
INICIATIVAS MATEMÁTICO COMPUTACIONALES DESDE LA UNIVERSIDAD DE BUENOS AIRES PARA CONTRIBUIR A LA TOMA DE DECISIONES EN EL CONTEXTO DEL COVID-19

MEHRNOOSH ARRAR *	LAOUEN BELLOLI †	ANA MARÍA BIANCO *	LEONARDO BOECCHI *
RODRIGO CASTRO †	GUILLERMO DURÁN * ‡	ROBERTO ETCHENIQUE §	NATALIA FERNÁNDEZ ¶
LUCIANA FERRER †	DIEGO GARBERVETSKY †	RODRIGO GOLDSMIT	CAROLINA GRILLO
JUAN KAMIENKOWSKI †	PABLO LACIANA †	ESTEBAN LANZAROTTI †	MARIO LOZANO **
RODRIGO MAIDANA ††	MAURICIO MENDILUCE	SOL MINOLDO ††	LEONARDO PEPINO †
E. PECKER-MARCOSIG †	EZEQUIEL PUERTA †	RODRIGO QUIROGA §§	GUILLERMO SOLOVEY *
MARINA VALDORA *	MARIANO ZAPATERO †		

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 citados 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_{ef}) (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 Epicontracts [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 compartimento 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:

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

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

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

donde

$$N = S(t) + I(t) + R(t) \text{ (población total constante)}$$

$$R_{ef}(t) = \beta(t)/\gamma \times S(t) \text{ (número reproductivo efectivo, adimensional)}$$

$\beta(t)$ 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:

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

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

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 $k = 0, \dots, M$ 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

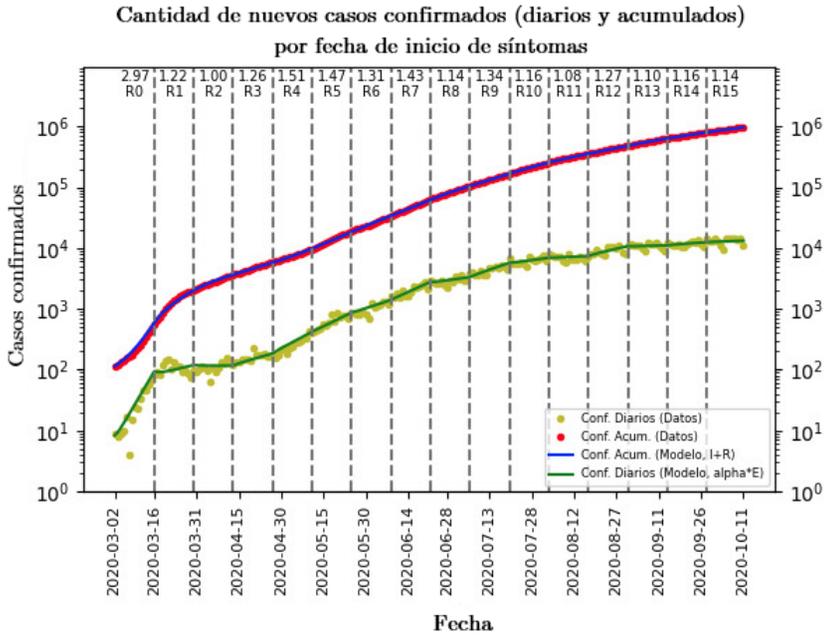


Figura 1: Evolución de $R_{ef}(t)$ para Argentina ajustando un modelo tipo SEIR.

