

# Revista Española de Comunicación en Salud

Año 2024, Suplemento 1, 1-77

ISSN 2444-6513 (versión impresa)  
ISSN 1989-9882 (versión electrónica)  
Depósito Legal: M-29853-2015

Editada por:



uc3m

Universidad  
**Carlos III**  
de Madrid

## **Editor-Jefe**

Dr. Daniel Catalán Matamoros, Universidad Carlos III de Madrid, España

## **Editores asociados**

Dr. César Leal Costa, Universidad Católica de Murcia, España

Dra. Andrea Langbecker, Universidad Carlos III de Madrid, España

Dr. José Borja Arjona Martín, Universidad de Granada, España

Dra. Elva Martín Batista, Centro de Biología Molecular Severo Ochoa, España

Dr. Rubén Mirón González, Universidad de Alcalá, España

Dra. María del Carmen Rodríguez García, Universidad de Almería, España

## **Asistente Editorial**

Dra. Natalia Martínez Medina, Universidad de La Sabana, Colombia

## **Comité editorial**

Dr. Miguel Muñoz-Cruzado y Barba, Universidad de Málaga, España

Dr. Carlos Elías Pérez, Universidad Carlos III de Madrid, España

Dra. Patricia Rocamora Pérez, Universidad de Almería, España

Dra. María Teresa Icart Isern, Universidad de Barcelona, España

Dr. Carlos Javier vander Hofstadt Román, Universidad Miguel Hernández, España

Dra. Rosario Gómez Sánchez, Universidad Católica de Murcia, España

Dra. Anna Sendra Toset, University of Ottawa, Canadá

Dra. Verónica V. Márquez Hernández, Universidad de Almería, España

## **Comité Científico**

Dr. José Luis Blanco Terrón, InCom-UAB, España

Dr. Carlos Cachán-Alcolea, Universidad Nebrija, España

Dr. José Cerdán, Århus University Hospital, Dinamarca

Dr. Luís Manuel Cibanal Juan, Universidad de Alicante, España

Dr. Antonio Python Cyrino, Universidade Estadual Paulista, Brasil

Dr. Jesús Díaz-Campo, Universidad Internacional de la Rioja, España

Dra. Gea Ducci, Università di Urbino Carlo Bo, Italia

Dra. Petya Eckler, University of Strathclyde, Reino Unido

Dr. Edgardo Escobar, ITMS-Telemedicina, Chile

Dr. Javier Galán Gamero, Universidad Carlos III de Madrid, España

Dr. Sergio Godoy Etcheverry, Pontificia Universidad Católica de Chile, Chile

Dra. Cristina González Díaz, Universidad de Alicante, España, España

Dr. Santiago Graiño Knobel, Universidad Carlos III de Madrid, España

Dra. Lilibian Gutiérrez-Coba, Universidad de La Sabana, Colombia

Dra. Melva Herrera-Godina, Universidad de Guadalajara, México

Dr. Jeong-Nam Kim, University of Oklahoma, Estados Unidos

Dr. Knut Tore Lappegård, Universidad de Tromsø, Noruega

Dra. Remedios López Liria, Universidad de Almería, España

Dr. Alessandro Lovari, Università di Cagliari, Italia

Dr. Carlos Macía Barber, Universidad Carlos III de Madrid, España

Dra. María de Fátima Martins, Universidade do Minho, Portugal

Dra. M<sup>a</sup> Carmen Neipp López, Universidad Miguel Hernández, España

Dra. Carmen Peñafiel Saiz, Universidad del País Vasco, España

Dr. Pedro Ramos Contreras, Universidad Autónoma de Nuevo León, México

Dra. Doreen Reifegerste, Universität Erfurt, Alemania

Dra. Patricia Rocamora Pérez, Universidad de Almería, España

Dra. Carmen Roper Padilla, Universitat Jaume I, España

Dr. Peter Schulz, University of Lugano, Suiza

## Indexación en bases de datos

La revista RECS se encuentra indexada en las siguientes bases de datos:

- En la base de datos Emerging Sources Citation Index (ESCI), incluida en la plataforma Web of Science.
- En la base de datos ISOC y en las herramientas de análisis DICE (Difusión y Calidad Editorial de las Revistas Españolas de Humanidades y Ciencias Sociales y Jurídicas) producidas por el CSIC.
- En las bases de datos ERIH Plus, CIRC y MIAR (Matriz de Información para el Análisis de Revistas).
- En la plataforma REDIB (Red Iberoamericana de Innovación y Conocimiento Científico).
- Incorporada a CUIDEN® es una Base de Datos Bibliográfica de la Fundación Index.
- Incluida en el catálogo LATINDEX. - Referenciada en Dialnet (Universidad de La Rioja).
- Indizada en Academic Search Premier (EBSCO).
- Indizada en DOAJ (Directory of Open Access Journals).

La revista RECS ha obtenido el certificado de revista excelente y el Sello de calidad FECYT en 2021, 2022 y 2023.

## Nota de copyright

Los textos publicados en esta revista están – si no se indica lo contrario – bajo una licencia Reconocimiento - Sin obras derivadas 3.0 España de Creative Commons. Puede copiarlos, distribuirlos y comunicarlos públicamente siempre que cite su autor y la revista y la institución que los publica y no haga con ellos obras derivadas. La licencia completa se puede consultar en: <http://creativecommons.org/licenses/by-nd/3.0/es/deed.es>

Los derechos de autor pertenecen al autor de la obra por el mero hecho de su creación:

- Los derechos de contenido moral son irrenunciables e inalienables.
- Los derechos de contenido económico, o derechos de explotación, pueden ser cedidos a terceros como ocurre con los trabajos publicados, en los que el autor cede parte o la totalidad de estos derechos a la editorial.

El autor puede autoarchivar sus artículos en un repositorio institucional, siempre que se cite su publicación en esta revista.

## Declaración de privacidad

Los nombres y direcciones de correo introducidos en esta revista se usarán exclusivamente para los fines declarados por esta revista y no estarán disponibles para ningún otro propósito u otra persona.

**Revista Española de Comunicación en Salud**, publicación semestral. Revista de acceso abierto y gratuito Suscripciones mediante envío de email a [recs@uc3m.es](mailto:recs@uc3m.es)

### Contacto:

Prof. Dr. Daniel Catalán Matamoros

Despacho 17.2.23

Universidad Carlos III de Madrid, Calle Madrid 133

28903 Getafe, Madrid España

e-mail: [recs@uc3m.es](mailto:recs@uc3m.es) - Web de RECS y envío de artículos: <http://www.uc3m.es/recs>

## Predicción multifuente y multimétodo para dar soporte a la toma de decisiones frente a la COVID-19: El proyecto PredCov

### Multi-source and multi-method prediction to support COVID-19 policy decision making: The PredCov Project

Editorial

La predicción de incidencia de SARS-CoV-2 resulta en la actualidad un proceso altamente complejo debido a que está determinado por un gran número de factores, como es la aparición de nuevas variantes del virus o el nivel de inmunidad a la enfermedad existente en la población. En una era post-pandemia, con una relajación de las medidas de distanciamiento social, la movilidad urbana y estacionalidad también juegan un papel importante en la forma en que la enfermedad se propaga. Otro factor significativo son las diversas intervenciones farmacéuticas y no farmacéuticas adoptadas en España, que han tenido un impacto en los hábitos de la población, y en la economía, existiendo sectores que han resultado más perjudicados o beneficiados por la pandemia.

El objetivo de este proyecto es desarrollar un sistema que permite combinar distintas fuentes de información para modelar futuros escenarios de propagación del COVID-19 (u otras enfermedades respiratorias similares) desde un enfoque amplio que combina aspectos epidemiológicos, sociales, económicos y de medios de comunicación.

Esto ha supuesto el desarrollo de metodologías de actuación innovadoras en el campo de la analítica de datos que permite poder hacer un seguimiento en tiempo real de pandemias y adelantar el éxito o fracaso de determinadas acciones políticas o sociales que se pueden emprender para disminuir su impacto. También se ha planteado simular escenarios alternativos de implementación de la vacuna contra la COVID-19 para dar soporte a los responsables de definir las políticas de vacunación más apropiadas.

El trabajo realizado en este proyecto abarca ocho líneas de investigación que se describen a continuación.

La primera línea ha realizado el diseño, desarrollo e implementación de una solución de digital de rastreo de contactos que mejora las propuestas realizadas durante la pandemia y mantiene garantías de privacidad. La solución desarrollada se basa en el uso de información de geolocalización masiva. Se ha llevado a cabo un análisis de movilidad de ciudadanos basado en datos de geolocalización GPS durante distintas fases de la pandemia en España e Italia, lo que ha permitido entender la modificación

de los patrones de movilidad y el efecto de las restricciones de distanciamiento social que fueron impuestas.

La segunda línea está centrada en realizar predicciones de incidencia de la COVID-19 en la Comunidad de Madrid mediante encuestas indirectas implantadas en el proyecto CoronaSurveys. En esta línea se ha estudiado aristas no descubiertas hasta ahora del Network Scale-Up Method (NSUM) y evaluando las campañas de vacunación y la efectividad de determinadas vacunas. Se ha conseguido un mapa de comparaciones basadas en simulaciones que ponen de manifiesto cuáles de los estimadores de incidencia de COVID-19 son los mejores y cómo es posible atajar los sesgos a los que están expuestos.

La tercera línea incluye el desarrollo de modelos de aprendizaje automático basados en la extracción de patrones de movilidad sobre datos de incidencia COVID-19 para realizar una estimación de la evolución de la pandemia. Se han desarrollado y validado modelos para la estimación de los flujos de tráfico en la ciudad de Madrid. También se han desarrollado modelos para la estimación de flujos de movimiento de personas entre zonas de Madrid usando información de movimientos en el servicio de BiciMAD. Estos modelos han sido utilizados para predecir la incidencia de COVID-19 en la Comunidad de Madrid.

La cuarta línea de investigación ha supuesto la mejora del simulador EpiGraph mediante un modelo social extendido que asigna a cada individuo una localización espacial tanto para su lugar de residencia como de trabajo. Estos valores, que han sido calculados para la ciudad de Madrid, totalizan 3.5 millones de individuos simulados.

También se ha extendido el simulador con un modelo de vacunación que incorpora distintos rangos de cobertura vacunal y se han simulado distintos escenarios de vacunación y uso de mascarillas teniendo en cuenta los resultados de otras líneas de este proyecto.

**Objetivo del Proyecto PredCov: desarrollar un sistema que permite combinar distintas fuentes de información para modelar futuros escenarios de propagación del COVID-19**

En la quinta línea los distintos modelos de predicción desarrollados en este proyecto se han integrado en un modelo combinado (ensemble forecast) que incluye aspectos novedosos, no contemplados por los modelos epidemiológicos actuales, como son la información de movilidad de la población extraída mediante la recopilación de trazas de Digital Contact Tracing, el rastreo de contactos, el análisis de los hábitos de la población y la capacidad de cumplimiento de las medidas de distanciamiento social. Las predicciones de este modelo permiten evaluar situaciones de riesgo (alta transmisibilidad del COVID-19) asociadas a eventos, a la percepción pública del riesgo, y a noticias falsas en temáticas de COVID-19.

La sexta línea de investigación realiza un análisis del papel de los medios de comunicación y redes sociales en la formación del debate y discurso de la población sobre la COVID-19, así como su influencia sobre la percepción pública del riesgo. Para llevarla a cabo se analizaron distintas las piezas informativas y posteriormente, mediante estudios de diseño experimental, se determinó la influencia de los medios de comunicación y redes sociales sobre la población durante una crisis de salud pública. De forma complementaria se desarrolló una encuesta nacional con una muestra representativa de todo el territorio y se elaboraron entrevistas en profundidad para conocer aspectos relevantes sobre el consumo de medios de comunicación y reticencia hacia la vacunación COVID-19.

Finalmente, en la octava línea de investigación se han analizado los efectos del shock de recolocación sectorial en función de la diversidad sectorial de la economía. Se han estudiado el gasto de consumo semanal de las tarjetas de crédito y débito, así como la evolución del empleo a través de la Muestra Continua de Vidas Laborales. Adicionalmente, se han presentado propuestas de rediseño de políticas de reducción del tiempo de trabajo utilizando microdatos de la muestra continua de la Encuesta de Población Activa para estimar las transiciones de los trabajadores en ERTE entre los sectores afectados y no afectados por la COVID-19, tanto en España como en la Comunidad de Madrid.

A modo de conclusión, el proyecto se ha completado cumpliendo con todas las expectativas y objetivos planteados inicialmente, los cuales han sido también ampliados de una forma importante. Este proyecto ha realizado una propuesta innovadora con un enfoque altamente multidisciplinar. Mediante la colaboración realizada en el proyecto se ha integrado información proveniente de múltiples fuentes de datos, muy diversas entre sí. Todos los resultados obtenidos han sido particularizados para la Comunidad de Madrid. El proyecto, además de los resultados y productos descritos anteriormente, ha servido para la creación de un equipo capaz de dar soporte a la toma de decisiones en la lucha contra la pandemia. Además, se ha creado una red de grupos de investigación e instituciones colaboradoras que permiten ampliar el alcance de los resultados producidos y enriquecer la información proporcionada en el proyecto. Hay que destacar que los resultados producidos son extrapolables a otras pandemias futuras y que los equipos del proyecto van a seguir manteniendo las líneas de investigación realizadas, por lo que seguirán

**El proyecto PredCov ha cumplido con todos los objetivos planteados y ha realizado una propuesta innovadora con un enfoque altamente multidisciplinar**

### David E. Singh

Investigador principal del proyecto PredCov (Multi-source and multi-method prediction to support COVID-19 policy decision making); Departamento de Informática, Universidad Carlos III de Madrid, España.

 [dexposit@inf.uc3m.es](mailto:dexposit@inf.uc3m.es)

## Estimación de la incidencia real de la COVID-19 en España

### Real COVID-19 incidence rate estimate in Spain

Aymar Cublier Martínez<sup>a</sup>, Diana Gómez-Barroso<sup>b,c</sup>, Concepción Delgado-Sanz<sup>b,c</sup>, Susana Monge<sup>b,d</sup>, Alberto Cascajo<sup>a</sup>, María Cristina Marinescu<sup>e</sup>, Amparo Larrauri<sup>b,c</sup>, Jesús Carretero<sup>a</sup>, David E. Singh<sup>a</sup>

<sup>a</sup> Departamento de Informática, Universidad Carlos III de Madrid, España

<sup>b</sup> Centro Nacional de Epidemiología, Instituto de Salud Carlos III, España

<sup>c</sup> CIBER en Epidemiología y Salud Pública (CIBERESP), España

<sup>d</sup> CIBER en Enfermedades Infecciosas (CIBERINFEC), España

<sup>e</sup> Instituto Químico de Sarria, España

#### Resumen

**Introducción:** Los modelos epidemiológicos han demostrado ser cruciales para apoyar la toma de decisiones de las autoridades sanitarias durante la pandemia de COVID-19, así como concienciar al público en general de las distintas medidas adoptadas por las autoridades (distanciamiento social, uso de mascarilla, vacunación, etc.). **Objetivos:** Describir la metodología para integrar diferentes fuentes de datos para generar una única serie temporal que proporciona tasas de incidencia reales de COVID-19 en España. **Metodología:** Esta serie considera tanto los casos notificados como los no notificados, es decir, aquellos que no han sido registrados por las autoridades sanitarias. **Resultados:** Este trabajo describe también cómo la información generada en este proyecto ha sido tratada y almacenada, presenta los datos de estimación de la incidencia real obtenidos, así como los organismos y equipos de investigación que la utilizan, además de los distintos canales de comunicación que han sido empleados para difundirla (página web, compartición de resultados con las autoridades sanitarias, y repositorio). **Conclusión:** Este trabajo integra información proveniente de múltiples fuentes de datos para el análisis y la predicción de la incidencia de la COVID-19. A través de un enfoque multidisciplinar, se ha logrado plantear respuesta a la problemática en la estimación de la incidencia real de casos de COVID-19.

Palabras clave: COVID-19; nowcasting; modelos epidemiológicos.

#### Abstract

**Introduction:** Epidemiological models have proven to be crucial in supporting the decision-making of health authorities during the COVID-19 pandemic as well as raising awareness among the general public of the different measures adopted by authorities (social distancing, mask usage, vaccination, etc.). **Objectives:** This work describes the methodology to integrate different data sources to generate a single time series that provides real incidence rates of COVID-19 in Spain. **Methodology:** This series considers both reported and non-notified cases, that is, those that have not been registered by health authorities. **Results:** This work also describes how the information generated in this project has been treated and stored, it presents the estimated real incidence data obtained, as well as the organizations and research teams that use it, and the different communication channels that have been used to disseminate it (webpage, sharing results with health authorities, and repository). **Conclusion:** This work integrates information from multiple data sources for the analysis and prediction of the incidence of COVID-19. Through a multidisciplinary approach, it has been possible to propose a response to the problem of estimating the real incidence of COVID-19 cases.

Keywords: COVID-19; nowcasting; epidemiological models.

## Introducción

Con la emergencia de la pandemia de COVID-19 se establecieron en España, y países de nuestro entorno, sistemas de vigilancia universal de la enfermedad basados en la declaración obligatoria de todo caso confirmado de COVID-19 de cualquier gravedad. Estos sistemas han proporcionado una información imprescindible para la toma de decisiones en salud pública durante el desarrollo de la pandemia. Por otra parte, se constató la necesidad de integrar el diagnóstico de COVID-19 en los sistemas históricos de vigilancia de gripe, planteando el desarrollo de sistemas de vigilancia integrada respiratoria (World Health Organization, 2022).

Siguiendo las recomendaciones internacionales de la Organización Mundial de la Salud (OMS) y del Centro Europeo para la Prevención y el Control de las Enfermedades (ECDC), en España se implementó el sistema centinela de vigilancia de infección respiratoria aguda (SiVIRA) (Instituto de Salud Carlos III, 2022b). El SiVIRA estuvo constituido por los sistemas centinela de vigilancia de infección respiratoria aguda leve en atención primaria (vigilancia de IRAs) y grave en hospitales (vigilancia de IRAG), para la vigilancia integrada de gripe, COVID-19 y virus respiratorio sincitial (VRS). El objetivo de la vigilancia centinela integral respiratoria es obtener información para guiar la toma de decisiones en el control y prevención de las enfermedades producidas por virus respiratorios en la población. Para ello, entre otros objetivos de vigilancia, se debe monitorizar la evolución de la incidencia de las enfermedades vigiladas, en el espacio y en el tiempo, dentro de los territorios vigilados.

Por lo tanto, existen dos fuentes principales de información de incidencia de COVID-19 en la población: 1) La vigilancia universal de COVID-19, de la que obtenemos datos de incidencia de la enfermedad en la comunidad y en hospitales, y 2) La vigilancia centinela de IRAs en Atención Primaria (SiViES), de la que obtenemos datos de incidencia por COVID-19 en la comunidad, y la vigilancia de IRAG en hospitales, de la que obtenemos datos de tasas de hospitalización por COVID-19 (SiVIRA).

El sistema universal de vigilancia de COVID-19 cuenta entre sus principales fortalezas el disponer de información relativa a toda la población, lo que asegura la representatividad de los datos de incidencia para cualquier grupo de estudio con suficiente potencia estadística, además de análisis estratificados entre diferentes subgrupos poblacionales de interés, como pudieran ser individuos con diferente estatus socioeconómico. Entre sus limitaciones se pueden citar los sucesivos cambios en las estrategias de diagnóstico, que hacen que la comparación de los datos generados entre diversos periodos y territorios sea limitada. Así mismo, es difícil estimar la verdadera carga de la pandemia en el ámbito hospitalario, por ejemplo, debido a la notificación de hospitalizaciones con prueba de COVID-19 positiva, que se hospitalizan por otra causa que no es COVID-19.

Los sistemas centinela SiVIRA se centran en la recogida de información en un número limitado de sitios centinela, seleccionados de forma que sean representativos del territorio vigilado. Al recoger datos de sólo una muestra de la población, son capaces de dar una información de mayor calidad. En SiVIRA, se estiman siempre incidencias de infección por COVID-19, resultantes de la consulta médica de pacientes con síntomas respiratorios en los diferentes ámbitos sanitarios. Además, la estrategia de diagnóstico está previamente definida, lo que permite una estimación más cercana a la realidad de la carga de enfermedad ocasionada por la enfermedad respiratoria. Son sistemas más eficientes y sostenibles, que se plantean con la resiliencia necesaria para dar respuesta a cualquier virus respiratorio emergente en el futuro. Entre sus principales limitaciones destacan que deben tener la suficiente cobertura para ofrecer una información representativa del territorio vigilado con una precisión aceptable, especialmente en los análisis de subgrupos de población con menor incidencia. En la actualidad, la cobertura de población vigilada en algunas comunidades autónomas debe mejorarse para conseguir una final consolidación del sistema centinela de vigilancia integral de infecciones respiratorias.

Este trabajo tiene como objetivo describir la metodología para integrar diferentes fuentes de datos para generar una única serie temporal que proporciona tasas de incidencia reales de COVID-19 en España. De este modo, se presenta una nueva metodología que combina la información de distintos sistemas de vigilancia (SiViES y SiVIRA) con el fin de estimar la incidencia real de la COVID-19 en España. Esta incidencia real incluye tanto los casos notificados (o recogidos a través de los sistemas de vigilancia) como los que no han sido notificados. Los resultados mostrados se corresponden a la incidencia real estimada para España.

**El sistema universal de vigilancia de COVID-19 cuenta entre sus principales fortalezas el disponer de información relativa a toda la población**

## Metodología

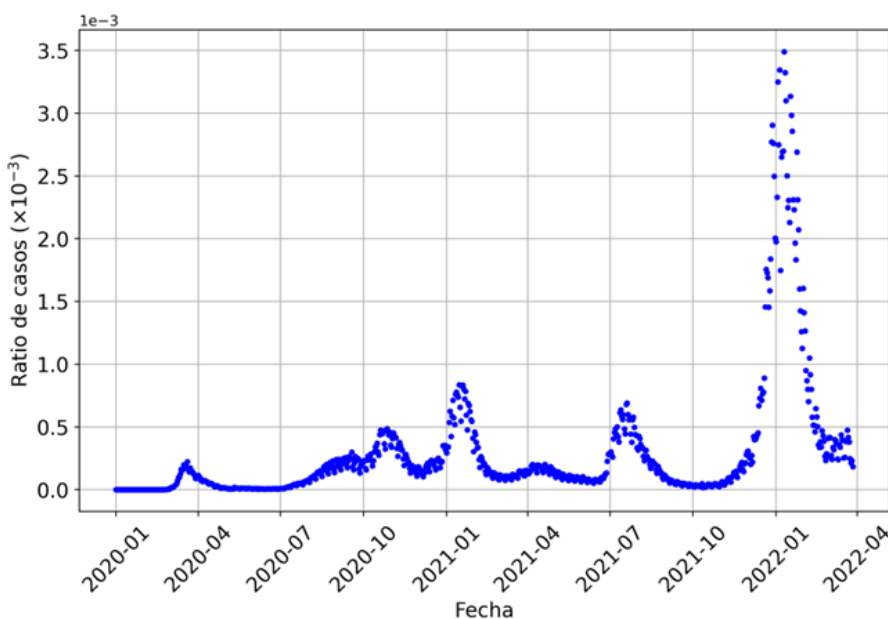
En esta sección presentamos una descripción de las fuentes de datos utilizadas para estimar el número de casos de COVID-19 en España. También se describe la metodología usada para integrar las diferentes fuentes de datos en una única serie temporal y escalarla para obtener niveles de incidencia que tengan en cuenta tanto los casos notificados como los no notificados.

