiten.

El 22 de julio del corriente año, se puso en funcionamiento un Centro de Telemedicina Universitario (CeTeU) en la Facultad de Ciencias Exactas y Naturales de la Universidad de Buenos Aires, con el fin de llevar adelante las tareas de identificación y acompañamiento de CEs. Su creación tuvo como objetivos cortar con la cadena de contagios asegurando que los potenciales infectados de Covid así como los ya confirmados permanezcan confinados en sus hogares; detectar de manera temprana posibles nuevos infectados; y asistir a las personas que están confinadas durante los días que dure su confinamiento. En particular, este centro de prevención brinda apoyo al Municipio de Vicente López y más recientemente incorporó a los Municipios de Tigre y Exaltación de la Cruz, todos ellos pertenecientes a la Provincia de Buenos Aires. Este Centro se enmarca en un programa de dicha Provincia que involucra a otros 34 centros universitarios más con la misma dinámica (todos ellos gestados desde Universidades públicas nacionales), 17 de ellos atendiendo al conurbano bonaerense y otros 17 atendiendo a municipios del interior de la Provincia.

Metodología y resultados

El CeTeU está integrado por un grupo integrado por más de 40 estudiantes, graduadas y graduados de dicha Facultad. Funciona de lunes a sábado en 4 turnos fijos de 4 horas, cada uno de los cuales es coordinado por una persona responsable de las tareas diarias. Cada integrante del centro cuenta con una computadora y auriculares de uso personal y exclusivo para realizar los llamados a través de un servicio de telefonía IP. Tanto las instalaciones físicas como las conductas individuales cumplen con las medidas sanitarias establecidas por un estricto protocolo de higiene y seguridad provisto por la unidad académica de la institución y aprobado por la Provincia. Toda la información recabada durante la identificación y los seguimientos de los CEs es registrada en una plataforma digital desarrollada por la PBA.

El trabajo diario de este centro se estructura de la siguiente manera:

³Se considera CE (según la definición de la PBA) a cualquier persona que haya permanecido a una distancia menor a 2 metros con un caso confirmado o sospechoso de COVID-19, mientras presentaba síntomas o durante las 48 horas previas al inicio de los síntomas, por un período de al menos 15 minutos.

Identificación:

- Comunicación telefónica con la persona sospechosa o confirmada de COVID-19.
- 2. Determinación de la fecha de inicio de síntomas.
- 3. Relevamiento de sus CE.
- 4. Ingreso de la información de los CE al aplicativo de la PBA.

Acompañamiento:

- 1. Comunicación telefónica con el CE.
- 2. Constatación de la fecha de último contacto con su vínculo epidemiológico.
- Consulta sobre la aparición de uno o más síntomas compatibles con COVID-19.
- 4. Relevamiento sobre las necesidades de asistencia alimentaria y/o insumos de higiene.
- 5. Clasificación de la evolución del seguimiento según: continúa en aislamiento, alta de seguimiento o pasa a ser un caso sospechoso por presencia de síntomas.

Además de las preguntas de rutina, las llamadas brindan un espacio de asesoramiento y contención a las personas durante el aislamiento estricto que deben afrontar, así como también, una oportunidad para relevar necesidades en la provisión de alimentos e insumos de higiene, entre otras. Esta información es debidamente recopilada y transmitida diariamente a los responsables de los municipios quienes, a su vez, la utilizan como fuente de información en su gestión interna.

Desde el inicio de funcionamiento del CeTeU y por el periodo de 16 semanas, se han realizado 31070 llamadas efectivas en total, de las cuales 6560 corresponden a personas sospechosas o confirmadas para hacer la identificación de sus CEs (un promedio de 71 personas por día), obteniendo 3931 personas con la clasificación de CE (un promedio de 42 personas por día). Esto muestra que, aproximadamente, cada 10 sospechosos contactados el centro identifica 6 CEs. Si bien los datos no son concluyentes, una explicación posible para esta relación es que una buena parte de los casos sospechosos son descartados debido a que sus hisopados resultaron ser negativos. Por otro lado, se acumuló un total de 20579 llamadas de seguimientos, con un promedio de 224 llamados

por día para esta tarea. En la Figura 15, se grafica el total por semana de los sospechosos contactados junto con el total por semana de CEs identificados, mientras que, en la Figura 16, se muestra el total por semana de llamados de seguimientos realizados.

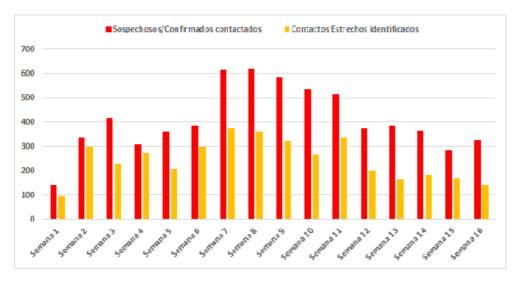


Figura 15: Número de sospechosos/confirmados y CEs identificados por semana del Centro.

Aunque todavía es difícil poder medir el impacto real que tiene la tarea que realiza el CeTeU, existen ciertos indicadores positivos. Un primer indicio se puede ver en la Figura 17, donde se grafica tanto la evolución de los casos confirmados en Vicente López, como en el resto de los 12 municipios correspondientes a la Región Sanitaria V (organización administrativa que utiliza el Ministerio de Salud de la PBA para agrupar los municipios). Los casos confirmados en Vicente López muestran una estabilización hacia fines del mes de julio (cuando el centro inició sus actividades), mientras que, el acumulado de los otros municipios de la Región Sanitaria V continuaron creciendo de manera sostenida.

Por otra parte, la evolución del R también muestra una disminución mayor en Vicente López que en el conjunto de los partidos de la misma Región Sanitaria, como puede apreciarse en la Figura 18.

Otros indicios son de tipo cualitativo, como la fuerte concientización que se realiza a través de cada llamado, ya sea explicando la necesidad del estricto aislamiento de los CEs y las medidas de prevención adecuadas que deben tomar. La respuesta social a estos llamados es, en la gran mayoría de los casos, positiva y de agradecimiento. Por otro lado, desde el Municipio de Vicente López también se ha manifestado un gran reconocimiento hacia la tarea rea-

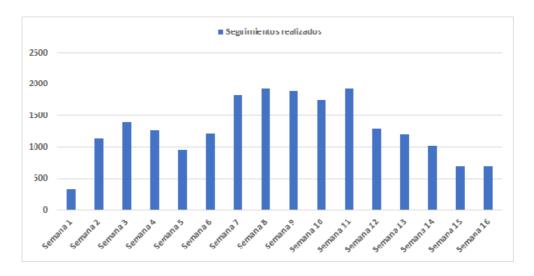


Figura 16: Número de seguimiento realizados a Contactos Estrechos por semana.

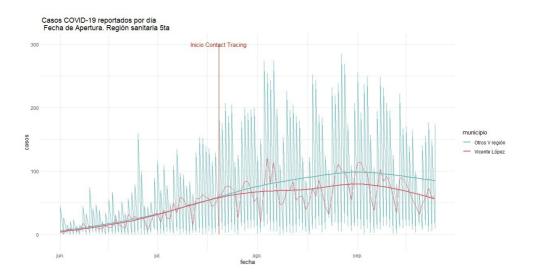
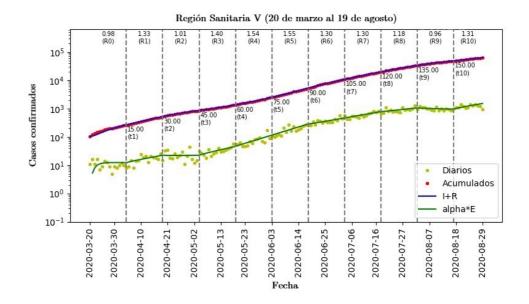


Figura 17: Evolución de casos confirmados a lo largo del tiempo para el municipio de Vicente López (en rojo) y para el conjunto de los otros municipios que corresponden a la misma Región Sanitaria que Vicente López (en verde), y sus respectivas curvas de ajuste.

Fuente: Antonio Milanese, Municipalidad de Vicente López.



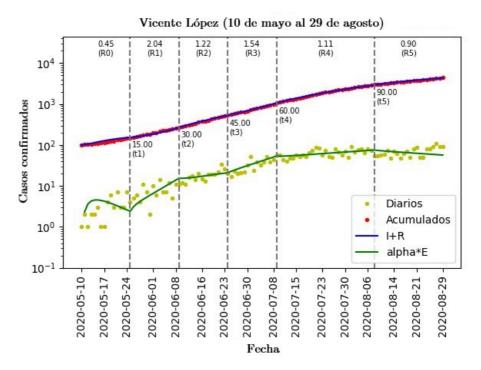


Figura 18: Evolución del R en Vicente López (abajo), y en el conjunto de los municipios de la Región Sanitaria V a la que pertenece Vicente López (arriba). Nota: Estos indicadores provienen del método de ajuste de parámetros basado en modelos de simulación descrito en la Sección 2.1 el cual es calculado automáticamente a diario por el sistema de análisis, control y visualización descrito en la Sección 2.4.

lizada en el centro y el aporte de la información suministrada. En el caso de Tigre, el comienzo es muy reciente, pero sin lugar a dudas la participación del Centro permite que el Municipio pueda identificar y acompañar una mayor cantidad de casos y sobre todo contactarse de manera más temprana.

Como conclusión, se puede afirmar que la experiencia del Centro de Telemedicina Universitario en la Facultad de Ciencias Exactas y Naturales de la Universidad de Buenos Aires es satisfactoria y está generando un impacto que próximamente podrá ser aún más cuantificable. Mientras tanto, es una herramienta útil en las tareas de prevención, siendo la primera línea de adquisición de datos utilizados para la gestión en la contención y mitigación de la enfermedad.

3.3. Una plataforma de simulación basada en agentes para el municipio de Marcos Paz, en la Provincia de Buenos Aires

Los modelos compartimentales descritos en la Sección 2.1 incurren en un nivel de abstracción que no permite estudiar dinámicas detalladas a nivel territorial. Los compartimientos, el número reproductivo R, y la estructura misma del modelo son una abstracción macroscópica, que representa el comportamiento emergente de redes de interacciones microscópicas complejas entre personas de diferentes edades que participan de manera muy diferenciadas del proceso epidémico. Estas redes de interacción se manifiestan en el territorio según la estructura socioeconómica de cada jurisdicción. A este nivel de detalle, tienen injerencia relevante aspectos tales como la ubicación de edificios públicos, comercios y plantas fabriles, vías de comunicación con otros municipios, ubicación y capacidad de hospitales, densidad poblacional en zonas residenciales y dinámica de aforo hacia zonas céntricas de mayor actividad, por mencionar algunas.

Analizar dinámicas sociales a este nivel, es de interés particular de municipios que deben tomar decisiones sobre reglamentaciones para manejar la epidemia, las cuales deben contemplar la realidad específica de la jurisdicción (un municipio sustentado en la actividad turística de veraneo, como el caso de Mar del Plata, desarrollará estrategias diferentes a un municipio mediterráneo centrado en la industria alimenticia, como el caso de Marcos Paz).

Es por ello, que se propone el desarrollo de una plataforma integral basada en modelos de simulación que permita asistir a la toma de decisiones con nivel de detalle territorial explícito, con foco particular en el Municipio de Marcos Paz en el segundo cordón del conurbano de la Provincia de Buenos Aires.

Metodología y resultados

La simulación basada en agentes individuales y espacialmente explícita [24, 23], permite ejercitar escenarios hipotéticos donde las intervenciones sobre la realidad toman la forma, por ejemplo, de restricciones en la cantidad de vehículos que se permite atravesar el municipio provenientes de otras jurisdicciones, cantidad de personal exceptuado residente al que se le permite realizar jornadas laborales fuera del municipio, cantidad y tamaño de los edificios públicos con autorización de aforo de ciudadanos y sus limitaciones de cantidad de personas compartiendo simultáneamente ciertos espacios, etc. Los agentes, además de formar grupos familiares ubicados acorde a la densidad de residencia indicada por los censos, cursan la epidemia atravesando los estados susceptible, expuesto, infectado asintomático, infectado sintomático (posiblemente leves o severos), recuperado, y fallecido.

Se desarrolló un modelo de simulación, junto a su infraestructura de visualización, para el Municipio de Marcos Paz [18], con capacidad de simular explícitamente los 65.000 habitantes del municipio y flujos de viajeros desde/hacia otros municipios. Mediante una interacción estrecha con la intendencia de Marcos Paz y su Ministerio de Salud realizamos fases incrementales de ajuste del modelo para capturar las principales dinámicas socioeconómicas que describen la movilidad de personas y la evolución de las métricas de la epidemia. En la Figura 19 se muestra un ejemplo de una "foto" de actividad simulada mediante nuestro modelo de agentes espacialmente explícito.

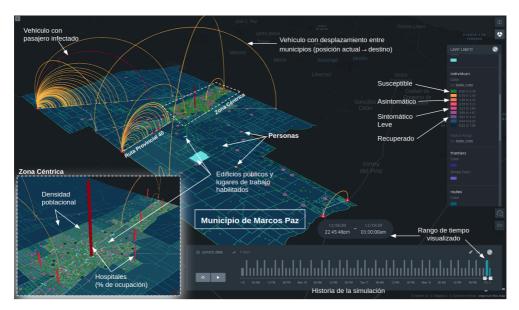


Figura 19: Visualización de simulación basada en agentes para el Municipio de Marcos Paz en la Provincia de Buenos Aires.

Esta herramienta muestra una gran riqueza para la comunicación, interacción y discusión de resultados y escenarios posibles con los tomadores de decisiones de políticas públicas. Si bien se encuentra en fases iniciales de desarrollo, su estructura abre las puertas para poder brindar un marco de trabajo para futuros análisis de problemáticas socioeconómicas que exijan una representación explícita a escala urbana en el Municipio de Marcos Paz.

4. Conclusiones generales y discusión

La pandemia puso en vilo a toda la sociedad debido principalmente a la elevada velocidad de dispersión del virus, influenciada por la alta movilidad que existe por los flujos comerciales y turísticos. Es por ello que tomó a la mayoría de las naciones por sorpresa, generando estragos en el sistema sanitario y económico mundial.

La comunidad científica internacional reaccionó rápidamente buscando métodos paliativos y vacunas, y desarrollando modelos para poder adelantarse o al menos estar al día con el desarrollo de la pandemia. En el caso de la Argentina, la comunidad científica respondió de forma muy rápida y proactiva generando soluciones de todo tipo, desde sistemas de detección, distintos tipos y mecanismos de testeos, tratamientos y diferentes propuestas para contener los contagios e ir entendiendo el comportamiento de la pandemia, hasta el desarrollo de la aplicación oficial para prevención y cuidado de la salud pública (aplicación Cuidar [5]), en la que participaron activamente científicas y científicos del grupo autor de este trabajo.

En este artículo hemos presentado los principales aportes que un grupo de científicos y científicas de los Institutos de Cálculo y de Ciencias de la Computación, pertenecientes a la Facultad de Ciencias Exactas y Naturales de la Universidad de Buenos Aires y al CONICET, junto a colegas de otras instituciones académicas, han realizado para aportar a la lucha contra la pandemia en Argentina. Se desarrollaron herramientas cuantitativas que están ayudando a entender el estado de la pandemia, así como también a ensayar escenarios para evaluar diferentes políticas de contención. Los trabajos desarrollados incluyeron políticas en territorio, tales como los desarrollos para hospitales y los centros de rastreos y de identificación de contactos (ésta última una tarea imprescindible pero que no fue implementada en la escala que hubiera sido necesaria a nivel nacional). También se brindó consultoría y apoyo científicotecnológico a municipios y estados provinciales, y se organizó un seminario semanal de aportes científicos en el marco de la pandemia, que incluyó entre

junio y diciembre la realización de 29 encuentros con diferentes especialistas de variadas disciplinas [11].

Si se tuviese que analizar el comportamiento de la Argentina ante la pandemia, se debería decir que la cuarentena temprana dictada por el gobierno nacional el 20 de marzo permitió un crecimiento lento de la curva de contagios (ver Figura 1), que a su vez dio pie a un fortalecimiento del sistema sanitario. Esto trajo como resultado favorable que, a diferencia de lo ocurrido en la mayor parte de los países de Europa, el sistema no se saturó y se evitaron potenciales muertes por colapso sanitario. En contraposición, este "aplanamiento" de la curva, donde se ve ese crecimiento lento pero permanente de los contagios, derivó en que el número acumulado de infectados a nivel nacional a lo largo de los meses fuese muy alto (ver Sección 2.5), y que lo mismo haya pasado con el número de fallecidos, convirtiendo a la Argentina en un país con una cifra de muertos cada 100 mil habitantes muy elevada. Estas cifras podrían haberse reducido con una estrategia nacional orientada a la supresión del virus, basada en el rastreo de contactos estrechos de los casos confirmados y sospechosos, y la búsqueda de contagiados asintomáticos mediante técnicas inteligentes de testeo. Algunas de estas acciones tuvieron buenos resultados locales (como puede verse, por ejemplo, en la Sección 3.2), pero la estrategia no llegó a expandirse a escala nacional ni a tiempo como para conseguir un descenso drástico de los contagios.

Esperamos encontrarnos cerca del fin de la pandemia, ya sea porque se consigue suprimir la misma a través de rastreos y testeos eficientes, porque se obtiene la inmunidad por medio de una campaña de vacunación, porque se consiguen tratamientos efectivos, o por una combinación de estas razones, pero es fundamental contar con información precisa para poder tomar las mejores decisiones para controlar la propagación del virus. Es posible que estas técnicas de modelado también puedan ser útiles para la organización de las campañas de vacunación y entrega de medicamentos, así como también, para el enfrentamiento a otro tipo de enfermedades de características similares.

En síntesis, consideramos que los aportes realizados por nuestros equipos de trabajo han sido muy valiosos para la lucha contra el Covid-19 en Argentina. Esto se evidencia en la receptividad que tuvieron las propuestas en diferentes gobiernos locales y provinciales, y en el fuerte impacto que buena parte de estos desarrollos tuvieron en los medios de difusión de alcance regional y nacional.

Contribución de los autores

Se detalla a continuación la participación principal de cada uno de los co-autores de la publicación en las diferentes secciones del trabajo.

Edición y recopilación general: Mehrnoosh Arrar, Guillermo Durán y Diego Garbervetsky.

Sección 2.1: Rodrigo Castro, Esteban Lanzarotti y Ezequiel Pecker-Marcosig.

Sección 2.2: Ana M. Bianco, Guillermo Solovey y Marina Valdora.

Sección 2.3.1: Laouen Belloli, Leonardo Pepino, Juan E. Kamienkowski y Luciana Ferrer.

Sección 2.3.2: Ana M. Bianco y Marina Valdora.

Sección 2.4: Rodrigo Castro, Pablo Laciana, Esteban Lanzarotti, Ezequiel Pecker-Marcosig y Mariano Zapatero.

Sección 2.5: Guillermo Durán, Roberto Etchenique, Mario Lozano, Rodrigo Maidana y Rodrigo Quiroga.

Sección 2.6: Roberto Etchenique.

Sección 3.1: Leonardo Boechi.

Sección 3.2: Guillermo Durán, Natalia Fernández, Diego Garbervetsky, Rodrigo Goldsmit, Carolina Grillo Vidal, Mario Lozano, Mauricio Mendiluce y Sol Minoldo.

Sección 3.3: Rodrigo Castro y Ezequiel Puerta.

Agradecimientos: Queremos agradecer a los gobiernos de la Provincia de Buenos Aires (colaboración formalizada mediante Convenio de Cooperación Institucional entre la FCEyN-UBA y el Ministerio de Salud de la PBA), de Marcos Paz (colaboración formalizada mediante Acuerdo de Cooperación Institucional entre la FCEyN-UBA y el Municipio), de Vicente López, Tigre y Exaltación de la Cruz, a las autoridades del Hospital del Cruce y su red de hospitales, a las autoridades de la Facultad de Ciencias Exactas y Naturales de la UBA y a las fundaciones Huésped y Fundar, por su fuerte apoyo en diferentes etapas de este trabajo. A Lauriano Alimenti, Carlos Bianco, Leandro Cahn, Inés Caridi, Jorge Cassará, Sebastián Ceria, Sebastián Crespo, Miguel Fernández, Julián Fernández Bonder, Juan Flo, Daniel Foguelman, Matías García Fallit, Mariano Goldman, Rafael Grimson, Daniela Hozbor, Matthieu

Jonckehere, Silvia Kochen, Ernesto Kofman, Nicolás Kreplak, Mauro Lizama, Mar Lucas, Franco Mársico, Antonio Milanese, Axel Molina, Nora Mores, Juan Pablo Pinasco, Ana Quaglino, Fernando Rago, Mauro Ramos, Martín Reydó, Daniela Rodríguez, Mateo Sagardoy, Lucio Santi, Soledad Santini, Fernando Schapachnik, Omar Sued, Pablo Terlisky, Gustavo Tieffenberg, Sebastián Uchitel, Federico Vázquez y Agustín Wydler, por sus contribuciones a los proyectos aquí presentados. También agradecemos especialmente a Jorge Aliaga, Daniel Feierstein, Axel Kicillof y Tomás Pueyo por las valiosas discusiones sobre análisis de datos y cuestiones sociales vinculados al Covid-19 en Argentina y el mundo, que nos ayudaron a entender más cabalmente la problemática de las situaciones planteadas por la pandemia.