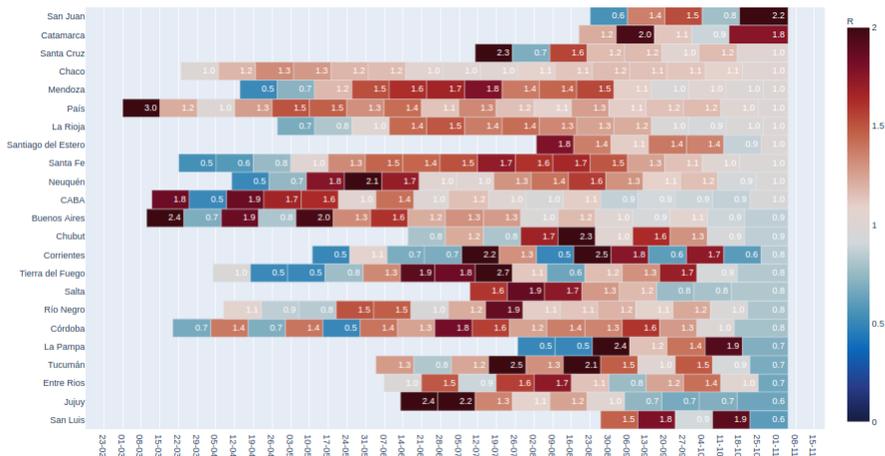


Figura 2: Mapa de calor temporal mostrando evoluciones del número reproductivo efectivo $R_{ef}(t)$ para las provincias argentinas (foto al 15/11/2020).
 Nota: Esta visualización es parte de los análisis automatizados disponibles en el sitio web descrito en la Sección 2.4.

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, $y_{(n-6)}, \dots, y_n$, y se ajustan los siguientes modelos:

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

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

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

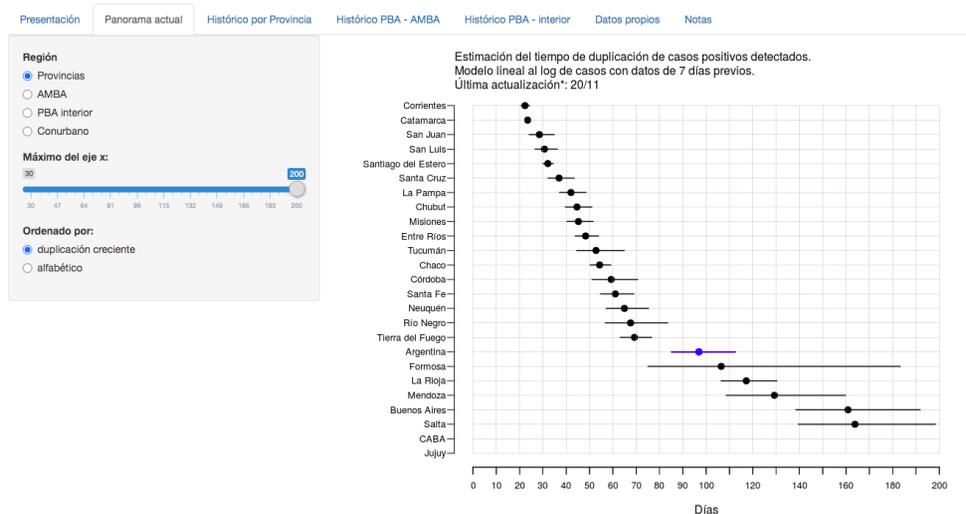


Figura 3: Captura de pantalla de la aplicación.

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 funcionamiento 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 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 construcción, 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-

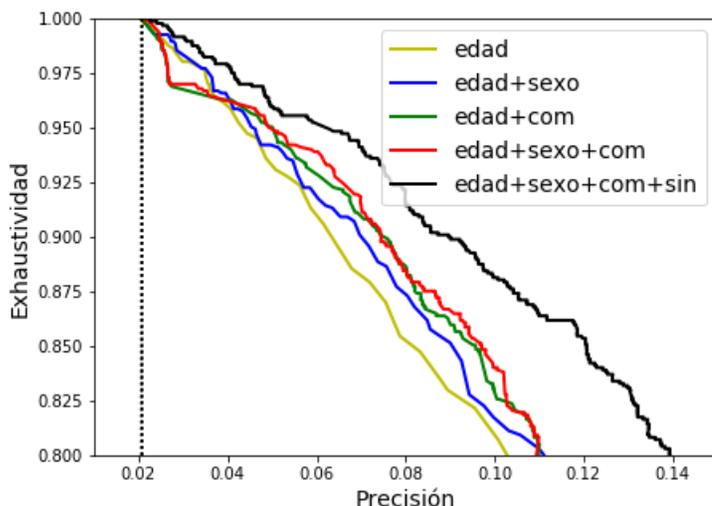


Figura 4: Curvas de exhaustividad y precisión en los datos de evaluación usando distintas listas de atributos de entrada.