### Fuentes de datos utilizadas

Los datos de casos diarios de COVID-19 en España se obtienen a partir del Instituto de Salud Carlos III (SiViES de ahora en adelante) y de la John Hopkins University, que a su vez los obtiene del Ministerio de Sanidad de España. Los datos SiViES consisten en casos diarios notificados de COVID-19 desde el 1 de enero de 2020 al 28 de marzo de 2022, cuando el SiViES discontinuó la publicación de datos para todas las edades y se centró en la población mayor de 60 años. La Figura 1 muestra la serie de datos de SiViES para España. Hemos escogido representar el ratio de nuevos casos diarios (número casos dividido por la población española) pues son los datos que utilizamos más tarde en el simulador epidemiológico.

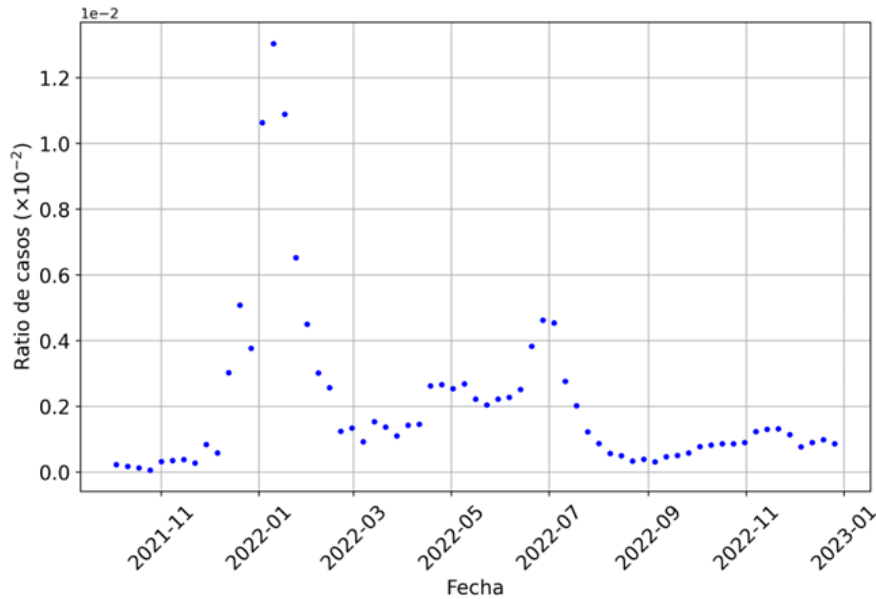
Efectivamente, la simulación es a escala del individuo, por lo que es más útil considerar una probabilidad de infección (dada por el ratio aquí representado) que un número de infecciones o una incidencia por 100,000 habitantes.

El Centro Nacional de Epidemiología (CNE) facilita datos adicionales de incidencia acumulada semanalmente (casos por 100.000 habitantes) de COVID-19. La serie temporal abarca desde el día 4 de octubre de 2021 hasta la actualidad. La Figura 2 muestra la serie de datos de SiVIRA para España.



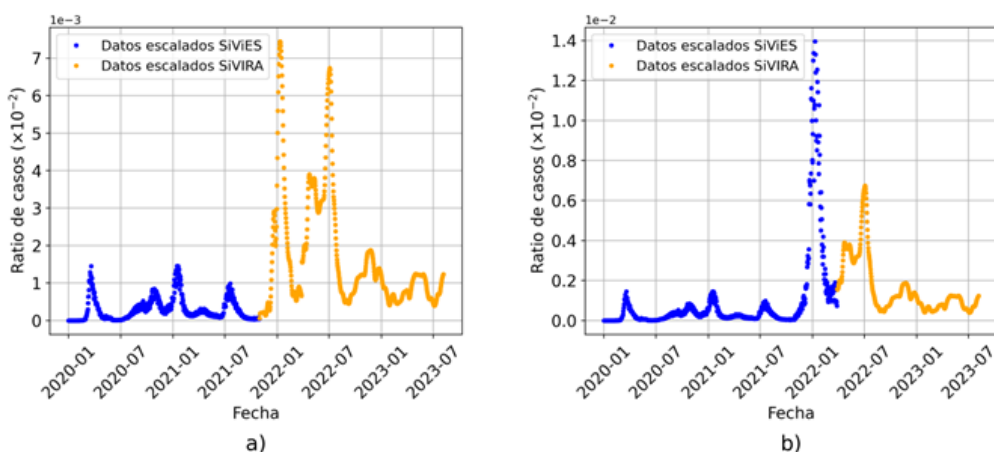
**Figura 1.** Ratio de casos notificados de COVID-19 del 1 de enero de 2020 al 28 de marzo de 2022 en España (fuente: SiViES).





**Figura 2.** Ratio de casos semanales de la semana 40 de 2021 (31 de octubre de 2021) a la semana 31 de 2023 (30 de junio de 2023) (fuente: SiVIRA).

Las estimaciones de casos de COVID-19 en Atención Primaria se obtiene a partir de información aportada por las comunidades autónomas (CCAA) que participan en el programa centinela de vigilancia para infecciones respiratorias. Inicialmente, estas CCAA eran Andalucía, Castilla y León, Cataluña, Extremadura y Comunidad de Madrid. Posteriormente, se incluyeron las comunidades de Asturias, Baleares, Cantabria, Comunidad Valenciana, Murcia, Ceuta y Melilla. En el tratamiento de los datos de esta serie temporal no se hizo ningún cambio significativo en la escala de los datos SiVIRA para que coincidieran con la curva anterior de ratio de casos nuevos diarios. Sin embargo, como los datos se notifican semanalmente, hemos utilizado un algoritmo de muestreo e interpolación para obtener una estimación de los casos diarios. De este modo, ambas series temporales (SiVIES y SiVIRA) tienen el mismo número de muestras.



**Figura 3.** Tasas de incidencia estimadas (ratio a 1) para España. Se muestran dos posibles escenarios para representar la onda Omicron en España (diciembre de 2021 a marzo de 2022): (a) utilizando casos notificados por SiVIRA, (b) utilizando casos notificados.

Se puede observar en la Figura 3 que los datos del SiViES muestran un ligero repunte de la incidencia acumulada en población mayor de 60 años a partir de marzo de 2022 (Instituto de Salud Carlos III, 2022a) mientras que los datos de SiVIRA fueron más precisos con respecto a este repunte. Como los datos del SiViES para población general finalizaron en marzo de 2022, se optó por combinar ambas series en una única serie temporal que muestra una estimación de la incidencia real de la COVID-19. La Figura 3 muestra los dos escenarios existentes para unificar ambas series temporales. Nótese que la diferencia entre ambas está en el modo de representar la onda Omicron en España (diciembre de 2021 a marzo de 2022) combinando ambas series. La discontinuidad observada se debe al cambio de factor de escala y al cambio de fuente de datos en el periodo de transición entre las series. Hay que tener en cuenta que la amplitud difiere entre los datos SiViES y los datos SiVIRA. La menor magnitud de la curva SiVIRA (Figura 3<sup>a</sup>) está relacionada con una metodología diferente utilizada para recopilar los datos. En este caso, sólo se reportan los casos sintomáticos, mientras que en los datos del SiViES (Figura 3b) los casos notificados incluyen también los casos asintomáticos obtenidos mediante pruebas individuales. Hemos elegido la serie completa SiViES correspondiente a la Figura 3b) porque considera un número más amplio de casos.

### Calibración del modelo

En la metodología propuesta, dividimos la serie temporal en seis periodos, que corresponden aproximadamente a las cinco ondas epidémicas de COVID-19 que ocurrieron entre el 1 de enero de 2020 y el 28 de marzo de 2022. Para cada periodo, obtenemos un factor de escala a partir de la literatura científica y de las autoridades sanitarias. Este factor de escala corresponde al porcentaje de casos detectados en cada periodo. Usando estos coeficientes, podemos aproximar los casos diarios reales que incluyen tanto los casos notificados como los que no lo son. La Tabla 1 muestra los factores de escala para cada periodo. La magnitud representada en la serie temporal generada es el ratio de casos diarios de COVID-19 a 1, es decir, el número de casos diarios dividido por la población española.

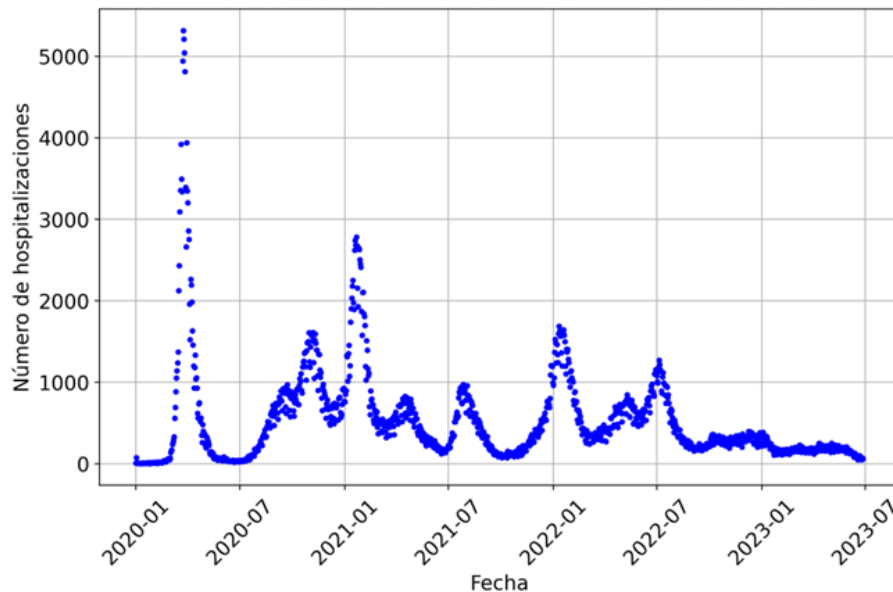
**Tabla 1.** Factor de escala para cada onda epidémica en España (del 1 de enero de 2020 al 31 de diciembre de 2022).

Periodo temporal	Factor de escala
2020-01-01 – 2020-05-31	6.67
2020-06-01 – 2020-12-31	1.72
2021-01-31 – 2021-06-30	1.54 <sup>1</sup>
2021-07-01 – 2021-10-31	1.43 <sup>1</sup>
2021-11-01 – 2021-11-30	4.00 <sup>1</sup>
2021-12-01 – 2022-03-28	4.00 <sup>1</sup>
2022-03-29 – 2022-12-31	9.89

<sup>1</sup>Valores aportados por el Ministerio de Sanidad y Consumo.

Considerando que la onda Ómicron alcanzó su mínimo en marzo de 2022, se actualizó el factor de escala de la tabla entre casos e infecciones para este período. El factor de escala correspondiente al último periodo (del 29 de marzo de 2022 al 31 de diciembre de 2022) ha sido calculado mediante la metodología descrita en Martínez et al. (2023). Para ello, ha sido necesario utilizar la serie temporal de hospitalizaciones por COVID-19 en España, que representamos en la Figura 4.

De este modo, una vez obtenidos los factores de escala para cada uno de los períodos de tiempo, los datos originales de las fuentes de datos son procesados con el fin de obtener la estimación de la incidencia real. Para cada uno de los días de la serie histórica -comenzando el 1 de enero de 2020- los valores se escalan utilizando los factores descritos en la Tabla 1. Desde el 1 de enero de 2020 hasta el 27 de marzo de 2022, se utiliza la serie temporal de SiViES y a partir del 28 de marzo de 2022, se utilizan los datos de la serie SiVIRA. La Figura 3b) muestra los valores resultantes de estimación de incidencia real.

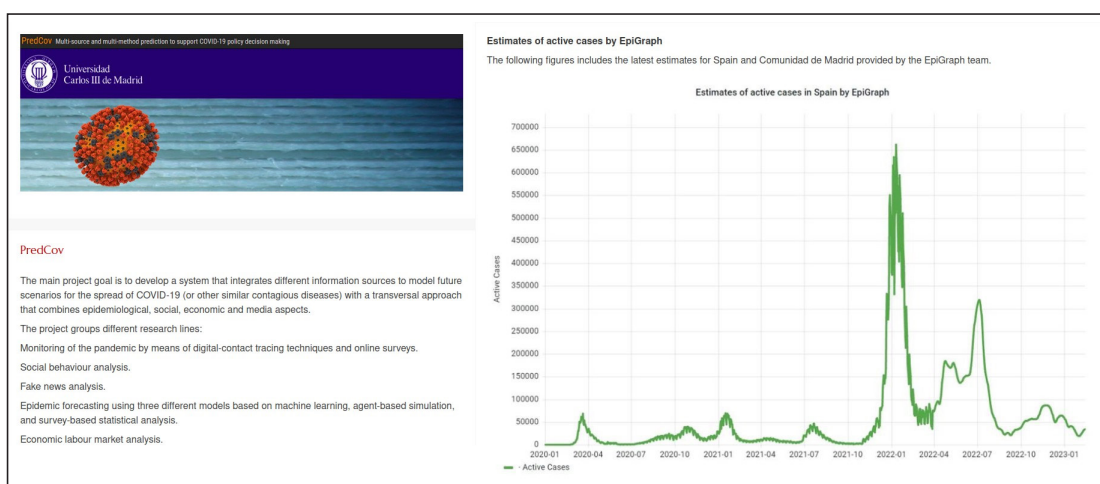


**Figura 4.** Número de hospitalizaciones por COVID-19 en España, del 1 de enero de 2020 al 29 de junio de 2023 (fuente: SiVIES).

## Resultados

Este trabajo se ha desarrollado en el contexto del proyecto Multi-source and multi-method prediction to support COVID-19 policy decision making (PredCov) (Universidad Carlos III de Madrid, 2022b) en el que propone la integración de distintos modelos de propagación de la COVID-19 así como distintas fuentes de datos. Se han utilizado tres canales de difusión de los datos producidos a través de: 1) el público general, 2) las autoridades sanitarias, y 3) modelos epidemiológicos con datos de estimación de incidencia real.

El primer canal de difusión está destinado al público general, al que se busca proporcionar predicciones de incidencia y concienciar sobre la necesidad de las políticas de distanciamiento social y vacunación impuestas por las autoridades. En este sentido, se ha desarrollado una página web que muestra las estimaciones de incidencia real para España y la Comunidad de Madrid (Universidad Carlos III de Madrid, 2022b). Los resultados también se han publicado en distintas redes sociales destinadas a la difusión y discusión de las herramientas, metodologías y aplicaciones creadas. La Figura 4 muestra una captura de la página web del proyecto, en la cual es posible acceder tanto a las estimaciones de incidencia real como las distintas predicciones de incidencia ofrecida por los otros equipos participantes en del proyecto.



**Figura 5.** Página web del proyecto PredCov. A la izquierda se muestra la página principal, mientras que a la derecha se muestra la sección que tiene la estimación de la incidencia real de COVID-19.

El segundo canal de difusión está orientado a las autoridades sanitarias, a las cuales se les proporciona información detallada del riesgo de incidencia de la COVID-19 o futuras pandemias en distintos escenarios que se han considerado (por ejemplo, abordando distintas estrategias de vacunación de COVID-19 o la aparición de nuevas variantes). En el contexto de este trabajo se ha estimado la incidencia real para España en general y la Comunidad de Madrid en particular, y se han elaborado distintos informes que se han enviado a autoridades sanitarias de la Comunidad de Madrid, el Ministerio de Sanidad y el Centro Europeo para la Prevención y el Control de las Enfermedades (ECDC). Respecto a este último organismo, los resultados se han publicado a través de las plataformas European Covid-19 Forecast Hub y European Covid-19 Scenario Hub, gestionadas por el ECDC. En los informes realizados también se evalúan distintas políticas sociales y económicas, así como el impacto de la potencial incidencia del COVID-19 en los recursos sanitarios, sociales y económicos de la Comunidad de Madrid y España.

El tercer canal de difusión consiste en alimentar a los modelos epidemiológicos del proyecto con los datos de estimación de incidencia real. Uno de los modelos que requiere de estos datos para realizar la calibración es EpiGraph (Universidad Carlos III de Madrid, 2022a), un simulador paralelo basado en agentes que modela la propagación de la COVID-19 utilizando un gran número de individuos -del orden de millones- mediante una red de interconexión realista basada en interacciones individuales reales extraídas de redes sociales y datos demográficos. Esta red de contactos incluye las características de cada individuo, sus relaciones en el trabajo o centros educativos, el hogar y durante el tiempo libre y sus desplazamientos diarios. Con este fin, los valores de incidencia están a disposición en un repositorio de GitHub (Universidad Carlos III de Madrid, 2022c) para los equipos interesados en participar. En este repositorio está disponible la siguiente información:

1. Gráficas semanales de incidencia, incluyendo:

- ISCIIRAW.png: datos originales de SiViES representados de forma gráfica.
- SiVIRARAW.png: datos originales de SiVIRA representados de forma gráfica.
- SiViES+SiVIRA\_2022-03-28\_SCALED.png: datos de incidencia escalada (ratio a 1).

2. Series temporales de incidencia semanal.

- TimeDependentSiVIRARAW.txt: serie temporal de datos originales provenientes del SiVIRA (ver la Figura 2). Esta serie incluye datos semanales desde la semana 40 de 2021.
- TimeDependentSiViES+SiVIRASCALED.txt: serie temporal de incidencia escalada (ratio a 1) representados en la Figura 3b). Esta serie incluye datos diarios desde la primera semana de 2020.
- TimeDependentHosp.txt: serie temporal de datos de hospitalizaciones (incluyendo todas edades). Esta serie incluye datos diarios desde la primera semana de 2020.

## Conclusiones

En este trabajo se aborda la problemática en la estimación de la incidencia real de casos de COVID-19. La infradetección de las fuentes actuales de información de incidencia tiene dos componentes: por una parte, el infradiagnóstico (aquellas infecciones que no llegan a hacerse pruebas y no son detectadas), y por otra, la infranotificación (aquellas infecciones que siendo diagnosticadas no llegan a notificarse a las autoridades sanitarias). Nótese que la primera componente tiene que ver con las políticas de detección desarrolladas y la realización de testeo por la población en general. La segunda componente está relacionada con el funcionamiento del sistema de vigilancia. En este trabajo se proponen soluciones para evitar esta infradetección combinando distintas fuentes de información.

La comunicación de los resultados se ha realizado utilizando múltiples canales que incluyen tanto la difusión al público en general (medios de comunicación, página web, etc.), informes enviados a distintos organismos regionales, nacionales y europeos, así como repositorios que permiten compartir la información generada con otros equipos.

Todos estos medios representan servicios desarrollados en el proyecto PredCov. Es importante destacar que todos los resultados obtenidos son generalizables a otras enfermedades contagiosas -de propagación similar a la COVID-19-, por lo que el impacto alcanzado en este proyecto puede ser también extendido a potenciales futuras pandemias.

**La infradetección de las fuentes actuales de información de incidencia tiene dos componentes: el infradiagnóstico y la infranotificación**

A modo de resumen, este proyecto ha realizado una propuesta innovadora con un enfoque altamente multidisciplinar, con grupos con un perfil muy diferente y con experiencia previa en investigación en temáticas de la COVID-19. Mediante la colaboración realizada en el proyecto PredCov se ha integrado información proveniente de múltiples fuentes de datos para el análisis y la predicción de la incidencia de la COVID-19. Todas estas fuentes de datos y resultados obtenidos han sido particularizados para España y la Comunidad de Madrid.

#### Contribuciones de los autores

Los autores participaron igualmente en la elaboración del manuscrito y aprobaron la versión final presentada.

#### Financiación

Este trabajo ha sido financiado mediante el Convenio firmado entre la Comunidad de Madrid (Consejería de Educación, Universidades, Ciencia y Portavocía) y la Universidad Carlos III de Madrid para la concesión directa de una ayuda para financiar la realización de actuaciones en materia de investigación sobre el SARS-COV 2 y la enfermedad COVID-19 financiado con los recursos REACT-UE del fondo europeo de desarrollo regional y el proyecto BCV-2022-1-0005 de la Red Española de Supercomputación.

#### Declaración de disponibilidad de datos

Los datos presentados en este estudio pueden ser solicitados al autor de correspondencia.

## Referencias bibliográficas

- Instituto de Salud Carlos III. (2022a). Evolución pandemia. <https://cnecovid.isciii.es/covid19/#evoluci%C3%B3n-pandemia>.
- Instituto de Salud Carlos III. (2022b). Metodología SiVIRA. sistemas y fuentes de información. <https://www.isciii.es/QueHacemos/Servicios/VigilanciaSaludPublicaRENAVE/EnfermedadesTransmisibles/Documents/GRIPE/Protocolos/Metodolog%C3%ADa%20SiVIRA,%20sistemas%20y%20fuentes%20de%20informaci%C3%B3n.%20Temporada%202022-23.pdf>.
- Martínez, A. C., Organero, M. M., Moriña, D., Barroso, D. G., & Singh, D. E. (2023). COVID-19 incidence estimates and forecast by metaprediction for the comunidad de Madrid. Technical report, Universidad Carlos III de Madrid.
- Universidad Carlos III de Madrid. (2022a). Epigraph, an agent-based epidemiological simulator <http://epigraph.uc3m.es>.
- Universidad Carlos III de Madrid. (2022b). Multi-source and multi-method prediction to support COVID-19 policy decision making (PredCov). <http://www.uc3m.es/ss/Satellite/GruposInvestigacion/es/TextoDosColumnas/1371351563042>.
- Universidad Carlos III de Madrid. (2022c). Repositorio de datos de incidencia de la COVID-19 para España y la comunidad de Madrid. <https://github.com/epigraph-forecast/IncidenceDataSpain>.
- World Health Organization. (2022). End-to-end integration of SARS-CoV-2 and influenza sentinel surveillance: revised interim guidance. Geneva. [https://www.who.int/publications/i/item/WHO-2019-nCoV-Integrated\\_sentinel\\_surveillance-2022](https://www.who.int/publications/i/item/WHO-2019-nCoV-Integrated_sentinel_surveillance-2022)

## Coronasurveys: encuestas indirectas en línea para monitorizar la evolución del COVID-19

## Coronasurveys: indirect online surveys to monitor the evolution of COVID-19

Antonio Fernández-Anta<sup>a</sup>, Jose Aguilar<sup>a,c,d</sup>, Juan M. Ramírez<sup>a</sup>, Rosa Elvira Lillo<sup>b</sup>, Sergio Díaz-Aranda<sup>a</sup>

<sup>a</sup> IMDEA Networks Institute, España

<sup>b</sup> Departamento de Estadística, Universidad Carlos III de Madrid, España

<sup>c</sup> Centro de Estudios en Microelectrónica y Sistemas Distribuidos, Universidad de Los Andes, Venezuela

<sup>d</sup> Grupo I+D+I en Tecnologías de la Información y las Comunicaciones, Universidad EAFIT, Colombia

### Resumen

**Introducción:** Seguir la evolución del COVID-19 ha sido esencial para la toma de decisiones sanitarias. Pero ello requiere de cifras fiables de los infectados, muertos y hospitalizados, muchas veces complicadas de tener por múltiples razones. El uso de técnicas de estimación rápidas y fiables, como las encuestas indirectas, es una opción. **Objetivos:** Presentar el proyecto *CoronaSurveys*, el cual monitorea el COVID-19 combinando encuestas indirectas con el método de ampliación de red (*Network Scale-up Method*, NSUM). **Metodología:** El sistema usa encuestas anónimas indirectas en línea para consultar sobre el COVID-19, y el método NSUM para estimar los casos. Las encuestas están en múltiples idiomas con preguntas para rastrear los casos activos, nuevos y de muertes, entre otros. **Resultados:** *CoronaSurveys* está operando desde marzo de 2020, y sigue recopilando datos, con más de cien mil respuestas en la actualidad (millones de muestras). El sistema ha hecho buenas estimaciones para España y el Reino Unido, entre otros países. **Conclusión:** Este sistema es adecuado en países con infraestructura sanitaria limitada o en situaciones de desconocimiento de la pandemia, como al inicio del COVID-19, porque su coste de implementación es pequeño, requiere dispositivos simples para usarlo, y de pocos participantes para obtener buenas estimaciones.