Esta iniciativa ha sido parcialmente apoyada por los proyectos "Herramientas matemático-computacionales para el control inteligente de niveles de aislamiento en cada municipio del conurbano de la Provincia de Buenos Aires" y "Proyección de tendencias y evaluación de escenarios de intervención para la epidemia COVID-19 en Argentina mediante Modelado y Simulación Computacional", ambos financiados por el Ministerio de Ciencia, Tecnología e Innovación Productiva de la República Argentina, y por el proyecto "Desarrollo de herramientas de optimización, estadística y ciencia de datos para la gestión, monitoreo y evaluación de políticas públicas", del programa PI-DAE de la Universidad de Buenos Aires. También ha recibido apoyo de ANID PIA/BASAL AFB18000, ISCI, Chile.

Referencias

- [1] A. Cori. Epiestim: a package to estimate time varying reproduction numbers from epidemic curves. *R package version*, páginas 1–1, 2013.
- [2] FCEyN y UBA- CONICET. Duplicación de casos de covid-19 en argentina, 2020. Disponible en https://institutodecalculo.shinyapps.io/tdcovid/.
- [3] S. Figar, V. Pagotto, L. Luna, J. Salto, M. Manslau, y A. Mistchenko. Community-level SARS-CoV-2 seroprevalence survey in urban slum dwellers of Buenos Aires city. *Argentina: a participatory research. medRxiv*, 2020.
- [4] Gobierno de la República Argentina . Fases de administración del aislamiento, 2020. Disponible en urlhttps://www.argentina.gob.ar/coronavirus/aislamiento/fases.

- [5] Gobierno de la República Argentina. Sistema y aplicación cuidar COVID-19, 2020. Disponible en https://www.argentina.gob.ar/jefatura/innovacion-publica/acciones-coronavirus/aplicacion-y-tableros-de-gestion.
- [6] R. K. Gupta, M. Marks, T. H. Samuels, A. Luintel, T. Rampling, H. Chowdhury, M. Quartagno, A. Nair, M. Lipman, I. Abubakar, et al. Systematic evaluation and external validation of 22 prognostic models among hospitalised adults with covid-19: An observational cohort study. European Respiratory Journal, 2020.
- [7] D. W. Hosmer Jr, S. Lemeshow, y R. X. Sturdivant. *Applied logistic regression*, volume 398. John Wiley & Sons, 2013.
- [8] Hospitales en Red, 2020. Disponible en https://hospitales.puntodev.com.ar.
- [9] N. Imai, I. Dorigatti, A. Cori, C. Donnelly, S. Riley, y N. Ferguson. Report 2: Estimating the potential total number of novel coronavirus cases in wuhan city, china. *Imperial College London (22-01-2020)*, 2020. Disponible en https://spiral.imperial.ac.uk/bitstream/10044/1/77150/12/2020-01-22-COVID19-Report-2.pdf.
- [10] Instituto de Cálculo. Estimación del IFR (infection fatality ratio) y de la población infectada por COVID-19 en Argentina. Sitio web del Instituto de Cálculo, UBA/CONICET. Disponible en http://www.ic.fcen.uba.ar/institucional/herramientas/.
- [11] Instituto de Cálculo. Sitio web del Instituto de Cálculo, UBA/CONICET
 Seminarios, 2020. Disponinle en http://www.ic.fcen.uba.ar/actividades-academicas/formacion/seminarios.
- [12] Instituto UBA-CONICET de Ciencias de la Computación. Análisis y control basado en datos COVID-19 en Argentina, 2020. Disponible en http://sedcovid.exp.dc.uba.ar/.
- [13] G. James, D. Witten, T. Hastie, y R. Tibshirani. An Introduction to Statistical Learning: with Applications in R. Springer, New York, NY, 2013.
- [14] Ministerio de Salud. Casos registrados en la república argentina covid-19. Disponible en http://datos.salud.gob.ar/dataset/covid-19-casos-registrados-en-la-republica-argentina.

- [15] Ministerio de Salud de la Provincia de Buenos Aires. Casos covid-19, 2020. Disponible en https://data.buenosaires.gob.ar/dataset/ casos-covid-19.
- [16] Ministerio de Salud de la Provincia de Buenos Aires. Información epidemiológica, 2020. Disponible en https://www.argentina.gob.ar/salud/ coronavirus-COVID-19/sala-situacion.
- [17] Ministerio de Salud de la Provincia de Buenos Aires. Regiones sanitarias, 2020. Disponible en https://www.gba.gob.ar/saludprovincia/regiones_sanitarias.
- [18] Municipio de Marcos Paz. Sitio web municipio, 2020. Disponible en https://www.marcospaz.gov.ar/.
- [19] S. Mustafa. Feature selection using sequential backward method in melanoma recognition. En 13th International Conference on Electronics, Computer and Computation (ICECCO, página 1–4, 2017.
- [20] V. Nagraj, N. Randhawa, F. Campbell, T. Crellen, B. Sudre, y T. Jombart. epicontacts: Handling, visualisation and analysis of epidemiological contacts. F1000Research, 7:566, 2018.
- [21] R. Pastor-Barriuso, B. Perez-Gomez, M. Hernan, M. Perez-Olmeda, R. Yotti, y J. Oteo. SARS-CoV-2 infection fatality risk in a nationwide seroepidemiological study. *Medrxiv*, 2020. Available:.
- [22] T. Pueyo. Coronavirus: The hammer and the dance. *Medium*, 2020. Disponible en https://tomaspueyo.medium.com/coronavirus-the-hammer-and-the-dance-be9337092b56.
- [23] P. Taillandier, B. Gaudou, A. Grignard, Q.-N. Huynh, N. Marilleau, y P. Caillou. Building, composing and experimenting complex spatial models with the GAMA platform. *Geoinformatica*, 23:299–322, 2019.
- [24] E. Vynnycky y R. White. An introduction to infectious disease modelling. *OUP Oxford*, 2010.
- [25] E. Williamson, A. Walker, K. Bhaskaran, S. Bacon, C. Bates, y C. Morton. Factors associated with covid-19-related death using opensafely. *Nature*, 584:430–436, 2020.
- [26] World Health Organization. Coronavirus disease (COVID-19). Octubre 2020. Disponible en https://www.who.int/

- docs/default-source/coronaviruse/situation-reports/20201005-weekly-epi-update-8.pdf.
- [27] L. Wynants, B. Van Calster, G. Collins, R. Riley, G. Heinze, y E. Schuit. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. *BMJ*, 369:1328, 2020.
- [28] A. Yacobitti, A. Lisandro Otero, A. Valeria Doldan Arruabarrena, J. Arano, S. Lage, y M. Silberman. Clinical characteristics of vulnerable population hospitalized with diagnosed of COVID-19 in the Southeast Region of Greater Buenos Aires, Argentina. J Gen Intern Med, 2020.
- [29] L. Yan, H.-T. Zhang, J. Goncalves, Y. Xiao, M. Wang, y Y. Guo. An interpretable mortality prediction model for covid-19 patients. *Nature Machine Intelligence*, 2:283–288, 2020.

Análisis de movilidad durante la pandemia de COVID-19 en las regiones del Biobío y Ñuble utilizando datos de teléfonos móviles

Sebastián Cancino * †

Helen De la Fuente *

Rodrigo Victoriano * †

Sebastián Astroza * †

Julio Covarrubias

Matías Durán

Juan Antonio Carrasco * †

Resumen

Mientras diferentes especialistas de la salud se encuentran en la búsqueda de un tratamiento farmacéutico efectivo contra el COVID-19, las autoridades de cada país se han enfocado en el diseño y la aplicación de una serie de medidas de confinamiento con el fin de reducir la movilidad. Estas medidas han presentado distintos niveles de efectividad en reducir la movilidad y los niveles de contagio, manifestando la necesidad de estudiar caso a caso dependiendo de las características de cada territorio. Este artículo tiene como objetivo describir los cambios en los patrones de movilidad durante la pandemia de COVID-19 para las regiones del Biobío y Nuble a través del uso de información proveniente de teléfonos móviles. Específicamente, se consideran los patrones de movilidad de las áreas urbanas del Gran Concepción y Chillán - Chillán Viejo, además de los patrones interurbanos e interregionales en ambas regiones. El análisis considera las variaciones que han existido en los patrones de viajes ante la aplicación de diversas medidas de confinamiento y reducción de la movilidad aplicadas entre marzo y octubre de 2020, lo que permite evaluar su efectividad y variabilidad a nivel temporal y territorial.

Palabras Clave: Movilidad, Confinamiento, Pandemia, COVID-19, Datos telefónicos.

^{*}Instituto de Sistemas Complejos de Ingeniería (ISCI), Santiago, Chile.

[†]Facultad de Ingeniería, Universidad de Concepción, Chile.

[‡]Facultad de Arquitectura, Urbanismo y Geografía, Universidad de Concepción, Chile.

[§]Empresa Nacional de Telecomunicaciones, Entel, Chile.

1. Introducción

Ante la ausencia de un tratamiento farmacéutico efectivo contra el COVID-19, las autoridades en todo el mundo se han enfocado en el diseño y la aplicación de una serie de medidas no-farmacéuticas para enfrentar la pandemia. Uno de los elementos centrales considerado en las políticas de respuesta ha sido el control de las aglomeraciones y la movilidad para buscar reducir la expansión del COVID-19 [6, 13]. También, se ha mostrado que las variaciones en la movilidad de las personas son de alta relevancia en el proceso de propagación del virus no sólo a escalas interurbanas [7, 12], sino que también a escalas más locales y urbanas [5, 15].

Con el fin de reducir los niveles de contagio del COVID-19, en diversos países se han implementado una serie de medidas de confinamiento y reducción de la movilidad, entre las cuales destaca el cierre de actividades educacionales y laborales [2, 11], los cordones sanitarios [9] y las cuarentenas [11]. Estas medidas han presentado distintos niveles de efectividad en reducir la movilidad y los niveles de contagio dependiendo del contexto nacional o local [11, 16]. Esta diversidad de resultados genera la necesidad de estudiar la relación entre movilidad y propagación del virus para cada territorio específico. Más aún, se ha observado que las medidas de reducción de movilidad tienen efectos distintos dependiendo de los grupos de población en los que se implementan [3, 11]. En el caso chileno, este efecto se ha manifestado principalmente a través de las diferencias en niveles de ingreso económico [5, 10], consistente con los altos niveles de segregación socio-espacial en Chile, lo que también realza el relevancia de los análisis focalizados territorialmente.

Considerando la importancia que han tomado los análisis de niveles de movilidad para el entendimiento del control efectivo del COVID-19, diversos estudios se han apoyado en una variedad de fuentes de datos de movilidad, entre los cuales destacan el uso de conteos vehiculares [6], el uso de información proveniente de motores de búsqueda y aplicaciones de redes sociales [3, 7, 9, 13] y el uso de datos de movilidad basados en información de teléfonos móviles [1, 12, 14]. Esta última fuente de información de movilidad ha sido altamente recomendada para estos propósitos. Dado su amplio alcance territorial y temporal, posee altos niveles de desagregación, pero sin infringir en la privacidad personal [4].

La aplicación de medidas de confinamiento y reducción de la movilidad en Chile se ha realizado de manera focalizada en el tiempo y el espacio, siendo guiadas por las oleadas de contagios que han existido en las distintas comunas del país de acuerdo al plan nacional de acción ante el COVID-19 [8]. De manera similar a otras regiones del país, las ciudades de las regiones del Ñuble y del Biobío han pasado por la aplicación de diversas de estas medidas sanitarias de manera heterogénea en el tiempo, donde principalmente han destacado la aplicación de cuarentenas y cordones sanitarios. Sin embargo, si bien se han hecho análisis que relacionan los niveles de movilidad con la aplicación de políticas públicas en Chile [7], incluso utilizando datos detallados de movilidad provenientes de teléfonos móviles [5, 10], hasta ahora no se han hecho análisis detallados de la situación específica de las regiones del Ñuble y del Biobío.

Este artículo tiene como objetivo describir los cambios en los patrones de movilidad durante la pandemia de COVID-19 en las regiones del Biobío y Ñuble a través del uso de información proveniente de teléfonos móviles. Específicamente, se consideran los patrones de movilidad de las áreas urbanas del Gran Concepción y Chillán - Chillán Viejo, además de los patrones interurbanos e interregionales en ambas regiones. Dentro del análisis se consideran, también, las variaciones que han existido en los patrones de viajes ante la aplicación de una diversidad de medidas de confinamiento y reducción de la movilidad que han sido aplicadas entre marzo y octubre de 2020, de manera de evaluar la efectividad y variabilidad de dichas medidas a escala temporal y a nivel territorial.

El resto del trabajo se estructura en las secciones descritas a continuación. La sección "Metodología" describe las fuentes de información, que en gran parte corresponden al uso de datos de teléfonos móviles, y el tratamiento necesario para su corrección y análisis. La sección "Resultados" presenta los análisis de patrones de movilidad realizados a nivel intercomunal e interregional y su relación con las políticas de confinamiento y reducción de la movilidad, aplicadas en los distintos períodos y territorios. Finalmente, la sección "Conclusiones y Discusión" presenta un resumen y discusión de los procedimientos y resultados obtenidos.

2. Metodología

Este estudio analiza el impacto de las diferentes medidas sanitarias tomadas por el Gobierno de Chile para responder a la crisis sanitaria producto de la pandemia de COVID 19 en el país. Estas medidas, tales como cuarentenas y cordones sanitarios, están enfocadas a reducir la movilidad de las personas, y por tanto, se analizan en este ámbito haciendo uso de datos provenientes de infraestructura de telecomunicaciones. El día 20 de julio del año 2020 las

autoridades sanitarias anunciaron el Plan Paso a Paso en el marco del plan de acción frente al COVID 19 del Gobierno de Chile¹, el cual pretende guiar

un desconfinamiento gradual de las comunas del país. Este plan se divide en

cinco fases y sus principales pasos son:

■ Paso 1 (Cuarentena): restricción en los desplazamientos, sólo 2 permisos para salir a la semana.

- Paso 2 (Transición): restricción en los desplazamientos durante fines de semana y festivos.
- Paso 3 (Preparación): se permiten viajes interregionales a comunas en Pasos 3, 4 y 5.
- Paso 4 (Apertura inicial): se permite retomar ciertas actividades que significan menor riesgo de contagio. Se evitan al máximo las aglomeraciones y contacto físico entre los asistentes. Traslados interregionales sólo entre localidades que estén en los Pasos 3, 4 y 5.
- Paso 5 (Apertura avanzada): Participar en eventos, reuniones sociales y recreativas de máximo 200 personas en lugares abiertos y 100 personas en lugares cerrados. Traslados interregionales sólo entre localidades que estén en los Pasos 3, 4 y 5.

Para estimar la movilidad se analizan los viajes de los habitantes de las regiones de Biobío y Ñuble durante el período de pandemia de COVID-19 desde marzo hasta octubre del 2020; se utilizaron datos estadísticos e información del uso de infraestructura de telecomunicaciones agrupadas a nivel de zona censal. Los datos de movilidad son recolectados por Entel Ocean y están disponibles públicamente a través del Ministerio de Ciencias de Chile ². La metodología para analizar los datos se detalla en [5]. Por otra parte, se estudia la movilidad interregional en el período desde marzo hasta julio de 2020. Una zona censal tiene alrededor de 2000-3000 habitantes y permite garantizar la anonimidad de la información de los individuos. Estos datos permiten entender la zona de hogar, en donde las personas se encuentran frecuentemente en horarios de no trabajo. Para cada día laboral (lunes a viernes), se determina el flujo desde cada zona hogar a otras zonas, durante horarios de trabajo en dos bloques: 10 hrs a 13 hrs (AM) y 14 hrs a 17 hrs (PM). Esta determinación del flujo considera el tiempo de permanencia de cada teléfono en cada zona censal, y

¹https://www.gob.cl/pasoapaso/

²Correspondiente al producto 51 del repositorio del Ministerio de Ciencia. Disponible en https://github.com/MinCiencia/Datos-COVID19

para asignarlos se utiliza la moda geográfica durante cada bloque horario. Para determinar la zona hogar se considera la moda geográfica mensual, durante un tiempo de pernoctación que va desde las 20 a las 23 hrs. Los flujos detectados pueden ser dentro de la misma comuna o hacia otras comunas (pero fuera de la zona hogar) y se interpretan como asociados a actividades laborales. Luego se toman promedios semanales para estimar la movilidad de cada comuna. Para el cálculo de la movilidad intraregión por comunas se utilizó una matriz origendestino para cada región, que contiene los flujos promedio semanales entre cada par de zonas censales de cada región, sin considerar flujos interregionales. Por el contrario, para el cálculo de la movilidad interregión, la matriz de origendestino utilizada contiene tanto los flujos interregionales como intraregionales.

2.1. Cálculo de la movilidad intraregión por comunas

A continuación, se detalla la metodología y las fórmulas utilizadas para el cálculo de la movilidad intraregional por comunas.

Notación

- i: zona censal de hogar
- j: zona censal de destino
- k: semana
- c: comuna
- r: región

El flujo promedio semanal f_{ijk} de teléfonos que van desde la i a la zona j en la semana k se estima tomando el máximo de teléfonos detectados entre ambas zonas para ambos bloques horarios $T1_{ijk}$ y $T2_{ijk}$ (Ecuación 1).

$$f_{ijk} = max \left(T1_{ijk}, T2_{ijk} \right) \tag{1}$$

Luego, se calcula el flujo semanal de salida fs_{ck} (i.e. la cantidad de celulares que son detectados fuera de su zona censal) para cada comuna como se muestra en la Ecuación 2.

$$fs_{ck} = \sum_{j} \sum_{i \in c, i \neq j} f_{ijk} \tag{2}$$

Por otra parte, el flujo hogar fh_{ck} (i.e cantidad de celulares que no son detectados fuera de su zona hogar) se calcula de la siguiente forma:

J. Covarrubias, M. Durán y J. Carrasco

$$fh_{ck} = \sum_{i \in c} f_{iik} \tag{3}$$

Notar que la suma de fs_{ck} y fh_{ck} , T_{ck} (Ecuación 4) representa la estimación del total de teléfonos detectados con zona hogar en la comuna c para la semana k. De esta manera, es posible calcular el flujo porcentual de salida s_{ck} como se muestra en la Ecuación 5.

$$T_{ck} = f s_{ck} + f h_{ck} \tag{4}$$

$$s_{ck} = \frac{f s_{ck}}{T_{ck}} \tag{5}$$

Para calcular la variación \overline{v}_{ck} de movilidad de cada comuna se considera su movilidad de salida porcentual durante un período base representada por b_c y se expresa la movilidad de cada semana k en términos relativos. Luego se resta 1 para expresar la movilidad en términos de disminución (Ecuación 6).

$$\overline{v}_{ck} = \frac{s_{ck}}{b_c} - 1 \tag{6}$$

Se consideran las 2 primeras semanas de marzo del 2020, antes de la declaración de la Fase 4 de pandemia por COVID-19 en Chile, como semanas "base". Así se tiene una aproximación para la movilidad usual de cada zona. Dicha elección se basó en que los casos detectados en aquella fecha correspondían a casos aislados y trazables.

2.2. Corrección por pérdida de teléfonos celulares

Producto de la pandemia de COVID-19, en Chile, como en muchas partes del mundo, el teletrabajo y la educación a distancia han tomado importancia como medidas para evitar la propagación del virus. Como consecuencia, aquellas personas con la posibilidad de trabajar o estudiar desde casa, comenzaron a utilizar redes Wi-Fi, en desmedro de las redes móviles, haciendo imposible rastrearlas a través de la infraestructura de telecomunicaciones. De esta forma, es importante notar que \overline{v}_{ck} representa una cota superior de la movilidad, ya que la variable fh_{ck} se encuentra subestimada, sobre todo en situaciones de confinamiento total. Para corregir esto, se calcula una cota inferior de movilidad, tomando como referencia la cantidad de teléfonos detectados T_{ck} en promedio para las semanas base, denotada como r_c . La variable ΔO (Ecuación 7) representa la diferencia de observaciones entre el período base y una semana k en particular. Notar que la corrección sólo aplica cuando las observaciones del período base son mayores a las de la semana en cuestión.

$$\Delta O = \max\left(r_c - T_{ck}, 0\right) \tag{7}$$

Luego, se calcula la cota inferior $v.lb_{ck}$ como se muestra en la Ecuación 8

$$\underline{v}_{ck} = \left(\frac{T_{ck}}{T_{ck} + \Delta O}\right) \times \overline{v}_{ck} \tag{8}$$

Finalmente, la variación de movilidad estimada para una comuna c en la semana k, con respecto al período de referencia, está dada por v_{ck} (Ecuación 9).

$$v_{ck} = \frac{\overline{v}_{ck} + \underline{v}_{ck}}{2} \tag{9}$$

2.3. Movilidad interregión

Con el objetivo de evaluar la efectividad de los cordones sanitarios impuestos por la autoridad en la interregión Biobío - Ñuble, se construye el siguiente indicador de los viajes desde una región a otra a nivel comunal durante cuatro feriados importantes en los que existieron cordones sanitarios entre ambas regiones: Semana Santa (10 al 12 de abril), Glorias Navales (12 de mayo), San Pedro y San Pablo (26 de junio) y Virgen del Carmen (16 de julio). Dado que estas medidas han sido instauradas principalmente durante fines de semana largos, se consideran para cada semana k, solamente los días viernes, sábado y domingo de dicha semana.