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

El algoritmo 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 pronóstico 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 pronóstico. 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 produjo 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/

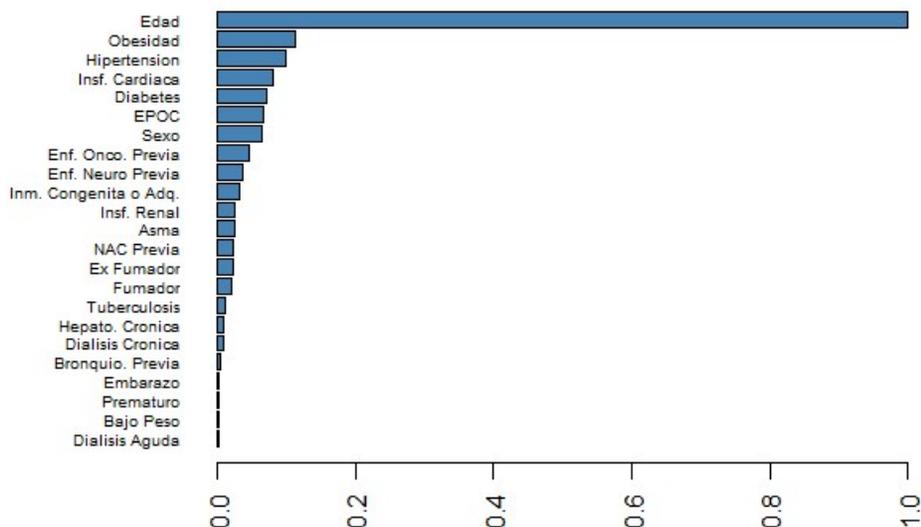


Figura 5: Importancia de los factores de riesgo a partir de 1000 árboles construidos usando remuestreo de los datos originales

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 $x \in \mathbb{R}^n$ representa al vector

de covariables. El modelo de regresión logística asume que $P(y = 1 | x) = \text{expit}(\beta T x)$, donde la función $\text{expit}(t) = 1/(1 + \exp(-t))$. Su inversa es la función $\text{logit}(p) = \log(p/(1-p))$ 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

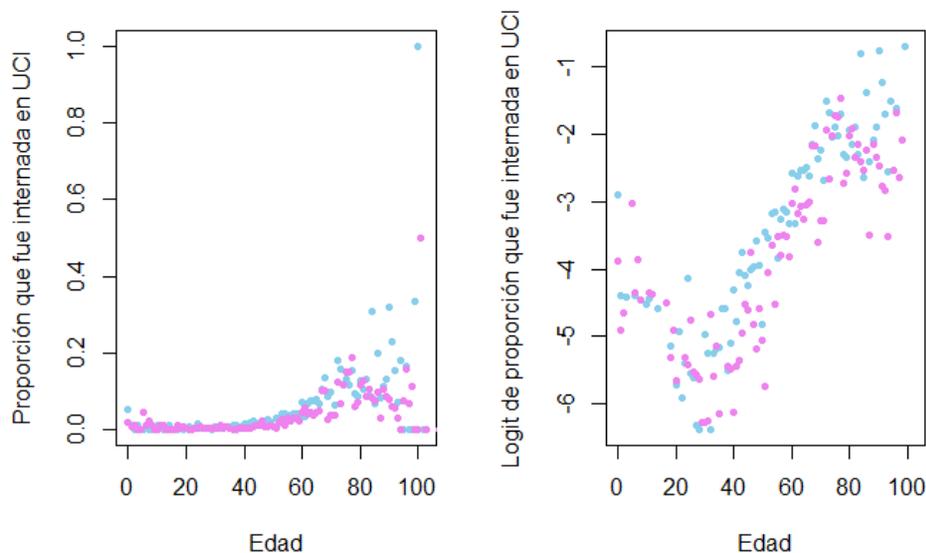


Figura 6: Proporción (panel izquierdo) y proporción en escala logit (panel derecho) de casos confirmados internados en UCI por edad y sexo (en celeste sexo masculino y en rosa femenino)

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).