Palabras clave: COVID-19; encuestas indirectas; métodos NSUM; inferencia estadística.

### Abstract

**Introduction:** Monitoring the evolution of COVID-19 has been essential for health decision-making. But this requires reliable numbers of the infected, dead and hospitalized, often difficult to have for multiple reasons. The use of fast and reliable estimation techniques, such as indirect surveys, is an option. **Objectives:** To present the *CoronaSurveys* project, which monitors COVID-19 by combining indirect surveys with the *Network Scale-up Method* (NSUM). **Methodology:** The system uses indirect online anonymous surveys to inquire about COVID-19, and the NSUM method to estimate cases. The system offers multi-language surveys with questions to track active cases, new cases, and fatalities, among others. **Results:** *CoronaSurveys* has been operating since March 2020, and continues to collect data, with over one hundred responses to date (over one million samples). The system has made good estimates for Spain and the United Kingdom, among other countries. **Conclusion:** This system is suitable in countries with limited health infrastructure or in situations of lack of knowledge about the pandemic, such as at the beginning of COVID-19, because its implementation cost is small and it requires simple devices to use it and few participants to obtain good estimates.

Keywords: COVID-19; indirect surveys; NSUM methods; Statistical inference.

## Introducción

**M**onitorizar las cifras que caracterizan la evolución del COVID-19 durante la pandemia fue esencial para la toma de decisiones estratégicas de muchos países, tales como cuándo confinar a la población, cerrar colegios, etc. Normalmente, en las primeras etapas de una pandemia hay poca capacidad para determinar los casos reales, porque no se tienen los recursos y la logística necesarios para realizar pruebas, por lo que los datos oficiales de las autoridades sanitarias pueden estar muy lejos del número real de casos (Maxmen, 2020; Ruppert et al., 2018). Esto es aún más real en los países de ingresos bajos y medianos. No tener acceso a datos fiables claramente afecta a la calidad de las decisiones tomadas por las autoridades, poniendo a la población en mayor riesgo. Incluso en etapas posteriores de la pandemia, puede seguir ocurriendo que los datos oficiales estén muy lejos de los reales por diferentes motivos, como descuido de las autoridades o pasividad de la población, entre otras razones (Nishiura et al., 2009).

En ese sentido, se requiere de otras estrategias y/o técnicas, más allá de las pruebas de laboratorio, que puedan estimar el número de casos y su evolución. Por ejemplo, técnicas que recopilen masivamente y fácilmente datos para hacer estimaciones razonablemente precisas, tales como las encuestas directas. Varias encuestas directas

para obtener datos de salud sobre el COVID-19 se han implementado, tales como las reportadas en los trabajos de Facebook y Oliver et al., 2020. Ahora bien, las encuestas directas necesitan una gran cantidad de participantes para lograr estimaciones fiables, además de que recopilan información de salud personal confidencial, lo que puede disuadir de responder a personas preocupadas por su privacidad.

Una alternativa a las encuestas directas son las indirectas, donde las preguntas que responde un participante no son sobre sí mismo, sino sobre sus contactos. Este enfoque tiene al menos tres ventajas importantes con respecto a las encuestas directas. Primero, la encuesta se puede diseñar de modo que no se recopile información personal del participante para preservar la privacidad de los participantes (es decir, es completamente anónima). Segundo, tiene un efecto multiplicador, ya que reduce el número de respuestas necesarias para lograr una cobertura de población específica, es decir, alcanzar una cobertura amplia en un período de tiempo corto. Tercero, permite

obtener estimaciones que converjan más rápido al verdadero valor del parámetro a estimar usando las técnicas que se conocen en la literatura como el método de ampliación de red (Network Scale-up Method, NSUM) (Bernard et al., 2010; Laga et al., 2021), que permiten realizar estimaciones con las respuestas recopiladas. Así, en el caso del COVID-19, las respuestas individuales actúan como conocimiento instantáneo de la situación actual de la pandemia desde un punto de vista personal. Cuando estas respuestas se analizan colectivamente, a través del tiempo y ubicaciones geográficas, se puede inferir una visión combinada de la pandemia. De esta manera, la pérdida de precisión debido a que los encuestados no siempre tienen información exacta sobre el estado de salud de los demás, puede compensarse con el aumento significativo de la cobertura (es decir, el número que cada encuestado reporta) inspirado en el “paradigma de la multitud” o la “sabiduría de los grupos”, por lo que ha sido utilizado con éxito en salud pública (Nishiura et al., 2009).

Particularmente, en el proyecto CoronaSurveys (<https://coronasurveys.org/>) se ha creado un sistema para estimar la cantidad de casos de COVID-19 en base a encuestas indirectas abiertas. El sistema usa el método NSUM a partir de encuestas anónimas indirectas en línea, para realizar las estimaciones (The CoronaSurveys research team, 2020 <https://coronasurveys.org/>). El sistema CoronaSurveys ha estado operando desde marzo de 2020, comenzando en sólo tres países inicialmente (España, Portugal y Chipre) y desplegando en unas semanas encuestas para todos los países del mundo. Este sistema cuenta con encuestas en 60 idiomas y permite reportar datos sobre la incidencia del COVID-19 en todos los países. El sistema sigue recopilando datos y ha recopilado más de cien mil de respuestas hasta el momento, reportando el estado de millones de personas. Todos los datos recopilados están disponibles para su uso abierto. Hasta donde sabemos, este es el sistema basado en NSUM de mayor escala jamás implementado y el único que ha recopilado datos continuamente durante un período de más de tres años mediante encuestas abiertas.

En este documento se comentan los aspectos más relevantes del proyecto CoronaSurveys. Para ello, se presenta antes que nada la metodología que está detrás del proyecto, posteriormente se analizan las estimaciones resultantes obtenidas, para culminar con una discusión sobre los resultados.

**En el proyecto CoronaSurveys (<https://coronasurveys.org/>) se ha creado un sistema para estimar la cantidad de casos de COVID-19 en base a encuestas indirectas abiertas**

## Metodología

El proyecto Coronasurveys ha desplegado encuestas indirectas en línea que se pueden contestar a través de un navegador web o una aplicación móvil. La encuesta consulta a los encuestados sobre la evolución del COVID 19, en particular, la encuesta original tenía sólo dos preguntas básicas: 1) ¿Cuántas personas hay en su entorno cuyo estado de salud conoce? y 2) ¿Cuántas de ellas fueron diagnosticadas o tuvieron síntomas del COVID-19?

La primera pregunta determinaba el alcance del participante (RI) y la segunda el número de casos conocidos (CI). Los datos se han venido agregando y mezclando diariamente, para hacer un pre-procesamiento para eliminar los valores atípicos, basado en los siguientes criterios: Si RI está fuera de 1,5 veces el rango inter-cuartil sobre el cuartil superior; Si CI/RI es  $> 1/3$  (para excluir a los participantes con contacto de casos excepcionalmente alto).

El cuestionario de CoronaSurveys se amplió con preguntas adicionales para mejorar su granularidad y estimar más parámetros de la pandemia (como fallecimientos), sobre las personas con síntomas. Algunas de estas preguntas son: 3) ¿Cuántas siguen enfermas?; 4) ¿Cuántas comenzaron con síntomas en los últimos 7 días?; 5) ¿Cuántas murieron?

Con estas preguntas adicionales, es posible rastrear los casos activos (Pregunta 3), los casos nuevos (Pregunta 4), y el número acumulado de muertes (Pregunta 5).

Desde el principio del proyecto, el mayor desafío ha sido conseguir participantes que rellenen la encuesta. Inicialmente se diseminó usando redes sociales en línea (Twitter, Facebook, WhatsApp, etc.), lo cual fue efectivo en los primeros meses de la pandemia. Más tarde, se usaron campañas de anuncios, principalmente en Facebook, en España y Brasil, con relativamente baja efectividad. Actualmente, desde julio de 2022 se ha utilizado la plataforma Prolific (<https://www.prolific.co/>), que cuenta con potenciales encuestados en diferentes países, a los que se les paga por completar encuestas. Los encuestados con la plataforma Prolific en varios países (inicialmente 11 países y ahora solamente España, Reino Unido y EEUU) rellenan una encuesta indirecta sobre lo que saben en su contexto sobre la evolución del COVID-19.

La parte superior de la Figura 1 muestra los países de los que se ha recolectado más información en el 2022, (con el mayor número de registros), con las cantidades recolectadas, debido principalmente a que, desde el proyecto, se han optimizado los fondos para obtener datos en países estratégicos. En concreto, España es el país con mayor participación, y casi duplica al siguiente país (Reino Unido). Por otro lado, la parte inferior de la Figura 1 muestra la participación por país en los últimos seis meses del 2022. Varios aspectos a destacar son los siguientes: i) prácticamente toda la participación de algunos de los países se concentra en esos 6 meses (por ejemplo, Reino Unido, Sudáfrica y Estados Unidos), ii) hay países en los que ha cesado la recogida (por ejemplo, Ucrania y Chipre) y, iii) España ya no es el país del que se recopila la mayor parte de la información, siendo este el Reino Unido.

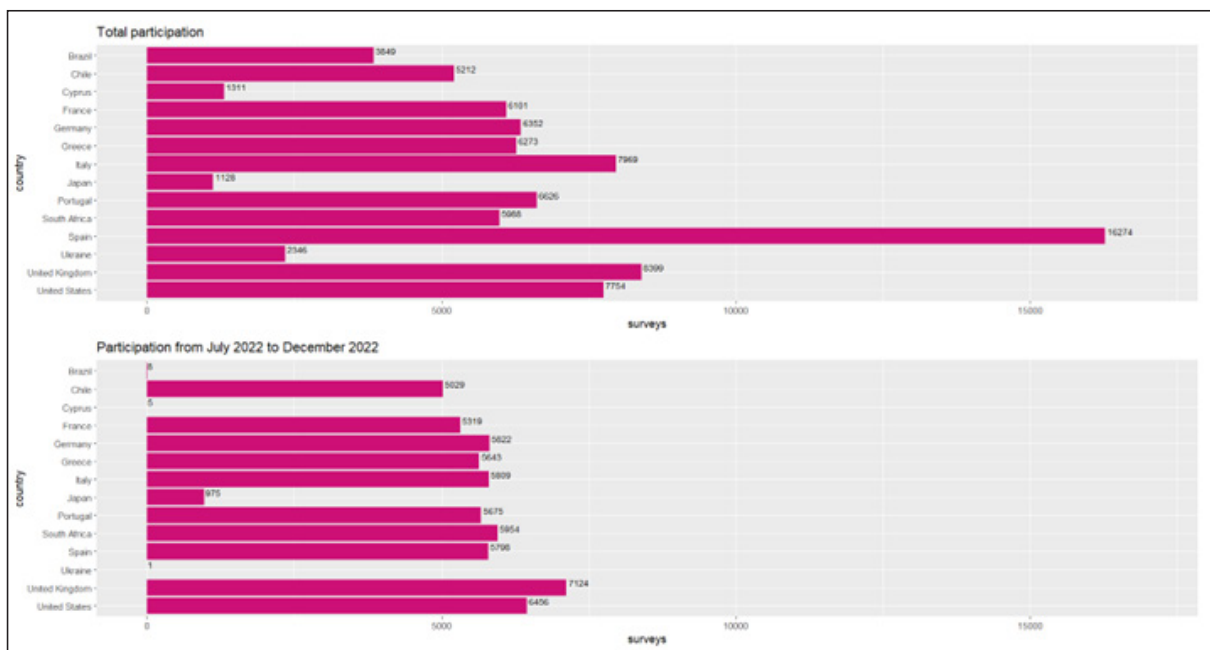


Figura 1. Participación por país.



Finalmente, con el conjunto de datos agregados por región, se puede realizar una estimación sólida de la prevalencia del COVID-19. Así, para una región  $i$ , obtenemos  $n_i$  respuestas, y cada respuesta  $j$  tiene una variable de alcance ( $RI_j$ ), que es el número de personas (red) cuyo participante  $j$  conoce (primera pregunta de la encuesta); y una variable de conteo ( $CI_j$ ) que es el número de personas conocidas por el encuestado con síntomas compatibles con COVID-19 (segunda pregunta de la encuesta). Entonces, una forma de estimar la proporción de personas infectadas con síntomas en la región  $i$  ( $I_i$ ) es de la siguiente manera:

$$I_i = \frac{\sum_1^{n_i} CI_i}{\sum_1^{n_i} RI_i}$$

Ha de dejarse claro que esta no es la única manera de hacer las estimaciones. El conjunto de métodos de estimación de este tipo son conocidos como métodos NSUM, un tema de mucha relevancia en la comunidad científica actualmente (Bernard et al., 2010; Laga et al., 2021). Básicamente, estos métodos surgen para poder tener información sobre fenómenos que son considerados socialmente tabús como por ejemplo, VIH, muertes en terremotos,...., donde los encuestados se sienten más cómodos si no responden sobre sí mismos.

## Resultados

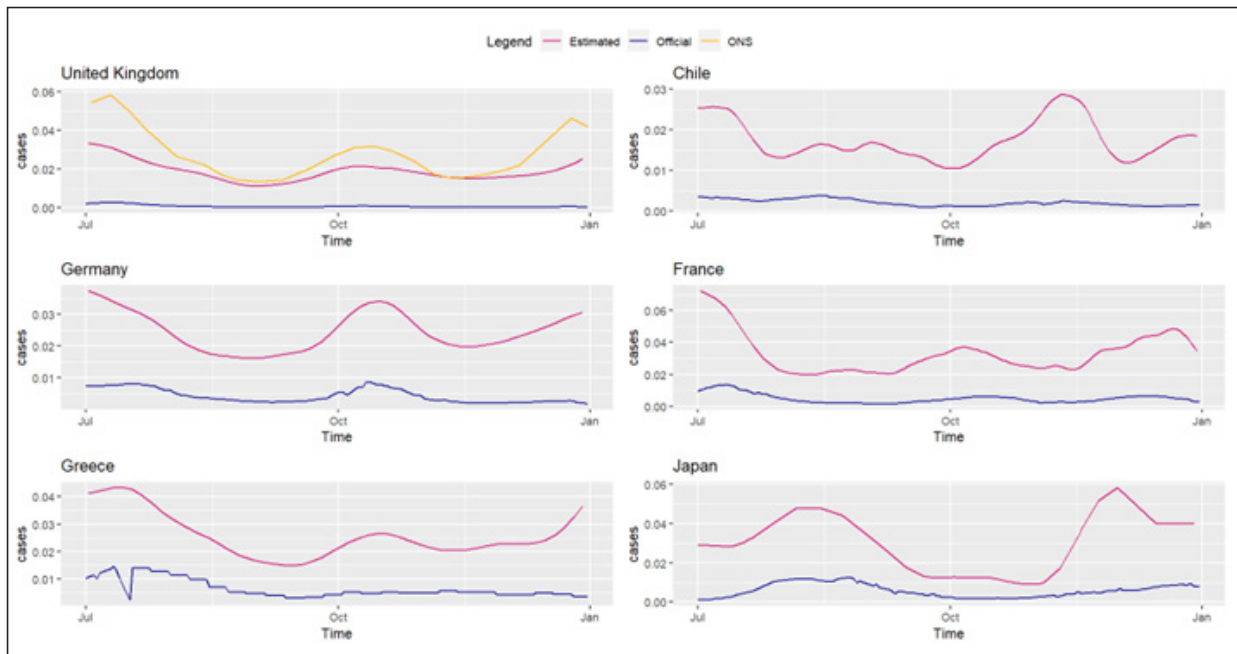
La Tabla 1 muestra un ejemplo del total de datos recopilados por país, donde la primera columna representa al país. Específicamente, se muestran tres variables: alcance, muertos y hospitalizaciones. Para algunos valores, las desviaciones típicas son muy grandes (por ejemplo, los valores de hospitalización para casi todos los países). Una desviación típica grande significa que los datos están muy dispersos con respecto a la media. La primera columna de la Tabla 1 describe el número de contactos (encuesta indirecta) declarados por el encuestado. Como se puede observar, se obtiene una media aproximada de 25 contactos por cada encuestado, que es un número muy por debajo del que Dunbar estableció como límite cognitivo de relaciones sociales estables que un ser humano puede mantener, el cual, según Dunbar, sería 150 (Dunbar, 1992). Dunbar (1992) propuso que en el mundo de las relaciones sociales existe una jerarquía de círculos definidos a partir de la intensidad y la frecuencia de las relaciones, tal que si el entramado social está bien estructurado, en cada uno de los círculos sucesivos se incluye más del doble de gente que en el anterior. En términos generales, esos círculos serían progresivamente 5, 15, 50, 150, 500 y 1500 individuos. Así, el círculo más pequeño estaría compuesto por cinco personas (nuestros seres queridos), después por 15 amigos íntimos, 50 amistades, 150 contactos importantes. 500 conocidos, y 1500 personas que quizás conozcas. De esta manera, 25 está entre el tamaño de amigos íntimos y el de amistades. Con esta media de contactos en CoronaSurveys y más de 100.000 respuestas recopiladas, el proyecto ha tenido hasta ahora un alcance de varios millones de personas en total.

Usando los conjuntos de datos recopilados en el proyecto, se han realizado varias pruebas para estimar la incidencia de COVID-19 en trabajos anteriores. Por ejemplo, el trabajo de Baquero et al., 2021 muestra la evolución del número de casos activos, casos nuevos diarios, y casos contagiosos, estimados con algunos de los métodos del proyecto Coronasurveys, utilizando datos de Portugal. Un segundo trabajo evaluó la calidad del conjunto de datos, analizando la precisión de varios métodos para estimar el número acumulado de casos de COVID-19 (García-Agundez et al., 2021). Se compararon las estimaciones de la encuesta con los resultados de un estudio serológico reportado en Pollán et al., 2020 para España, realizado entre el 27 de abril y el 11 de mayo de 2020. Particularmente, el coeficiente de determinación ( $R^2$ ) de las estimaciones en todas las regiones (comunidades autónomas) de España en el periodo de tiempo del estudio serológico arrojó 0,89, resultado bastante bueno. Una conclusión interesante de este trabajo es que un valor de 50 respuestas por región permitió realizar una estimación razonable de casos, muy alineada con los datos oficiales, los cuales tienen un coste mucho mayor para su obtención. Incluir más respuestas podría aumentar la precisión, pero los números se mantienen razonablemente estables.

**Tabla 1.** Algunas variables del conjunto de datos.

País	Media de Alcance	Desviación típica de alcance	Media de Infectados	Desviación típica de Infectados	Media de Hospitaliz.	Desviación típica de Hospitaliz.
Chile	25.32	4.11	0.31	1.01	2.29	9.67
Francia	14.48	2.59	0.39	1.05	3.84	15.21
Alemania	24.47	3.83	0.52	1.08	15.34	27.47
Grecia	29.33	4.74	0.63	1.44	0.95	1.47
Italia	31.28	4.90	0.61	1.28	1.03	3.85
Japón	14.58	2.82	0.47	1.10	0.28	1.12
Portugal	32.80	5.05	0.24	0.98	22.13	24.12
Suráfrica	33.36	5.27	0.51	1.34	82.01	46.33
España	29.97	4.85	0.35	1.22	0.78	1.29
Inglaterra	29.95	4.84	0.56	1.46	8.46	15.31
Estados Unidos	27.82	4.96	0.38	1.13	7.61	20.28

Por otro lado, la Figura 2 muestra las estimaciones de casos usando el conjunto de datos de CoronaSurveys del último semestre del 2022 para alguno de los países, comparado con los números oficiales (Ojo et al., 2020). En esa base de datos, Japón tiene pocos datos recopilados, y Reino Unido muchos datos recopilados. En general, se puede ver que las estimaciones siguen bastante bien el comportamiento de los datos oficiales en algunos países (como Alemania). En el Reino Unido, también están muy bien alineados con las estimaciones de la Oficina de Estadísticas Nacional (ONS por sus siglas en inglés Office for National Statistics, <https://www.ons.gov.uk/>). El país con mayor diferencia en cuanto a las estimaciones realizadas es Chile.

**Figura 2.** Estimación de casos de COVID.

Finalmente, recientemente se han comparado las estimaciones con los datos oficiales en el Reino Unido y Australia (ver Tabla 2). Las estimaciones de vacunación están muy cerca de los valores oficiales. Las tasas de vacunación en Australia y Reino Unido se han estimado en 76,50 % (73,70 % - 79,29 %) y 78,86 % (intervalo de confianza (IC) del 95 %: 77,00 % - 80,72 %), mientras que los valores oficiales (OWID) son 84,95 % y 79,71 %, respectivamente. En el caso de la mortalidad y hospitalizaciones en el último mes, los valores oficiales se encuentran dentro del IC de las estimaciones para el caso de Australia. En concreto, la tasa de mortalidad es del 0,34% (0,00% - 0,22%) y la oficial es del 0,006%, y la tasa de hospitalización es del 1,02% (0,36% - 1,68%) y la oficial es del 1,327%. Además, en el caso del Reino Unido, los valores oficiales de ONS están dentro del IC de las estimaciones del número de casos, casos en los últimos 7 días y casos en las últimas 24 horas. Para el resto de variables, las diferencias nunca son grandes en los casos en los que existen diferencias entre los valores oficiales y las estimaciones (posiblemente por subregistro en los datos oficiales).