El flujo interregional desde una comuna c hacia una región r se calcula como la suma de los flujos entre las zonas i pertenecientes a la comuna c y las zonas j pertenecientes a la región r, tal como se muestra en la Ecuación 10.

$$fr_{crk} = \sum_{i \in c} \sum_{j \in r} f_{ijk} \tag{10}$$

Luego, se calcula el flujo porcentual en relación al total de teléfonos detectados en la comuna c durante la semana k. Con el objetivo de hacer comparables los flujos porcentuales de salida entre las distintas comunas, se aplica un factor de ajuste por población, tal como se muestra en la Ecuación 11.

$$sr_{crk} = \frac{fr_{crk}}{T_{ck}} \times p_c, \tag{11}$$

Finalmente, este indicador se agrega para los pares de regiones de interés, sumando sobre las comunas correspondientes, y se normaliza por el total de viajes estimados para el período base, para así expresar la movilidad en términos porcentuales con respecto a dicho período.

2.4. Supuestos

El cálculo de la movilidad como aquí se describe requiere asumir algunos supuestos importantes. En primer lugar, la escala en la que los datos son capturados no permite la detección de viajes dentro de una misma zona censal. Sin embargo, debido a que la metodología está diseñada para detectar viajes de trabajo, se asume que la mayoría de personas trabajan en una zona censal distinta a la de su zona censal de hogar. Por otro lado, el tomar el máximo entre los flujo T1 y T2 (mañana y tarde), entrega una mejor caracterización de los flujos siempre y cuando la mayoría de personas trabajen en una jornada laboral completa. Finalmente, al realizar un ajuste de observaciones con respecto a las dos primeras semanas de marzo implica asumir que no ha habido grandes movimientos (cambios de domicilio) desde esta fecha. Esto puede no ser cierto, puesto que durante estas semanas aún podían existir personas de vacaciones, haciendo aumentar las observaciones de referencia en sitios turísticos.

3. Resultados

En esta sección se muestra la evolución temporal de la movilidad a nivel comunal en el Área del Gran Concepción y Región del Ñuble y a nivel interregional entre Ñuble y Biobío.

3.1. Movilidad a nivel comunal en el Gran Concepción

Las Figuras 1 y 2 presentan la movilidad de las comunas del Gran Concepción entre los meses de marzo y octubre, y separadas entre comunas centrales y comunas periféricas respectivamente. En el gráfico es posible apreciar que la mayor baja de movilidad ocurre luego de la declaración de Fase 4 de la pandemia de COVID-19 en el país, cuyo principal hito corresponde al cierre de establecimientos educacionales, con lo que la movilidad se ve reducida en valores que van desde un 29 % en Talcahuano, hasta un 37 % en Chiguayante. Luego de esto, la autoridad sanitaria decretó la medida de cuarentena total en las comunas de Hualpén y San Pedro de la Paz la que entró en vigencia el día lunes 6 de abril y se extendió por dos semanas marcadas por círculos rojos en la Figura 1. La medida también incluyó un cordón sanitario alrededor de ambas comunas. Como consecuencia, los niveles de movilidad fueron los más bajos registrados durante el período de estudio, con una reducción de un $40\,\%$ para San Pedro de la Paz y un 34% en Hualpén. El valor de San Pedro de la Paz durante su primera semana de cuarentena corresponde, también, a la movilidad más baja registrada de alguna comuna del Gran Concepción entre los meses de marzo y octubre. Durante estas dos semanas, las demás comunas del Gran Concepción presentan pequeñas tendencias al alza y valores de movilidad mayores a los alcanzados en la semana inmediatamente después a la declaración de Fase 4.



Figura 1: Evolución temporal de movilidad (flujo de salidas de la zona hogar) en comunas centrales del Gran Concepción.

Durante el mes de mayo se observa una tendencia al alza en todas las comunas del Gran Concepción, en mayor medida en las comunas periféricas. En el mes de junio es posible apreciar una leve baja de movilidad en las primeras semanas; la tendencia general es al alza si se considera la totalidad del mes. Durante este período no existieron medidas sanitarias dictadas por la autoridad, al menos durante los días de semana.

Los meses de julio y agosto presentan una notoria tendencia al alza en todas las comunas del Gran Concepción. Como respuesta, la autoridad implementó una serie de cordones sanitarios en distintos puntos de la ciudad con el objetivo de reducir la movilidad. Los primeros cordones sanitarios fueron implementados en las comunas de Coronel y Lota durante el mes de julio, mientras que durante el mes de agosto entran en vigencia el cordón sanitario que rodeó la comuna de Talcahuano y los cordones sanitarios entre las comunas de Concepción y Penco, y entre las comunas de San Pedro de la Paz y Coronel. Si bien estos cordones sanitarios producen una leve reducción de movilidad en las comunas, sus efectos son momentáneos y no logran revertir la tendencia al alza.

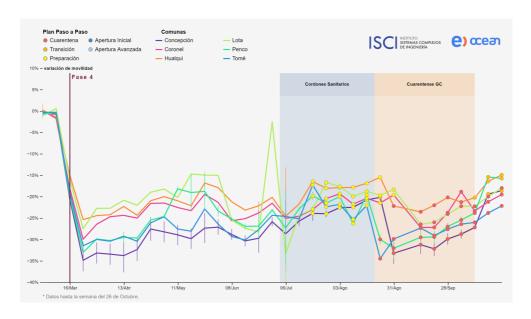


Figura 2: Evolución temporal de movilidad (flujo de salidas de la zona hogar) en comunas periféricas del Gran Concepción.

Como consecuencia del aumento de contagios, durante la cuarta semana de agosto las comunas de Penco y Tomé retrocedieron a la Fase 1 (Cuarentena) del Plan Paso a Paso, mientras que las comunas de Concepción y Hualpén retrocedieron a la Fase 3 (Transición). El efecto de la cuarentena en Penco y Tomé es inmediato, llegando a niveles de reducción de movilidad de 30 % y 34 % durante la primera semana de cuarentena. Por otra parte, las comunas de Concepción y Hualpén no muestran una disminución de movilidad en esta semana, principalmente debido a que la fase de transición se enfoca a disminuir la movilidad durante los fines de semana. Durante esta semana la mayoría de las comunas del Gran Concepción alcanzan los valores más altos de movilidad desde el inicio de la pandemia, destacando Hualpén y Hualquí, ambos con sólo un 15 % de reducción con respecto a las dos primeras semanas de marzo.

En la semana siguiente retroceden a cuarentena total todas las comunas del Gran Concepción, con excepción de Lota y Coronel donde la cuarentena comenzó a regir una semana después. Esta medida tiene un efecto inmediato en la movilidad. Durante la primera semana de aplicación de la cuarentena las comunas de Concepción, Talcahuano, Chiguayante, Coronel y Lota registran sus niveles de movilidad más bajos desde el inicio de la pandemia, mientras que en Hualpén y San Pedro de la Paz la reducción de movilidad no logra superar los niveles alcanzados durante la primera cuarentena de marzo en ambas comunas. En general los impactos de la cuarentena son desiguales territorial-

mente, ya que comunas centrales presentan impactos importantes, tales como Concepción, Chiguayante, Talcahuano, San Pedro, Hualpén, lo que contrasta con Penco, Tomé, Lota y parte de Coronel, que no tienen un impacto significativo al comparar con las semanas anteriores a la cuarentena. Durante el mes de septiembre y las primeras semanas de agosto, y con todas las comunas del Gran Concepción aún bajo cuarentena total, la movilidad muestra una sostenida tendencia al alza, particularmente en las comunas periféricas del Gran Concepción. En la cuarta semana de octubre, las principales comunas del Gran Concepción entran en Fase 2 (Transición) llevando la movilidad hasta los niveles que existían previo a la cuarentena, y en algunos casos mayores, continuando en aumento hasta el final del mes.



Figura 3: Evolución temporal de movilidad (flujo de salidas de la zona hogar) en distintas comunas de la Región de Ñuble.

3.2. Movilidad a nivel comunal en la Región de Ñuble

Para analizar la movilidad en la Región de Ñuble, se consideró una escala temporal idéntica a la Región del Biobío, es decir, a inicios de la Fase 4. Esto involucra como hito principal el cierre de los establecimientos educacionales, trayendo consigo una fuerte reducción de la movilidad en las principales comunas de la región que varía del 25 % al 45 %, como señala la Figura 3.

Posteriormente, durante el primer periodo de cuarentena en Chillán y Chillán Viejo instaurado en el mes de abril y en conjunto con la medida de cordón sanitario, se mantiene la disminución de la movilidad, pero con un

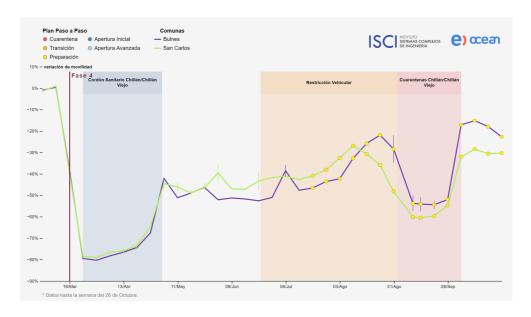


Figura 4: Evolución temporal de movilidad hacia la intercomuna de Chillán y Chillán Viejo desde distintas comunas de la Región de Ñuble.

leve crecimiento en el transcurso de las semanas. Una vez terminada esta cuarentena, entre los meses de mayo y junio, la movilidad llega rápidamente a niveles cercanos al $20\,\%$ para Chillán, Chillán Viejo y San Carlos y al $30\,\%$ para Bulnes.

Respecto a la restricción vehicular aplicada en los meses de junio, julio y agosto, se observa una leve reducción de los flujos de movilidad al comenzar esta medida. Sin embargo, iniciado julio, se visualiza una tendencia al alza estadísticamente similar a meses sin ninguna medida. Es probable que el efecto de esta medida no se refleje en los datos analizados debido a que la metodología aplicada mide primordialmente los viajes laborales en días de semana y la restricción vehicular se aplicó fuera de los horarios punta. Un segundo aspecto a destacar en este periodo es el gran aumento de movilidad en Bulnes, lo que podría estar asociado a un eventual paso a la Fase 4 (Apertura Inicial) del Plan Paso a Paso, el que se vio mermado debido a un aumento de brotes en esta comuna, provocando un estancamiento en la Fase 3 (Preparación).

En el segundo período de cuarentena en las ciudades de Chillán y Chillán Viejo, la movilidad se acerca al 30% de reducción máxima, cifra menor a la observada en el primer periodo de cuarentena (aproximadamente de un 40%). En contraste, se observa drásticamente una tendencia al alza una vez que las comunas avanzan al desconfinamiento.

Al observar en detalle la efectividad del cordón sanitario y cuarentenas en la intercomuna de Chillán y Chillán Viejo (Figura 4), es posible notar que

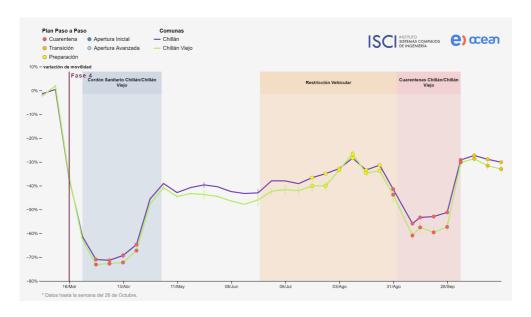


Figura 5: Evolución temporal de movilidad hacia el centro de Chillán desde el resto de Chillán y Chillán Viejo.

estas medidas ayudan a mantener la reducción de viajes hacia la intercomuna, así, ciudades como Bulnes y San Carlos se acercan a valores del $50\,\%$ o más. Sin embargo, al finalizar las cuarentenas coincide con un aumento cada vez mayor de la movilidad, principalmente por la dependencia del sector terciario presente en la ciudad de Chillán.

La Figura 5 refleja el desplazamiento interno hacia el centro de Chillán, el cual muestra un aumento de la movilidad registrándose el período más alto a mediados de agosto, con valores que rondan el 25 % de reducción en comparación a las semanas base. Por otra parte, la segunda cuarentena genera un impacto importante, con registros mayores al 50 % de disminución de flujos, pero que no alcanzan al periodo de abril a mayo, en que existió cuarentena y cordón sanitario de manera paralela.

3.3. Movilidad en la interregion Biobío - Ñuble

El grado de movilidad en la interregión se muestra en las Figura 6 y 7, en donde se analizaron las comunas más próximas a los límites regionales y cercanas a las capitales de cada región. En cuanto a la escala temporal, se analizan los principales feriados y festivos hasta el mes de julio, en los cuales se implementaron cordones sanitarios en las ciudades capitales de cada región (Gran Chillán y Gran Concepción).

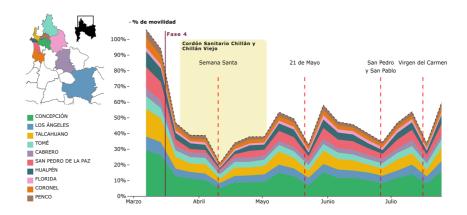


Figura 6: Evolución temporal de movilidad (flujo de salidas de la zona hogar) desde distintas comunas de la Región del Biobío hacia la Región de Ñuble considerando fines de semana.

La Figura 6 muestra que, a modo general en cada fecha festiva, la movilidad disminuye considerablemente desde la región del Biobío hacia Nuble y desde la región del Biobío hacia Nuble. Así mismo, luego del inicio de la Fase 4 de la pandemia, la movilidad tiende a estabilizarse en torno a un $40\,\%$ con respecto al periodo base. Durante el fin de semana del feriado de semana santa, la movilidad llega aproximadamente al 20% producto de la implementación del cordón sanitario. Luego de esa fecha, la movilidad comienza a aumentar, en especial luego del fin del cordón sanitario de Chillán y Chillán Viejo, llegando a niveles cercanos al 55 %. El siguiente cordón sanitario, correspondiente al feriado del 21 de mayo, logra disminuir la movilidad hasta llevarla aproximadamente al nivel del 35 %. El cordón sanitario correspondiente a la celebración de San Pedro y San Pablo produce efectos menores reducción de movilidad (en relación a la tendencia de las semanas anteriores), llegando a niveles de aproximadamente un 40 %. Finalmente, durante el fin de semana correspondiente a la celebración de la Virgen del Carmen, la movilidad vuelve a disminuir producto del cordón sanitario, llegando hasta aproximadamente un 35 %. La comuna que más contribuye al flujo interregional desde Biobío a Nuble es Concepción, la que también presenta los mayores impactos al aplicarse los cordones sanitarios. Otras comunas del Gran Concepción, tales como Talcahuano y Tomé, también presentan disminuciones de movilidad en estas fechas. Por otro lado, la segunda comuna que más contribuye al flujo interregional es Los Ángeles, la que no tuvo cordones sanitarios en estas fechas y no presenta disminuciones notorias de movilidad en los mencionados fines de semana.

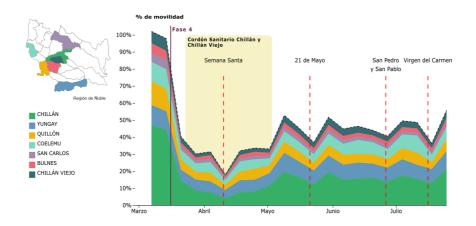


Figura 7: Evolución temporal de movilidad (flujo de salidas de la zona hogar) desde distintas comunas de la Región de Ñuble hacia la Región del Biobío considerando fines de semana.

La Figura 7 muestra la movilidad desde comunas de la Región de Ñuble hacia la Región del Biobío. De manera similar a la gráfica anterior, se evidencia una rápida disminución de movilidad interregional luego del inicio de la Fase 4 de la pandemia, llegando a valores en torno al 30 % de movilidad. En el fin de semana del feriado de semana santa, la movilidad se reduce hasta un 20 % aproximadamente. Luego del fin del cordón sanitario en Chillán y Chillán Viejo, la movilidad aumenta llegando al 50 %, mientras que en el fin de semana largo correspondiente al 21 de mayo, esta disminuye aproximadamente hasta el 35 %. Luego de esto, la movilidad se mueve en torno a valores cercanos al 40 %, con una leve disminución en el fin de semana correspondiente a la celebración de San Pedro y San Pablo. Para el último festivo estudiado, correspondiente a la celebración de la Virgen del Carmen, la movilidad disminuye hasta un 35 % aproximadamente. La comuna que más aporta al flujo interregional es Chillán, seguido de Yungay y Quillón.

4. Conclusiones y Discusión

Luego de comparar cada una de las medidas sanitarias se observa que los impactos más significativos en la movilidad de las personas están relacionados con el cierre de establecimientos educacionales. Además, se experimentan bajas importantes de movilidad al comienzo de cada cuarentena, perdiendo su efectividad con el paso del tiempo. Los efectos de la cuarentena para el Gran Concepción son notorios en comunas como San Pedro de la Paz o Hualpén, pero no logran mantener una baja de movilidad sostenida en el tiempo, pues sólo durante el periodo en que las cuarentenas son aplicadas. Si bien las causas de este comportamiento en la movilidad de las personas a lo largo de la cuarentena se encuentran fuera del alcance de este estudio, es probable que una multiplicidad de factores contribuya a ello, tales como la necesidad de las personas de abastecerse, imposibilidad de trabajar desde el hogar, o incluso la fatiga y estrés que producen los períodos largos de cuarentena.

Al analizar la movilidad entre las regiones de Nuble y Biobío se muestran un efectos notorios de los cordones sanitarios en la reducción de movilidad interregional durante fines de semana largos, principalmente entre las comunas de Concepción y Chillán. La excepción son las comunas que no tienen un cordón sanitario en estas fechas, ya que no presentan disminuciones significativas de movilidad. Por otra parte, los cordones sanitarios implementados durante días de semana entre comunas del Gran Concepción no muestran efectos claros en reducción de la movilidad.

En cuanto al comportamiento de los datos durante la restricción vehicular aplicada en la Región de Ñuble, no se observa una disminución de la movilidad, ya sea por los flujos registrados hacia la intercomuna Chillán-Chillán Viejo como de las personas que se desplazan dentro de estas mismas comunas.

Para Chillán y Chillán Viejo, al ser comunas conurbadas, ambas ciudades presentan efectos muy similares mostrando reducciones fuertes en los períodos de cuarentena. Por otra parte, estos impactos en la disminución de la movilidad se ven también reflejados en sus comunas aledañas como San Carlos y Bulnes, debido a los servicios ofrecidos por la intercomuna.

Es necesario destacar que los métodos presentados en este estudio poseen limitaciones principalmente asociadas a los datos de movilidad utilizados. En primer lugar, la muestra de teléfonos corresponde a una única compañía telefónica, lo que puede inducir ciertos sesgos (socioeconómicos y/o territoriales) dependiendo de las características de los segmentos de clientes a los que apun-

ta dicha compañía. Por otra parte, la metodología utilizada está diseñada para captar viajes de trabajo. Al considerar la moda geográfica para determinar los movimientos de las personas, viajes cortos (en términos temporales) durante los bloques de estudio no son considerados dentro de la métrica de movilidad.

A pesar de las limitaciones mencionadas, los resultados y análisis presentados en este artículo ponen de manifiesto la relevancia de contar con datos de telefonía móvil y el desarrollo de metodologías de análisis que apoyen la toma de decisiones en circunstancias como la pandemia del COVID 19. En particular, los datos de telefonía móvil permiten un grado de granularidad espacio-temporal que facilita el entendimiento de los efectos de las distintas medidas sanitarias en la movilidad de las personas. Lo anterior se ha manifestado en el trabajo conjunto del equipo de Movilidad de la Iniciativa COVID Analytics del Instituto de Sistemas Complejos de Ingeniería con autoridades nacionales y regionales. En el caso particular de las regiones de Ñuble y Biobío, estos análisis han sido relevantes para la implementación y evaluación de medidas sanitarias por parte de las Intendencias Regionales, a través del vínculo con la Secretaría Regional del Ministerio de Ciencia, Tecnología e Innovación de la zona Centro-Sur.

Agradecimientos: Los autores agradecen apoyo de ANID PIA/BASAL AFB180003 y a Digital Entel Ocean por facilitar el acceso a los datos de movilidad y el desarrollo metodológico del equipo de Movilidad de la Iniciativa COVID Analytics del Instituto de Sistemas Complejos de Ingeniería, disponible en https://covidanalytics.isci.cl/movilidad/equipo/. Esta investigación ha sido financiada por el Instituto Sistemas Complejos de Ingeniería (ANID PIA AFB180003).

Referencias

- [1] H. Allcott, L. Boxell, J. Conway, B. Ferguson, M. Gentzkow, y B. Goldman. Economic and health impacts of social distancing policies during the coronavirus pandemic. *SSRN Electronic Journal*, 2020.
- [2] K. A. Auger, S. S. Shah, T. Richardson, D. Hartley, M. Hall, A. Warniment, K. Timmons, D. Bosse, S. A. Ferris, P. W. Brady, A. C. Schondelmeyer, y J. E. Thomson. Association Between Statewide School Closure and COVID-19 Incidence and Mortality in the US. *JAMA*, 324(9):859–870, 09 2020.