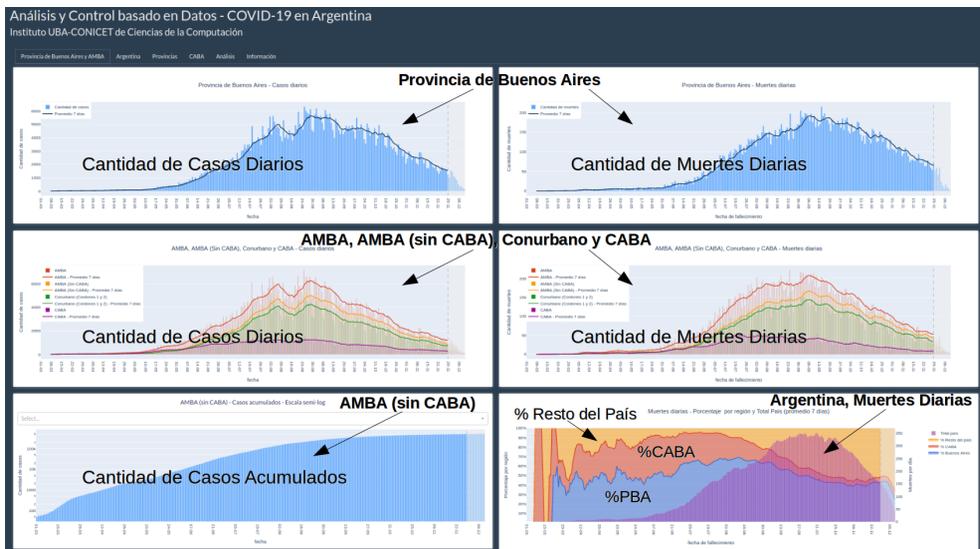


Figura 7: Vista del sitio mostrando casos confirmados diarios y fallecimientos diarios para la Provincia de Buenos Aires, AMBA, AMBA sin incluir la Ciudad Autónoma de Buenos Aires (CABA), Cordones 1 y 2 del Conurbano Bonaerense y CABA.

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.

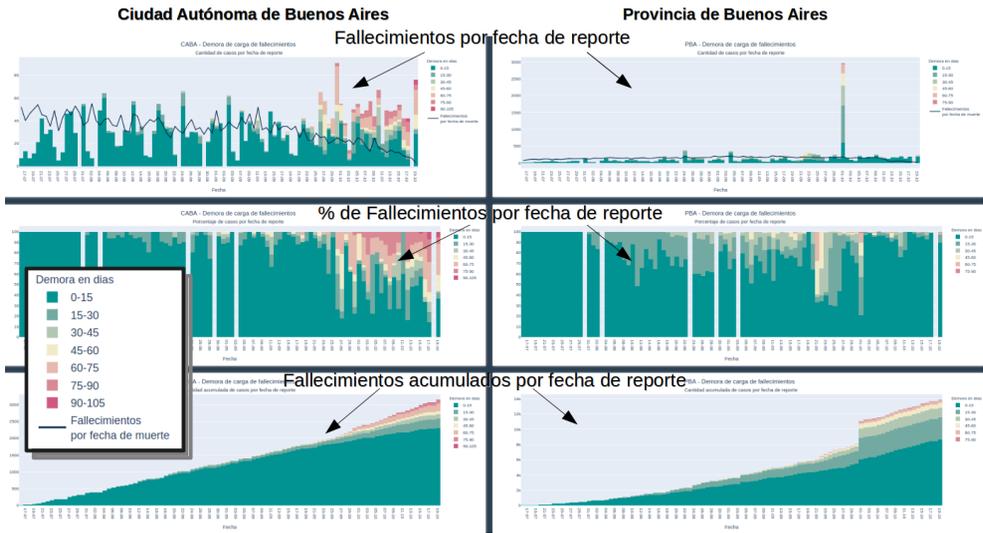


Figura 9: Vista del sitio mostrando la carga de fallecimientos y sus demoras. Nota: Izquierda: Ciudad Autónoma de Buenos Aires; Derecha: Provincia de Buenos Aires. Arriba: Cantidad de fallecimientos cargados por día; Centro: Porcentajes (normalizado a 100 %) de demoras sufridas por los fallecimientos 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 testing, 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 etario y sexo de España. Estimado en [21]

0-9		10-19		20-29		30-39		40-49	
M	F	M	F	M	F	M	F	M	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	F	M	F	M	F	M	F	M	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 entrega 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 $1/\text{IDR}$, 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.