**Tabla 2.** Métricas de incidencia de COVID-19 en % (95% de IC) obtenidas desde encuestas indirectas e informes oficiales para Australia y el Reino Unido. (1) Personas de 12 años o más que hayan recibido al menos una/dos/tres dosis al 31 de agosto de 2022. (2) Sólo datos de Inglaterra, 5 semanas.

	Australia		Reino Unido		
	Encuestas Indirectas	OWID	Encuestas Indirectas	OWID	ONS
Casos (en el mes pasado)	12.43 (10.26 - 14.60)	1.585	8.67 (7.39 - 9.96)	0.305	9.663
Tasa Vacunación	76.50 (73.70 - 79.29)	84.95	78.86 (77.00 - 80.72)	79.71	93.6/88.2/70.2 (1)
Mortalidad (en el mes pasado)	0.34 (0.00 - 0.72)	0.006	0.43 (0.13 - 0.73)	0.004	0.005 (2)
Hospitalizaciones (en el mes pasado)	1.02 (0.36 - 1.68)	1.327	0.81 (0.40 - 1.22)	0.158	0.044 (2)
Casos (últimas 24 horas)	2.03 (1.10 - 2.96)	0.069	1.30 (0.78 - 1.82)	0.023	1.458
Casos (últimos 7 días)	2.71 (1.64 - 3.78)	0.211	1.30 (0.78 - 1.82)	0.073	1.116

Otro tema interesante que se ha abordado desde CoronaSurveys es la posibilidad de acceder a datos fiables en momento de repuntes de la pandemia como fue el caso de China donde los datos oficiales no ofrecían garantías de ser fiables (Ramírez et al., 2023).

En la web del proyecto <https://coronasurveys.org/> se puede tener acceso a toda la información, documentación y resultados que se han generado desde su existencia en el 2020. Son particularmente interesantes los gráficos que continuamente se han ido generando con las respuestas recibidas, y los trabajos que se han publicado relacionados tanto con la evolución de la pandemia como con los métodos empleados o la eficacia de las vacunas. Por último, resaltar el enfoque internacional del proyecto, y el equipo de investigadores que han hecho y están haciendo posible la innovación en la forma de recoger y analizar datos en momentos críticos como ha sido la pandemia del COVID-19.

## Discusión

Durante la pandemia, se ha evidenciado la necesidad de tener datos fiables de los casos confirmados y muertes para determinar su evolución. Pero también se ha podido ver que obtenerlos a veces ha sido muy complicado. En particular, cuando no se tienen todos los recursos para establecer los casos reales de infección. Así, se requieren de enfoques de estimación alternativos, como el que se ha usado en el proyecto Coronasurys para monitorizar y estimar los posibles casos de personas infectadas con COVID-19. Esto es aún más relevante en países con

un mal sistema de salud. Particularmente, las encuestas anónimas indirectas abiertas proporcionan estimaciones relativamente cercanas a los valores reales, sin embargo, los métodos de estimación a partir de ellas siguen siendo un tema de investigación para garantizar estimaciones más precisas.

Si bien aquí solo se presenta uno de los métodos de estimación de NSUM sobre la cantidad de personas infectadas con COVID-19, el equipo de investigación ha seguido probando y diseñando nuevos métodos de estimación, aplicándolos a nuevas encuestas en otros ámbitos. El grupo está explorando otros enfoques, algunos basados en la teoría de grafos, otros usando técnicas de aprendizaje automático, así como agregando preguntas de control a las que interesan para validar los datos.

En el grupo de investigación se es consciente de las debilidades y sesgos de las estimaciones usando encuestas indirectas, ya que normalmente los métodos NSUM se basan en tres condiciones: todos en la población tienen la misma probabilidad de conocer a alguien en una subpoblación, para cada persona en la red de un encuestado, el encuestado conoce cada subpoblación a la que pertenece la persona, y los encuestados pueden recordar completamente a todos en su red social en el momento de la encuesta.

Estos supuestos se cumplen raramente en la práctica de manera precisa, produciéndose en general sesgos de 4 tipos:

- Error de transmisión, que es producido por el hecho de que el participante no conoce el estado de todos sus conocidos, es decir, no saben todo acerca de las personas en sus redes sociales, violando la condición 2;
- Error de recuerdo, que ocurre cuando los participantes recuerdan incorrectamente el número de personas en algunas subpoblaciones, violando la condición 3. El encuestado puede estar sobreestimando o subestimando;
- Efecto barrera, es el sesgo que se produce porque la probabilidad de saber algo de una subpoblación depende del individuo. Esto significa que es más probable que un participante conozca más de ciertas subpoblaciones que de otras, violando la condición 1.
- Finalmente, sesgo de respuesta, que es un error que se produce por una declaración deliberadamente errónea de los participantes.

Todo lo anterior puede conducir a una estimación inexacta. Propuestas para manejar estos sesgos son un tema interesante de investigación de interés para nuestro grupo.

Finalmente, creemos que un sistema de encuestas como el proyecto CoronaSurveys es especialmente adecuado en países de bajo ingreso, ya que el costo de preparar e implementar una encuesta es extremadamente pequeño, los participantes pueden usar dispositivos muy simples para completar la encuesta (ya que está basada en la web), y el número de los participantes que deben tener información sobre el tema de estudio (por ejemplo, de una pandemia) es bastante bajo. También ha servido para homogeneizar conceptos a nivel internacional cuando la definición de caso era diferente por país, e incluso comunidad autónoma. Particularmente concluimos que las encuestas indirectas abiertas anónimas, en combinación con el método NSUM, ofrecen una opción económica y flexible para monitorear epidemias.

**Concluimos que las encuestas indirectas abiertas anónimas, en combinación con el método NSUM, ofrecen una opción económica y flexible para monitorear epidemias**

### Contribuciones de los autores

Los autores participaron igualmente en la elaboración del manuscrito y aprobaron la versión final presentada.

### Financiación

Este trabajo fue parcialmente financiado por la subvención CoronaSurveys-CM, financiada por IMDEA Networks y Comunidad de Madrid, España, subvenciones COMODIN-CM y PredCov-CM financiadas por la Comunidad de Madrid y la Unión Europea a través del Fondo Europeo de Desarrollo Regional (FEDER), subvenciones TED2021-131264B-I00 (SocialProbing) y PID2019-104901RB-I00 financiadas por el Ministerio de Ciencia e Innovación - Agencia Estatal de Investigación, España (MCIN/AEI/10.13039/501100011033) y la Unión Europea (NextGenerationEU/PRTR), y donaciones individuales al Proyecto CoronaSurveys <https://coronasurveys.org>.

### Declaración de disponibilidad de datos

Los conjuntos de datos presentados en este estudio se pueden encontrar en línea en un repositorio abierto: <https://github.com/GCGImdea/coronasurveys/data/aggregate/>.

### Conflictos de interés

Los autores declaran que no hay conflicto de interés.

## Referencias bibliográficas

- Baquero, C., Casari, P., Fernandez Anta, A., García-García, A., Frey, D., Garcia-Agundez, A., ... Sanchez, I. (2021) The CoronaSurveys System for COVID-19 Incidence Data Collection and Processing, *Frontiers in Computer Science.*, <http://dx.doi.org/10.3389/fcomp.2021.641237>
- Bernard, H. R., Hallett, T., Iovita, A., Johnsen, E. C., Lyerla, R., McCarty, C., et al. (2010). Counting Hard-To-Count Populations: the Network Scale-Up Method for Public Health. *Sex. Transm. infections* 86 (Suppl. 2), ii11–ii15. <http://dx.doi.org/10.1136/sti.2010.044446>
- Dunbar, R. (1992). Neocortex size as a constraint on group size in primates. *J. Hum. Evol.* 22, 469–493, [https://doi.org/10.1016/2560047-2484\(92\)90081-J](https://doi.org/10.1016/2560047-2484(92)90081-J).
- Facebook Data for Good. COVID-19 symptom survey {request for data access. <https://dataforgood.fb.com/docs/covid-19-symptom-survey-request-for-data-access/>. Accessed: 2020-05-20.
- García-Agundez A., Ojo O., Hernandez H., Baquero C., Frey D., Georgiou C., ... Fernandez Anta A. (2021). Estimating the COVID-19 Prevalence in Spain with Indirect Reporting via Open Surveys. *Frontiers in Public Health*, 9. <http://dx.doi.org/10.3389/fpubh.2021.658544>
- Laga I, Bao L., & Niu X. (2021) Thirty Years of The Network Scale-up Method, *Journal of the American Statistical Association*, 116(535), 1548-1559, <http://dx.doi.org/10.1080/01621459.2021.1935267>
- Maxmen, A. (2020). How Much Is Coronavirus Spreading under the Radar?. *Nature* 10. <http://dx.doi.org/10.1038/d41586-020-00760-8>. Available at: <https://www.nature.com/articles/d41586-020-00760-8>
- Nishiura N., Klinkenberg D., Roberts, M., & Heesterbeek, J. (2009) Early epidemiological assessment of the virulence of emerging infectious diseases: a case study of an influenza pandemic. *PLoS One*, 4(8), 2009. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0006852>.
- Ojo O., García-Agundez A., Girault B., Hernandez H., Cabana E. Garcia, A., ... Fernandez A. (2020). CoronaSurveys: Using Surveys with Indirect Reporting to Estimate the Incidence and Evolution of Epidemics, arXiv:2005.12783 v2 Presented in KDD Workshop on Humanitarian Mapping, San Diego, California USA, August 24, 2020
- Oliver N., Barber X., & Roomp, K. (2020). The covid19 impact survey: Assessing the pulse of the COVID-19 pandemic in Spain via 24 questions. *CoRR*, abs/2004.01014.
- Pollán, M., Pérez-Gómez B., Pastor-Barriuso, R., Oteo, J., Hernán, M.A., Pérez- Olmeda, M., et al. (2020) Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. *Lancet*. 396, 535–44. [http://dx.doi.org/10.1016/S0140-6736\(20\)31483-5](http://dx.doi.org/10.1016/S0140-6736(20)31483-5)

Ruppert, E., Grommé, F., Upsec-Spilda, F., & Cakici, B., (2018). Citizen Data and Trust in Official Statistics. *Economie Statistique/Economics Stat.* (505-506), 171–184. <http://dx.doi.org/10.24187/ecostat.2018.505d.1971>

The CoronaSurveys research team. (2020). CoronaSurveys: Monitoring the incidence of COVID-19 via open surveys. <https://coronasurveys.org/>. Accessed: 2020-05-20.

Ramírez, J., Diaz Aranda, S., Aguilar, J., Ojo, O., Lillo, R., & Fernández Anta, A. (2023). A Snapshot of COVID-19 Incidence, Hospitalizations, and Mortality from Indirect Survey Data in China in January 2023. medRxiv 2023.02.22.23286167 Presented in the 6th International workshop on Epidemiology meets Data Mining and Knowledge discovery. Long Beach, California, USA, August, 2023.

## Rastreo digital de contactos: datos de geolocalización a gran escala como alternativa al fracaso de las aplicaciones basadas en Bluetooth

### Digital Contact Tracing: Large-scale Geolocation Data as an Alternative to Bluetooth-based Apps' Failure

Francisco Caravaca<sup>a</sup>, José González-Cabañas<sup>b</sup>, Roberto Boris Martínez<sup>a</sup>, Ángel Cuevas<sup>a,b</sup>, Rubén Cuevas<sup>a,b</sup>, Martin Maier<sup>c</sup>, María Calderón Pastor<sup>a,e</sup>, Jorge Pleite Guerra<sup>d</sup>

<sup>a</sup> Departamento de Ingeniería Telemática, Universidad Carlos III de Madrid, España

<sup>b</sup> Instituto UC3M-Santander de Big Data, España

<sup>c</sup> Institut National de la Recherche Scientifique, España

<sup>d</sup> Departamento de Tecnología Electrónica, Universidad Carlos III de Madrid, España

<sup>e</sup> Departamento de Ingeniería de Sistemas Telemáticos, Universidad Politécnica de Madrid, España

#### Resumen

**Introducción:** Las aplicaciones móviles de rastreo de contactos desplegadas actualmente han fracasado como solución eficaz en el contexto de la pandemia COVID-19. Ninguna de ellas ha conseguido atraer al número de usuarios activos necesario para lograr una operación eficiente.

**Objetivo:** ampliar la definición de los sistemas digitales de rastreo de contactos. **Metodología:** Se ha propuesto un protocolo para el rastreo de contactos utilizando información de proveedores de localización. Dicho protocolo ha sido implementado, y se ha medido su rendimiento. **Resultados:** la solución ha sido implementada, la cual además de ser eficiente garantiza la privacidad de los usuarios

**Conclusión:** La solución presentada permitiría el rastreo de contactos, y con suficiente eficiencia para que escale correctamente.

Palabras clave: COVID-19; rastreo de contactos; Facebook; Google; geolocalización; privacidad.

#### Abstract

**Introduction:** Currently deployed mobile contact-tracking applications have failed as an effective solution in the context of the COVID-19 pandemic. None of them have succeeded in attracting the number of active users required for efficient operation. **Objectives:** expand the definition of digital contact tracing systems. **Methodology:** A protocol for contact tracking using location provider information has been proposed. The protocol has been implemented, and its performance has been measured.

**Results:** the solution has been implemented, which in addition to being efficient guarantees the privacy of users. **Conclusion:** The presented solution would allow contact tracing, and with sufficient efficiency to scale correctly.

Keywords: COVID-19; contact tracing; Facebook; Google; geolocation; privacy.

## Introducción<sup>1</sup>

Existe un consenso generalizado de que cualquier estrategia para luchar eficazmente contra pandemias, tras la experiencia del COVID-19, requiere un rastreo eficiente de los contactos estrechos de las personas infectadas. Estudios recientes concluyen que el rastreo manual no es lo suficientemente rápido y recomiendan el uso de sistemas digitales de rastreo de contactos capaces de utilizar información de localización a gran escala (Ferretti et al., 2020). Un elemento clave del éxito de un sistema digital de rastreo de contactos es su adopción, es decir, el porcentaje de personas que lo utilizan de forma activa y eficaz.

Singapur fue uno de los primeros países en implantar un sistema digital de localización de contactos en el caso del COVID-19 a principios de 2020. Este país optó por implantar una aplicación (app) móvil que utiliza la tecnología Bluetooth (BT) para identificar cuándo dos usuarios han estado cerca. Si uno de esos usuarios da positivo en la prueba COVID-19, el otro es identificado como posible contagio. El 20% de la población de Singapur instaló la app móvil. Pero esto no fue suficiente. De hecho, un representante del Ministerio de Sanidad de Singapur declaró que necesitarían que tres cuartas partes de los ciudadanos instalaran la app para que la estrategia de rastreo digital de contactos tuviera éxito (StraitTimes, 2020).

Aunque no está claro cuál es la tasa de adopción a partir de la cual una app BT de rastreo de contactos se convierte en eficaz para controlar una pandemia, algunos estudios preliminares sugirieron que para mitigar la pandemia del COVID-19 sería necesaria la adopción por parte del 60% de la población de un país (Ferretti et al., 2020; Hinch et al., 2020). Algunos estudios de simulación muestran que si la adopción es inferior al 20%, el beneficio de una aplicación de localización de contactos BT es muy pequeño, pero podemos observar un impacto significativo con una tasa de adopción superior al 40%. Una vez más, nos referimos a la tasa de personas que utilizan activamente la aplicación, y no al número de instalaciones.

Las aplicaciones nuevas de rastreo de contactos basadas en BT tienen un problema importante, tienen que conseguir desde cero el alto índice de adopción necesario en un breve periodo de tiempo. Por lo que sabemos, ni los investigadores ni las instituciones públicas o privadas han propuesto una estrategia convincente para alcanzar la tasa de adopción requerida. Por el momento, parece que el éxito de cualquier app de localización de contactos BT depende únicamente de la responsabilidad de las personas, y en el caso del COVID-19 eso no ha sido suficiente.

A pesar del problema descrito y del fracaso de la aplicación de Singapur, la mayoría de los países occidentales, especialmente en Europa, también optaron por aplicaciones móviles que utilizan la tecnología BT como sistema de rastreo de contactos. En concreto, la mayoría de estos países optaron por utilizar el protocolo descentralizado de rastreo de proximidad con preservación de la privacidad (DP-3T). El principal objetivo del diseño de DP-3T es ofrecer garantías de privacidad total. En concreto, se pretende garantizar que las aplicaciones de rastreo de contactos que utilicen este protocolo no puedan ser mal utilizadas en el futuro para prácticas intrusivas en la privacidad de los usuarios, como la publicidad o la vigilancia masiva. A pesar de que los autores de este trabajo creemos que el protocolo DP-3T ofrece garantías de privacidad muy sólidas, algunos autores han expuesto que la solución no ofrece garantías de privacidad plenas.

Para apoyar a las autoridades sanitarias (HAs) dispuestas a desplegar aplicaciones de rastreo de contactos, Google y Apple desarrollaron el denominado sistema Google-Apple Exposure Notification (GAEN) (Apple and Google, 2021) inspirado en el protocolo DP-3T. GAEN se integró en iOS y Android. El sistema operativo (OS) se limita a registrar los encuentros del usuario con BT y ofrece esta información a la aplicación móvil, que implementa el algoritmo para identificar los contactos de riesgo. A pesar de este esfuerzo, por lo que sabemos, ninguna de las apps de rastreo de contactos existentes ha contribuido significativamente a mitigar la transmisión del virus asociado al COVID-19.

Por ejemplo, los primeros datos de las autoridades sanitarias suizas indicaron que sólo el 12% de las personas infectadas informaban de que son seropositivas a través de la aplicación (Salathé et al., 2020). En España, esta cifra se reducía a aproximadamente el 2% en la práctica, a pesar de un documento basado en un estudio piloto

**Un elemento clave del éxito de un sistema digital de rastreo de contactos es su adopción, es decir, el porcentaje de personas que lo utilizan de forma activa y eficaz**

<sup>1</sup> Una parte del contenido de este artículo ha sido publicada en la revista Electronics <https://doi.org/10.3390/electronics10091093>. Toda la parte de implementación y evaluación de la solución propuesta son totalmente originales.



realizado en La Gomera (Islas Canarias) que planteaba expectativas mucho mayores sobre la eficacia de la aplicación (Rodríguez et al., 2021). Por último, un informe sobre la aplicación en el Reino Unido (Inglaterra y Gales) (Wymant et al., 2021) presentaba resultados bastante positivos sobre la contribución de la aplicación. Sin embargo, al leer el informe en detalle encontramos los resultados bastante decepcionantes. Aunque el informe afirma que el número de usuarios activos oscila entre el 24,2% y el 33,2%, no discute por qué el número de usuarios activos se redujo en gran medida de 16,5M a 13M durante diciembre de 2020 y enero de 2021, lo que implica en realidad un 21% de usuarios activos. En realidad, se trata de una cuestión muy importante, ya que justo en medio de uno de los peores periodos de la pandemia en el Reino Unido, el número de usuarios activos disminuyó casi un 20%. Esto refleja claramente la insatisfacción de los usuarios con la aplicación. Además, el informe muestra la opacidad de este tipo de soluciones para proporcionar datos útiles a las HAs. Los autores se basaron en modelos para estimar diferentes métricas para analizar la eficiencia de la app. Una vez más, los resultados son decepcionantes.

Por ejemplo, los autores afirman:

Nuestro análisis sugiere que se evitó un número relativamente grande de casos de COVID-19 mediante el rastreo de contactos a través de la app del NHS, que oscila entre aproximadamente 200 mil y 900 mil dependiendo de los detalles del método, en comparación con los 1,9 millones de casos que realmente se produjeron (Wymant et al., 2021, p. 410).

La gran variación registrada indica claramente que no es posible evaluar con precisión la eficacia de las aplicaciones basadas en BT y que buscan poner la privacidad por encima de la eficiencia para controlar las infecciones.

Además, las pruebas científicas ponen de manifiesto que la transmisión aérea de COVID-19 es irrefutable (Scientific Brief: SARS-CoV-2 and Potential Airborne Transmission, 2020; Prather et al., 2020; Lednicky et al., 2020), otra limitación importante de las actuales aplicaciones de rastreo de contactos BT. Las apps están diseñadas para identificar el contacto a corta distancia entre dos individuos, es decir, a menos de dos metros de distancia. Sin embargo, la transmisión aérea implica que es posible el contagio entre dos personas a distancias mayores. Por lo tanto, las aplicaciones existentes de rastreo de contactos por vía aérea pueden pasar por alto una fracción importante de contactos que deberían identificarse como contactos de riesgo.

Por último, soluciones como DP-3T, diseñadas con el objetivo principal de ofrecer una privacidad total, presentan otras carencias importantes en la lucha contra una pandemia. Algunas de ellas son: 1) Incluso si la tasa de adopción fuera alta, la mayoría de las apps desplegadas requieren que los usuarios infectados declaren voluntariamente su condición de positivos a través de la app, dejando una tarea muy importante como es el control de una pandemia en manos de la decisión de los individuos. Por ejemplo, un primer estudio realizado en Suiza demuestra que 1/3 de los usuarios de la app que dieron positivo no utilizaron la app para informar de su caso (Salathé et al., 2020). 2) No se puede evaluar el rendimiento y la eficiencia de la app de rastreo de contactos, ni siquiera cuántos usuarios infectados se han detectado a través de la app, como reconocen los autores del protocolo DP-3T (Salathé et al., 2020). Y 3), no son capaces de proporcionar información de contexto agregada y no invasiva de la privacidad, que podría ser de gran valor para mejorar nuestros conocimientos sobre los patrones de transmisión de COVID-19 u otros virus. Por ejemplo, en este documento, consideramos que revelar estadísticas agregadas del tipo lugares (restaurantes, instalaciones deportivas, transporte público, hospitales, etc.) que los usuarios infectados visitaron mientras podían contagiar puede ser útil para identificar sesgos estadísticos en el tipo específico de lugares que pueden revelar puntos calientes para la transmisión del virus.

Dado el contexto descrito, el principal objetivo de este trabajo es ampliar la definición de los sistemas digitales de rastreo de contactos teniendo en cuenta los siguientes elementos clave: 1) evitar soluciones que requieran una adopción masiva desde cero, como ha demostrado la experiencia; 2) las soluciones de rastreo de contactos deben diseñarse para considerar como referencia una distancia de transmisión aérea superior a dos metros; 3) orientar el diseño de las soluciones estableciendo como objetivo principal la eficiencia en la lucha contra la pandemia (es decir, salvar vidas y mitigar el impacto en la economía) en lugar de la privacidad. Por supuesto, la solución propuesta debe cumplir las leyes de protección de datos y privacidad vigentes en el país donde se implante.

En este artículo, proponemos un sistema alternativo de trazado digital de contactos basado en los tres elementos clave anteriores como principios fundamentales de diseño:

1) Alta tasa de adopción: Proponemos utilizar información de localización en tiempo real de (literalmente) miles de millones de personas de todo el mundo que ya está disponible en bases de datos de grandes empresas BigTech como Facebook (FB), Google, Apple, operadores de telecomunicación, etc. En este documento nos referiremos a estas empresas como proveedores de localización o *Location Providers* (LPs) por sus siglas en inglés. Algunos de estos LPs, principalmente Google y Facebook, tienen una tasa muy grande de usuarios activos, más del 50%, en muchos países occidentales.