J. Covarrubias, M. Durán y J. Carrasco COVID-19 EN LAS REGIONES DEL BIOBÍO Y ÑUBLE

- [3] G. Bonaccorsi, F. Pierri, M. Cinelli, A. Flori, A. Galeazzi, F. Porcelli, A. L. Schmidt, C. M. Valensise, A. Scala, W. Quattrociocchi, y F. Pam-Economic and social consequences of human mobility restrictions under covid-19. Proceedings of the National Academy of Sciences, 117(27):15530-15535, 2020.
- [4] C. O. Buckee, S. Balsari, J. Chan, M. Crosas, F. Dominici, U. Gasser, Y. H. Grad, B. Grenfell, M. E. Halloran, M. U. G. Kraemer, M. Lipsitch, C. J. E. Metcalf, L. A. Meyers, T. A. Perkins, M. Santillana, S. V. Scarpino, C. Viboud, A. Wesolowski, y A. Schroeder. Aggregated mobility data could help fight covid-19. Science, 368(6487):145–146, 2020.
- [5] A. Carranza, M. Goic, E. Lara, M. Olivares, G. Y. Weintraub, J. Covarrubia, C. Escobedo, N. Jara, y L. J. Basso. The social divide of social distancing: Lockdowns in santiago during the COVID-19 pandemic. SSRN Electronic Journal, 2020.
- [6] A. Cartenì, L. Di Francesco, y M. Martino. How mobility habits influenced the spread of the covid-19 pandemic: Results from the italian case study. Science of The Total Environment, 741:140489, 2020.
- [7] C. Cuadrado, M. J. Monsalves, J. Gajardo, M. P. Bertoglia, M. Najera, T. Alfaro, M. Canals, J. Kaufmann, y S. Peña. Impact of small-area lockdowns for the control of the covid-19 pandemic. medRxiv, 2020.
- [8] G. de Chile. Plan de acción paso a paso nos cuidamos, 2020.
- [9] H. Fang, L. Wang, y Y. Yang. Human mobility restrictions and the spread of the novel coronavirus (2019-ncov) in china. Journal of Public Economics, 191:104272, 2020.
- [10] N. Gozzi, M. Tizzoni, M. Chinazzi, L. Ferres, A. Vespignani, y N. Perra. Estimating the effect of social inequalities in the mitigation of covid-19 across communities in santiago de chile. medRxiv, 2020.
- [11] S. Hsiang, D. Allen, S. Annan-Phan, K. Bell, I. Bolliger, T. Chong, H. Druckenmiller, L. Y. Huang, A. Hultgren, E. Krasovich, P. Lau, J. Lee, E. Rolf, J. Tseng, y T. Wu. The effect of large-scale anti-contagion policies on the covid-19 pandemic. *Nature*, 584(7820):262–267, Aug 2020.
- [12] J. S. Jia, X. Lu, Y. Yuan, G. Xu, J. Jia, y N. A. Christakis. Population flow drives spatio-temporal distribution of covid-19 in china. *Nature*, 582(7812):389–394, Jun 2020.

- [13] M. U. G. Kraemer, C.-H. Yang, B. Gutierrez, C.-H. Wu, B. Klein, D. M. Pigott, L. du Plessis, N. R. Faria, R. Li, W. P. Hanage, J. S. Brownstein, M. Layan, A. Vespignani, H. Tian, C. Dye, O. G. Pybus, y S. V. Scarpino. The effect of human mobility and control measures on the covid-19 epidemic in china. *Science*, 368(6490):493–497, 2020.
- [14] S. Lai, I. Bogoch, N. Ruktanonchai, A. Watts, X. Lu, W. Yang, H. Yu, K. Khan, y A. J. Tatem. Assessing spread risk of wuhan novel coronavirus within and beyond china, january-april 2020: a travel network-based modelling study. medRxiv: the preprint server for health sciences, página 2020.02.04.20020479, Feb 2020.
- [15] H. Tian, Y. Liu, Y. Li, C.-H. Wu, B. Chen, M. U. G. Kraemer, B. Li, J. Cai, B. Xu, Q. Yang, B. Wang, P. Yang, Y. Cui, Y. Song, P. Zheng, Q. Wang, O. N. Bjornstad, R. Yang, B. T. Grenfell, O. G. Pybus, y C. Dye. An investigation of transmission control measures during the first 50 days of the covid-19 epidemic in china. *Science*, 368(6491):638-642, 2020.
- [16] D. Zhu, S. R. Mishra, X. Han, y K. Santo. Social distancing in Latin America during the COVID-19 pandemic: an analysis using the Stringency Index and Google Community Mobility Reports. *Journal of Travel Medicine*, 07 2020. taaa125.

ESTIMACIÓN DE LOS CASOS DE COVID-19 Y BROTES EN TIEMPO REAL A TRAVÉS DE LLAMADAS TELEFÓNICAS

Ezequiel Álvarez

Daniela Obando *

SEBASTIÁN CRESPO **

Enio García *

Nicolás Kreplak **

Franco Mársico **

Resumen

Uno de los principales problemas, para controlar la propagación de la epidemia de COVID-19, es el retraso en la confirmación de los casos. Tener información sobre la evolución de la epidemia o los brotes, es crucial en la toma de decisiones para las políticas de Salud Pública. El presente trabajo propone un algoritmo que estima el número de casos de COVID-19, en tiempo real, utilizando los datos de las llamadas telefónicas a una línea de COVID-19. Modelando las llamadas como fondo (proporcional a la población) más señal (proporcional a los infectados), se ajustó el indicador en la Provincia de Buenos Aires (Argentina) con un coeficiente de determinación $R^2 > 0.85$. El resultado permitió estimar el número de casos dado el número de llamadas de un distrito específico, días antes que los resultados del laboratorio estuviesen disponibles. El algoritmo fue validado con datos reales y permitió hacer seguimiento temprano. Además, se realizó un seguimiento detallado de las incertidumbres en las estimaciones, ya que la alarma utiliza la importancia de los observables como indicador principal para detectar una anomalía. Se presenta en detalle el caso de Villa Azul, Quilmes, donde resultó crucial para controlar un brote a tiempo. Las herramientas fueron diseñadas urgentemente con los datos disponibles, muchas de ellas, están actualmente en desarrollo y, por lo tanto, tienen ciertas limitaciones con posibilidad de mejoras.

Palabras Clave: Vigilancia epidemiológica, Vigilancia molecular, Sistemas de alarma temprana, COVID-19.

^{*}International Center for Advanced Studies (ICAS), ICIFI-CONICET y ECyT-UNSAM, Campus Miguelete, Buenos Aires, Argentina.

^{**}Ministerio de Salud de la Provincia de Buenos Aires, La Plata, Buenos Aires, Argentina.

1. Introducción

La epidemia de COVID-19 está causando daños globales en prácticamente todos los aspectos de la sociedad mundial, desde principios de 2020. Aunque se está haciendo un gran esfuerzo en muchos campos de la ciencia para mitigar sus efectos, la enfermedad se está propagando continuamente y, en muchas regiones, una segunda ola está causando grandes preocupaciones. Las dificultades para controlar la epidemia se deben, en parte, a la combinación crucial de ser altamente contagiosa [7], tener un largo período de incubación [15], durante el cual, es posible el contagio unos días antes de la aparición de los síntomas [10], tener casos leves o asintomáticos [7] y, también, porque el diagnóstico puede tardar días desde que se contacta con el sistema sanitario. En particular, esto último, hace que los brotes se propaguen y la epidemia evolucione mientras se procesan los resultados de laboratorio. Este efecto es más importante en los países de ingresos bajos y medianos debido a problemas operativos y logísticos, generalmente causados por las desigualdades tecnológicas y económicas [1, 21].

En las catástrofes masivas los Sistemas de Alerta Temprana (SATs) juegan un papel clave para la mitigación de daños [9, 14, 20], disminuyendo los tiempos de respuesta por parte de los equipos de emergencia y mejorando su efectividad. La principal estrategia de los SATs en la vigilancia de enfermedades infecciosas es la incorporación de la información producida lo más cercanamente posible al momento de la infección [8, 13]. En este caso, la aparición de los síntomas y su detección por parte de la percepción de salud individual y comunitaria, es la primera señal detectable de un brote. Los sistemas de alerta temprana basados en la vigilancia sindrómica se han aplicado en la vigilancia epidemiológica para la identificación y confirmación temprana de los brotes [11, 12, 16, 18, 19]. Una de las principales características del sistema de alerta temprana es la utilización de la información sanitaria proporcionada por la población para activar las alarmas locales. Hoy en día, con el amplio uso de aplicaciones de telefonía celular y líneas telefónicas específicas del sistema de salud, constantemente se generan importantes bases de datos con información sobre la vigilancia sindrómica [5]. La geolocalización juega un papel principal en la definición espacial y temporal de los brotes detectados por los sistemas de alerta temprana [4].

En la Provincia de Buenos Aires, Argentina, la línea telefónica 148 de COVID-19 es uno de los primeros contactos entre una persona que se cree

infectada y el Sistema de Salud. El equipo capacitado de Atención Médica recibe y responde a las preguntas de las personas generando, simultáneamente, una base de datos de vigilancia sindrómica. Si la persona tiene síntomas que podrían indicar una infección por COVID-19, se le indica que siga el protocolo correspondiente. Es importante señalar que, esa base de datos de síndromes, se utilizó como insumo para la estimación de casos y la detección de brotes en la provincia de Buenos Aires.

En este trabajo, se presenta un método para mitigar los efectos de la epidemia mediante la estimación del número de casos de COVID-19 sin tener que esperar las confirmaciones de laboratorio. Esto proporciona al sistema de salud una herramienta para reaccionar con anticipación y evaluar las políticas de salud pública actuales o futuras. En la Sección 2 se describen los datos de la línea COVID y se presentan los detalles del modelo matemático para estimar el número de casos utilizando los datos de las llamadas telefónicas. En la Sección 3.1 se muestra cómo funciona el modelo en la Provincia de Buenos Aires y cómo puede ser usado para rastrear la epidemia en-vivo. En la Sección 3.2 se presenta la Alarma de brote temprano y sus detalles a través del caso de Villa Azul, Quilmes. Las limitaciones y mejoras actuales del modelo se discuten en la Sección 4 y, finalmente, se presentan las conclusiones en la Sección 5.

2. Métodos

En esta sección, se describe el modelo matemático implementado para relacionar las llamadas telefónicas a una línea COVID-19 con los casos confirmados por el laboratorio por distrito y por día. En los siguientes párrafos s esboza el funcionamiento de la 148 línea COVID y, luego, se describen los detalles del modelo.

2.1. COVID - línea 148 en la provincia de Buenos Aires

La Provincia de Buenos Aires (PBA) es la provincia más poblada de Argentina, con más de 17 millones de habitantes. Alrededor de 13 millones de personas viven en el área metropolitana que rodea la Ciudad de Buenos Aires. Es importante destacar que, los 4 millones restantes, viven en una vasta área y con baja densidad de población conocida como el interior de la provincia. Esta heterogeneidad demográfica conduce a un sistema de salud hiper-centralizado. Con el fin de atender la creciente demanda de asistencia médica causada por el COVID-19, las autoridades de salud pública implementaron, en febrero de 2020, un teléfono COVID específico al que se accede marcando 148. El obje-

tivo de esta línea COVID es atender todas las inquietudes de la comunidad relacionadas con la pandemia, lo que incluye preguntas, dudas, informes de síntomas y referencia al sistema de atención sanitaria, entre otros.

La línea COVID creció en la cantidad de personal a medida que la epidemia se propagó en la PBA. La solicitud de llamadas pasó de unos pocos cientos por día en marzo hasta aproximadamente 20 mil por día a finales de agosto. Hasta finales de junio, el sistema no se saturó y se atendieron todas las llamadas que requerían asistencia. Lo cual permite considerar que, durante este periodo, un indicador procedente de esta línea COVID sería relativamente insesgado. Esto es especialmente cierto si se compara con otros indicadores como las pruebas o el procesamiento en laboratorio, que estaban cambiando su comportamiento considerablemente a medida que la epidemia se propagaba. El periodo comprendido entre el 1 de abril al 26 de junio de 202, la línea COVID mostró ser relativamente estable a los cambios importantes.

Cuando la gente llama a la línea COVID 148, entra en un menú de voz automático en el que una de las opciones corresponde a síntomas similares a los de COVID. Cuando el usuario entra en esta opción, su llamada es atendida por un operador capacitado y, aplicando un breve cuestionario sobre su experiencia, indica si la llamada no pasa el umbral para ser registrada o si corresponde a una de las dos categorías registradas: contacto cercano y caso sospechoso. Si la llamada corresponde a alguna de estas categorías, el operador registra sus datos y, en particular, el distrito desde el que llama. En la Figura 1 se muestra el flujo de trabajo de la línea COVID. En la fase inicial de la implantación del sistema, el registro no incluía información fiable sobre la dirección exacta del usuario. Este hecho crucial llevó a desarrollar el sistema, explicado en la Sección 2.2, restringiendo la información sobre el usuario sólo a su distrito. Se prevén futuras actualizaciones del sistema que proporcionen una localización más precisa de la llamada, sin embargo, el trabajo actual, se restringe al distrito del llamante y sólo una vez que su llamada fue tomada por un operador entrenado por COVID.

2.2. Modelo matemático para estimar los casos de llamadas telefónicas a la línea 148 COVID

En esta sección se presenta el modelo matemático para estimar los nuevos infectados usando los datos de la llamada telefónica, aplicado a la PBA. La estructura mantiene las mismas líneas que [2], pero con diferentes propósitos y diferentes filtros en el conjunto de datos.

El conjunto de datos consideró llamadas de muchos distritos y durante un rango de tiempo determinado a una línea COVID. Cada una de estas llamadas

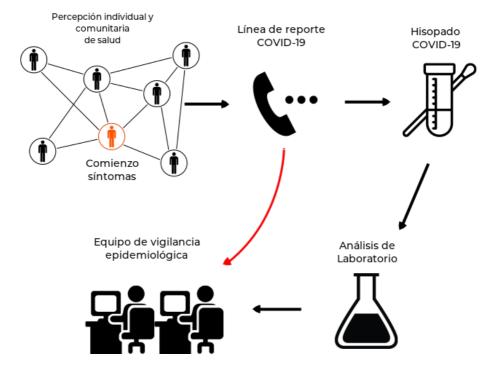


Figura 1: Flujo de trabajo de la línea 148 de COVID.

Nota: A medida que las personas llaman a la línea COVID, según su percepción de la salud, los operadores formados en COVID-19 determinan si corresponden a casos sospechosos o de contacto cercano. En tal caso, su historial pasa al equipo de vigilancia epidemiológica y se ordena un hisopado para análisis de COVID-19. Algunos días después, el resultado del hisopado se añade al registro correspondiente. El flujo de información ordinario se esquematiza con flechas negras. El algoritmo descrito en este documento funciona con la primera parte de la información que se entrega en el flujo de trabajo a medida que los operadores determinan que el caso se corresponde con los criterios correspondientes (flecha roja).

puede ser:

fondo: personas con síntomas similares, pero no infectadas

señal: personas infectadas con COVID-19.

En este caso, el "fondo" se corresponde a aquellas llamadas no atribuibles a casos reales de COVID-19 y que, por lo tanto, pueden ser caracterizadas como "ruido". En cambio, la "señal" es lo que realmente se quiere medir/detectar y se corresponde a las llamadas relacionadas con casos de COVID-19.

Bajo supuestos razonables de homogeneidad en el espacio y el tiempo, es posible modelar que las llamadas de fondo en cada distrito y ventana de tiempo son proporcionales a la población total del distrito y a la duración de la ventana de tiempo. Mientras que, las llamadas de $se\tilde{n}al$, son proporcionales al número total de personas infectadas en el distrito cuyo registro se abre en la correspondiente ventana de tiempo, aunque su confirmación de laboratorio pueda estar disponible en un momento posterior. Por lo tanto, si se divide todo el conjunto de datos en trozos correspondientes en el espacio a los distritos en la PBA, y en el tiempo a las ventanas de tiempo de $\Delta t^{(j)}$ días, que pueden ser elegidos arbitrariamente.

De esta manera, se definen los parámetros: N_p , es la población del distrito correspondiente; N_I , es el número de infectados confirmados en el mismo distrito cuyo registro fue abierto durante la ventana de tiempo correspondiente; n_c es el ajuste al número total de llamadas; y N_c , no presente en la ecuación, es el número total de llamadas realmente realizadas. Se debe observar, por lo tanto, que este conjunto de ecuaciones (una para cada ventana j), puede extenderse en función de la longitud de la ventana de tiempo elegida. Una vez planteado este conjunto de ecuaciones $j = 1, \ldots, k$, es posible ajustar los mejores valores de los coeficientes θ_p, θ_I que minimizan la distancia cuadrada entre n_c y N_c . Es importante considerar que sólo hay dos coeficientes (θ_p, θ_I) que deben ajustarse a todas las ecuaciones diferentes para cada ventana. Se plantea, entonces, la siguiente ecuación para todos los periodos j:

$$n_c^{(j)} = \theta_p \Delta t^{(j)} N_p^{(j)} + \theta_I N_I^{(j)} \tag{1}$$

Este ajuste funciona mejor si todas las ventanas corresponden a períodos en los que los métodos de prueba no han cambiado drásticamente, como puede ser, por ejemplo, si el número de pruebas diarias se modifica considerablemente, o si se consideran nuevos síntomas como umbral para la prueba, entre otros. La razón de exigir esto, es tener un equilibrio coherente entre el número de infectados notificados y el número de llamadas en todas las ventanas. Con este

objetivo, es mejor ajustar los parámetros cada vez que se produzcan cambios importantes en los métodos de prueba y de notificación.

Una vez ajustados los parámetros $\theta_{p,I}$ en la Ecuación 1, incluyendo su incertidumbre del ajuste, es posible estimar el número de nuevos infectados en una determinada ventana como:

$$n_I^{(j)} = \frac{1}{\theta_I} (N_c^{(j)} - \theta_p \Delta t^{(j)} N_p^{(j)})$$
 (2)

Obsérvese que, el lado derecho, requiere datos que se obtienen en el mismo día y,por lo tanto, se puede estimar el número de casos n_I en-vivo, sin necesidad de esperar los resultados del laboratorio. A su vez, el algoritmo permite estimar el número total de casos nuevos en cada tramo, pero no determinar cuál de las llamadas corresponde a los nuevos casos. La incertidumbre en la estimación de n_I se calcula aplicando la fórmula de propagación de errores en la Ecuación 2. Si las variables están correlacionadas, como por ejemplo θ_p y θ_I , se debe tener esto en cuenta, sin embargo, en el caso utilizado en este estudio, no considera esta correlación en comparación con otros términos. Para los parámetros $\theta_{p,I}$ se utilizó la incertidumbre proveniente del ajuste, para N_c , se utilizó la incertidumbre de Poisson; y para N_p se debe decidir si se añade una incertidumbre sistemática o sólo se utiliza Poisson, como se realizó en este trabajo. Como se discute más adelante, las incertidumbres en las estimaciones juegan un papel central en el diseño de la alarma de brotes tempranos y, por lo tanto, deben ser manejadas con cuidado, especialmente las sistemáticas si están presentes.

Para aplicar este algoritmo en la PBA, se utilizó el conjunto de datos de las llamadas telefónicas a la línea 148 COVID. Se trabajó con todas las llamadas telefónicas que entraron en la línea COVID y que alcanzaron el umbral de ser un contacto cercano o un caso sospechoso. La razón de la granularidad seleccionada para las alarmas es porque el distrito desde el que llama el usuario está registrado por el operador. Aunque la dirección también está registrada, la mayoría de las veces, en la práctica poseen muchas ambigüedades, palabras mal escritas, u otros errores no intencionados dan como resultado que sólo alrededor de 50 %-70 % de los casos puede ser reconstruido correctamente. Se consideró el conjunto de datos de las llamadas entre el 1 de abril y el 26 de junio, ya que, posteriormente, el centro de llamadas se sobrecargó, produciendo sesgos intratables. A lo largo de este período, se ajustaron los datos varias veces en diferentes conjuntos de datos, obteniendo resultados bastante similares y con un coeficiente de determinación siempre satisfactorio $\mathbb{R}^2 > 0,85.$ En particular, a medida que aumentaban los casos, se obtuvieron estimaciones más precisas para θ_I , como es de esperar.

Para mostrar la robustez de las hipótesis, se muestra cómo funciona este modelo con los datos del 1 de mayo al 26 de junio, divididos en dos ventanas de tiempo de igual duración. Se consideraron todos los distritos en la PBA cuyo número de llamadas en estos trozos fue superior a 100. Después de este filtrado, se contó con 43 trozos, es decir, 43 puntos de datos. Después de realizar el ajuste indicado en la Ecuación 1 se obtiene:

$$\theta_p = (5, 16 \pm 1, 59) \times 10^{-6}$$
llamadas por habitante por día

 $\theta_I=0,69\pm0,05$ llamadas por persona infectada

Vale la pena notar que, los valores precisos de estos coeficientes ajustados, tienen una fuerte dependencia del proceso de filtrado de llamadas y de la arquitectura del sistema de llamadas. En particular, estos valores difieren de los presentados en [2], porque se consideró un nivel diferente de filtrado para obtener el distrito de cada usuario. El ajuste de este conjunto de datos arroja un coeficiente de determinación $R^2=0,91$, que indica la solidez de las hipótesis implicadas. La Figura 2 muestra la comparación entre los datos y el ajuste para el número de llamadas telefónicas, tal como se plantea en la Ecuación 1.