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. Algunos estudios indicaron que, entre el 10 % y el 20 % de las infecciones son suficientes para detener la dinámica de crecimiento, en lugar del 60-70 % 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 seroprevalencia, 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 (f_2) 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 β_{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

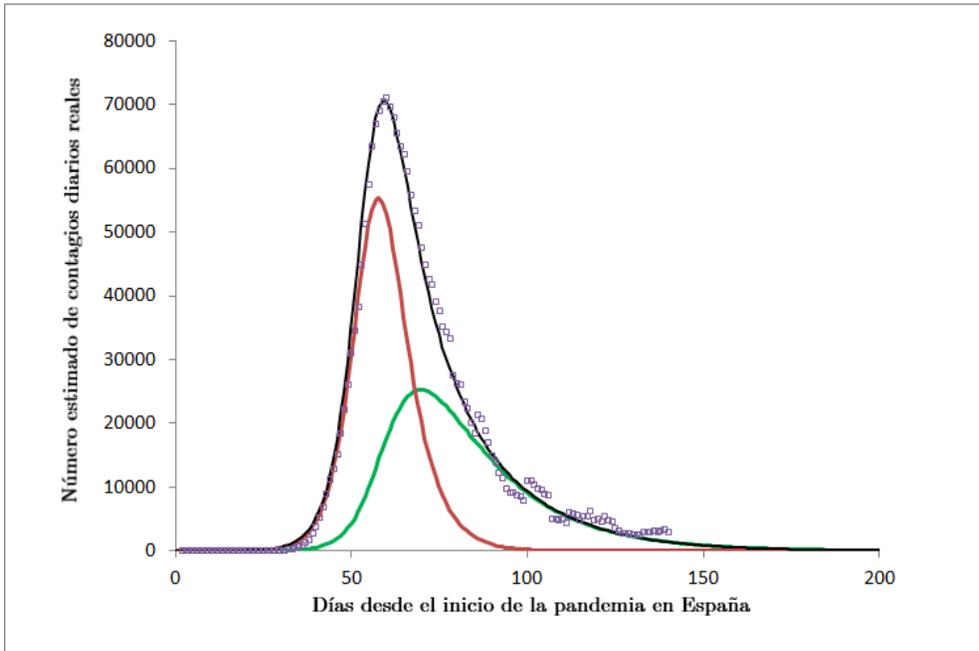


Figura 10: Patrón de inmunidad de “cagazo” para España.

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.

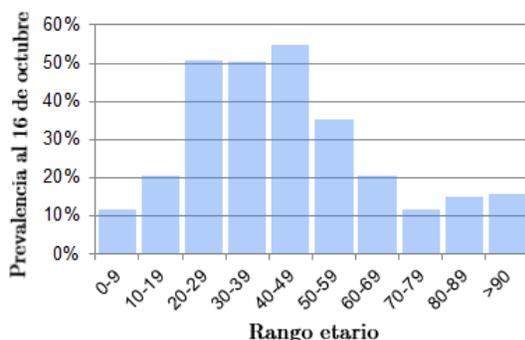


Figura 11: Prevalencia estimada para los diferentes rangos etarios en la Ciudad de Buenos Aires.

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.

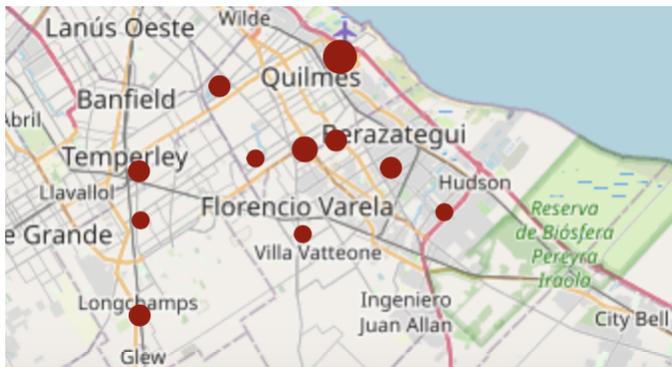


Figura 12: Localización de los hospitales de la región sudeste del Gran Buenos Aires, cuyo hospital central es el Hospital del Cruce de Florencio Varela.

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 CONICET, 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.