2) Identificación de contactos en el rango de transmisión aérea: Para geolocalizar a los usuarios tanto en exteriores (GPS.gov, 2017) como en interiores (Google, 2020) con una precisión de pocos metros, estas empresas BigTech utilizan una combinación de técnicas que se basan en múltiples señales, como la información de localización GPS, la potencia de la señal WiFi SSIDs, las señales de la red celular, etc.

3) Requisitos legales y éticos: Estamos interesados en realizar el rastreo de contactos sólo para los individuos que han dado positivo de COVID-19. La identidad de los individuos infectados es información sensible manejada por la Autoridad Sanitaria o *Health Authority* (HA) (por sus siglas en inglés) de cada país, que también es responsable de ejecutar la estrategia de rastreo de contactos. Por lo tanto, la HA tiene la identidad de los individuos infectados mientras que el LP tiene los datos para realizar el rastreo de contactos de esos individuos. Proponemos un sistema que permite llevar a cabo el rastreo de contactos utilizando los datos de los LP sobre aquellos individuos que dieron positivo en las pruebas, tal y como informaron las HA. Incluso las leyes de protección de datos más restrictivas, como la GDPR (EU, 2016), prevén explícitamente excepciones en las que los datos personales pueden utilizarse para controlar epidemias y su propagación (véase el artículo 6 del GDPR (EU, 2016), considerando 46). Sobre esta base jurídica podría ser posible un acuerdo para realizar un intercambio de datos entre los LP y los HA. Sin embargo, para proporcionar mayores garantías de privacidad, proponemos una arquitectura sencilla y un protocolo de comunicación que permitan el intercambio de información entre un LP y una HA limitando significativamente la capacidad de (1) las HAs para obtener el grafo de contactos de un individuo, es decir la estructura y conexión de contactos de los distintos usuarios y (2) los LPs para conocer la identidad de los individuos infectados.

**Hay pocos trabajos que expongan el fracaso de las aplicaciones de rastreo de contactos desplegadas y propongan soluciones alternativas que no dependan de nuevas aplicaciones móviles**

Hay pocos trabajos en la literatura que expongan el fracaso de las aplicaciones de rastreo de contactos desplegadas y propongan soluciones alternativas que no dependan de nuevas aplicaciones móviles (Mokbel, Abbar, & Stanojevic, 2020; Reichert, Brack, & Scheuermann, 2020; Nakamoto, Wang, Guo, & Zhuang, 2020; Rahman, Khan, Khandaker, Sellathurai, & Salan, 2020). Creemos que sería importante realizar pruebas piloto con las más prometedoras para medir su eficacia. Hasta donde sabemos, nuestro trabajo es el primero que propone una solución que proporciona altas garantías de privacidad para implementar el rastreo de contactos aprovechando los datos de geolocalización que están disponibles en las bases de datos de las BigTechs.

Este trabajo es un propuesta prometedoras pero es importante hacer patente que no tenemos pruebas de si nuestro sistema resolverá el problema del rastreo de contactos. Sin embargo, creemos que es una alternativa técnicamente sólida que merece la pena explorar. Además, sirve al propósito principal de este trabajo: animar a la comunidad investigadora a revisar el diseño de soluciones digitales de rastreo de contactos para crear futuras medidas de mitigación más efectivas y eficientes frente a futuras oleadas de COVID-19 y otras pandemias.

Finalmente, hemos desarrollado una implementación de la solución propuesta y hemos hecho pruebas de rendimiento que demuestran que su implementación técnica es viable y cumpliría los requisitos de inmediatez necesarios en la identificación de contactos estrechos en el contexto de una pandemia.

### *Contexto de la solución propuesta*

#### Razonamiento de la solución

Proponemos una solución novedosa de rastreo de contactos que puede utilizar potencialmente datos de geolocalización de miles de millones de usuarios para encontrar a personas que han estado en contacto con individuos con test positivo. Nos referimos a ellos como contactos de riesgo. La información de geolocalización es propiedad de empresas BigTech denominadas Proveedores de Localización o *Location Provider* (LP) en este documento, y la información de los usuarios que han dado positivo es propiedad de las Autoridades

Sanitarias o *Health Authority* (HA). El núcleo de nuestra solución puede describirse como sigue: Las autoridades sanitarias envían a los proveedores de localización los ID de los usuarios infectados. Los LP utilizan la información de localización que poseen para encontrar los contactos de riesgo de los ID recibidos (de acuerdo con las directrices proporcionadas por expertos en epidemiología) y devuelven la lista de IDs de contactos de riesgo a la HA. Por último, las HAs se ponen en contacto con los contactos de riesgo para informarles del protocolo de prevención que deben seguir.

A efectos prácticos, proponemos utilizar el número de teléfono móvil de las personas como ID de usuario en nuestra solución. Los LPs conocen el número de teléfono móvil de una parte importante de los usuarios que utilizan sus servicios ya que se usa para aspectos de seguridad como el doble factor de autenticación, la recuperación de claves, etc., y es razonable suponer que los HAs registran el teléfono móvil de los usuarios infectados para comunicarse con ellos.

Desgraciadamente, el intercambio directo de datos en claro entre las HAs y los LPs presenta importantes problemas de privacidad. En concreto, los LPs no deben recibir IDs en claro de los individuos infectados y los HAs no deben poder vincular las identificaciones de los contactos de riesgo con su correspondiente usuario infectado. Nuestra solución aborda este reto permitiendo la realización de la tarea de rastreo de contactos con fuertes garantías de privacidad. Para ello, definimos una arquitectura y un protocolo de comunicación que implican, además de a los LPs y las HAs, a dos actores más: un proveedor de identidades (IDP por sus siglas en inglés) y una autoridad independiente de terceros (ITPA por sus siglas en inglés).

### ¿Por qué utilizar datos de geolocalización?

Adopción: La principal limitación del rastreo de contactos basado en aplicaciones móviles es la necesidad de conseguir una alta tasa de usuarios activos. Este es un importante cuello de botella que hasta ahora ha hecho fracasar todos los intentos en esta línea.

Nuestra solución evita este cuello de botella utilizando datos de geolocalización a gran escala ya disponibles y propiedad de empresas BigTech. Para comparar explícitamente la penetración de los datos de BigTechs frente a las aplicaciones móviles de BT, la tabla 1 muestra para 18 países para los que hemos encontrado datos sobre el número de instalaciones de aplicaciones de localización de contactos: 1) la tasa de penetración de los smartphones, el sistema operativo Android (Dimoco, 2020; StatCounter Global Stats, 2020; Demographics of Mobile Device Ownership and Adoption in the United States, 2020) y los usuarios activos mensuales o *monthly active user* (MAU) comunicados por FB; (Facebook, 2020) y 2) la tasa de penetración de la aplicación móvil de rastreo en base al número de instalaciones, así como una estimación de su penetración en términos de usuarios activos. Son datos obtenidos en los años 2020 y 2021. La lista de fuentes que hemos utilizado para informar sobre el número de instalaciones de aplicaciones móviles puede consultarse aquí (FDVT, 2021). Nótese que, hasta donde sabemos, Suiza es el único país que informa del porcentaje de usuarios activos de su app, 63% a 21 de diciembre de 2020 (Swiss Federal Statistical Office, 2020). Para tener una estimación de la fracción de usuarios activos de otros países que comunican el número de instalaciones, aplicamos la proporción suiza al número total de instalaciones.

Según nuestra estimación, ninguno de los países alcanza una tasa de adopción significativa cercana al 40% para las aplicaciones móviles de rastreo de contactos, y sólo 5 países superan el 20%. Por el contrario, la penetración de Facebook supera el 50% en todos los países excepto Alemania (45,5%). Del mismo modo, la penetración de Android es superior al 40% en todos los países excepto EE.UU. (32%) y Suiza (39%). Nótese que la penetración de Android sólo representa un límite inferior de la penetración de Google. Google tiene otras aplicaciones muy populares, como Gmail y Google Maps, que son muy utilizadas por los usuarios de iOS.

Precisión: Las grandes empresas tecnológicas utilizan sofisticadas técnicas que combinan señales de GPS, WiFi y redes celulares para geolocalizar a los usuarios con gran precisión tanto en exteriores como en interiores (GPS.gov, 2017; Google, 2020). Google afirma ser capaz de geolocalizar a los usuarios con una precisión de 1 a 2 metros utilizando algoritmos de multilateración basados en la señal Wifi de 3 puntos de acceso (Google, 2020).

Por lo tanto, las altas tasas de penetración y precisión de localización de las BigTechs las convierten en una fuente de datos que puede ser suficiente para implementar soluciones eficientes de rastreo de contactos. Trabajos de investigación recientes, que utilizan datos de LPs con una penetración mucho menor que FB o Google, también respaldan esta hipótesis (Aleta et al., 2020).

**Tabla 1.** Penetración en porcentaje de smartphones, Android, Facebook, e instalaciones de aplicaciones de trazo de contactos y número de usuarios activos estimados en 18 países. La población de cada país para calcular la penetración se ha obtenido de la base de datos del Banco Mundial (Data World Bank, 2020).

Países	Smartphone	Android	Facebook	BT apps móviles	
				Instalaciones	Usuarios activos estimados
Alemania	90	61	45.50	34.5	21.7
Australia	105	44	71.42	27.6	17.4
Austria	117	78	50.25	9	5.7
Bélgica	68	41	65.00	12.2	7.7
Croacia	71	59	50.84	2	1,3
Dinamarca	115	55	71.03	34.8	21.9
España	90	71	62.05	11.5	7.2
Estados Unidos	81	32	69.90	2.5	1.6
Finlandia	140	97	59.65	45.3	28.5
Francia	79	51	58.35	9.5	6
Holanda	82	48	63.09	25	15.8
Irlanda	78	42	65.54	40.5	25.5
Italia	84	62	57.80	21.1	13.3
Letonia	96	69	52.45	9.1	5.7
Portugal	104	78	67.47	1	0.6
Reino Unido	85	40	66.64	36.05 *	21.7*
Rep. Checa	84	66	53.32	14	8.8
Suiza	97	39	52.38	33.4	21.1

\* El número de usuarios activos en la aplicación móvil de rastreo de contactos para el caso de Reino Unido corresponde sólo a Inglaterra y Gales.

### Otros beneficios

La solución propuesta permite supervisar el rendimiento de la misma. Además, las ubicaciones geográficas pueden asociarse a categorías específicas denominadas puntos de interés (Point of Interest o POIs por sus siglas en inglés). Por ejemplo, una ubicación determinada puede asociarse a un restaurante, una estación de tren o un hospital. Nuestra solución aprovecha esto para proporcionar una distribución estadística de los POIs visitados por los usuarios infectados frente a los POIs visitados por la población general. La comparación de estas distribuciones puede ayudar a identificar sesgos estadísticos en los POIs más visitados habitualmente por los usuarios infectados, que podrían ser focos de infección.

### Requisitos de privacidad

Por un lado, los expertos en privacidad y las Autoridades de Protección de Datos (DPAs por sus siglas en inglés) han mostrado su preocupación por el uso de información de geolocalización para el rastreo digital de contactos. Básicamente argumentan que puede facilitar a los gobiernos, a través de sus agencias de seguridad, la aplicación de una vigilancia masiva debido a la escalabilidad que ofrecen las tecnologías digitales. Por lo tanto, nuestra solución debería limitar la capacidad de las autoridades sanitarias para deducir masivamente la información del grafo de contactos de las personas utilizando los datos recibidos de los LP. Además, debería proporcionar disposiciones de privacidad para permitir revelar los ataques selectivos dispuestos a inferir el grafo de contactos de individuos concretos.

Por otro lado, las empresas BigTech disponen de medios para inferir la identidad de los individuos infectados. Pueden aprovechar los datos de geolocalización, pero también otras fuentes de información como correos electrónicos, publicaciones en redes sociales o consultas en motores de búsqueda de su propiedad. Por ejemplo, pueden detectar a un usuario que visitó un centro de pruebas después de visitar su sitio web y luego permanece en su casa durante un periodo similar a una cuarentena obligatoria. Por lo tanto, creemos que las propuestas, como la nuestra, que aprovechan los datos de geolocalización de las empresas BigTech no suponen ningún riesgo adicional para la privacidad de los usuarios infectados porque las empresas BigTechs ya cuentan con las capacidades necesarias para inferir con alta probabilidad cuando un usuario ha sido infectado. A pesar de ello, deben ofrecerse garantías de privacidad adecuadas. En particular, nuestra solución no debería proporcionar a los LPs información explícita sobre la identidad de los usuarios infectados. También debería limitar la capacidad de los LPs para inferir tales identidades a partir de la información recibida de las HAs.

### Cumpliendo con los requisitos de privacidad

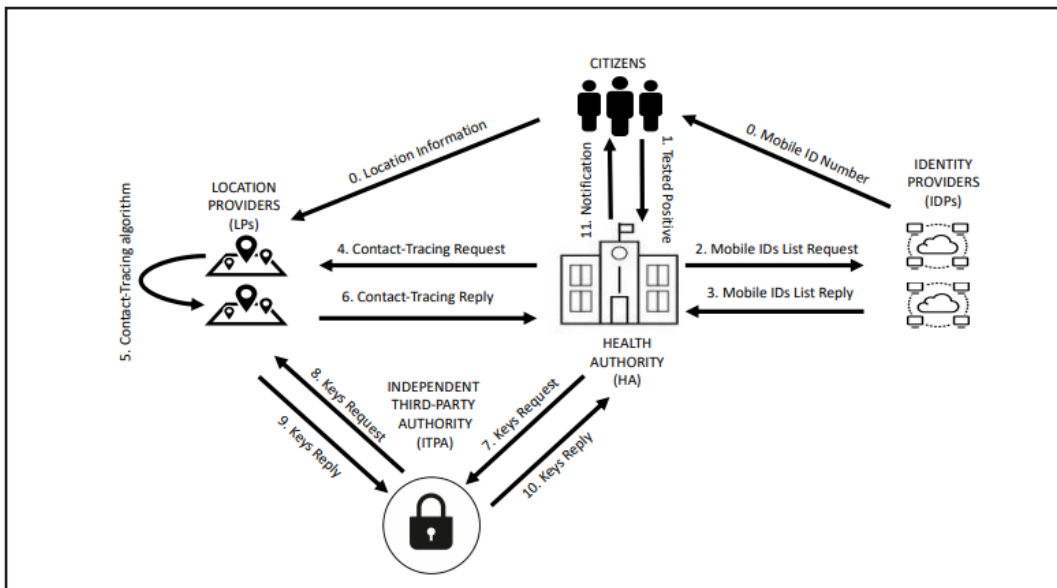
Para cumplir los requisitos de privacidad definidos, aprovechamos los siguientes principios: K-anonymity, criptografía básica y auditoría basada en no repudio.

**K-anonymity:** En nuestra solución, la HA envía una lista de IDs de usuario al LP, y éste responde con los contactos de riesgo de dichos IDs de usuario. Aprovechando los principios del K-anonymity, la HA mezcla en su solicitud  $M$  IDs de usuarios infectados y  $N$  IDs aleatorios reales (es decir, números de teléfono móvil aleatorios asociados a usuarios reales) donde  $M \lll N$ . Esto sirve para anonimizar la identidad de los usuarios infectados y dificultar la capacidad de los LPs para inferir fácilmente los ID pertenecientes a los usuarios infectados. Los IDs aleatorios utilizados por la HA son proporcionados por un proveedor de identidades (IDP) para garantizar que se trata de ID existentes. En nuestra solución, los IDP están representados por operadores de redes móviles.

Además, la HA debe agregar los IDs en grupos. Hay dos tipos de grupos: Los grupos infectados incluyen exclusivamente IDs de usuarios infectados; los grupos aleatorios incluyen IDs de usuarios aleatorios o una mezcla de usuarios aleatorios e infectados. Los mensajes de la HA al LP incluyen  $K$  grupos de los cuales sólo  $L$  son grupos infectados, donde  $L \lll K$ . Tras la recepción de un mensaje de solicitud de la HA, el LP calcula los contactos de riesgo de cada ID de usuario. A continuación, agrega en la respuesta los contactos de riesgo de todos los identificadores de usuario en un único grupo. Este proceso de agregación se basa en el concepto de K anonymity para evitar que la HA relacione los IDs de contactos de riesgo recibidos con un individuo concreto. Obsérvese que cuanto mayor sea el tamaño de los grupos, mayores serán las garantías de privacidad.

**Criptografía:** Una HA honesta sólo está interesada en los IDs de contactos de riesgo asociados a los grupos infectados. Para impedir que las HAs accedan a los ID de contacto de grupos aleatorios, el LP cifra la lista de contactos de cada grupo (incluida en la respuesta al HA) utilizando una clave diferente por grupo. Por lo tanto, la HA recibe los IDss de contactos de todos los grupos cifrados. Para recuperar las claves de los grupos infectados, la HA tiene que enviar una solicitud a un intermediario al que denominamos Autoridad Independiente de Terceros (ITPA por sus siglas en inglés). En esta solicitud, la HA indica el número total de grupos en la consulta, así como el ID de los grupos infectados. A su vez, la ITPA solicita las claves de todos los grupos al LP y reenvía a la HA sólo las claves asociadas a los grupos infectados. Por último, utilizando las claves recibidas, la HA obtiene los IDs de los contactos de riesgo asociados únicamente a los grupos infectados, completando así el procedimiento de rastreo de contactos.

**Auditoría basada en no repudio:** Nuestra solución se basa en el concepto de responsabilidad civil para garantizar los derechos de privacidad de los usuarios. Obsérvese que se trata de un planteamiento ampliamente adoptado en el ordenamiento jurídico de las democracias avanzadas. Por ejemplo, un Estado no puede impedir que alguien conduzca por encima del límite de velocidad, pero quien lo haga es responsable de ello. En el caso de la privacidad, un Estado no puede impedir que una empresa BigTech aplique prácticas intrusivas en la privacidad, pero sí castigarla en caso de que un proceso de auditoría revele el uso de esas prácticas. Por lo tanto, una HA o un LP que utilice los datos que recibe para fines distintos del rastreo de contactos será responsable de ello. Por ejemplo, una HA malintencionada puede llevar a cabo un ataque selectivo (véase la sección Ataques Potenciales y Contramedidas) para desvelar el grafo de contactos de un individuo y filtrarlo a otros poderes públicos. Esto sería un delito equivalente a filtrar el historial médico de un individuo objetivo a otras ramas del gobierno. Nuestra solución recoge las pruebas de no repudio necesarias para que la entidad auditora correspondiente las utilice para desvelar cualquier posible ataque por parte de una HA.



**Figura 1.** Propuesta de protocolo para rastreo de contactos y arquitectura.

### Protocolo para el rastreo de contactos utilizando información de proveedores de localización

En esta sección, describimos los pasos del protocolo de comunicación, incluida la secuencia de mensajes intercambiados por los cuatro actores implicados en nuestra solución: Autoridad Sanitaria (AS), Proveedor de Localización (LP), Proveedor de Identidad (IDP) y una Autoridad Independiente de Terceros (ITPA).

- Paso 0: Este paso se refiere al contexto básico en el que se basa nuestra solución. Por un lado, los LPs registran información histórica sobre la ubicación de los usuarios que ejecutan su sistema operativo, aplicaciones móviles, etc. También almacenan el número de teléfono móvil de una parte importante de los usuarios. Por otro lado, los IDPs, es decir, los operadores de telefonía móvil, proporcionan a los usuarios números de teléfono móvil que sirven como IDs de usuario en nuestra solución.

- Paso 1: La HA obtiene los IDs de los usuarios que han dado positivo en una ventana de tiempo determinada (por ejemplo, un día).

- Paso 2: La HA activa el proceso de rastreo de contactos solicitando al IDP una lista de  $N$  IDs de usuario (es decir, números de teléfono móvil reales). El valor de  $N$  lo decide la HA y puede variar de una solicitud a otra.

Hay algunas observaciones a tener en cuenta: (1) Este mensaje incluye un identificador único denominado *Transaction ID* que se incluirá en todos los mensajes restantes del proceso; (2) El mensaje se firma con la clave privada de la HA. Obsérvese que en el resto del proceso todas las entidades firman con su clave privada los mensajes que envían.

- Paso 3: El IDP responde a la petición de la HA con una lista de  $N$  IDs de usuario aleatorios.

- Paso 4: La HA crea  $K$  grupos. Como se ha explicado anteriormente, sólo  $L$  de estos grupos son grupos infectados y  $K - L$  son grupos aleatorios. Los grupos resultantes se incluyen en un mensaje *Contact-Tracing Request* que se envía al LP. Es importante señalar que los IDs de usuario incluidos en un grupo infectado no pueden estar presentes en otros grupos infectados en esta solicitud ni en solicitudes pasadas o futuras.

- Paso 5: Tras la recepción de la solicitud de rastreo de contactos, el LP ejecuta el algoritmo de rastreo de contactos para identificar los IDs de contactos de riesgo de cada ID de usuario incluido en la solicitud. Los IDs de contacto de riesgo de todos los usuarios de un grupo se agregan para eliminar cualquier vínculo entre un ID de usuario y un ID de contacto de riesgo.

Además, el LP recopila los POIs visitados por cada ID de usuario en una ventana temporal definida en el pasado (por ejemplo, los últimos 10 días). A continuación, calcula la distribución de los tipos de POIs visitados por los ID de usuario incluidos en cada grupo, así como la distribución global de los tipos de POIs visitados por todos los ID de usuario incluidos en la solicitud.

La información asociada a cada grupo, es decir, la lista de IDs de contactos de riesgo y la distribución de tipos de POI, se cifra con una clave independiente por grupo.

Por último, el LP agrega la información cifrada por grupo junto con la distribución de tipos de POI para todos los IDs de usuario y crea un mensaje *Contact-Tracing Reply* que se envía a la HA.

Tres observaciones importantes a tener en cuenta son: (1) El LP debe mantener un registro de la clave utilizada para cifrar cada grupo; (2) El algoritmo de rastreo de contactos implementado por el LP así como el número de días para la identificación de los POIs visitados debe ser definido por los epidemiólogos y está fuera del objetivo de este trabajo; (3) el LP almacena todos los mensajes *Contact-Tracing Request* recibidos con fines de auditoría.

- Paso 6: Tras la recepción del *Contact-Tracing Reply* la HA necesita descifrar la información asociada a los grupos infectados, es decir, la lista de contactos de riesgo y el tipo de distribución de POIs. Para ello, envía un mensaje *Keys Request* a la ITPA incluyendo el número total de grupos incluidos en la *Contact-Tracing Request* y los identificadores de los grupos infectados.

- Paso 7: La ITPA envía al LP el mensaje *Keys Request* pero sólo incluye el *Transaction ID*.

- Paso 8: Tras la recepción del mensaje *Keys Request*, el LP envía al ITPA un mensaje *Keys Reply* que incluye las claves de todos los grupos.