3. Resultados

3.1. Seguimiento de la epidemia a través de estimaciones de modelos

El modelo matemático descrito en la Sección 2.2, proporciona un marco para estimar con muchos días de antelación el número de casos confirmados por laboratorio por día, en función de la distribución espacio-temporal de las llamadas telefónicas a la línea COVID. Se trata de un logro convincente, porque la información de la llamada telefónica está disponible en tiempo real, mientras que la confirmación de los casos por parte del laboratorio puede requerir desde unos pocos días hasta una semana desde que los pacientes informan de sus primeros síntomas. A lo largo de esta sección se muestra cómo utilizar este sistema para tener una estimación de la evolución de la epidemia en tiempo real, junto con los resultados de los casos reales en la PBA.

Como este sistema fue desarrollado y aplicado en el momento, no hubo tiempo para la validación. Sin embargo, la obtención de un $R^2 > 0,85$ en el ajuste fue una señal satisfactoria de que el modelo funcionaba bien hasta el

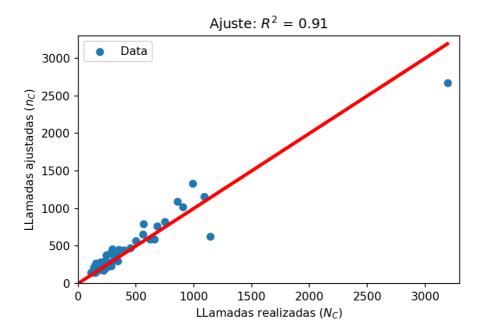


Figura 2: Llamadas realizadas versus ajustadas durante el período analizado (1 de mayo al 26 de junio dividido en dos ventanas de tiempo).

Nota: El número de llamadas total proviene del número de casos de COVID-19 confirmados por el laboratorio usando el ajuste en la Ecuación 1. El punto de datos superior derecho corresponde al distrito de La Matanza, cuya población de 1,7 millones de habitantes es al menos tres veces mayor que la del resto de los distritos.

momento. Con el paso de los meses, se tuvo la posibilidad de comparar en una ventana de tiempo amplia la estimación del modelo con el número de casos medidos y confirmados por el laboratorio por distrito.

3.2. Alarma de brotes tempranos

A lo largo de esta sección, se detallará un subproducto del modelo de la Sección 2 para detectar brotes de COVID-19 considerablemente antes de producirse la confirmación del laboratorio. Se describe, brevemente, el funcionamiento del modelo y, luego, se proporciona el detalle a través de la descripción de un caso real ocurrido a mediados de mayo en Villa Azul y Villa Itatí en la PBA.

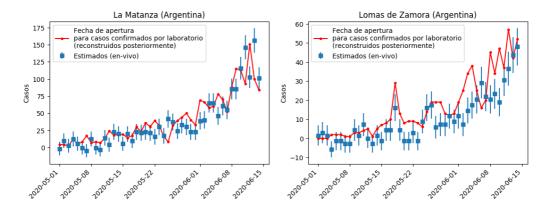


Figura 3: Comparación de datos reales versus estimación del modelo con una barra de error que representa un IC del 95 % para dos distritos ejemplo de la Provincia de Buenos Aires (La Matanza y Lomas de Zamora).

Nota: La línea roja correspondiente a los casos reales confirmados con su apertura de registro en la fecha correspondiente, se reconstruye muchos días después. En las fechas en las que la línea roja supera la estimación suele ser debido a que los operativos de DETECTAR (pruebas puerta a puerta [17]), fueron llevados a cabo. En general, el modelo arroja una muy buena estimación para monitorear la epidemia en todos los distritos afectados de la PBA.

3.2.1. Identificación de la formación de un brote

Con la estimación en tiempo real de los casos por día en cada distrito, es de interés desarrollar una herramienta estadística y automática que pueda disparar una alarma cuando un posible brote esté gestándose. Tener una alarma temprana sobre este tipo de características epidemiológicas, es una herramienta crucial para evitar su propagación y consecuencias drásticas.

Para detectar un posible brote hay muchos indicadores que deben ser analizados simultáneamente. Por un lado, es importante tener una estimación del número absoluto y relativo diario de casos y, por otro lado, también es importante tener una estimación de la variación diaria de estos observables. Para tener un indicador cuantitativo objetivo del potencial de un brote en una región determinada, es esencial tener una evaluación correcta de las incertidumbres en todas las estimaciones del modelo. A lo largo del sistema implementado como Alarma de brotes tempranos, se ha considerado la detección de anomalías una anomalía como un indicador central. Aquí la significancia se define como la distancia a cero del valor central del indicador, medida en unidades de su incertidumbre. O, en otras palabras,

$$significancia = \frac{Valor\ central}{Incerteza} \tag{3}$$

El algoritmo desarrollado, calcula diariamente la estimación del número total de casos nuevos en cada distrito en la PBA. Dado que, en la franja temporal estudiada, especialmente antes de junio, el número de casos estimados por día de muchos distritos estaba por debajo de ~ 5 - 10, consideramos incluir la estimación de los casos de los últimos dos días. Esto reduciría la relativa incertidumbre de Poisson debido a las pequeñas cifras. Se calculó que, el número de casos estimados en valor absoluto, y también en relación con 100 mil habitantes, era igualmente sensible para todos los distritos.

Un tercer y decisivo observable, que señala el nivel de peligro de un brote, es el aumento diario de los casos estimados. Dada la estimación diaria proporcionada por el modelo matemático, es posible reconocer una curva de rápido incremento de muchas maneras. Se ha considerado un ajuste lineal a la estimación de casos de los últimos 3 días y, la pendiente de esta recta, como una estimación del valor central del incremento diario. También se utilizó la significación como el indicador más relevante para decidir el nivel de peligro de cada distrito. En este caso, el cálculo de la barra de error en la pendiente de la línea, incluye todas las incertidumbres de la estimación de cada día incluidas en el cálculo de la incertidumbre de ajuste a través de los residuos de mínimos cuadrados. Se utilizaron 3 días para ajustar una línea porque es el tiempo mínimo necesario para ver un aumento consecutivo de dos días, mientras que todavía se está muy por delante de los resultados del laboratorio. Además, tres días es también una buena ventana de tiempo para las características específicas de COVID-19.

Esta alarma de brotes tempranos, ha proporcionado a la administración del Ministerio de Salud de la PBA herramientas muy importantes para identificar posibles brotes durante el aumento de la curva epidémica. Dado que la granularidad del algoritmo es muy pobre (distritos), el sistema debe complementarse con otros indicadores independientes, en particular los que pueden ayudar a proporcionar una localización más precisa del brote. Esto se hacía normalmente llamando manualmente a los casos registrados, y luego enviando a agentes de DETECTAR [17], para verificar si de hecho las condiciones in situ serían las previstas. La alarma de brotes tempranos ha indicado muchos brotes que han sido controlados desde mediados de abril a mediados de junio. En particular, se describirá en la Sección 3.2.2 el muy especial caso de Villa Azul, Quilmes (ha sido cubierto en los titulares de las noticias durante varias semanas [3]), y se proporcionarán los detalles de cómo la Alarma de brotes tempranos indicó el distrito de Quilmes.

3.2.2. Estudio de caso: Villa Azul - Quilmes

En esta sección se muestran los detalles de uno de los brotes indicados por la Alarma de brotes tempranos a mediados de mayo en el distrito de Quilmes. Este caso fue el primer gran brote en un barrio de bajos recursos en la PBA y tuvo un gran impacto en las noticias [3], no sólo por su magnitud sino también por su detección temprana, que llevó a un estricto bloqueo y aislamiento del brote para controlar su propagación a los barrios cercanos.

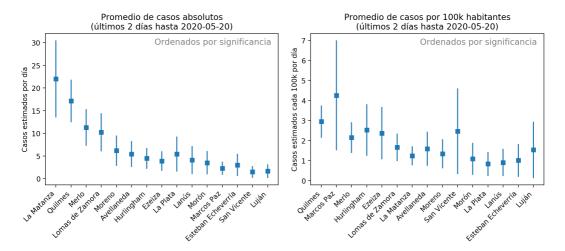


Figura 4: Estimación de casos por día usando el modelo de las llamadas telefónicas de la línea COVID de los últimos dos días.

Nota: Obsérvese que los distritos no están ordenados por su valor central, sino por su significado, que se define como la tasa entre el valor central y la incertidumbre. Es por eso que las barras de error son cruciales para proporcionar una alarma de brote temprano. Los resultados se muestran en valor absoluto (izquierda) y relativo a cada 100 mil habitantes (derecha).

Mientras que Quilmes está en la posición superior cuando se escala a relativo por cada 100 mil habitantes. El 20 de mayo la alarma indicaba un gran número de casos estimados en el distrito de Quilmes, en particular Quilmes tenía la estimación más alta en número de casos por habitantes de los dos últimos días, medida a través de la significación del indicador. En la Figura 4, se muestra el escenario del 20 de mayo en el que Quilmes es casi tan grande como La Matanza en valor absoluto, con $\sim 1/3$ de su población. Además, el indicador del ajuste del aumento diario, también indicaba que Quilmes era el distrito de mayor importancia (véase la Figura 5). Este último indicador, sobre el ajuste de incremento diario a los últimos 3 días, puede ser visualizado en la Figura 6a, donde se trazó la estimación diaria de los últimos 7 días en

Quilmes. El ajuste se obtuvo usando los 3 últimos puntos de datos en rojo. Dado que todos estos indicadores apuntan al distrito de Quilmes, el equipo de vigilancia se encargó de localizar las llamadas telefónicas y observó un exceso procedente de Villa Azul, un barrio de bajos ingresos de Quilmes y próximo al distrito de Avellaneda.

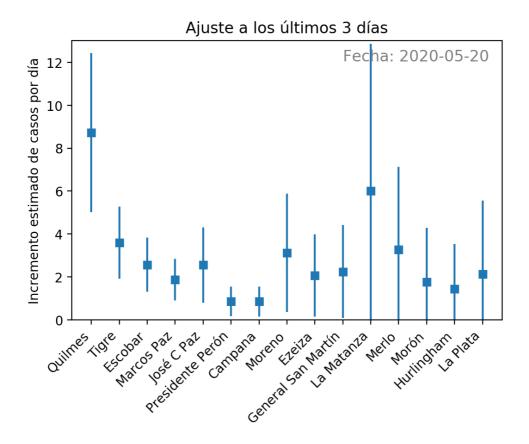


Figura 5: Pendiente de un ajuste lineal a la estimación de casos por día de los últimos tres días.

Nota: La barra de error corresponde a la inclusión de la incertidumbre en cada estimación por día y en la determinación de la pendiente en el ajuste. Esta gráfica es crucial para ordenar los distritos según la significación en esta variable. En la Figura se ve, el 20 de mayo, a Quilmes en la posición superior e indica una potencial alarma temprana de un brote, como fue consecuentemente confirmado por otros indicadores unos días después.

Estas observaciones, indicadas de antemano por la Alarma de brotes tempranos, debían ser verificadas por un indicador complementario independiente. Al día siguiente un operativo DETECTAR [17], fue enviado a Villa Azul, donde se verificó la situación aguda y se inició de inmediato el rastreo puerta a puerta con resultados de laboratorio urgentes. Como los primeros resultados estaban confirmando el brote en Villa Azul, la Administración de la PBA decidió un estricto encierro y aislamiento durante 14 días desde el 24 de mayo [3].

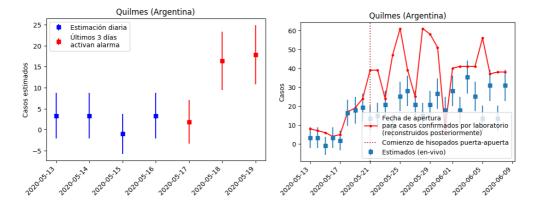


Figura 6: Izquierda: Visualización de la alarma de brote temprano el 20 de mayo para Quilmes, según lo dispuesto por el equipo de Atención de la Salud en la provincia de Buenos Aires. Derecha: Casos confirmados por el laboratorio, el día en que el Sistema de Salud aterrizó en Villa Azul para comenzar a probar puerta por puerta, y un rango más amplio de fechas para captar el panorama general del caso.

Nota: El estricto encierro en Villa Azul sin permiso de entrada ni salida duró desde el 24 de mayo hasta el 8 de junio. Como se puede ver en la trama, durante la prueba puerta por puerta, la línea sólida va por encima y no se correlaciona con los casos estimados por llamadas telefónicas, como se esperaba.

3.2.3. Descripción epidemiológica y operativa de Villa Azul

Villa Azul (Quilmes) y Villa Itatí (Avellaneda) son dos barrios adyacentes de bajos recursos. El último análisis demográfico indica que, Villa Azul, tiene una población de 3.128 habitantes y Villa Itatí de 15.142. La alta densidad de edificación y vivienda, y las pequeñas calles ponen a la población en estrecho contacto. Estas características, hacen que estos barrios sean susceptibles de una rápida difusión [6]. Teniendo esto en cuenta, la detección temprana de los brotes implica un desafío principal en estos casos complejos en los que la detección y el bloqueo de la propagación deben hacerse cuando se reportan los primeros casos. En particular, la alarma temprana descrita anteriormente para el brote, ocurrió en Villa Azul permitiendo una rápida respuesta del equipo

del Sistema de Salud para mitigar y controlar su propagación a Villa Itatí.

Una vez que se produjo el estricto cierre y aislamiento, el equipo de atención social entregó agua y alimentos. No se permitió a las personas salir de la casa durante todo el aislamiento. Los equipos sanitarios de vigilancia activa comenzaron con un monitoreo de los síntomas puerta a puerta. Se analizaron los casos con manifestaciones clínicas relacionadas con COVID-19. Los casos confirmados fueron aislados dentro de sus casas en los casos en que esto era posible (si había una habitación vacía, por ejemplo) y en los casos en que no lo era, las personas fueron enviadas a un centro fuera del hospital.

4. Perspectivas y alcance

El desarrollo del modelo matemático para estimar el número de casos de COVID-19 se hizo con urgencia y adaptándolo a los datos disponibles. No hubo tiempo de solicitar cambios en la adquisición ni en el procesamiento de los datos. Por supuesto, el algoritmo y el sistema pueden ser mejorados en muchas direcciones.

Una de las mayores debilidades del algoritmo es la gran granularidad, que corresponde a los distritos. Las poblaciones de los distritos en el área relevante son en promedio 500 mil personas. Esta cuestión, se traduce en que la alarma de brotes tempranos deja de funcionar una vez que la densidad de casos es tal que hay más de unos pocos brotes en cada distrito. Situación que ocurrió a finales de junio en la PBA. Para una futura implementación, e está llevando a cabo una solución a este problema, obteniendo una dirección de confianza del operador entrenado por COVID que toma la llamada. Una solución más estable sería obtener esta información de la compañía telefónica, sin embargo, las regulaciones muchas veces bloquean esta posibilidad.

Por otro lado, el algoritmo tiene un beneficio muy importante que es su imparcialidad. Dado que la línea COVID funciona 24 horas los 7 días de la semana y con una metodología bastante equitativa todo el tiempo, la estimación del algoritmo no depende de la disponibilidad de pruebas o de instalaciones de prueba sobrecargadas, entre otras. Por supuesto, el sistema tiene ligeros sesgos que pueden provenir, por ejemplo, de diferentes señales de fondo debido a las diferentes características de los distritos, o al comportamiento social estacional a medida que pasan los meses. Algunos de estos sesgos pueden resolverse reajustando el modelo de vez en cuando, otros ajustando diferentes modelos en diferentes regiones. Lo importante es que el algoritmo proporciona información sobre las llamadas de fondo que varían en el espacio y el tiempo.

Se podrían hacer más estudios para comprender y extraer las propiedades del fondo, como puede ser su estacionalidad, las variaciones según las regiones, los anuncios públicos o las noticias, etc.

El punto crucial del modelo matemático, es que reconoce las anomalías debidas a comportamientos colectivos. Por lo tanto, se encuentra que el modelo matemático y los algoritmos de Alarma de brotes tempranos pueden ser útiles para muchas otras enfermedades epidemiológicas, como por ejemplo el Dengue, y otros eventos, como las catástrofes naturales, entre otros. Actualmente, se está trabajando en la mejora de este sistema en muchos aspectos, incluyendo también los algoritmos de aprendizaje automático, avances que serán publicados en un trabajo futuro.

5. Conslusiones

En el presente trabajo, se muestra la creación de un algoritmo de vigilancia sindrómica basado en la correlación entre las llamadas telefónicas a una línea COVID, la población de los distritos y los casos reportados. Este algoritmo funciona entendiendo que las llamadas telefónicas a una línea COVID son una parte debido a que las personas no infectadas tienen síntomas similares (fondo) y otra parte debido a las personas infectadas ($se\tilde{n}al$). Observando que el fondo tiene que ser proporcional a la población del distrito, mientras que la señal es proporcional a los casos reportados, hemos ajustado nuestra suposición. El coeficiente de determinación para la Provincia de Buenos Aires (PBA) es siempre $R^2 > 0.85$ para las diferentes muestras, lo que indica la solidez de nuestra hipótesis. Además, hemos validado nuestro modelo con datos reales.

A lo largo del manuscrito se describió el modelo, sus estimaciones y cómo se calcularon sus barras de error. También se ha mostrado cómo las estimaciones, que se obtienen en tiempo real, pueden ser utilizadas para abordar las políticas de Salud Pública sin necesidad de esperar los resultados de laboratorio, que requieren muchos más días para converger. El algoritmo funcionó en la PBA de abril a junio, ya que durante este tiempo el centro de llamadas entrenado por COVID no se sobrecargó. Por lo tanto, la estimación fue relativamente imparcial.

Se muestra cómo esta estimación puede ser usada para crear una Alarma de brotes tempranos. Además, se describe cómo la construcción de indicadores que tienen que ver con los casos diarios, y el incremento diario de casos, puede indicar brotes con antelación. La variable estadística relevante en este caso es la significación, ya que es una medida real de cuán lejos de cero están los

indicadores. Es importante destacar que este sistema puede detectar un brote y, en particular, ejemplificamos su aplicación en la detección de brotes en Villa Azul, Quilmes.

Las limitaciones de la Alarma de brotes temprano fueron discutidas en este documento, muchas de ellas debido a las características de los datos disponibles en el momento de su desarrollo (urgente). Se han señalado muchas maneras de mejorar su sensibilidad y precisión, en las que se está trabajando actualmente. Esta alarma también sería útil, no sólo para otras enfermedades epidemiológicas, sino también para eventos que produzcan cambios en el comportamiento colectivo, como la epidemia de Dengue, las catástrofes naturales u otras.

El algoritmo y el modelo matemático presentados han sido una de las principales herramientas en el tablero del sistema PBA Health Care durante la epidemia, y sus versiones actuales y de actualización siguen siendo utilizadas para rastrear la epidemia y detectar brotes.

Agradecimientos: Agradecemos el fantástico trabajo realizado por el centro de llamadas COVID 148, en particular, a R. Vaena, P. Rispoli y L. H. Molinari por útiles conversaciones. E.A. y F.M. agradecen a la Dra. I. Caridi por las discusiones. E.A. agradece a CONICET, UNSAM, CAF y Easytech-Green por el apoyo financiero y logístico durante esta investigación.

Conflictos de interés: Los autores declaran no tener ningún conflicto de interés.

Referencias

- [1] O. Ahmed. Mitigating the spread of covid-19 in low-income countries. Scientific Research and Essays, 15(2):41–44, 2020.
- [2] E. Alvarez y F. Marsico. Covid-19 mild cases determination from correlating covid-line calls to reported cases. arXiv preprint arXiv:2008.01176, 2020.
- [3] Buenos Aires Times. Villa in Buenos Aires Provinc seaamid surge in COVID-19 cases. led off mayo 2020. Disponible https://www.batimes.com.ar/news/argentina/ villa-in-buenos-aires-province-isolated-due-to-covid-19. phtml.