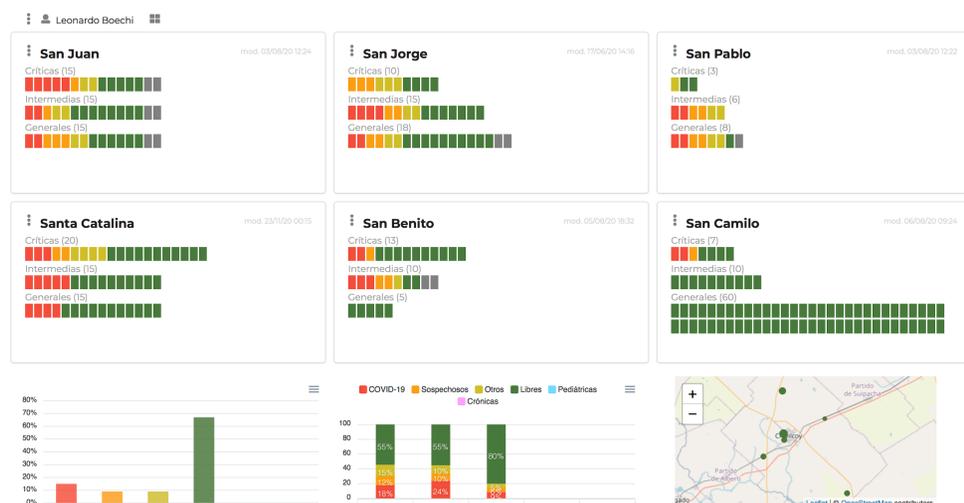


Figura 13: Panel Principal de Camas.

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:

1. 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.
3. 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.

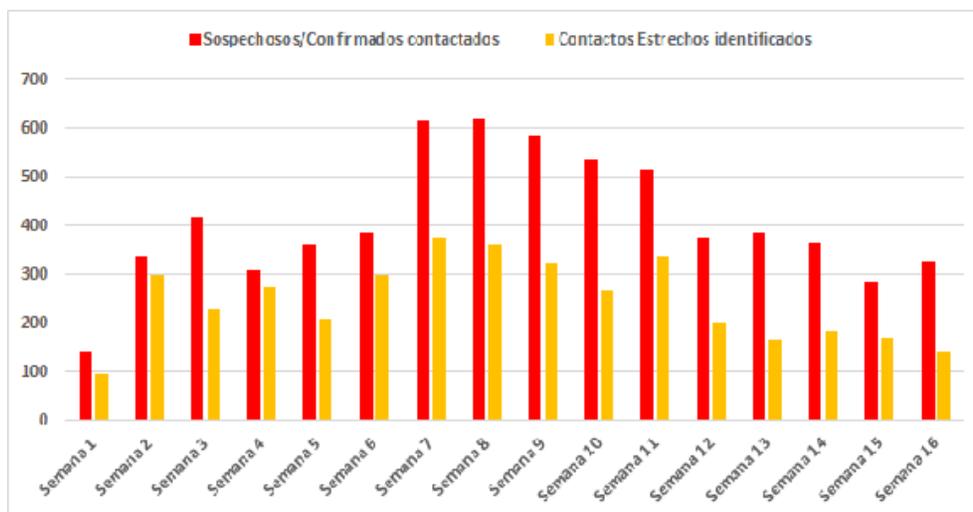


Figura 15: Número de sospechosos/confirmados y CEs identificados por semana del Centro.

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-

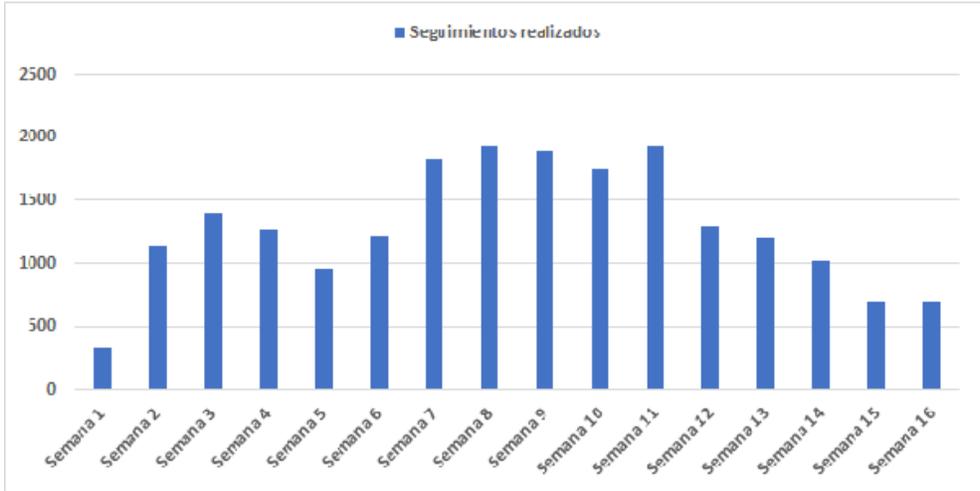


Figura 16: Número de seguimiento realizados a Contactos Estrechos por semana.