- Paso 9: El ITPA comprueba si el número de claves de la respuesta recibida coincide con el número real de grupos comunicados por la HA. Si los números coinciden, el ITPA genera un mensaje *Keys Reply* para la HA que sólo incluye las claves de los grupos infectados. En caso contrario, el mensaje *Keys Reply* incluye un error indicando que el número de grupos comunicado no coincide con el número de claves facilitado por el LP.

- Paso 10: Tras la recepción del mensaje *Keys Reply*, la HA descifra la información sobre los contactos de riesgo y el tipo de distribución de POIs incluida en el mensaje *Contact-Tracing Reply* para los grupos de usuarios infectados.

- Paso 11: La HA se pone en contacto con los contactos de riesgo.

### *Ataques potenciales y contramedidas*

Como se ha explicado anteriormente, nuestra solución está diseñada para impedir que tanto los LPs como las HAs se comporten de manera inadecuada y accedan a información que no están autorizadas a obtener.

A continuación, explicamos en detalle las contramedidas que ofrece nuestra solución para evitar: 1) que los LPs intenten deducir los IDs asociados a los individuos infectados; y, 2) que las HAs intenten obtener el grafo de contactos de los ciudadanos.

### *Inferencia de la identidad de los usuarios infectados por parte del LP*

Un LP malicioso podría intentar desvelar la identidad de los usuarios infectados basándose en la información recibida en los mensajes *Contact-Tracing Request*, lo que se conoce como ataque de reidentificación. Para ello, podrían utilizar una obtener la única solicitud o combinar solicitudes posteriores para identidad de los usuarios infectados.

Para evitar los ataques de re-identificación, la HA tiene que reutilizar los IDs que ya han sido utilizados incluyéndolos en grupos aleatorios de peticiones posteriores. De lo contrario, si los ID aleatorios sólo se utilizan una vez y se descartan, el LP podría inferir con una probabilidad muy alta que los ID repetidos en diferentes consultas pertenecen a individuos infectados.

Además de la reutilización de IDs, nuestra solución se basa en el principio de K-anonymity. El número de IDs aleatorios,  $N$ , en los mensajes de solicitud es varias veces mayor que el número de IDs de usuarios infectados,  $M$ . La complejidad para realizar un ataque de reidentificación crece con el cociente  $N/M$ . Además, nuestra solución permite introducir un alto nivel de aleatoriedad en los mensajes de petición para evitar que los LPs puedan inferir patrones que permitan identificar grupos que incluyan IDs de usuarios infectados: 1) el número de IDs de usuarios infectados y aleatorios difiere de mensaje a mensaje; 2) el número de grupos en un mensaje difiere de mensaje a mensaje; y, 3) la longitud de los diferentes grupos dentro del mismo mensaje también debería diferir. Además, la HA podría enviar de vez en cuando mensajes que no incluyan ninguna identificación de usuario infectado.

Más allá de las medidas técnicas, el principal argumento para apoyar nuestra solución es que los LPs potentes como Google o Facebook dispuestas a identificar a los ciudadanos infectados ya pueden hacerlo fácilmente con la información que poseen. Por lo tanto, las medidas de privacidad adoptadas en nuestra solución ofrecen garantía suficientes para evitar que aumente el riesgo de un posible ataque de reidentificación por parte de los LP con respecto a ataques que pueden implementar a día de hoy.

### *Inferencia por parte de la HA del grafo de contactos de un usuario-id*

Nuestra solución no puede evitar de antemano que una HA maliciosa obtenga el grafo de contactos de un individuo concreto. Por ejemplo, una HA puede realizar un ataque dirigido utilizando dos veces el mismo ID en dos grupos infectados diferentes. Los contactos de riesgo comunes en los dos grupos pueden revelar el grafo de contactos del individuo objetivo. Una persona podría volver a infectarse y aparecer en dos solicitudes de rastreo de contactos diferentes de la misma HA, aunque estas solicitudes no deberían producirse en un corto periodo de tiempo. La segunda solicitud debería aparecer al menos unas semanas o meses después de la primera solicitud de rastreo de contactos que contenga el ID de esta persona.

Una opción eficiente para evitar que la HA pueda obtener el grafo de contactos de usuarios re-infectados, es obligar a la HA a eliminar la información de los grupos de contacto descifrados al cabo de un periodo de tiempo en el cuál es altamente improbable que un usuario vuelva a reinfectarse. Por ejemplo, cada semana o cada dos semanas. De esta forma, la HA no podrá obtener el grafo de una persona que se re infecta pasados unos meses.

Además, nuestra solución conserva las pruebas de no repudio necesarias para demostrar que se ha producido un ataque de este tipo. La entidad auditora sólo tiene que comprobar si el HA ha utilizado el mismo ID dos veces (o más) en grupos de usuarios infectados en el mismo mensaje o en mensajes diferentes durante la ventana temporal previa al borrado de datos. La entidad auditora puede recuperar todos los mensajes *Contact-Tracing Request* del LP. Del mismo modo, la entidad auditora recupera de la ITPA, para cada mensaje *Contact-Tracing Request*, cuáles son los grupos infectados declarados por la HA. Con esa información, la entidad auditora podría identificar los ataques de la HA. La capacidad de auditoría descrita proporciona garantías de privacidad basadas en la responsabilidad innegable, una técnica muy utilizada en las democracias desarrolladas.

Por último, nuestra recomendación es ejecutar el proceso de auditoría descrito una vez al día para detectar cualquier HA maliciosa poco después de que haya implementado un ataque.

### *Implementación*

El protocolo que forma parte de la solución descrita se implementa sobre HTTPS, en concreto, los dos proveedores: Proveedor de Identidad (IDP) y Proveedor de Localización (LP), y también la Autoridad Independiente de Terceros (ITPA) se desarrollan como servidores HTTPS que reciben peticiones POSTs de la Autoridad Sanitaria (HA), o de los otros servidores. El uso de HTTPS proporcionará cifrado de extremo a extremo de los datos transmitidos, los clientes también comprobarán la autenticidad de los certificados de los servidores. Además, todos los mensajes son firmados y comprobados por todas las partes implicadas.

El repositorio del código está disponible en <https://github.com/fcaravaca/DigitalContactTracing>.

### *Estructura*

Los servidores de la implementación (IDP, LP e ITPA) son contenedores Docker que contienen el propio servidor (será un servidor Express) y una base de datos MySQL. El fichero de configuración permite la posibilidad de lanzar varios contenedores docker de cada tipo de servidor.

### *Encriptación de contactos*

Los contactos de los grupos se envían a la Autoridad Sanitaria cifrados, en concreto con AES-256 (CBC como modo de funcionamiento). El Proveedor de Localización cifrará cada uno de los grupos por separado, con diferentes Claves y Vectores de Inicialización (IVs) generados aleatoriamente. Estas Claves e IVs se almacenan en la base de datos del Proveedor de Localización y sólo se enviarán a una Autoridad Independiente de Terceros (ITPA). La ITPA recibirá el número de grupos y sólo enviará las claves e IVs de los grupos infectados a la Autoridad Sanitaria.

### *Firmas*

Cada uno de los actores involucrados en el protocolo tendrá que firmar sus mensajes para permitir que el otro extremo verifique la autenticidad de los mensajes enviados. Esto se hace con un par de claves Privada y Pública, el emisor utilizará su clave privada para firmar la información, y el receptor la verificará utilizando la clave pública del emisor.

Esta solución utiliza RSA con SHA256. La información se convertirá en una cadena base64, a la que se aplicará el hash SHA256. Esta información se cifrará utilizando la clave privada del remitente. El receptor calculará el hash



SHA256 de la información recibida y descifrará la firma utilizando la clave pública del remitente. A continuación, el receptor comprobará si la firma y la información son equivalentes, validando así la autenticidad del remitente. La firma también puede utilizarse con fines de no repudio.

Cada uno de los servicios tendrá acceso a su clave privada, y a las claves públicas de los otros sistemas, para firmar y validar las firmas respectivamente.

### Formato de los mensajes

Los mensajes entre los servicios siguen la misma estructura básica, en forma de JSON:

```
{
  "información Base64(información),
  "firma": Base64(RSA(SHA256(Base64(información)), privateKey)),
  "id": remitente
}
```

La información es un JSON, cuyo formato varía entre cada mensaje, si bien, en cada uno de los mensajes siempre hay un campo que corresponde al *id transaction*. La información se envía calculando su representación Base64, esto es debido a que la representación del objeto JSON puede variar de extremo a extremo, por lo que es más seguro firmar y enviar la información codificada en Base64, aunque los datos transmitidos aumentarán.

### Auditoría

La auditoría se puede realizar con la información guardada en las Bases de Datos de las ITPAs, LPs e IDPs. Estas tablas incluyen los IDs de las transacciones, la HA que solicita la transacción y algunos de los parámetros relacionados con las mismas. La entidad auditora debería solicitar la información a los LP, IDP e ITPA, para después comprobar el estado de cada una de las bases de datos y ver las solicitudes realizadas desde un HA específico, ver la frecuencia de las solicitudes, o comprobar si el HA solicita información a un IDP cada vez que realiza una solicitud a un LP.

Podrían utilizarse otros datos de auditoría, guardando los datos brutos de las solicitudes realizadas desde la HA al LP. Esta información podría combinarse con la base de datos ITPA, para ver entonces si un identificador de un usuario se utiliza en varios grupos infectados, ya que la HA podría estar intentando inferir el gráfico de contactos de un usuario.

### Mensajes de transacciones completas

En esta sección mostramos el contenido de los mensajes compartidos entre las partes implicadas en una simulación sencilla, para comprobar los contactos de 10 personas infectadas. La HA incluirá 10 teléfonos por cada uno de los teléfonos infectados, por lo que enviará 110 contactos en diferentes grupos. En particular, esta HA decide hacer 2 grupos para los teléfonos infectados, por lo tanto esto significa que en promedio los grupos deben tener una longitud de 5 teléfonos, y habrá 20 grupos de teléfonos al azar. Por lo tanto, un total de 22 grupos serán enviados al Proveedor de Localización.

Aquí mostramos el contenido de las peticiones y respuestas decodificando el base64 del campo info.

Resumen de los parámetros de simulación:

```
Phones infected (M = 10) = +34 600 000 001, +34 600 000 002, +34 600 000 003,
+34 600 000 004, +34 600 000 005, +34 600 000 006, +34 600 000 007, +34
600 000 008, +34 600 000 009 , +34 600 000 010
```

```
L = 2
```

```
N/M = 10 -> N = 10
```

```
K = 22 as K - L = 20, to maintain the same group size in all K groups
```

La HA establece un nuevo *transaction ID* con el valor cc92e749-4480-4f1d-9db6-978efcd46523, y envía un mensaje al IDP solicitando 100 números de teléfono aleatorios.

```
{
  "id": "HA",
  "info": {
    "transaction_ID": "cc92e749-4480-4f1d-9db6-978efcd46523",
    "amount": 100
  },
  "signature": "I7e7zF7Dwi/u20YqQu8CSdWLZm3mdH0gxTIVKeZGTks/v5XNBqOvH3BO
6GrB1HKcKNYJmY4mgz5NEHJfNimRdDcY/Y5RsLq41DDH1SMuY6K9uCCGgPkiW3fekfdIv
n1+7bwp3zLoEvfOdFIOY2RBhTfa3TDRvke26I7WlyqI9RtVBytdduYTGATVPmWaek/5i
tw6CWPkLelOaf9lbzhRRiZAYj3L7M6D6FWNOLYMJDdIdFMRzTq7NyhFQlQ6mdjg1gUJeb
gWSJxclylzl0hcJGpkyRVTQDQsfzDovQ5conImUcY9UAi/ZNBvQeqzmQVC3ILKGeijMJC
R1pSQq6wGQ=="
}
```

Respuesta de IDP a HA, que devuelve los 100 teléfonos aleatorios.

```
{
  "info": {
    "transaction_ID": "cc92e749-4480-4f1d-9db6-978efcd46523",
    "amount": 100,
    "ids": [
      "+34 653 788 444", "+34 648 411 356", "+34 603 539 920", "+34 649
785 675", "+34 663 729 249", "+34 696 432 331", "+34 603 278
344", "+34 682 887 384", "+34 605 283 029", "+34 650 437 443",
"+34 623 465 678", "+34 603 997 526", "+34 685 327 541", "+34
632 726 211", "+34 611 010 619", "+34 669 470 430", "+34 601
175 472", "+34 683 395 773", "+34 650 757 568", "+34 651 397
644", "+34 662 262 889", "+34 650 545 033", "+34 677 323 626",
"+34 673 550 058", "+34 630 476 478", "+34 617 817 103", "+34
672 261 416", "+34 666 891 280", "+34 628 257 589", "+34 670
695 441", "+34 648 951 489", "+34 643 088 158", "+34 632 070
640", "+34 614 093 031", "+34 640 922 608", "+34 640 439 978",
"+34 684 187 530", "+34 613 848 622", "+34 600 669 193", "+34
614 035 237", "+34 673 195 847", "+34 617 674 407", "+34 698
818 999", "+34 658 500 365", "+34 648 217 882", "+34 680 477
943", "+34 672 391 058", "+34 685 116 085", "+34 604 456 764",
"+34 634 022 350", "+34 663 709 522", "+34 639 383 419", "+34
686 091 747", "+34 637 293 314", "+34 688 197 010", "+34 639
884 733", "+34 629 404 805", "+34 618 506 098", "+34 635 722
271", "+34 648 088 325", "+34 621 851 577", "+34 614 437 919",
"+34 696 116 590", "+34 692 536 429", "+34 691 483 831", "+34
602 657 187", "+34 657 763 786", "+34 677 778 487", "+34 668
296 886", "+34 631 319 526", "+34 633 183 965", "+34 646 668
115", "+34 662 066 763", "+34 692 687 369", "+34 656 674 921",
"+34 628 398 123", "+34 663 186 294", "+34 663 953 368", "+34
674 459 789", "+34 668 683 752", "+34 649 132 555", "+34 682
772 220", "+34 639 669 893", "+34 646 046 481", "+34 665 777
823", "+34 640 628 251", "+34 658 464 281", "+34 681 813 791",
"+34 626 986 647", "+34 691 164 273", "+34 637 169 576", "+34
635 382 402", "+34 634 279 625", "+34 658 723 255", "+34 643
013 653", "+34 663 884 025", "+34 679 732 955", "+34 600 346
094", "+34 657 118 433", "+34 667 194 644"
    ]
  },
  "id": "IDP",
  "signature": "W4uZzTh9DAu0N8w+R8qobHrrtUib0FcKezIGbfvpa7ajUDP1GaRB9hm2
qo8hDoteDsGviSjMBJ7c104PoiMeOY3Iw3D88lp6wAZf1j2L34YfqFpQJHfXUyJNSJKX
lCd/LsWiC4mWS3yXDoDtm455gBs2EInTFe8/yso5Qhxp020heosa4YQS0i/PRywKIP9h
45geSVvSniKq7K0Q9r+pKa5yBigV8p1EJKtQ85h+H0dkx7UGX/c67Xdk+p0cYv/bA11o
wISApiUKRGZ5c5IpMhvsFQ61ApoRcCwz85zHqb4PhZ33ia5t/ijoZnc6PJkek8h0KisnC
2CEonuki9A=="
}
```

```
}

```

Tras recibir los telefonos, la HA genera los grupos y asigna a cada uno de ellos un ID de grupo aleatorio. El tamaño medio de los grupos es de 5, pero la cantidad en cada grupo será aleatoria. Los IDs de los grupos infectados son 1e85ada3-bc1e-4442-bbd4-20e3b2e20691 y 2c43ac0e-f072-4aad-aeae-6988671b5746, con longitudes de 7 y 3 respectivamente.

```
{
  "id":"HA",
  "info":{
    "transaction_ID":"cc92e749-4480-4f1d-9db6-978efcd46523",
    "groups":[
      {
        "group_id":"6c62ab11-8866-4dfc-9c66-5459b68d722e",
        "ids":["+34 650 757 568", "+34 600 669 193", "+34 663 709 522", "+34
          674 459 789"
        ]
      },
      {
        "group_id":"607ac506-588e-4a0d-ac8f-c6118fa608f5",
        "ids":["+34 605 283 029", "+34 628 257 589", "+34 648 217 882", "+34
          635 722 271", "+34 633 183 965", "+34 668 683 752"
        ]
      },
      {
        "group_id":"6513caal-86e7-4d3d-8f4a-e18090be2606",
        "ids":["+34 603 278 344", "+34 672 261 416", "+34 662 066 763", "+34
          682 772 220", "+34 640 628 251", "+34 600 346 094"
        ]
      },
      {
        "group_id":"f21c85eb-319f-4934-8053-b0f873893223",
        "ids":["+34 683 395 773", "+34 613 848 622", "+34 673 195 847", "+34
          634 022 350", "+34 686 091 747", "+34 688 197 010", "+34 649 132
          555", "+34 658 464 281", "+34 626 986 647"
        ]
      },
      {
        "group_id":"100a569f-2137-4711-b74c-7e25ae53dabb",
        "ids":["+34 649 785 675", "+34 673 550 058", "+34 614 437 919", "+34
          691 483 831", "+34 643 013 653"
        ]
      },
      {
        "group_id":"7749d184-c758-4b1d-935b-cfc3016bac7e",
        "ids":["+34 648 411 356", "+34 650 545 033", "+34 672 391 058", "+34
          629 404 805", "+34 602 657 187"
        ]
      }
    ]
  },
}
```

```

{
  "group_id": "03fbc052-5134-4703-be20-c407c0835e9a",
  "ids": [ "+34 651 397 644", "+34 614 035 237", "+34 668 296 886", "+34
    631 319 526", "+34 646 668 115", "+34 663 953 368", "+34 657 118
    433"
  ]
},
{
  "group_id": "8303b685-f91d-4e1d-b110-9e799c175a4c",
  "ids": [ "+34 650 437 443", "+34 670 695 441", "+34 680 477 943", "+34
    618 506 098", "+34 692 687 369"
  ]
},
{
  "group_id": "19d4189a-a579-4634-b413-8a7e4f3e2468",
  "ids": [ "+34 696 432 331", "+34 617 817 103", "+34 639 884 733", "+34
    639 669 893"
  ]
},
{
  "group_id": "d2691a90-012d-40f9-91f2-6abf79f01443",
  "ids": [ "+34 603 539 920", "+34 677 323 626", "+34 696 116 590", "+34
    677 778 487", "+34 658 723 255"
  ]
},
{
  "group_id": "01e71438-d681-440f-9ebe-2f719157e08a",
  "ids": [ "+34 682 887 384", "+34 666 891 280", "+34 646 046 481"
  ]
},
{
  "group_id": "cac5d3ed-d5e1-4ca4-9d97-c57ea390ba99",
  "ids": [ "+34 601 175 472", "+34 684 187 530", "+34 658 500 365", "+34
    604 456 764", "+34 657 763 786"
  ]
},
{
  "group_id": "1e85ada3-bc1e-4442-bbd4-20e3b2e20691",
  "ids": [ "+34 600 000 002", "+34 600 000 004", "+34 600 000 005", "+34
    600 000 006", "+34 600 000 008", "+34 600 000 009", "+34 600 000
    010"
  ]
},
{
  "group_id": "54b5d3e4-5c4c-497c-98c2-fd880a0aa04b",
  "ids": [ "+34 663 729 249", "+34 630 476 478", "+34 698 818 999", "+34
    663 186 294", "+34 665 777 823", "+34 637 169 576"
  ]
},

```

```

{
  "group_id": "2c43ac0e-f072-4aad-aeae-6988671b5746",
  "ids": [ "+34 600 000 001", "+34 600 000 003", "+34 600 000 007"
  ]
},
{
  "group_id": "26debb37-f758-4f0c-af1f-ecb8c29d5b27",
  "ids": [ "+34 603 997 526", "+34 643 088 158", "+34 639 383 419", "+34
    692 536 429", "+34 635 382 402", "+34 667 194 644"
  ]
},
{
  "group_id": "85cdf393-613b-41bf-92fe-046318523af0",
  "ids": [ "+34 632 726 211", "+34 614 093 031", "+34 656 674 921", "+34
    681 813 791", "+34 663 884 025"
  ]
},
{
  "group_id": "cf3b88bc-5961-49c6-a6e4-f925491e3f7e",
  "ids": [ "+34 653 788 444", "+34 662 262 889"
  ]
},
{
  "group_id": "e960bb6e-660b-4ff6-8e0f-a840ce01a260",
  "ids": [ "+34 669 470 430", "+34 640 439 978", "+34 617 674 407", "+34
    621 851 577", "+34 628 398 123"
  ]
},
{
  "group_id": "a61a1144-08d7-4a95-9306-96918e0be333",
  "ids": [ "+34 623 465 678", "+34 648 951 489", "+34 685 116 085", "+34
    637 293 314"
  ]
},
{
  "group_id": "5960a1a5-db1a-42bd-bb67-adfcff3f85c9",
  "ids": [ "+34 611 010 619", "+34 640 922 608", "+34 648 088 325", "+34
    634 279 625"
  ]
},
{
  "group_id": "daf8579e-a23e-4fa7-9b46-885158930741",
  "ids": [ "+34 685 327 541", "+34 632 070 640", "+34 691 164 273", "+34
    679 732 955"
  ]
}
},
"signature": "FviNsY6T+jauxiC/hcKOjpoH6WTiKfIyFhCIT6LENDL+IglMB3PxzAPT

```

```

DeBe1WHTv1NMQVEew+0okHW8Qb0BbQZZ01PSrU1jwwLGLLmnrhYdgdbqlAuMP29Sovs2Z
PUBPrf770kHH2qXvpt8H0xZd1Y5LB0d58XXMW2Isn7qIJ6cnf3PsVHJZ/c8SvN1ntM6yQ
X10Mtr3rekPy3knkAb0bo0G1K3FMO/gNSoLryEYJGtmK1P5copdZ0grszam4PUJSbz2zn
n8cGRxPny2sBcFu6j4CXVC/ZPyInYxEB0e0FmHiopEsqEmp9PCka845kh0ShDvpBtxPOW
YEhmGA/xyw=="
}