- [4] D. Chen, J. Cunningham, K. Moore, y J. Tian. Spatial and temporal aberration detection methods for disease outbreaks in syndromic surveillance systems. *Annals of GIS*, 17(4):211–220, 2011.
- [5] V. Diwan, D. Agnihotri, y A. Hulth. Collecting syndromic surveillance data by mobile phone in rural india: implementation and feasibility. *Global health action*, 8(1):26608, 2015.
- [6] W. Finch y M. Hernández Finch. Poverty and covid-19: rates of incidence and deaths in the united states during the first 10 weeks of the pandemic. Frontiers in Sociology, 5:47, 2020.
- [7] M. Gao, L. Yang, X. Chen, Y. Deng, S. Yang, H. Xu, y X. Gao. A study on infectivity of asymptomatic sars-cov-2 carriers. *Respiratory Medicine*, 106026, 2020.
- [8] J. Ginsberg, M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski, y L. Brilliant. Detecting influenza epidemics using search engine query data. *Nature*, 457(7232):1012–1014, 2009.
- [9] K. Goniewicz y F. M. Burkle. Disaster early warning systems: the potential role and limitations of emerging text and data messaging mitigation capabilities. *Disaster medicine and public health preparedness*, 13(4):709–712, 2019.
- [10] X. He, E. H. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y. C. Lau, J. Y. Wong, Y. Guan, X. Tan, et al. Temporal dynamics in viral shedding and transmissibility of covid-19. *Nature medicine*, 26(5):672–675, 2020.
- [11] K. Hope, D. Durrheim, E. d'Espaignet, y C. Dalton. Syndromic surveillance: is it a useful tool for local outbreak detection? *Journal of Epidemiology & Community Health*, página 374–374, 2006.
- [12] R. Katz, L. May, J. Baker, y E. Test. Redefining syndromic surveillance. Journal of epidemiology and global health, 1(1):21–31, 2011.
- [13] G. Krause, D. Altmann, D. Faensen, K. Porten, J. Benzler, T. Pfoch, y H. Claus. Survnet electronic surveillance system for infectious disease outbreaks, germany. *Emerging infectious diseases*, 13(10), 2007.
- [14] U. Kyriacos, J. Jelsma, M. James, y S. Jordan. Monitoring vital signs: development of a modified early warning scoring (mews) system for general wards in a developing country, 2014.

- [15] H. Lee, K. Kim, K. Choi, S. Hong, H. Son, y S. Ryu. Incubation period of the coronavirus disease 2019 (covid-19) in busan, south korea. *Journal* of Infection and Chemotherapy, 26(9):1011–1013, 2020.
- [16] J. Lombardo, H. Burkom, E. Elbert, S. Magruder, S. H. Lewis, W. Loschen, J. Sari, C. Sniegoski, R. Wojcik, y J. Pavlin. A systems overview of the electronic surveillance system for the early notification of community-based epidemics (ESSENCE II). *Journal of urban health*, 80(1):i32–i42, 2003.
- [17] Ministerio de Salud. Detectar dispositivo estratégico de testeo para coronavirus en territorio argentino. *Argentina.gob.ar*, 2020. Disponible en https://www.argentina.gob.ar/coronavirus/detectar.
- [18] J. Pavlin. Investigation of disease outbreaks detected by "syndromic" surveillance systems. *Journal of Urban Health*, 80(1):107–114, 2003.
- [19] M. Stoto, M. Schonlau, y L. Mariano. Syndromic surveillance: is it worth the effort? *Chance*, 17(1):19–24, 2004.
- [20] G. Texier, L. Pellegrin, C. Vignal, J.-B. Meynard, X. Deparis, y H. Chaudet. Dealing with uncertainty when using a surveillance system. *International Journal of Medical Informatics*, 104:65–73, 2017.
- [21] L. Verhagen, R. Groot, C. Lawrence, J. Taljaard, M. Cotton, y H. Rabie. Covid-19 response in low-and middle-income countries: Don't overlook the role of mobile phone communication. *International Journal of Infectious Diseases*, 99:334–337, 2020.

IMPACTOS DEL COVID-19 EN EL CONSUMO ELÉCTRICO CHILENO

Rodrigo Moreno * †

Miguel Sánchez * †

Carlos Suazo * ‡

Matías Negrete * §

Daniel Olivares * §

Diego Alvarado * †

Carlos Sepúlveda * †

HÉCTOR OTÁROLA * ‡

Leonardo Basso * ¶

Resumen

Este artículo identifica los impactos del COVID-19 en la demanda eléctrica de distintos tipos de consumidores en Chile, incluyendo demandas a nivel residencial, comercial e industrial. De forma inédita, se determina la evolución de la demanda residencial y comercial con un gran nivel de granularidad temporal para 32 comunas de la ciudad de Santiago. A su vez, se analiza la evolución del consumo industrial por sector económico. Se demuestra que los impactos del COVID-19 en los consumos son altamente dependientes de los tipos de consumidores. Se estima, por ejemplo, que solamente debido a las políticas sanitarias, los consumos residenciales experimentaron un alza de hasta un 17 % durante junio con respecto al mismo periodo de 2019. Por el contrario, el comercio y la industria presentaron caídas importantes, con algunos sectores industriales como Hoteles y Restaurantes que presentaron una baja de hasta un 75 % en sus consumos con respecto al año 2019.

Palabras Clave: Demanda eléctrica, Medidores inteligentes, Política energética, COVID-19.

^{*}Instituto Sistemas Complejos de Ingeniería (ISCI), Santiago, Chile.

[†]Departamento de Ingeniería Eléctrica, FCFM, Universidad de Chile, Chile.

[‡]SPEC Energy Consulting, Santiago, Chile.

[§]Departamento de Ingeniería Eléctrica, Pontificia Universidad Católica de Chile, Chile.

[¶]Departamento de Ingeniería Civil - Transporte, FCFM, Universidad de Chile, Chile.

1. Motivación

El COVID-19 tuvo sus inicios en noviembre de 2019 en la provincia de Hubei, China. Desde entonces, un gran crecimiento en el número de contagios diarios y el colapso de los sistemas sanitarios tomó la atención global. Más tarde, en marzo de 2020, fue declarado como pandemia por la Organización Mundial de la Salud (OMS) [18].

Desde ese entonces, la situación sanitaria llevó a diversos estados a tomar políticas restrictivas de movilidad para la contención de la pandemia. Europa, luego de China, fue una de las primeras zonas afectadas masivamente por el COVID-19. Así, fue declarada por la OMS como centro global de la pandemia el 13 de marzo de 2020 [3]. Por ejemplo, en Italia se decretó una cuarentena nacional el 9 de marzo, limitando el movimiento de personas y permitiendo exclusivamente la movilidad relacionada con necesidades básicas, como trabajadores de la salud u otros trabajos esenciales. En España se decretó estado de alarma el 14 de marzo, pero más tarde (29 de marzo), se endurecieron las medidas en respuesta a las altas tasas de contagio, obligando a permanecer en casa a todos los trabajadores no esenciales. En la misma línea, Reino Unido declaró estado de cuarentena el 23 de marzo, cerrando todas las actividades comerciales no esenciales.

En el caso de Chile, en línea con las políticas de otras jurisdicciones, se decretó estado de excepción constitucional de catástrofe el 18 de marzo. Las primeras cuarentenas fueron adoptadas a partir del 26 de marzo. Estas cuarentenas fueron implementadas por zonas, siguiendo criterios determinados por la autoridad. Estas zonas fueron aumentando en la medida que la cifra de contagios aumentaba. Por ejemplo, en la provincia de Santiago, donde se concentra la mayor población de Chile, fue declarada la cuarentena total el día 15 de mayo [12].

En este contexto, la intensidad de las políticas sanitarias han impactado en el consumo energético de los distintos países. La International Energy Agency reporta en [9] cómo ha variado la demanda eléctrica con respecto al año 2019 en algunos países de Europa, China e India. De este modo, Italia se reporta como el país que presentó la caída más abrupta en demanda eléctrica, mostrando una disminución por sobre el 25 % durante abril. Por otro lado, en [15] se reportan variaciones a la baja en la demanda en los siete Regional Transmission Operators (RTOs) en Estados Unidos. De hecho, las jurisdicciones del New York Independent System Operator (NYISO) y de Midcontinent Inde-

pendent System Operator (MISO) fueron aquellas que mostraron una mayor disminución (en torno al 10%) en sus consumos con respecto al año 2019.

En Chile, las medidas de cuarentena y la reducción de movilidad también han afectado la actividad económica (la actividad económica del segundo trimestre del 2020 se contrajo un 14.1 % en comparación al mismo periodo del 2019 [4]) y, con esto, la demanda eléctrica nacional también fue afectada, como se muestra en la Figura 1. El consumo eléctrico durante el periodo de mayo a agosto sufrió un retroceso en torno al 3.5 % con respecto al 2019.

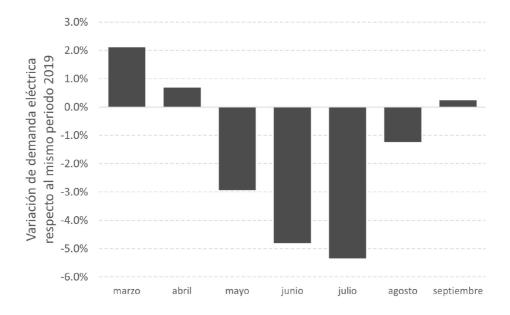


Figura 1: Variación de la demanda eléctrica nacional con respecto al mismo período del 2019.

Nota: Marzo fue el mes que se registró el primer contagio en Chile [5].

Un aspecto importante a considerar es que, si bien las tendencias globales (a nivel nacional) de los consumos eléctricos son a la baja, en los distintos segmentos (consumo residencial, comercial e industrial) podrían existir tendencias opuestas. Por ejemplo, los clientes residenciales pueden presentar una tendencia creciente en sus consumos debido a las medidas de confinamiento. Además, ciertos segmentos del comercio o la industria pueden presentar niveles variados de impacto producto de las políticas para combatir el COVID-19. Por esta razón, resulta interesante describir los efectos de la pandemia en la demanda eléctrica para distintos tipos de consumidores.

Las contribuciones de este artículo se resumen en los siguientes dos puntos:

■ Determinar el impacto de la pandemia del COVID-19 en los consumos eléctricos desagregados para distintos tipos de clientes y por sec-

tor económico en Chile, incluyendo clientes residenciales, comerciales e industriales.

 Identificar los impactos en los clientes residenciales por zonas/comunas de Santiago y su potencial correlación con el nivel socioeconómico de la población.

Este artículo se estructura de la siguiente manera. La sección Revisión bibliográfica presenta estudios relacionados con el trabajo realizado. La sección Metodología describe los procedimientos, las fuentes de información y los casos de estudio llevados a cabo en los análisis. La sección Resultados y discusión cuantifica los impactos de la pandemia del COVID-19 en los consumos eléctricos para clientes regulados (residenciales y comerciales) y clientes industriales. Además, se analizan posibles causas de los impactos considerando variables socioeconómicas y climáticas. Finalmente, la sección Conclusiones resume los principales resultados y aprendizajes del estudio.

2. Revisión bibliográfica

Un gran número de estudios han analizado el impacto del COVID-19 en los sistemas eléctricos. Estos se pueden agrupar en documentos que i) describen las variaciones de la demanda eléctrica, ii) desarrollan metodologías o análisis estadísticos que determinan el impacto de la pandemia en los mercados eléctricos, iii) analizan los impactos a mediano y largo plazo del COVID-19 en los mercados eléctricos, y iv) estudian el impacto de la pandemia en los consumos eléctricos industriales según su rubro económico.

En el primer grupo de estudios descriptivos se encuentra [3], el cual analiza el impacto de las distintas políticas sanitarias aplicadas en Europa para la contención del COVID-19. Ejemplifican con seis países cuyas medidas adoptadas tienen distinta profundidad. De este modo, los países con políticas más restrictivas como España e Italia presentan grandes disminuciones de sus consumos en días laborales, mientras que en países con medidas más blandas, como el caso de Suecia, se observa que el consumo ha aumentado durante los fines de semana. Asimismo, en [1] analizan los cambios en las variaciones de consumo en Ontario, distinguiendo cambios en los patrones de demanda entre el periodo pre y post pandemia.

La referencia [16] profundiza en el análisis del caso español. Para esto, analizan la disminución de la demanda, la disminución de las emisiones de CO_2 , y cambios en los precios considerando los datos del operador del sistema español. Además, se analizan posibles cambios en la demanda residencial, considerando

la variación de demanda de un único cliente particular ubicado en Córdoba, correspondiente a un hogar de cuatro personas. Asimismo, estudian cómo los centros de salud han cambiado sus patrones de consumo debido a la pandemia.

En línea con lo anterior, en [17] estudian el comportamiento de 17 hogares en Queensland, Australia. De este modo, identifican que las alzas de consumos residenciales se deben a un mayor uso de dispositivos electrónicos y elementos de cocina.

En los estudios que desarrollan metodologías para analizar los impactos del COVID-19 en los mercados eléctricos, se puede encontrar que en [13] se propone una red neuronal para estudiar la elasticidad de la demanda por electricidad y petróleo según el porcentaje de personas infectadas por la pandemia. En [2] se estudia el efecto de la pandemia sobre la demanda eléctrica en Estados Unidos. Uno de los principales desafíos es que las series de tiempo de los consumos mezclan los efectos propios de la pandemia con otros efectos como las condiciones climáticas. Como solución a esta problemática, se propone aplicar modelos de regresión para separar el efecto de variables climáticas y así poder aislar los efectos de la pandemia. Esta metodología es aplicada a Regional Utilities que poseen la información de todos los clientes de su zona de servicio.

Por otro lado, [15] plantea un análisis interdisciplinario para analizar los impactos reales de la pandemia en Estados Unidos. Para esto, se cuenta con la base de datos Coronavirus Disease and Electricity Market Data Aggregation (COVID-EMDA) que contiene información de la demanda eléctrica, información sanitaria, información climática e índices de movilidad, entre otros. A partir de ésta, con una metodología backcasting, se estima la demanda eléctrica en ausencia del COVID-19. Así, comparado con este benchmark se cuantifica el impacto de la pandemia en la demanda eléctrica.

Otro grupo de estudios analizan las perspectivas futuras en los mercados eléctricos. En [19] se analizan los impactos en la demanda destacando una reducción de la demanda global, pero un aumento de la demanda residencial. Además, explica impactos en el mercado de generación y el aumento de la participación de generación renovable. De la misma forma, en [8] analizan los efectos de corto y largo plazo en el mercado estadounidense a partir del consumo a nivel nacional. Así, en el corto plazo, se espera una disminución de las emisiones de CO_2 debido a que el menor consumo ha desplazado generación a partir de combustibles fósiles. Sin embargo, en el largo plazo, una menor demanda reduciría los incentivos de invertir en tecnologías renovables. Finalmente, [10] plantea las oportunidades asociadas al COVID-19 en la transición sustentable de los mercados eléctricos de Noruega y Finlandia.

Finalmente, otros estudios han analizado los cambios de demanda eléctrica

en distintas industrias según su rubro económico. En el mercado brasilero, en [6], se analiza el impacto de la pandemia en el consumo eléctrico de distintos sectores económicos. Así, se deduce que los sectores más afectados han sido la industria automotriz, la de servicios y la textil, con caídas de consumo en torno al 25 %. En [11] se analiza el impacto del COVID-19 en distintos sectores económicos y en las distintas provincias de Canadá, destacando los impactos en la industria del petróleo y su consumo eléctrico asociado. La industria de manufactura de vehículos registró caídas de hasta un 60 % del consumo en Ontario.

En resumen, los últimos estudios relativos al impacto del COVID-19 en los sistemas eléctricos analizan cómo la demanda cambia sus patrones de consumo, principalmente, desde una perspectiva sistémica. Cabe destacar que en [16] [17] analizan el impacto del sector residencial a partir del comportamiento de un conjunto de hogares. Además, algunos trabajos como [6] y [11] desagregan el impacto del COVID-19 en distintas industrias según su sector económico. En esta línea, este artículo analiza el impacto de las cuarentenas asociadas al COVID-19 en la demanda de energía eléctrica desagregada por tipo de cliente (residencial, comercial e industrial) en Chile. Además, se analiza el impacto del COVID-19 en los consumos eléctricos de distintos niveles socioeconómicos de Santiago, Chile.

3. Metodología

En este capítulo se describen los procedimientos metodológicos para analizar el impacto del COVID-19 en los consumos eléctricos de distintos tipos de clientes. Dado el origen de los datos, la metodología se divide en dos partes principales: el estudio de clientes regulados, es decir, los clientes residenciales y pequeños comercios, y el estudio de clientes industriales.

3.1. Análisis de consumos regulados: residenciales y comerciales

A continuación, se detalla la metodología para estudiar el impacto de las políticas asociadas al COVID-19 en la demanda de clientes regulados, es decir, en la demanda de clientes pequeños, formalmente denominados residenciales y comerciales.

Para el estudio de estos segmentos se utilizó la información de 230 mil equipos de medición inteligente en clientes de la empresa Enel Distribución instalados en la ciudad capital, Santiago (en 32 comunas). Estos datos se

encuentran protegidos por un Non-Disclosure Agreement (NDA) y han sido suficientemente agregados en este artículo con el fin de proteger la privacidad de los consumidores.

Los datos de consumos provenientes de los equipos de medición inteligente contienen la siguiente información para cada cliente:

- Datos de ubicación: En este ítem se incluyen las coordenadas georeferenciales, dirección, y la comuna a la cual pertenece el cliente.
- Datos comerciales: Incluye el tipo de facturación (que se utiliza para diferenciar entre clientes residenciales y de pequeño comercio) y el tipo de tarifa o plan que tiene el cliente con la empresa.
- Consumos eléctricos: La energía consumida con granularidad horaria desde enero de 2019 hasta septiembre de 2020 en kWh.

Aparte del estudio general de las 32 comunas, los casos de estudio más detallados se focalizan en 5 comunas representativas de las diferentes realidades socioeconómicas. Estas comunas son: Las Condes, Santiago, La Florida, La Cisterna y Renca. En el Anexo se presentan los números de medidores por comuna y por tipo de cliente utilizados para los experimentos.

De este modo, se analizan los impactos del COVID-19 en los clientes regulados a través de los siguientes cuatro casos de estudio:

3.1.1. Caso i: Tendencias en los consumos semanales de clientes residenciales y comerciales durante la pandemia

Se estudia la variación de demanda eléctrica semanal con respecto a la primera semana de marzo de 2020 (semana 10). Durante esta semana, el Gobierno de Chile no había decretado estado de emergencia aún, por lo tanto, es una semana que no fue afectada por la pandemia. El consumo de cada medidor inteligente se agrega (suma) a nivel semanal y luego las muestras se promedian dentro de una comuna para cada semana. Así, para cada semana, se determina el consumo promedio para un cliente residencial y un cliente comercial en una comuna específica. De este modo se analizan las variaciones en el tiempo de dicho consumo promedio para las cinco comunas representativas.

3.1.2. Caso ii: Separación de los efectos de la pandemia y de la estacionalidad del clima en las tendencias de consumos residenciales

Se estudian los cambios de consumo eléctrico en clientes residenciales cuya demanda es poco sensible a las bajas temperaturas. De esta manera, en este segmento particular de clientes no sensibles a temperaturas, las alzas de consumo del año 2020 con respecto al año 2019 se pueden asociar principalmente a los efectos de la pandemia del COVID-19 y de las políticas sanitarias tomadas por las autoridades.

Para identificar estos clientes, se analizan sus consumos promedios durante semanas frías (semana del 22 de abril y el 6 de mayo de 2019) y semanas cálidas durante el 2019 (semana del 4 de marzo de 2019). En este artículo, un cliente se considera poco sensible a las temperaturas, si su consumo promedio durante semanas frías es similar al de semanas cálidas (estableciendo una cierta tolerancia para pequeñas diferencias que se pueden considerar despreciables).

Para este conjunto de clientes poco sensibles a las temperaturas, se analiza la variación del consumo mensual promedio del 2020 con respecto a su consumo mensual promedio del mismo mes del 2019. Los porcentajes de variación se atribuyen a los efectos de las políticas sanitarias para enfrentar el COVID-19.

3.1.3. Caso iii: Cambios conductuales (a nivel horario) de consumos residenciales

A diferencia de los casos anteriores, en este caso de estudio se analizan las variaciones de los consumos hora a hora durante semanas representativas incluyendo el periodo pre-pandemia (representado por la semana del 2 de marzo), el período inicial de cuarentenas intensivas (representado por las semana del 20 de abril), el período de cuarentenas totales (representado por la semana del 22 de junio), y el periodo de apertura gradual (representado por la semana del 21 de septiembre). Esto se realiza con el fin de analizar los cambios en los patrones de consumo hora a hora, los que correlacionan de forma importante con los hábitos diarios de los consumidores (hora de despertar, jornada laboral, comidas, descansos, etc). En esta línea, se analizan, además, los adelantos/retrasos temporales de la demanda peak intradiaria. Este análisis se realiza solamente para dos de las cinco comunas representativas (las comunas de Las Condes y Renca).

3.1.4. Caso iv: Visión general del impacto de la pandemia en las32 comunas de la Región Metropolitana

Se analiza el panorama general del impacto de la pandemia en la región. Para esto se cuantifica la variación de la demanda comunal promedio durante semanas representativas con respecto a la semana del 2 de marzo (semana 10). Las semanas representativas utilizadas en este análisis son las mismas que se describen en el caso iii.

Finalmente se discute acerca de las diferencias del impacto del COVID-19 en las distintas comunas según su nivel socioeconómico.

3.2. Metodologías para el consumo industrial según sector económico

Se cuantifica el impacto de la pandemia en los consumos eléctricos de clientes industriales cuyo medidor es directamente accesible y cuyos datos son publicados por el Coordinador Eléctrico Nacional (operador del sistema eléctrico chileno). Estos corresponden a 2500 grandes clientes que informan sus consumos al operador del sistema. En este artículo, estos consumos se agrupan según el rubro económico del cliente, de acuerdo con los registros del Servicio de Impuestos Internos (dicho servicio define 17 rubros para los clientes eléctricos en cuestión).

De esta forma se compara el consumo eléctrico promedio mensual de cada sector económico durante el 2020 (entre los meses de enero y septiembre) con respecto a los mismos meses de 2019. Este análisis se hace para cada rubro económico. Así, se indica cuales son las industrias que más han visto afectados sus consumos por motivos de la pandemia del COVID-19.