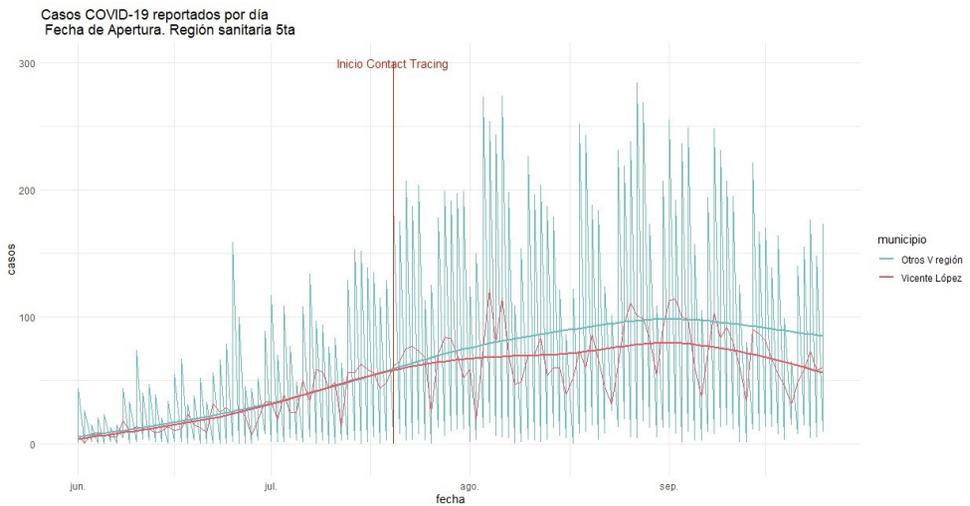


Figura 17: Evolución de casos confirmados a lo largo del tiempo para el municipio de Vicente López (en rojo) y para el conjunto de los otros municipios que corresponden a la misma Región Sanitaria que Vicente López (en verde), y sus respectivas curvas de ajuste.

Fuente: Antonio Milanese, Municipalidad de Vicente López.