```

El proveedor local ejecuta el algoritmo de detección de contactos estrechos y envía los contactos de cada grupo cifrados.

```

{
  "id": "LP1",
  "info": {
    "transaction_ID": "cc92e749-4480-4f1d-9db6-978efcd46523",
    "groups": [
      {
        "group_id": "6c62ab11-8866-4dfc-9c66-5459b68d722e",
        "contact_ids": "sUUEPK+NeCbPZLHTNim9oy8JcJZRyZ4m7mq1wGMLVd3uds1a
Zeyk9YWZtS0syYHJv0zAqjGWzcNBDu00qbs+yyuVNE5b/Yu8owY81kQL6GIHTIW
d3SKOoFhM0Y24fiXMOOrX31uZwiZB7YHb85Ty7fZpG80fjxjp/aTCQ/vX2SLsv6Q
uuKXtehCfdPRQMfeI+M4CRh7wjHZ6n22pHp4yKZ3VE3XM2N09n5IUKgbtAmki70
ACGa8R3znGmtwcfykNU"
      },
      {
        "group_id": "607ac506-588e-4a0d-ac8f-c6118fa608f5",
        "contact_ids": "UA2guYJVDDXqclP8r5uq4A0WRqgJQmLHVW4EV+VPmacv0Z5K
eA+IlLewtyOCHyEcZ4DooQlPznyWB8pTDpTzT6SyonAjaLTcTKShHNAFHFqr0tb
8LVdjoGtU0/nzetuDZtuaLMsC0FuoJDFRE6ZE/A=="
      },
      {
        "group_id": "6513caa1-86e7-4d3d-8f4a-e18090be2606",
        "contact_ids": "ZwoFGVVARyBzrRBZTvDlYcYZbr8QFV8nssU/+lmnrIJZ41ca
6hPQ9h5Tgl4pcbqo7lp43LdDyPKK1ng8rPGYmdKQIRsRAZSeCun0aZ/HYwuIUuz
Fvc5/R5VgIvlpw75SiTP0Be3Qfq1GsC4ZJ5Uhtv1JJKUkn8k8qpd389j3HS4IB2
V81Uqts/CX8BmWCnrp8alr1ANV+syoCN8fAGyiucYP1Xmg/Gq/L3ZuyN2ufG54W
MTS4KebQ6E0q38nud3HHivM09fxkSVrsVT8216ESA=="
      },
      {
        "group_id": "f21c85eb-319f-4934-8053-b0f873893223",
        "contact_ids": "InjL6sLDyjjv9DcoRJR0nQSiJYErNhKQHIUf2XgXh8nk4vs2
uOAFk3DaG4PRIIZ1diEDf4cgsdt5f6wqXyXHhN7xfXvD8QFAdJHR/F3MEgj03KB
Lji9Opk6Nf8QUn9DEk8mutrMnxr5FW90+yfNqEcCHA7TxMxagnB6M8exPae1zkj
OGBbZ9jH7R7b6ID1wIbSon99LurtySQekPAISZWRsVHyPHcui81FRckqZBcV1Df
bFOBA+NXZSjHyU+6j50de6yfxUXTU0s4ii8gRoGLAaZDuahywpbAfJH4juXSiM3
Bo/LkaOmSZqchtHuwaFR9Bt1jXhFafrFbypMknCdsuHOzpl1HF7wvxEDdpAmUwp
DF9pCDrnPrFwir9A001NqU5Fi5Ft+JHf9cvTWBKSobfUJVSbiQSNPurkuzkmps9
v20crmctYg0+bkkySRBGgKuv91z9E4TMLaw2UHiDo0zCMxd8J2IkDntDmyQ/2wo
ccl+F0b8vR3kJVsFH0vJSFV02DpmZC5nruhhf8c8pMp4txQ3yUr7NlkMkWlcTdq
gAbEV0Tfmfu7HG190mQwKgANHVOOqIDup3Qn7RATOIzKtA=="
      }
    ]
  }
}

```

```

},
{
  "group_id": "100a569f-2137-4711-b74c-7e25ae53dabb",
  "contact_ids": "hRLwG5I+xLtDB0g+Me0cxH/gQ5FxSBT4/bpLSiol575Sp5oK
dNqFwFkTaGG9yajwCHoptU+tVW0Lp/rxmwtSLV2unzdT3inodGIIs9mbfxuhU6
7z7nbPwaZ2wvXSBbrudusQakbTHx7NfoqeDGGWOB7KSZZE01RAzjnakjFAI6fg1
bRZL22TE4ymyVneSXZOMbVmslqBvb1PjVNbflAipkpP125MRMpmnanQ0hSw1oLn
HF96xuvHfiAvcj7ooGFn/TUxuIkRrfV9Vh4VsC9qg=="
},
{
  "group_id": "7749d184-c758-4b1d-935b-cfc3016bac7e",
  "contact_ids": "RzNuT90hIXX6iB7BvVMMwNv9Ma4uIGl+92MkAFKHM/ln7ub7
+cy30tWZeGct8ocgaqyeA7VukDhZBScwAQt4N//2XWeiAMjykBnOlhosxAutsXO
10oBvDmulyATgILb8va4mOGUeIXTWakBuo5lBeSvXYy7mzbgrsP0l3fhVrS0kb3
cqlupE1/Z+4nqSJYq+qg6UNY9jFMLTTmc5dPyt+Kwe2Szg88blhDmpCzxoan0="
},
{
  "group_id": "03fbc052-5134-4703-be20-c407c0835e9a",
  "contact_ids": "WNY+YPvgn4Ed/1A/Xx00H0LUkQSKfWI1hQ7rv8nq2bCHXGR7
6BfQUpNcCKtuD3IBslgJPW5U1tU8DDgKBFSQGCZPGsn1DtbXt8oqvHgHHw0UUL
UXkDqTbzRoL7zkXQkiBbeLqz0x6wrVaeEDn48MTIngEQ04bPe0P/mbk71X8mwDB
xm8VL+/P0nsAmRh1S1DeZ/r+Ut83nmuxA2XRimkQFGYLV7WB3BUI3/zceD6WrSU
L4dhvJJoVRuEsYuuzyy"
},
{
  "group_id": "8303b685-f91d-4e1d-b110-9e799c175a4c",
  "contact_ids": "S4qXjgbXGtTfTRKDbXI9PiktosJ9D59FIGE3rgv1sC6Sp6hf
XbkfA1zGUDAR00nAcpsLEEj+fvUAXE2BIxSA7Qm30FSbx86IAvsBlMMBRA9qUFy
8GcXBC3NRy96gCfYEUwPvc5saTYgMX8b/MQU1rskPLBshqYLEC8iSuSUFm6n9in
n9nbaQtqxAUCMOU9/+I5K8tdrEXw32r141e02ENO5hinzOMuImh9+Mkp5ymeXJQ
h1Fpp92EkDgnQ03EKXV"
},
{
  "group_id": "19d4189a-a579-4634-b413-8a7e4f3e2468",
  "contact_ids": "FSv+GtwB1tyy+1YuYKc+pW2DLOsWqpEhXpAYtwQrz7ySNnjM
K5wCP3qjYQ+JH8wKHlr62AjS8aPDoMkkL+LJMznxA6gNRf0YdkQYcNQ0Ro5dXSh
pT7OPE3MsnB4LcZ4tsUQhz9A39WawxmyTg1LSmw=="
},
{
  "group_id": "d2691a90-012d-40f9-91f2-6abf79f01443",
  "contact_ids": "fH80TsqZ2MYqV/FRMyB5CRf77ShprDZYbciUmY5qNVRD0SjC
EpM+UpKWzCZDMtu2lznvlDaSjv3sBhMZmRlpuhm0BqDx8mxn1bH/Klkj8zvwZo1
qk2qlmo86alyqEfsFVIVG6rSIXqGREcf0X+ADGDEaxCuh7j4enETQx49/qMvPAS
O+I9M1ebKL9RhkFWl3LCixghsDbQ29G+bItyBcq4YKWOBVLK4o3HRi6r2AxG0ka
+KDxqOaTw9FzXQVgvt5ZpGYzhmG4X1kBWx2RrAvg=="
},
{
  "group_id": "01e71438-d681-440f-9ebe-2f719157e08a",
  "contact_ids": "07GUI8Un3bsYm6fM+jPZFL2RhqaE0oUcmErEnYknIbT/OZVR

```

```

i+yRwAr4v4/EbcIUjQQQ9yIiHX3K00CyYRr0JvUpYmwXfJedDYyV6PZnTDSr3q
PFxaxfHVq+qHYfEfyZcl9cAqXGV9giF4ldWn3oR180ymYlGp5YueBPeX7o9Los9
loS040SNAuFvi2Iw4Qa71MUAH20JaaDBwksTG/yp48o7f/g98eBvv6en4mh2U="
},
{
  "group_id": "cac5d3ed-d5e1-4ca4-9d97-c57ea390ba99",
  "contact_ids": "UWD2qCZWG4RA766BHa8ZiXe1foxk0NsxcbUf41eYrHUPQXql
v96CauQW9NoSng/iDFsvAFd9F7zbVViwYdVJMEQFANypvsnfGegpGpKktZKDZ9N
YCX9CSUMRHe/TISk0Qjzr3zLNZKeD2JzvG9ZsGipCqv/B/fvAnr1A5w9nz8iv6E
1L+1l8jXlm+AWujLtGe10s5HlkvQ4trPL1qc9x9sg9fgyI1seX47h87PCbMYpor
2KRRKAp3wzPL8q40S2+yFgE5vnuhPnfaxxGwlh1xA=="
},
{
  "group_id": "1e85ada3-bc1e-4442-bbd4-20e3b2e20691",
  "contact_ids": "F/00zIj7tKxjGLVYOJaUfjLJmZ6+kr0wHbMWzgzH6KwV82tG
VkW+zoUB1p1Ctpclo7yud49wUdUly8sg+QMNZ84tAYCzWNbT6vjLwi7fL3h7wo0
cPoVOZibsnq/osD6999Zns3kVHPBXfv2W8hvg5lUAdP/9PLa8G9yeXUMn4o3IBZ
8vCg0gIcEb87rgszIHiftQfdcSiclMiPi/8wCWYi2eAHigQP8pLEoKN8yUdBTzS
CdJr7HETgkuw+D6G5kGCw1xNTulGem2i/jhcfbT/sJHWhzLPXrtJ/0kg38rwbRS
ISmnK4vLXhVcOigGCGUTaDBVvYI7bHlFgCK3qWsFYQfQwpSLyn45JYJ4eysHNrg
="
},
{
  "group_id": "54b5d3e4-5c4c-497c-98c2-fd880a0aa04b",
  "contact_ids": "D60jicPbCrzIq8Vo9wZ/WjTypIv2xxMbPQhCRR1/GSAqTvJ9
R4vT1fNB0iI0BBpG5IOEWBFK9eyBq9aQ9a6rQrG/TtjvyIpjLaZ2XWK5ppIcvLU
qq76MQyN5f00vSec11lTDik2nxAcU7yTjn32P+t/wDzENy5No02nGFJvLjwkXbN
bz1uncazLoszvlhG0+MzE21hsYStEVJjSsYn9M+OqNYMQ9RfI2cpZ5gPQJ/DWzE
4LSvLpUCr+DNC0/eGJ+FwFYiViWZzzkNaIhPh3WlVINsS0XPHDY9HUuJopW6f7c
xUcGLG8Bh+BYlHPb9EYR1of8kgvNZ9LeRNTIuRSqMQ=="
},
{
  "group_id": "2c43ac0e-f072-4aad-aeae-6988671b5746",
  "contact_ids": "qej+knFvl7LAtPrA2IzBDFbF69qcZ2r65Bz+NTMEHl5dbu++
+eYTUqykvRg5lgie3h3lkrpa/f0FTiTsQTPzJqwLfy88ajxq0mfx9Ap2EgpMswV
d3CPGZLWtk1JBWgksDaJaIZW8mksdQPZE41Z9mg=="
},
{
  "group_id": "26debb37-f758-4f0c-af1f-ecb8c29d5b27",
  "contact_ids": "cBU2kBm0bS30t6X35hQczki6xKN05lIqoFjXUJCpk8nTdiyy
FOmAM+0tnpZa2JDe1pZh+K/Hb0tpQDpwfQyQL4g3rHFIZRjd0TcZNRQWpB1e2aw
xSMhHMOb5ly2iMb56pSJJ5rpjA16jRik4j96KqOTtTtdYgvPTCD4QtnHSQu4Kq6
igUyue0A+M5phK5qXkvD+hz5/M1kAyaaxE6zRkTHyL+4+32XuyfJ2UbnJzK6096
7A6Lm05ULQHBLXYirdUyAUoIfhDH7V+KdJnDCmVr7Hrv/pAovUaI82tgEAAaXGd
Ji3Qcy7Jm+3imMs/NCbd"
},
{
  "group_id": "85cdf393-613b-41bf-92fe-046318523af0",

```



```

"contact_ids": "SZ050yUALZuwE8ELq8Uc6nmaiAKU2cogIXj3ZYWF8KVN5bx4
UPKH108AImJ0mkm5QKEunn2w1kp61J+EUUjVgWWCe1PmyPpvOXWOp1dBj9AKSr+
LkrmY2lWk0zkZlZAxXFuoTKd1OpzH/teKRjboy/SDSyvwfY1d69l9M5zkFnpPEN
5qP/V2LvlQYfiANu+ZQzTRvexd2qNng33e3JWM0+5D5xy43UH+X78uRba0lu0yn
c5Gywev3+Jj3CZHNGE3"
},
{
"group_id": "cf3b88bc-5961-49c6-a6e4-f925491e3f7e",
"contact_ids": "G9etHXY4n4dmci6NMJi9D/jAiTRmz6rrJkvbU1UKtN11s1Mw
7vpk3A/kVnt3+b3dCfqSvnJ0u2EIgeYHUIUZPnh8iVMm9eWJbUf7WRcISVPcdrr
kcqgb7I3pB5YtRvUZ4KdRsbx20YY745HHDnN+zA=="
},
{
"group_id": "e960bb6e-660b-4ff6-8e0f-a840ce01a260",
"contact_ids": "NW1A8Qw22LLziJaoMqjJaOf277WxytGM/3VPXl0cQ10/+TXl
LgUzIZZod7WI+RHGZSDVa7x29cXyfmr5JPYUVDLVYKhc+Km01TJR+yxzm1Cu0xQ
WY1W7ZehDlLqt0UgXhXGQ5hzmKDsQ5TISNC3Jq9+m00oY3ZA6gfkXAAmt+udh5Z
tcbb2IKwVx4oAzyR4WiQoiKgZ7UjrjDxn5y+IyTgyDPPiJ0DOSYPiH18GYXLf5L
YI2EmbAxVwEXZ8vvWUba0jFo1YHKVoGzSidhkipBQ=="
},
{
"group_id": "a61a1144-08d7-4a95-9306-96918e0be333",
"contact_ids": "6woWdnRrnJcAtoHnbQ0wSkPjIFqN+zYkWMJjEjBuyM8nMovl
vEJ8JT6fKA2CLnaM2jvseMA2VyF6lixusvm7tD+IrpcXenAUFUhxju/h3/Xj/xj
3v6b+o3bQ1dQMjLpAJXpH+9m/2KSVtL7Ps+mOyw=="
},
{
"group_id": "5960a1a5-db1a-42bd-bb67-adfcff3f85c9",
"contact_ids": "Ve983FBn/05TfV89sI6Pq4V2xFc9+pvuhwB586Io3FEBwPMk
x21szaDRwSTYqrMSaupyoexNiqn/mYiAekMes+QUa6u1ZcFavM1lgM6/nXw=="
},
{
"group_id": "daf8579e-a23e-4fa7-9b46-885158930741",
"contact_ids": "YNHYzblXgm44Tn+WAXAZyv8zLhTSo2uRF4SM8uCVpwHi476
j5MB76U7CrP71L6Iiy8uCKGkkqa/mvfSNzch43jZ0T2sXps062Mm065NNpLznE7
p0L+r40a0oGniPjq3cKpFv+jL3azFnxoVgXGY2E6HboXPYuDmBzniZageFBNE4
M0wyLCdx3LYpsbTERoIoNo3mfvJl2ykacwhKFzyGj+NJO70ZhNTNta53R5ZbU=="
}
]
},
"signature": "PKUx4vVALPBWF5Gn7tWB01dMMAWVDuyKQxumgPcjZH1qgEqXqZs30k+K
NdlwmoDecmYyxkoSuet8z+K5Y6XM70DdghGwwIIJvJwnmd17C3JUM+jyHmVikPlhLQgFP
SX0oZHJBBK7frSQo9I4vavnNnLujQG9ZcrZy/Ap7JMimligNit5Yw/EBlmiIYVIblvip1
EvqA7/kBpXF20tfgKlma9uMR5fA3+bzYDcjDiMyJ0qGVcht4yFQ709mwVk8CF1h8HFOr
IsvH70Nu7V9r8JJMYiI7YWgdprMxGZ/vcTYM+dFg02RVFSspybUVAh8Z1iPvRDVHLXNqW
u0QEzEAfrA=="
}

```

La HA envía una petición de clave a la ITPA, especificando el ID de la transacción, el número total de grupos, cuales son los grupos infectados, y cuál era el ID del LP que se consultó.

```
{
  "id": "HA",
  "info": {
    "transaction_ID": "cc92e749-4480-4f1d-9db6-978efcd46523",
    "total_groups": 22,
    "infected_groups": [
      "2c43ac0e-f072-4aad-aeae-6988671b5746",
      "1e85ada3-bc1e-4442-bbd4-20e3b2e20691"
    ],
    "LP_ID": "LP1"
  },
  "signature": "azOpxZlXkLUwx6lqW/HEJJeGvQIHCeGKuc5gmOS3r2/Xuft8nbKTZnl6/Pbbxdl5XdJBxarigMVCshWxUn8SC+oRaMUEg+nKJzF03lLa+RljNq/xWlf7JU3yLZFR6ljHCZJJueg1Qg0/hIqnbYpEay6dDLgFw9CeuQfs3qkohWbds6WA0xbtnCvLJrT0ekcteCl12ZT1M6qui2zZpBItWsCS8OSqkRmxPyBKNkXQ2TLvwwLuSIyGq+wEicJlKXPsSEXc1tdYdmIrmV9iWcIymqc7pzFMA6x/Np0514XEOfbpmDotrQZpiW3nERMAYRPMq6ejXD2sJZAQWL6DyShmZpQ=="
}
```

Recibida la información, la ITPA solicitará al LP las claves para descifrar los contactos.

```
{
  "info": {
    "id": "ITPA",
    "transaction_ID": "cc92e749-4480-4f1d-9db6-978efcd46523"
  },
  "signature": "wl0dseb8LDJgfuQZWUQ1fumb7UD0k5KvDwp4P2ZYNXATDQxRnmK9UNK8WSc5uaf5viRcoJ7TnLdyH1ILrUeOahZxf+K0ghUz98PMYtALW095buXwgxTShhyImT/r6pZY+0A8nczH7kP22t5SrcYppAH88nN52oUrTQoodY92Riy2pmILX1ARhXfhp5EDBfdTpvYvPKM3J9x/j6DSZ7MSvnqmm6bIjqTdrDcyk4J+7rA3wqowPtQGwBnNqLSP+WLZ49hxRawP4g8bkQs9g3UIL9ktdagmhme7bHrkn9pnKtaBX4d+N1bugbtsnWAUtoxpMqdIwRbg8Uho6uR1/uVN9w=="
}
```

El LP responderá con la Clave y el IV.

```
{
  "info": {
    "transaction_ID": "cc92e749-4480-4f1d-9db6-978efcd46523",
    "keys": [
      {
        "group_id": "01e71438-d681-440f-9ebe-2f719157e08a",
        "aes_key": "dbf0d880a8d8de24d09eff2d3359c8416286aacb9f666f93cf33d46158899a06",
        "iv": "96b0874c1806d8dd441fcaa8f4491ff6"
      },
      {

```

```
"group_id": "03fbc052-5134-4703-be20-c407c0835e9a",
"aes_key": "c7646cb2117fd29a73871a9cfa80f8020d02ee414c79bf783cbe
ccaba4a2e6ba",
"iv": "764cc4012f81d90738cb5fb3df110af4"
},
{
"group_id": "100a569f-2137-4711-b74c-7e25ae53dabb",
"aes_key": "92fb488e762acde51d0d5ce9cf68c3e862e07db0e25d235e4074
74739abaa9ba",
"iv": "b764f39f9966ab02e839a819d3c8e5b2"
},
{
"group_id": "19d4189a-a579-4634-b413-8a7e4f3e2468",
"aes_key": "c0c40b5f5f0a350eedd10a8856082ab0df3a069fa7474401ef5
feb425773215",
"iv": "defa401a2a175c7f7a6604b43aa03eba"
},
{
"group_id": "1e85ada3-bc1e-4442-bbd4-20e3b2e20691",
"aes_key": "2068c64ee3a05c3661afbe97747e44169785de604cf6503cc65b
0ea0b0777a27",
"iv": "63a07db34447eeca94f111955a10f85a"
},
{
"group_id": "26debb37-f758-4f0c-af1f-ecb8c29d5b27",
"aes_key": "541f1f0d9051c8d4f1a3ba140a4142c014d1409b09a2cbe6cf38
e42d679bee14",
"iv": "e1a2db8b1c67b9554cb609fbf5a5189a"
},
{
"group_id": "2c43ac0e-f072-4aad-aeae-6988671b5746",
"aes_key": "be463056c7ce14fcd3dce92b274c97893f2b669726024aa2a484
1bfa64a3122a",
"iv": "b2c3856d2afd93eb3b4e98f657abbb80"
},
{
"group_id": "54b5d3e4-5c4c-497c-98c2-fd880a0aa04b",
"aes_key": "fd42f6d85edc941550905bd7cdd32c67e3a42f408a2d1549d26d
2b3635895809",
"iv": "a30c6d0098f7217ccf9e7269e7a07c56"
},
{
"group_id": "5960a1a5-db1a-42bd-bb67-adfcff3f85c9",
"aes_key": "03f633c90ed401585d0a014456da123c200768c1ff0cdc399fd4
aeb8783810a7",
"iv": "be614bb14a3cd5fab10620c32ff1d61a"
},
{
"group_id": "607ac506-588e-4a0d-ac8f-c6118fa608f5",
```

```

    "aes_key": "9f86747f3327e4e86c4544f682d83b0e949b6e4ad437913aa779
      603096cd9806",
    "iv": "0df4ec0fc5fe5077ccb7a18fe6bd6a35"
  },
  {
    "group_id": "6513caa1-86e7-4d3d-8f4a-e18090be2606",
    "aes_key": "0355f32bbe5ab055a90db35329c82012f9e89aa6c66bb0473c8c
      93c411e3276f",
    "iv": "1b42acdbef8fe02d1f537b431698d495"
  },
  {
    "group_id": "6c62ab11-8866-4dfc-9c66-5459b68d722e",
    "aes_key": "1e0e465138a67734b9adde87204a398e00ea5133afc2ade34feb
      3713c5493a5e",
    "iv": "13b7b1a65a76a1dd2fa54f9483328373"
  },
  {
    "group_id": "7749d184-c758-4b1d-935b-cfc3016bac7e",
    "aes_key": "ca7e03095a81e63256f77f0d90b6b0ee54956492978138988805
      d852faa2e0a1",
    "iv": "7ea6c1272ffbe1a5676e2cd9bc277f62"
  },
  {
    "group_id": "8303b685-f91d-4e1d-b110-9e799c175a4c",
    "aes_key": "249a64a7c553422a67310a8b5f0130b38eeb4333f705402bd6e8
      e6de228fbd7a",
    "iv": "44640c138e3d9f27b1880b129a01a480"
  },
  {
    "group_id": "85cdf393-613b-41bf-92fe-046318523af0",
    "aes_key": "e3e5278ff69a25bec2af00c1f69120ac72f37da45270656b977f
      45457499e18e",
    "iv": "ed79a52af5f9400fdca5210d45cd0064"
  },
  {
    "group_id": "a61a1144-08d7-4a95-9306-96918e0be333",
    "aes_key": "8e76896d4f453a45d11d667460d7d6bdadadaaa08a48d90f9caf85
      7cfb5d2ef727",
    "iv": "7d0c8a508afbb240626c53aecca6f7bf"
  },
  {
    "group_id": "cac5d3ed-d5e1-4ca4-9d97-c57ea390ba99",
    "aes_key": "4f271c40151d5982e7958947c44ab3dbb8c503572f88213b5eee
      84d80b4acc2a",
    "iv": "65b72d8c1fff5d9b2b46ccad7769fc71"
  },
  {
    "group_id": "cf3b88bc-5961-49c6-a6e4-f925491e3f7e",
    "aes_key": "0ece50f67905f47f850273562af971b5c24ad04fb5766f7170dd