4. Resultados y discusión

Este capítulo muestra y discute los resultados en términos del impacto de las políticas implementadas para combatir el COVID-19 en la evolución de los consumos eléctricos de los clientes regulados de la Región Metropolitana (i.e. residenciales y de pequeño comercio) y el impacto en los consumos industriales de los distintos sectores económicos a nivel nacional. El periodo de estudio va desde la semana del día 2 de marzo (semana 10) hasta la semana del 21 de septiembre de 2020 (semana 39).

4.1. Tendencias del consumo regulado: residenciales y comerciales

En esta subsección se describen los resultados de los cuatro casos de estudios para los clientes regulados.

4.1.1. Caso i: Tendencias en los consumos semanales de clientes residenciales y comerciales durante la pandemia

La Figura 2 muestra la variación de los consumos regulados durante el período de estudio. En áreas de colores se destacan las distintas políticas sanitarias tomadas por las autoridades del Gobierno de Chile, las que se describen en la Tabla 1. Se observa que los consumos residenciales han sufrido una alza con respecto a la semana 10 (2 de marzo). Este resultado es contrario a las tendencias sistémicas (Figura 1), que muestran una baja en la demanda eléctrica durante los meses de mayo a julio.

Tabla 1: Políticas sanitarias aplicadas por el Gobierno de Chile durante la pandemia.

antiene desde el a el fin del hori-				
cia con una cua-				
comunas del sec-				
tor oriente de Santiago, y termina				
ento de éstas. Al				
eríodo, sólo dos				
cen en una cua-				
ntiago y Ñuñoa)				
odo, las autori-				
dades aumentan gradualmente el				
número de comunas en cuarentena.				
e período las 32				
io están en cua-				
anas las autori-				
entena a toda la				
ana, permitien-				
s esenciales.				
del 29 de junio,				
producto de las mejoras en los índi-				
os, las autorida-				
confiamiento por				
_				

La Figura 2 muestra que desde la aplicación de los cierres de colegios en

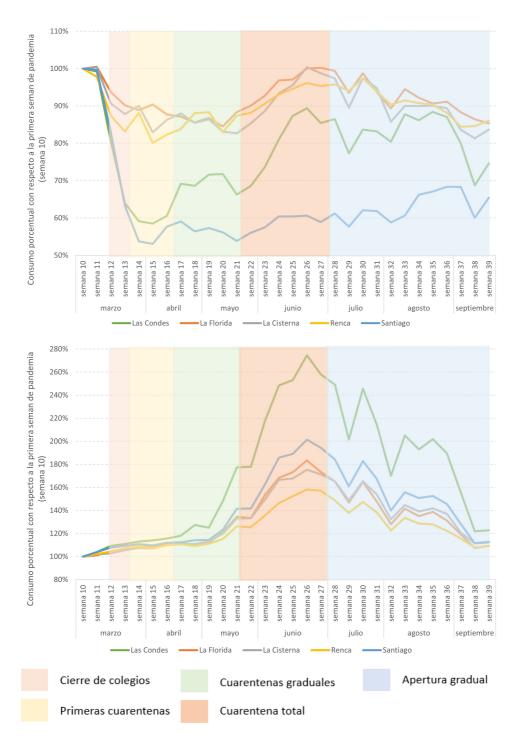


Figura 2: Variación de demanda eléctrica regulada comercial (gráfico superior), y residencial (gráfico inferior) durante las semanas de pandemia en 5 comunas de Santiago.

la semana 12 (16 de marzo) los pequeños consumos comerciales tuvieron una reducción paulatina de sus demandas en comparación a la semana 10 (2 de marzo). Las reducciones tuvieron su peak en la semana 15 (6 de abril). De esta forma, en la comuna de Santiago y Las Condes se alcanzaron niveles de reducción de $47\,\%$ y $42\,\%$ respectivamente en comparación a la semana 10. El resto de las comunas exhibió disminuciones que oscilan en torno al $15\,\%$ de disminución durante la semana 15.

Entre la semana 15 y la semana 20 (11 de mayo), los consumos eléctricos de pequeños comercios exhiben una recuperación coincidente con el periodo de cuarentenas graduales. Por ejemplo, si Las Condes consumía un 58 % durante la semana 15 (con respecto a la semana 10), esta cifra se eleva a un 72 % durante la semana 20. Esta recuperación en los consumos eléctricos se ve fuertemente afectada por la política de cuarentena total aplicada durante la semana 21 (18 de mayo). Las Condes, durante dicha semana, redujo sus consumos en un 66 % con respecto a la semana 10. Asimismo, los pequeños comercios de la comuna de Santiago son los que más vieron afectados su demanda eléctrica durante la semana 21 (18 de mayo), consumiendo en torno al 57 % del valor que exhiben en la semana 10.

Asimismo, gran parte de los consumidores de pequeños comercios muestran un peak de consumo en torno a la semana 26. Esta semana coincide con la época de más bajas temperaturas del invierno (ver Figura 3).

Finalmente, durante julio y agosto, los consumos mantienen una tendencia relativamente estable. Los consumidores de Santiago fueron quienes más redujeron su demanda, consumiendo en torno al $63\,\%$ en promedio, durante los meses de julio y agosto, con respecto a lo que consumían durante la semana 10.

La Figura 2 también muestra las tendencias de los consumos residenciales. Durante las primeras semanas de pandemia, de la mano con el cierre de los colegios (semanas 12-14), los consumos de Las Condes y Santiago aumentaron en aproximadamente un 10 % con respecto a la semana 10. Por otro lado, consumos residenciales en comunas como Renca y La Cisterna presentaron un menor nivel de respuesta, en torno al 5 %. A partir de finales de abril (semana 18), el efecto de la pandemia se combina con las menores temperaturas en Santiago que también influyen en el comportamiento de los consumos residenciales, aumentando sus demandas de forma significativa.

El peak de consumos ocurre en la semana 26 (22 de junio). Durante esta semana, los clientes de Las Condes aumentaron su consumo 2.7 veces (en promedio) en comparación a su consumo de la semana 10, mientras que, en Renca los consumidores aumentaron la demanda 1.6 veces.

Posterior a la semana del peak, la política de apertura gradual y el alza

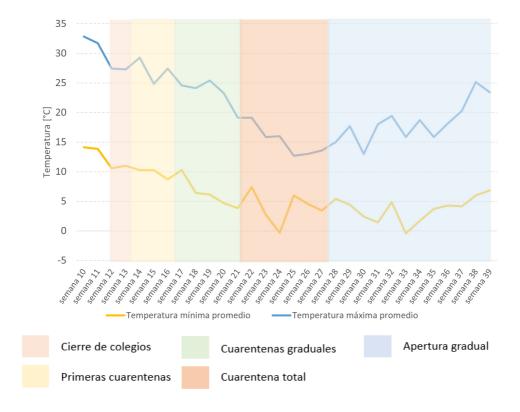


Figura 3: Variación de las temperaturas máximas y mínimas semanales (promedios de los siete días).

Nota: En colores se describen las distintas políticas sanitarias implementadas por la autoridad.

en las temperaturas coinciden con una caída en los consumos residenciales. Luego de la llegada de la primavera en la semana 39, el alza de consumos de los clientes de Las Condes alcanzó un 23% (con respecto a la semana 10), mientras que las demás comunas se ubicaron en torno al 10% con respecto a la primera semana de marzo.

Un aspecto relevante a destacar es que las variaciones anteriores corresponden a promedios comunales. No obstante, existe una dispersión importante dentro de cada comuna. En este sentido, la Figura 4 presenta las dispersiones en las variaciones para dos comunas: Renca (arriba) y Las Condes (abajo). Por ejemplo, se ilustra que existe un 20 % y 7,5 % de clientes residenciales en Las Condes y Renca, respectivamente, que presentan más del doble del consumo en la semana 21 (semana del 18 de mayo) con respecto a la semana 10 (semana del 2 de marzo). Esta dispersión aumenta en los meses de invierno, así, es posible observar que la cola de la distribución de la semana 26 es más

grande que la de la semana 21, debido a que una porción los clientes aumentan sus consumos por las bajas temperaturas.

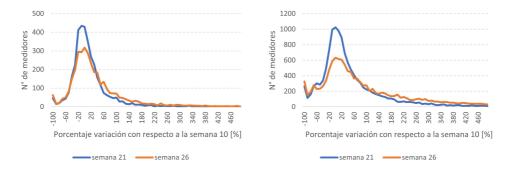


Figura 4: Histogramas de variación porcentual de los consumos con respecto a la semana 10 en las comunas de Renca y Las Condes.

Nota: El gráfico a la izquierda muestra el caso Renca y el gráfico a la derecha muestra el caso de Las Condes para las semanas 21 (18 de mayo) y 26 (22 de junio). Estos histogramas barren el rango desde -100% hasta 490% en intervalos de 10%.

Del mismo modo es posible observar también una dependencia importante entre los consumos de las distintas comunas y las políticas sanitarias. Esto se puede observar más claramente en las primeras semanas de la pandemia (semanas 12-16), donde todas las comunas estudiadas presentan un cambio importante de comportamiento (entre un 7% en La Cisterna y un 16% en Las Condes con respecto a la semana 10) debido a las primeras medidas de confinamiento tomadas por el Gobierno de Chile. Estas semanas son previas al aumento de demanda producto de las bajas temperaturas.

Finalmente cabe destacar que las mayores alzas en los consumos coinciden con los sectores socioeconómicos de más alta renta. La Figura 5, elaborada a partir de datos del Instituto Nacional de Estadísticas [7], muestra la distribución de los salarios brutos en 32 comunas de la Región Metropolitana en términos mensuales. Se puede observar que comunas del sector oriente, donde se ubica Las Condes, presentan los ingresos salariales más altos (2.300 USD mensuales en promedio). Otras comunas como Santiago y La Florida presentan un salario medio (en torno a los 1000 USD en promedio). Finalmente, comunas como La Cisterna o Renca presentan los menores ingresos de la región (en torno a los 650 USD en promedio).

De esta forma, las mayores alzas se pueden observar en la comuna de Las Condes (comunas de ingresos mayores), mientras que, los menores cambios se observan en Renca y La Cisterna (comunas de ingresos menores). A modo de hipótesis, la diferencia de respuestas en las distintas comunas puede explicarse

por una posible menor respuesta al confinamiento [14] y por una menor respuesta a las bajas temperaturas producto de un menor nivel de electrificación de las soluciones de calefacción en las comunas de menores rentas.

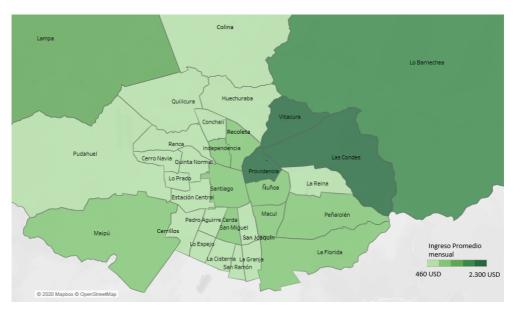


Figura 5: Distribución de los ingresos promedios mensuales de acuerdo al Instituto Nacional de Estadísticas [7].

4.1.2. Caso ii: Separación de los efectos de la pandemia y de la estacionalidad del clima en las tendencias de consumos residenciales

El alza de los consumos eléctricos residenciales indicados en la Figura 2 se puede explicar, principalmente, por dos efectos: las cuarentenas y las temperaturas que son cada vez más bajas a medida que se acerca el invierno. Aunque estos efectos pueden actuar combinadamente (i.e., una familia puede consumir más electricidad durante días fríos, en situación de pandemia, producto que existen más personas habitando el hogar por un mayor número de horas al día), en esta sección se busca separar estos dos efectos para tener una mejor noción del impacto de la pandemia. Es así cómo se analizan las alzas de los consumos eléctricos del año 2020 con respecto a los consumos durante el mismo periodo del año 2019 en una submuestra de consumidores poco sensibles a los cambios de temperatura.

De esta manera se calculan las alzas promedios de estos clientes en el conjunto de comunas representativas. Estos resultados se resumen en la Tabla 2, donde se puede observar que durante abril estas comunas experimentaron alzas del 9 % con respecto al mismo período del año 2019. Estas alzas sufrieron una leve alza durante mayo, alcanzando el 12 % en comparación al 2019. Este mes coincide con la aplicación de la política de cuarentenas graduales. Durante los meses de junio a agosto se observan los impactos más fuertes con aumentos de hasta un 17 % con respecto al mismo período de 2019. Durante estos meses, las políticas sanitarias implementadas fueron la cuarentena total y la apertura gradual (desde el 29 de julio). Finalmente, durante septiembre, gran parte de la Región Metropolitana había abierto sus actividades parcialmente, de este modo, el alza de consumos para este mes alcanzó el 7 % con respecto al mismo mes en 2019, siendo el menor de todos los valores del periodo.

Tabla 2: Variación de demanda eléctrica 2020 poco sensible a los cambios de temperatura con respecto al mismo mes 2019.

Mes	Incremento de consumo [%]
Marzo	9
Abril	8
Mayo	12
Junio	17
Julio	13
Agosto	10
Septiembre	7

Es importante mencionar que los resultados anteriores presentan variaciones promedios, no obstante, existe una dispersión asociada. En este sentido, la Tabla 3 muestra la proporción de clientes que más que duplicaron sus consumos en distintas comunas y meses. Se revela, por ejemplo, que comunas como Santiago y Las Condes presentan un 8 % y 10 %, respectivamente, de clientes que han más que duplicado su consumo producto del confinamiento durante el mes de mayo (con respecto al mismo mes 2019). En otras comunas como Renca y La Cisterna, estos clientes representan aproximadamente el 6 % de la muestra analizada. El periodo más crítico comprende los meses desde junio a agosto, donde la cantidad de clientes que duplicaron su consumo oscila entre el 10 % y el 17 % con respecto a igual periodo en el año 2019 en las comunas consideradas.

4.1.3. Caso iii: Cambios conductuales (a nivel horario) de consumos residenciales

Es interesante observar el cambio en el comportamiento de los consumos residenciales que se observan en los perfiles horarios de la Figura 6 para los

Comuna	mar	abr	may	jun	jul	ago	\mathbf{sept}
Las Condes	4%	7 %	10%	17%	16%	17%	11%
La Florida	4 %	5 %	7 %	17%	16%	15%	12%
La Cisterna	2%	4%	6%	10 %	15%	14%	17%
Renca	2%	4 %	6%	15%	14%	13%	10%
Santiago	4 %	7 %	8%	14%	14 %	14%	10%

Tabla 3: Proporción de la demanda 2020 que presenta aumentos de consumo igual o mayor a dos veces con respecto al mismo periodo 2019.

clientes residenciales de Las Condes y Renca. Primero, para ambas comunas durante la situación pre-pandemia (semana 10, correspondiente a la semana del día 2 de marzo e indicada en color azul) se observa la presencia de un peak pronunciado a las 10:00pm durante todos los días. Asimismo, se observa un peak menor entre las 8:00am y 9:00am durante los días laborales.

Entrando en la pandemia, durante la semana 17 (a partir del 20 de abril, indicada en gris), el peak diario aumenta y se adelanta 2 horas, hacia las 8:00pm. También, aparece un segundo peak diario a las 2:00pm (más asociado a un día de fin de semana como se puede observar en el perfil de consumo azul correspondiente a la semana 10), aumentando considerablemente el consumo de energía durante la tarde. Finalmente, desaparece el peak que se observaba en semanas previas durante la mañana (en torno a las 9:00am) en días laborales. En lugar de este peak matutino, los consumos exhiben un aumento paulatino hasta el peak de las 2:00 pm.

En la semana de máximo consumo, semana 26 (a partir del 22 de junio, indicada en amarillo), se observa un efecto combinado de la pandemia más las bajas temperaturas. Esto explica un aumento considerable en los niveles de demanda. El peak de consumos durante estas semanas ocurre a las 9:00 pm.

Finalmente, la semana 39 (a partir del 21 de septiembre, indicada en verde) se caracteriza por una disminución en los niveles de consumo, esto coincide con el alza de las temperaturas y la política de apertura paulatina. Además, en Las Condes durante septiembre se recupera el peak matutino (a las 10:00 am) que existía durante el periodo pre-pandemia.

Una diferencia relevante a destacar entre las comunas es la intensidad con que se presentan los cambios mencionados. Mientras en Las Condes estos cambios son más marcados, en Renca los cambios son más leves, lo que podrían explicarse por una posible menor respuesta al confinamiento (información consistente con informes de movilidad [18]) y por una menor respuesta a las bajas temperaturas producto de un menor nivel de electrificación de las soluciones de calefacción. Esto también es coincidente con los distintos niveles

socioeconómicos de las comunas que se observa en la Figura 5, ya que comunas de mayores ingresos podrían presentar un mayor acceso a soluciones de calefacción mediante equipos eléctricos y una mayor probabilidad de aplicar teletrabajo.

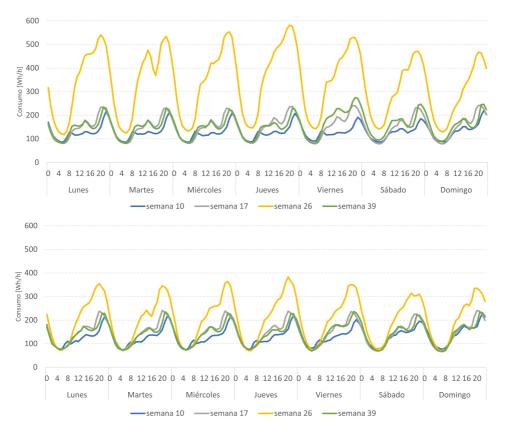


Figura 6: Consumo promedio horario de clientes residenciales en cuatro semanas y en dos comunas: Las Condes (gráfico superior) y Renca (gráfico inferior).

4.1.4. Caso iv: Visión general del impacto de la pandemia en las 32 comunas de la Región Metropolitana

Las Figuras 7 y 8 presentan una panorámica general de cómo ha aumentado la demanda residencial en todas las comunas de Santiago en el periodo de estudio. Este aumento a sido sustancial, llegando al 91 % (en promedio para las 32 comunas) durante la semana 26 (22 de junio) con respecto al periodo prepandemia (2 de marzo). Estas figuras refuerzan que la mayor diferencia en los consumos se ha presentado en la zona oriente de la capital, en comunas como Vitacura, La Reina, Providencia y Las Condes. En Vitacura, por ejemplo, el aumento de la demanda residencial en la semana 26 alcanza un 236 % con

respecto al consumo de la semana 10 (2 de marzo). En cambio, durante la misma semana, Cerro Navia fue la comuna con menos variaciones, alcanzando un alza del $46\,\%$ en comparación a la semana 10.

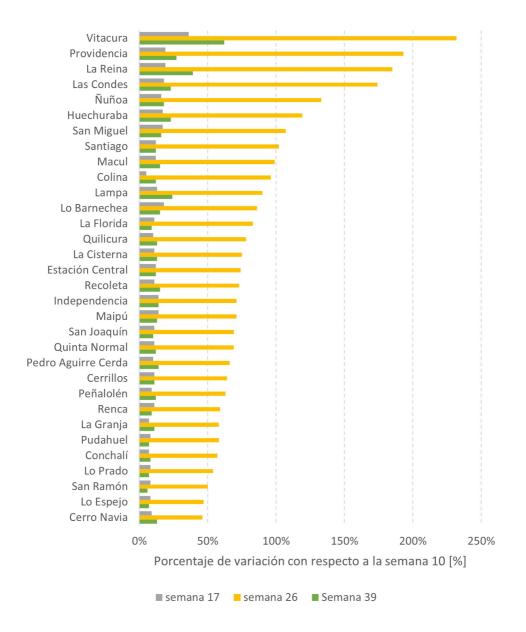


Figura 7: Variación de consumo de clientes residenciales en semana 17 (semana del 20 de abril), 26 (semana del 22 de junio) y 39 (semana del 21 de septiembre), todas con respecto a la semana 10 (semana del 2 de marzo).

C. SEPÚLVEDA, H. OTÁROLA Y L. BASSO

La diferencia de cambios en los consumos eléctricos en las comunas es coincidente con los niveles socioeconómicos, tal como se ha observado y discutido anteriormente.

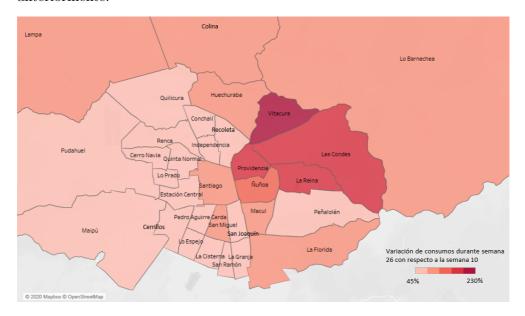


Figura 8: Distribución geográfica de los aumentos de consumos durante la semana 26 (semana del 22 de junio) con respecto a la semana 10 (semana del 2 de marzo) por efecto combinado de la estacionalidad y la pandemia.