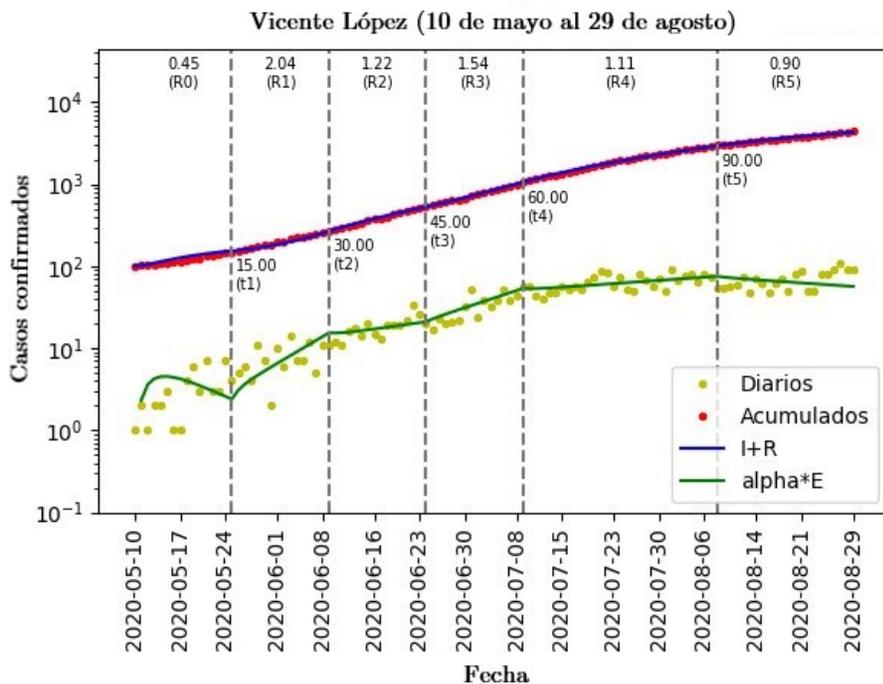
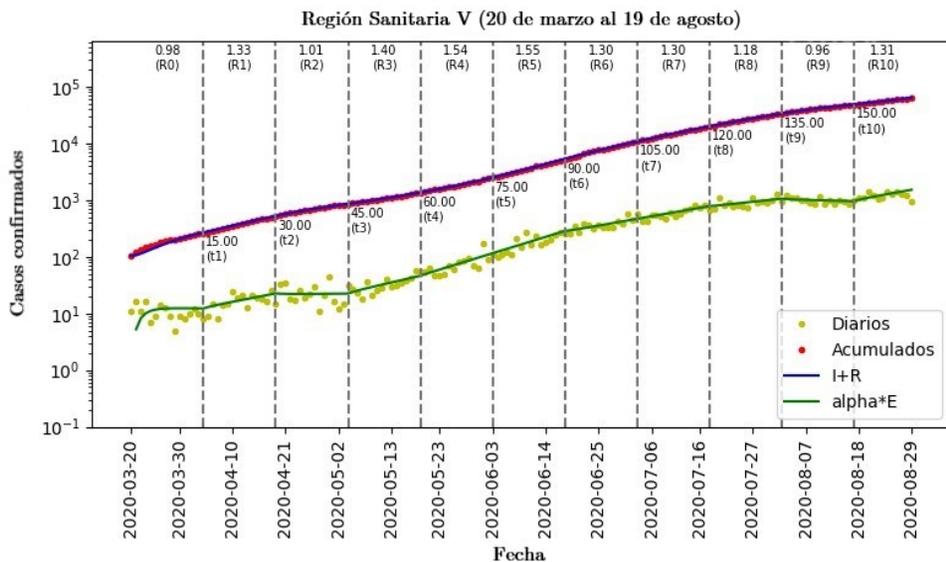


Figura 18: Evolución del R en Vicente López (abajo), y en el conjunto de los municipios de la Región Sanitaria V a la que pertenece Vicente López (arriba). Nota: Estos indicadores provienen del método de ajuste de parámetros basado en modelos de simulación descrito en la Sección 2.1 el cual es calculado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

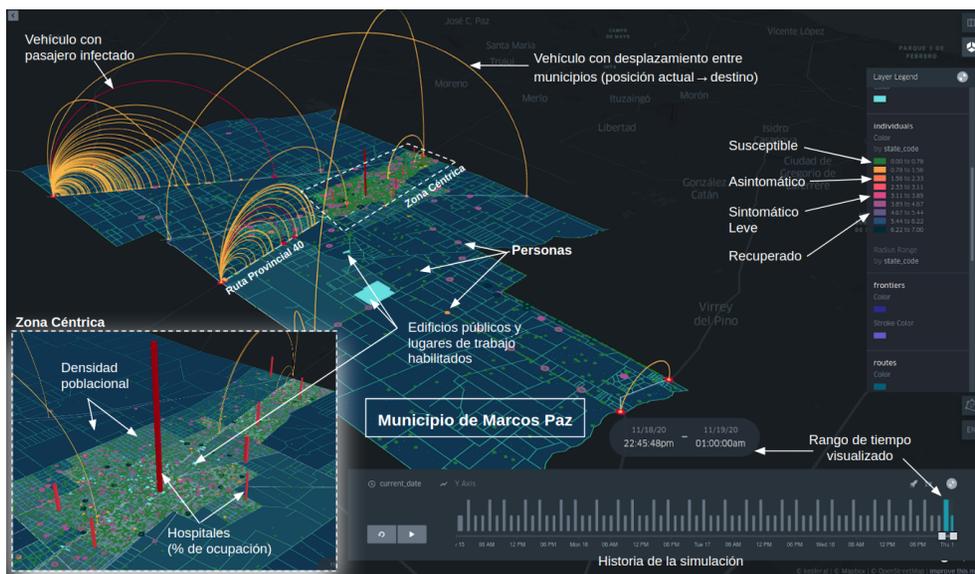


Figura 19: Visualización de simulación basada en agentes para el Municipio de Marcos Paz en la Provincia de Buenos Aires.

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 científico-tecnológico 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 “aplanaamiento” 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. Kamienskowski 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 PIA-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](https://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.