4.2. Tendencias en el consumo industrial según sector económico

La Figura 9 muestra las variaciones del consumo eléctrico nacional a nivel mensual durante los últimos tres años. Si bien, el primer trimestre del año 2020 muestra un crecimiento relativo a 2019 cercano al 2.8 %, en línea con las proyecciones de la autoridad, desde abril a la fecha el consumo eléctrico muestra una caída importante alcanzando un retroceso de hasta -5.3 % durante el mes de julio. Una consecuencia directa de las medidas de la cuarentena a nivel nacional ha sido una disminución de la actividad económica, que según cifras oficiales, cayó en promedio un -13 % en el periodo abril-agosto 2020 [4].

La situación anterior, se ha manifestado de distintas formas en el consumo eléctrico de los principales rubros de la economía tal como se muestra en la Tabla 4. Es posible observar que, dentro de los rubros con más impactos en la reducción de su consumo eléctrico, destacan actividades como Hoteles y Restaurantes, junto con la Construcción. Estas fuentes laborales fueron las más afectadas a nivel nacional entre otros, por el cierre de gran parte de

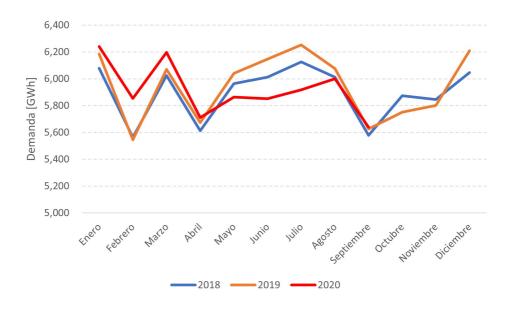


Figura 9: Demanda eléctrica mensual del Sistema Eléctrico Nacional [5].

sus actividades, un menor desplazamiento de la población en general, y la prohibición de viajes domésticos e internacionales.

La restricción de libre tránsito de los habitantes bajo cuarentena ha afectado también de manera importante al consumo eléctrico de otros rubros como la Intermediación Financiera (Bancos, Compañías de Seguro, Sociedades de Inversión, entre otros), las Industrias Manufactureras, el Comercio en general y la Enseñanza (tanto primaria, como secundaria y universitaria).

Una de las principales actividades económicas del país, la Explotación de Minas y Canteras, mostró un crecimiento relevante de su consumo durante la primera mitad del año, a pesar de la expansión de las medidas sanitarias a lo largo de Chile. No fue sino hasta el mes de julio, con la imposición de cuarentenas en las ciudades de Iquique y Antofagasta, que las principales faenas mostraron leves caídas en sus niveles de consumo. No obstante, a nivel anual, dicho sector económico sigue mostrando un crecimiento en su consumo respecto de los valores observados el año 2019. En esta misma línea, sectores que han mostrado un crecimiento en el consumo eléctrico con respecto al 2019 son los rubros Servicios Sociales de Salud, Administración Pública y Defensa, Actividades de Servicios Comunitario, y Pesca como se puede observar en la Tabla 4.

Tabla 4.	Variaciones	de consumo	2020	por rubro	con	respecto	al	año s	2019

Rubro		feb	mar	abr	may	jun	jul	ago	sep
Hoteles y Restaurantes	3 %	5%	-13 %	-13 %	-76 %	-75 %	-75 %	-74 %	-74 %
Construcción	-15 %	-13 %	-17 %	-9%	-42 %	-56 %	-68 %	-68 %	-73 %
Actividades Inmobiliarias,	-3%	-4%	14 %	-21 %	-15 %	-11%	-31 %	-21 %	-39 %
Empresariales y de Alquiler	-3 70	-4 /0	14 /0	-21 /0	-10 /0	-11 /0	-31 /0	-21 /0	-33 /0
Industrias Manufactureras	-6%	-4%	-16 %	-27%	-54 %	-57%	-53 %	-41 %	-39 %
no Metálicas	-0 70	-470	-10 /0	-21 70	-04 /0	-01 70	-00 70	-41 /0	-00 70
Consejo de Administración	-9%	-4%	-14 %	-9%	-46 %	-54 %	-53 %	-48 %	-30 %
de Edificios y Condominios	0 70	170	11.70	0 70	10 70	0170	30 70	10 70	90 70
Industrias Manufactureras	-5%	0%	5 %	12%	-23 %	-23 %	-31 %	-29 %	-24 %
Metálicas	0 70	0 70	0 70	1270	20 70	2070	91 70	20 70	2170
Comercio al por Mayor y Menor, Rep.	-18%	-9%	-5 %	19%	-47 %	-28 %	-21 %	-18 %	-16%
Veh. Automotores/ Enseres Domésticos									
Enseñanza	-1 %	-8%	-16 %	-17 %	-27 %	-36 %	-28 %	-26 %	-11 %
Industrias Manufactureras	-2%	3%	3 %	9%	-9 %	-20 %	-15 %	-16 %	-6%
no Metálicas									
Intermediación Financiera	-5 %	10 %	4 %	4 %	-8 %	-14 %	-15 %	-13 %	-5 %
Explotación de Minas	9 %	14%	7%	10%	1 %	2 %	-2 %	2 %	-3%
y Canteras	0 70	1170	. , 0	1070	170	- / 0	- 70	- / 0	0,0
Suministro de Electricidad	7%	15 %	10 %	12%	-4 %	-7%	-21 %	-11%	-3%
y Agua	. , 0	1070	1070	1270	170	. , ,	2170	1170	0,70
Agricultura, Ganadería,	1%	1%	4 %	10%	-2 %	-7%	-11 %	-7%	-2%
Caza y Silvicultura	170	170	1,0	1070	270	. , ,	1170	. 70	- 70
Servicios Sociales	6 %	8%	3 %	9%	-7%	-5%	-5 %	-3 %	1%
y de Salud	0 70	0,0	0 70	0 70	• /0	9 70	570	-3 70	170
Adm. Pública y Defensa, Planes	3 %	-5%	-2 %	6%	-6 %	-4%	-21 %	1%	3 %
de Seg. Social Afiliación Obligatoria	0,0	0,70	- / 0	0 70	0 70	170	2170	170	0,0
Otras Actividades de Servicios	28 %	35 %	9 %	16%	7%	19%	-6%	-3%	7%
Comunitarias, Sociales y Personales									
Pesca	20%	20%	-1 %	11 %	11%	8 %	2%	8%	8%

5. Conclusiones y Trabajo Futuro

Este artículo analiza el impacto de las cuarentenas asociadas al COVID-19 en la demanda de energía eléctrica desagregada por tipo de cliente (residencial, comercial e industrial) en Chile. Además, se analiza el impacto del COVID-19 en los consumos eléctricos de distintos niveles socioeconómicos de Santiago, Chile.

Este análisis demuestra que el impacto del COVID-19 en el consumo eléctrico varía ampliamente según el tipo de consumidor. Con respecto a consumidores regulados, por ejemplo, los pequeños comercios mostraron una baja importante en sus consumos. En este sentido, las comunas de Santiago y las Condes alcanzaron caídas superiores al $40\,\%$ con respecto a la situación inmediatamente pre-pandemia. Esto es consistente con la desaceleración de la economía chilena observada durante el mismo periodo de estudio.

Los clientes residenciales, por otro lado, mostraron incrementos en sus consumos desde el inicio de la pandemia con la implementación de las primeras políticas sanitarias. El período de mayores alzas ocurrió entre junio y agosto coincidiendo con los meses más fríos y con las políticas sanitarias más restrictivas. Particularmente, la semana de mayores alzas fue aquella del 22 de junio donde se alcanzó un alza del 91 % en promedio para Santiago con respecto a la primera semana de marzo. Esta alza es producto del efecto combinado de la pandemia y de las bajas temperaturas de invierno. Con respecto a las alzas únicamente producto de la pandemia, estas alcanzan un 17 % en el mes de junio con respecto al mismo mes del año 2019. Cabe mencionar que estos resultados varían entre las distintas comunas. De hecho, se observa que las comunas con mayores aumentos de consumo corresponden a aquellas con mayores rentas. Asimismo, al interior de una comuna los resultados también muestran una amplia dispersión en la variación de consumos.

Con respecto al impacto en los consumos eléctricos industriales, el sector con mayor disminución fue el Hotelero y Restaurantes con disminuciones de hasta el 75 % en los meses de mayo-septiembre con respecto al mismo periodo del año anterior. Otros rubros con caídas significativas en sus consumos eléctricos fueron el de Construcción, Actividades Inmobiliarias, Industrias Manufactureras No Metálicas, y Consejo de Administración de Edificios y Condominios. Otros sectores como la Pesca o Minería evidenciaron alzas en sus consumos durante gran parte del periodo de estudio, mostrando un menor nivel de impacto producto de las políticas sanitarias.

Una conclusión importante de este trabajo y con impactos en la política pública es que el confinamiento puede aumentar la demanda eléctrica residencial de forma importante, imponiendo un potencial costo adicional para las familias. Esto podría presentar problemas particulares para el caso de familias vulnerables, quienes, aparte del aumento de costos de servicios básicos, se pueden encontrar en situaciones financieras complejas, producto de la crisis económica asociada a la pandemia. Este es un problema que las autoridades deben analizar y resolver con especial cuidado.

Entre las posibles aplicaciones y extensiones de este estudio, se pueden identificar:

- La mejora de políticas públicas, en particular, focalizando aquellas asociadas a pobreza energética y familias vulnerables.
- La mejora de prácticas operativas (incluyendo prácticas y metodologías de predicción de demanda y despacho económico) de manera de operar el sistema de forma más económica y segura frente a situaciones producidas por las medidas que toma la autoridad para controlar la pandemia.

- La identificación de zonas de la red eléctrica con un mayor o menor nivel de estrés producto de la pandemia, lo que podría adelantar o atrasar, respectivamente, el plan de inversiones de las empresas eléctricas.
- La identificación de dependencias entre distintos sectores (e.g., energía, transporte, economía) con el fin de entender de mejor manera las interrelaciones entre dichos sectores y así los efectos globales de la pandemia.
- Mejorar la caracterización de los consumos eléctricos, incluyendo los impactos de la pandemia (o eventualmente otros fenómenos) en las categorías de clientes.

Finalmente, es importante recalcar que este tipo de estudios detallados de los consumos eléctricos no es solamente relevante en el contexto de la pandemia. En general, un monitoreo y análisis profundo de los datos de consumos energéticos tiene el potencial de informar y mejorar decisiones y procesos tanto en el ámbito público como privado. Ejemplos de esto último incluyen decisiones de política pública, diseño regulatorio, estudios tarifarios de redes eléctricas, inversión y operación del sistema, entre otros.

Agradecimientos: Los autores agradecen a Enel Distribución, SPEC Energy Consulting y el Ministerio de Energía por facilitar la información necesaria para este artículo. Asimismo, agradecen a la iniciativa COVID Analytics del Instituto Sistemas Complejos de Ingeniería ISCI. Esta investigación ha sido financiada por el Instituto Sistemas Complejos de Ingeniería (ANID PIA/BASAL AFB180003).

Anexo

La Tabla 5 muestra la cantidad de equipos de medición inteligente distribuidos en la Región Metropolitana. Inicialmente se cuenta con 326 mil equipos, sin embargo, por motivos de confiabilidad de los datos, finalmente, se consideran 230 mil.

Tabla 5: Cantidad de equipos de medición inteligente por tipo de cliente y comuna.

C	Daridan dal	Pequeño
Comuna	Residencial	comercio
Cerro Navia	6581	195
Lo Espejo	1713	20
San Ramón	3080	129
Lo Prado	7128	180
Conchalí	11917	530
Pudahuel	5097	74
La Granja	2301	70
Renca	3510	130
Peñalolén	5050	221
Cerrillos	6334	197
Pedro Aguirre Cerda	2709	56
Quinta Normal	9833	520
San Joaquín	5649	167
Maipú	9468	193
Independencia	6622	480
Recoleta	11769	806
Estación Central	12144	508
La Cisterna	6960	891
Quilicura	3645	83
La Florida	9936	260
Lo Barnechea	484	26
Lampa	1385	96
Colina	1008	95
Macul	5641	264
Santiago	26686	3033
San Miguel	8262	351
Huechuraba	1070	88
Ñuñoa	10633	682
Las Condes	12510	380
La Reina	5768	135
Providencia	9860	1112
Vitacura	3522	107
Total	218275	12079

Referencias

- [1] A. Abu-Rayash y I. Dincer. Analysis of the electricity demand trends amidst the COVID-19 coronavirus pandemic. *Energy Research & Social Science*, 68:101682, 2020. Disponible en http://www.sciencedirect.com/science/article/pii/S2214629620302577.
- [2] D. Agdas y P. Barooah. Impact of the covid-19 pandemic on the us electricity demand and supply: an early view from data. *IEEE Access*, 8:151523–151534, 2020.
- [3] A. Bahmanyar, A. Estebsari, y D. Ernst. The impact of different COVID-19 containment measures on electricity consumption in Europe. *Energy Research & Social Science*, 68:101683, 2020. Disponible en http://www.sciencedirect.com/science/article/pii/S2214629620302589.
- [4] Banco Central de Chile. Índice mensual de actividad económica, imacec, 2020. Disponible en https://www.bcentral.cl/en/web/banco-central/area/statistics/imacec.
- [5] CEN. Balance de transferencias. Coordinador Eléctrico Nacional (CEN), 2020. Disponible en https://www.coordinador.cl/mercados/docume ntos/transferencias-economicas/ventas-mensuales/2020-ventas-mensuales/.
- [6] Câmara de Comercialização de Energia Electrica. COVID-19: consumo de energia deve recuar, 2020. Disponible en https://bit.ly/36Psfd.
- [7] I. N. de Estadísticas. Encuesta suplementaria de ingresos, 2018. Disponible en https://www.ine.cl/estadisticas/sociales/ingresos-y-gastos/encuesta-suplementaria-de-ingresos.
- [8] K. Gillingham, C. Knittel, J. Li, M. Ovaere, y M. Reguant. The short-run and long-run effects of COVID-19 on energy and the environment. *Joule*, 4(7):1337 1341, 2020. Disponible en http://www.sciencedirect.com/science/article/pii/S2542435120302737.
- [9] IEA. Covid-19 impact on electricity. International Energy Agency (IEA), 2020. Disponible en https://www.iea.org/reports/covid-19-impac t-on-electricity.
- [10] W. Kanda y P. Kivimaa. What opportunities could the COVID-19 outbreak offer for sustainability transitions research on electricity and mobility? *Energy Research & Social Science*, 68:101666, 2020. Disponible en

- http://www.sciencedirect.com/science/article/pii/S221462962 0302413.
- [11] A. Leach, N. Rivers, y B. Shaffer. Canadian electricity markets during the covid-19 pandemic: An initial assessment. *Canadian Public Policy*, 46(S2):S145–S159, 2020.
- [12] Ministerio de Salud. Ministerio de Salud decreta cuarentena total para la ciudad de Santiago y seis comunas aledañas, 2020. Disponible en https://www.minsal.cl/ministerio-de-salud-decreta-cuarentena-total-para-la-ciudad-de-santiago-y-seis-comunas-aledanas/.
- [13] N. Norouzi, G. Zarazua de Rubens, S. Choupanpiesheh, y P. Enevoldsen. When pandemics impact economies and climate change: Exploring the impacts of COVID-19 on oil and electricity demand in China. *Energy Research & Social Science*, 68:101654, 2020. Disponible en http://www.sciencedirect.com/science/article/pii/S2214629620302292.
- [14] M. Olivares, M. Goic, G. Weintraub, J. Covarrubia, C. Escobedo, y L. Basso. El impacto de las dos primeras semanas de cuarentena masiva en la Región Metropolitana. *Instituto de Sistemas Complejos (ISCI)*, 2020. Disponible en https://isci.cl/wp-content/uploads/2020/06 /Reporte-3-ISCI-movilidad-11-de-junio.pdf.
- [15] G. Ruan, D. Wu, X. Zheng, H. Zhong, C. Kang, M. A. Dahleh, S. Sivaranjani, y L. Xie. A cross-domain approach to analyzing the short-run impact of COVID-19 on the US electricity sector. *Joule*, 4(11):2322 2337, 2020. Disponible en http://www.sciencedirect.com/science/article/pii/S2542435120303986.
- [16] I. Santiago, A. Moreno-Munoz, P. Quintero-Jiménez, F. Garcia-Torres, y M. Gonzalez-Redondo. Electricity demand during pandemic times: The case of the COVID-19 in Spain. *Energy Policy*, 148:111964, 2021. Disponible en http://www.sciencedirect.com/science/article/pii/S030 1421520306753.
- [17] S. Snow, R. Bran, M. Glencorss, y N. Horrocks. Drivers behind residential electricity demand fluctuations due to COVID-19 restrictions. *Energies*, 13:5738, 2020.
- [18] WHO. Who director-general's opening remarks at the media briefing on COVID-19. World Health Organization, marzo 2020. Disponible en https://www.who.int/dg/speeches/detail/who-director-general-

C. SEPÚLVEDA, H. OTÁROLA Y L. BASSO

 $s-opening-remarks-at-the-media-briefing-on-covid-19---11-m\\ arch-2020.$

[19] H. Zhong, Z. Tan, Y. He, L. Xie, y C. Kang. Implications of COVID-19 for the electricity industry: A comprehensive review. *CSEE Journal of Power and Energy Systems*, 6(3):489–495, 2020.

Instituto Sistemas Complejos de Ingeniería

INSTITUTO SISTEMAS COMPLEJOS DE INGENIERÍA



El Instituto Sistemas Complejos de Ingeniería (ISCI)

trabaja en un continuo que va desde el desarrollo de metodologías de frontera hasta aplicaciones innovativas en Sistemas Complejos de Ingeniería, en los cuales interactúan infraestructura, comportamiento humano y gestión.

GRUPOS DE INVESTIGACIÓN

- Recursos Naturales: Minería, Forestal y Acuicultura
- Transporte
- Localización y Logística
- Consumer Analytics
- Smartcities
- Data Science
- Organización Industrial
- Energía







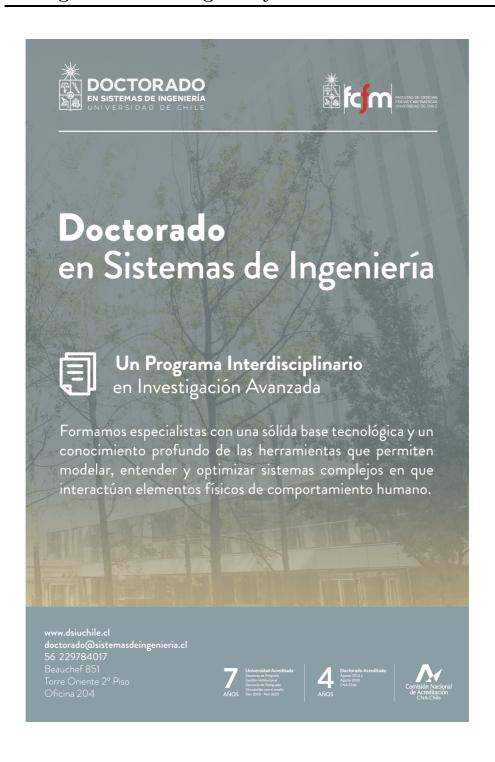








contacto@isci.cl www.isci.cl







MAGÍSTER EN GESTIÓN DE OPERACIONES







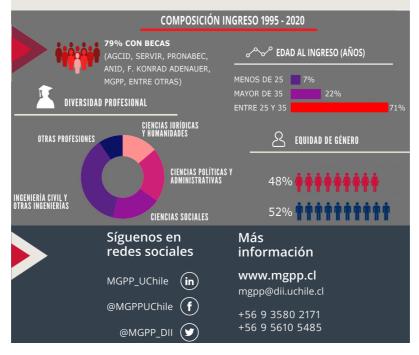


MAGÍSTER EN GESTIÓN Y POLÍTICA PÚBLICAS

PROGRAMA LÍDER EN CALIDAD Y COMPROMISO PÚBLICO

Referente en la formación de profesionales que cuentan con los conocimientos, las competencias y las habilidades para liderar los procesos de diagnóstico, comprensión, resolución y priorización de problemas públicos, vinculando el conocimiento con las necesidades de los países y resguardando los criterios de eficiencia, eficacia, impacto y legitimidad ciudadana de las acciones públicas.

REFERENTE EN AMÉRICA LATINA Y EL CARIBE EN LA FORMACIÓN EN GESTIÓN Y POLÍTICAS PÚBLICAS









¿Por qué elegir el **MBAMIN?**





ejecutivos claves de la industria minera y destacados académicos con postgrados en las mejores universidades del mundo.



Énfasis en los desafíos de productividad y competiti-vidad que enfrentan los productores mineros, y en las palancas claves de creación de valor sostenible.





Incorporación de nuevas Desarrollo de habilidades tecnologías, innovación y herramientas de punta adaptativa frente a los para la toma de decisiones. cambios y la innovación.

- En el **MBAMIN** podrás
- » Potenciar y desarrollar tu red contactos.
- Asistir a clases con un enfoque participativo.
- Participar del curso de especialización en el extranjero (optativo).

Abril

PROGRAMA WEEKEND clases cada dos semanas 09:00 - 21:00 hrs

MBA Versión Industria Minera | Ingeniería Industrial Campus República | República 701 | Santiago | Chile mbamin@dii.uchile.cl +562 2978 0554



Duración: 3 semestres académicos

www.mbe.cl

Contacto: mbe@dii.uchile.cl | +56 22 978 4835





EDUCACIÓN EJECUTIVA