```

```

        66aea4c0e170",
        "iv":"bbb3920f27327b543ddfd8fc9e4169de"
    },
    {
        "group_id":"d2691a90-012d-40f9-91f2-6abf79f01443",
        "aes_key":"aacf4153249b2da7d5c6f98374b47118e5f28e6ac4a34a5092c0
            eb0d1b767285",
        "iv":"eb81d0420dfccc435b38bf67ac00fca2"
    },
    {
        "group_id":"daf8579e-a23e-4fa7-9b46-885158930741",
        "aes_key":"894220016aab3e1d77abade8e2fddc1ab7f3044f8cbfa56303f9
            82157995479e",
        "iv":"58ea2a0f7e0ddc1c4b9ffceca6253e88"
    },
    {
        "group_id":"e960bb6e-660b-4ff6-8e0f-a840ce01a260",
        "aes_key":"2890ac60629f156055fac9aa6cf2da21301e5660e156564f8b36
            c1a72fd44a26",
        "iv":"313f5e5cedfb4466ec8cfb363a788ba6"
    },
    {
        "group_id":"f21c85eb-319f-4934-8053-b0f873893223",
        "aes_key":"3ec64b5c8b62770f76e98916ec41128462f7b7a26a5c9a12653c
            df924cf02726",
        "iv":"33445f835d1612f0fe5925b076288a32"
    }
]
},
"id":"LP1",
"signature":"lKY4EsbTX3nX7rtMNXWTQbTWB35ameUTerHEHs9VB9TGtQ83btCCH9yL
AHx4ePgVewm+olIhP2oVxfQPr7lc420p9d0DryS74UTiRl0f72PD4kZLpinH3y75QuVeJ
AeQO40z9pcI8HI7Owu8oTcT/wEEe67Q5EIxnqrEA+YVXdeicf4D90nrhIuLsLDRwk5bZ
nXTWms1DZlqW8rH2W8jDPJYZX6e0K7rFXRGZ2TZqbnsN/3YFBD0yKeSjiC1TtvAIpx9IZ
JAKeilLcaLE4wwGoX+cjkSvH8RKwSk5EZPXpoT/WcJYu0eIDGRL8IhgGSTNag4AJl6TRqJ
cEVSVDdfow=="
}

```

La ITPA comprobará que el número de grupos coincide, y devolverá sólo las claves para desencriptar la información de los grupos infectados.

```

{
  "info":{
    "transaction_ID":"cc92e749-4480-4f1d-9db6-978efcd46523",
    "keys":[
      {
        "group_id":"1e85ada3-bc1e-4442-bbd4-20e3b2e20691",
        "aes_key":"2068c64ee3a05c3661afbe97747e44169785de604cf6503cc65b
            0ea0b0777a27",

```

```

    "iv":"63a07db34447eeca94f111955a10f85a"
  },
  {
    "group_id":"2c43ac0e-f072-4aad-aeae-6988671b5746",
    "aes_key":"be463056c7ce14fcd3dce92b274c97893f2b669726024aa2a484
      1bfa64a3122a",
    "iv":"b2c3856d2afd93eb3b4e98f657abbb80"
  }
]
},
"id":"ITPA",
"signature":"HA469U9kYE+PadUL40fnei2kD6gZrYc61bDXid6Su9C4WX+L3n6xiUJu
WIMRAE2T8w4CNRh3k5ZtrJ7TzLP23Fhka0/QCxtSV0Z5qDXukFHRhzVgr/FQnIDE14qx
V8Jjk2Fcay5BiRrUfaGPeztAoiCm6tagd1W0pJ4RjXWk92g6wch0vkz1JkicfEE/m1ngF
YPT6KpK/Oxkx2ShX23ZIY4xCRPkpmYJzY0faqIOQgd6DwxP09wkBnoh5H2CY7f4fOP/cD
bxuhAE2Z46dFmLDFUgJEqKrih1mKPrFOIEkoE1u65eTyYGw23ReHxzo1qy7mshbCeqrXE
Y2ei1XSvUA=="
}

```

Una vez que la Autoridad Sanitaria reciba las claves de los grupos infectados, podrá descifrarlas.

Las tablas 2, 3 y 4 muestran el estado de las tablas de la transacción cc92e749-4480-4f1d-9db6-978efcd46523.

**Tabla 2.** Base de datos del IDP.

transaction_id	health_authority	phones_requested	creation_date
cc92e749-4480-4f1d-9db6-978efcd46523	HA	100	2022-10-03 09:32:45

**Tabla 3.** Base de datos del ITPA.

transaction_id	ha	lp	transaction_timestamp	total_groups	infected_groups	result
cc92e749-4480...	HA	LP1	2022-10-03 09:32:45	22	2	Transaction OK

**Tabla 4.** Base de datos del Proveedor de Localización.

transaction id	ha	group id	aes key	iv	creation date
cc92e749-4480...	HA	01e71438-d681-440f-9ebe-2f719157e08a	dbf0d88...	96b0874...	2022-10-03 09:32:45
cc92e749-4480...	HA	03fbc052-5134-4703-be20-c407c0835e9a	c7646cb...	764cc40...	2022-10-03 09:32:45
cc92e749-4480...	HA	100a569f-2137-4711-b74c-7e25ae53dabb	92fb488...	b764f39...	2022-10-03 09:32:45
cc92e749-4480...	HA	19d4189a-a579-4634-b413-8a7e4f3e2468	c0c40b5...	defa401...	2022-10-03 09:32:45
cc92e749-4480...	HA	1e85ada3-bc1e-4442-bbd4-20e3b2e20691	2068c64...	63a07db...	2022-10-03 09:32:45
cc92e749-4480...	HA	26debb37-f758-4f0c-af1f-ecb8c29d5b27	541f1f0...	e1a2db8...	2022-10-03 09:32:45
cc92e749-4480...	HA	2c43ac0e-f072-4aad-aeae-6988671b5746	be46305...	b2c3856...	2022-10-03 09:32:45
cc92e749-4480...	HA	54b5d3e4-5c4c-497c-98c2-fd880a0aa04b	fd42f6d...	a30c6d0...	2022-10-03 09:32:45
cc92e749-4480...	HA	5960a1a5-db1a-42bd-bb67-adfcff3f85c9	03f633c...	be614bb...	2022-10-03 09:32:45
cc92e749-4480...	HA	607ac506-588e-4a0d-ac8f-c6118fa608f5	9f86747...	0df4ec0...	2022-10-03 09:32:45
cc92e749-4480...	HA	6513caa1-86e7-4d3d-8f4a-e18090be2606	0355f32...	1b42acd...	2022-10-03 09:32:45
cc92e749-4480...	HA	6c62ab11-8866-4dfc-9c66-5459b68d722e	1e0e465...	13b7b1a...	2022-10-03 09:32:45
cc92e749-4480...	HA	7749d184-c758-4b1d-935b-cfc3016bac7e	ca7e030...	7ea6c12...	2022-10-03 09:32:45
cc92e749-4480...	HA	8303b685-f91d-4e1d-b110-9e799c175a4c	249a64a...	44640c1...	2022-10-03 09:32:45
cc92e749-4480...	HA	85cdf393-613b-41bf-92fe-046318523af0	e3e5278...	ed79a52...	2022-10-03 09:32:45
cc92e749-4480...	HA	a61a1144-08d7-4a95-9306-96918e0be333	8e76896...	7d0c8a5...	2022-10-03 09:32:45
cc92e749-4480...	HA	cac5d3ed-d5e1-4ca4-9d97-c57ea390ba99	4f271c4...	65b72d8...	2022-10-03 09:32:45
cc92e749-4480...	HA	cf3b88bc-5961-49c6-a6e4-f925491e3f7e	0ece50f...	bbb3920...	2022-10-03 09:32:45
cc92e749-4480...	HA	d2691a90-012d-40f9-91f2-6abf79f01443	aacf415...	eb81d04...	2022-10-03 09:32:45
cc92e749-4480...	HA	daf8579e-a23e-4fa7-9b46-885158930741	8942200...	58ea2a0...	2022-10-03 09:32:45
cc92e749-4480...	HA	e960bb6e-660b-4ff6-8e0f-a840ce01a260	2890ac6...	313f5e5...	2022-10-03 09:32:45
cc92e749-4480...	HA	f21c85eb-319f-4934-8053-b0f873893223	3ec64b5...	33445f8...	2022-10-03 09:32:45

### *Análisis de rendimiento de la solución propuesta*

Nuestra solución requiere que su implementación pueda escalar a un número elevado de teléfonos. Por ello se ha ejecutado la simulación con diferentes parámetros para comprobar si el sistema puede escalar, es decir, cuánto tarda la ejecución del algoritmo en diferentes circunstancias. Para cada uno de los escenarios, ejecutamos el sistema y registramos el tiempo varias veces, y luego calculamos el tiempo medio de ejecución.

En realidad, el Rastreo Digital de Contactos llevará más tiempo, ya que estas simulaciones no incluyen el tiempo necesario para realizar el rastreo real de contactos en los proveedores de localización: obtener los contactos cercanos asociados a cada uno de los teléfonos. Por lo tanto, el tiempo total necesario para ejecutar este sistema es el siguiente:

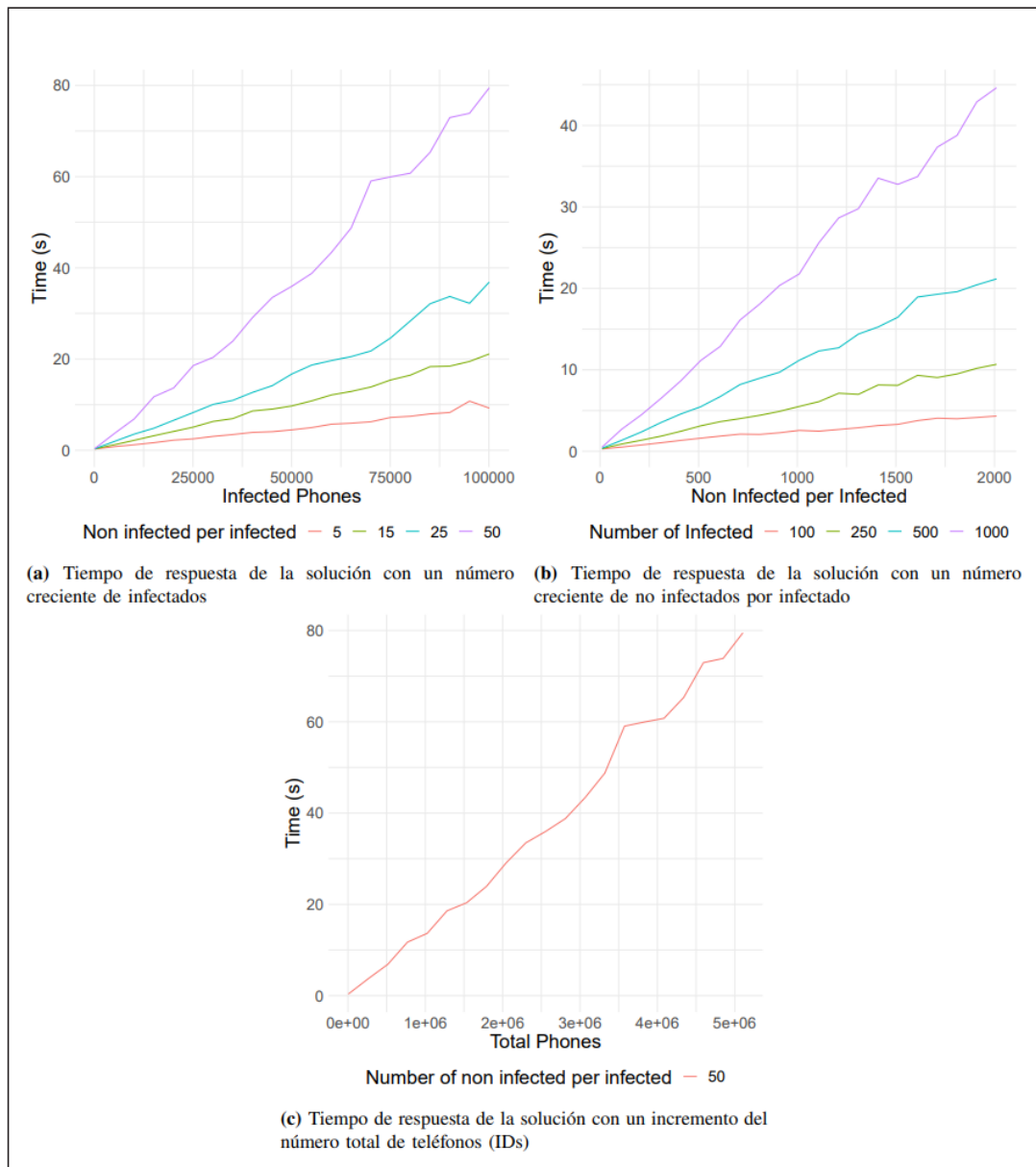
$$TiempoTotal = t_0 + CTR_{(M+N)} + IDR_N + NetworkDelays \quad (1)$$

Donde  $t_0$  representa el tiempo que necesitan los sistemas para preparar, enviar y procesar las diferentes solicitudes del sistema, a excepción de la Solicitud de Rastreo de Contactos, de  $M + N$  teléfonos en el Proveedor Local, y el tiempo que necesita el IDP para recuperar  $N$  teléfonos reales. Este tiempo cambiará en función del estado de la red (*Network Delays*). Por lo tanto, el tiempo representado en los siguientes escenarios es el  $t_0$ , a excepción del escenario  $C$  que también incluye los *Network Delays*. Aunque el  $t_0$  cambiará dependiendo de las capacidades del dispositivo que ejecute cada uno de los servicios.

Para simplificar las mediciones sólo hay un IDP y un LP que la HA puede contactar. En un entorno real, sería mejor consultar a varios LPs.

### Número de contactos a rastrear

La figura 2 muestra diferentes escenarios. El primero (2a) muestra cómo el tiempo necesario para ejecutar el algoritmo aumenta el número de teléfonos infectados, con diferentes números de  $N / M$ . La figura 2c representa el mismo escenario cuando  $N / M = 50$  para mostrar el número total de teléfonos: se tardarán 80 segundos en procesar una solicitud de 100000 personas infectadas con más de 50000000 teléfonos solicitados a un Proveedor de Localización.



**Figura 2.** Evaluacion de rendimiento de las solución propuesta.

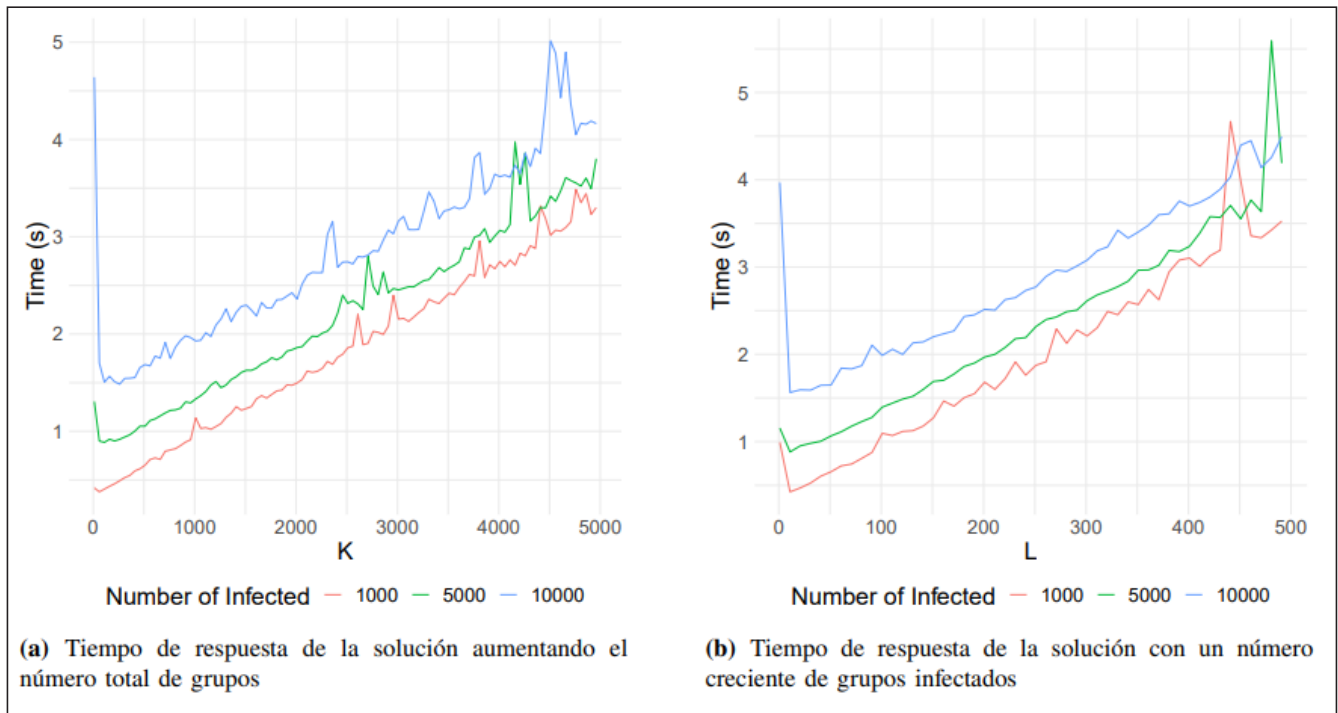


### Número de grupos

Ejecutamos diferentes simulaciones, con diferentes valores de  $M$  (1000, 5000, 10000), y  $N = 10 \cdot M$  para ver el efecto en el rendimiento. Las figuras 3a y 3b muestran cómo el aumento de la cantidad de grupos se traduce en un mayor tiempo de ejecución de los valores.

La figura 3a varía entre  $K = 2$  y  $K = 5000$ , mientras que la figura 3b va desde  $L = 1$  hasta cerca de  $L = 500$ .

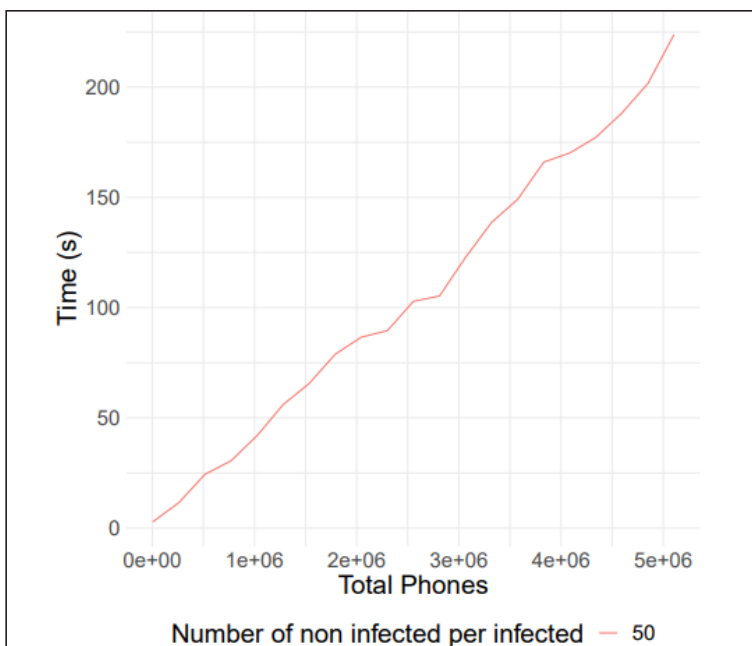
Como se observa en las figuras, al aumentar el número de grupos se tarda más en procesar las solicitudes, con la excepción de sólo dos grupos en total.



**Figura 3.** Rendimiento con diferente número de grupos para  $N = 10 \cdot M$ .

### Transacción en un entorno completo no contenido

Desplegamos el sistema: 1 LP, 1 IDP, y 1 ITPA en máquinas en diferentes redes y recursos limitados (por ejemplo, raspberry pi y VPS) para ver cómo se comporta el sistema en una situación real. Como se ve en la Figura 4, el sistema escala lo suficientemente bien incluso con varios millones de teléfonos en una sola transacción.



**Figura 4.** Tiempo de respuesta de la solución en un entorno real.

Los resultados sugieren que este sistema de comunicación aumenta linealmente con el número de teléfonos solicitados. Aunque el tiempo requerido podría considerarse muy alto, es necesario entender que el sistema no necesita tener un tiempo de ejecución muy bajo. Además, el tiempo para realizar el rastreo de contactos real en el LP podría requerir aún más tiempo.

## Conclusiones

El único enfoque de rastreo digital de contactos utilizado para luchar contra la pandemia COVID-19 en la mayoría de países desarrollados consistió en la utilización de aplicaciones móviles que aprovechan la tecnología Bluetooth para identificar encuentros de proximidad. En este artículo, hemos destacado la principal limitación de este enfoque: la falta de una adopción suficiente de este tipo de aplicaciones móviles, lo que ha hecho fracasar todos los intentos en este sentido.

Debido a la importancia que pueden tener las soluciones digitales de rastreo de contactos para ayudar a combatir futuras pandemias, es obligación de los investigadores, las autoridades de salud pública y las empresas tecnológicas explorar alternativas para encontrar una solución eficaz de rastreo de contactos. Como ejemplo de este esfuerzo de exploración, en este artículo proponemos una primera solución alternativa prometedora para el rastreo de contactos que invita a las autoridades sanitarias y a las empresas BigTechs a cooperar juntas.

Proponemos utilizar datos de geolocalización escalables y precisos ya existentes, que probablemente sirvan para construir una solución digital eficiente de localización de contactos. La alternativa presentada a las actuales aplicaciones de localización de contactos se basa en la alta tasa de adopción ya disponible de la información de localización en tiempo real procedente de miles de millones de ciudadanos de todo el mundo. Esta información se almacena en bases de datos de grandes empresas tecnológicas que ya cuentan con un enorme número de usuarios activos. Además, esta solución tiene en cuenta las ubicaciones interiores y exteriores, abordando también la transmisión aérea demostrada en el caso del COVID-19. Esta propuesta define una arquitectura que aprovecha tales datos y proporciona suficientes garantías de privacidad a los ciudadanos.

Por último, hemos implementado nuestra solución sobre HTTPS con el objetivo de ver si nuestra solución es escalable en términos de volumen de datos y tiempo de respuesta incluso en el caso de manejar peticiones incluyendo millones de identificadores. Esta implementación proporciona cifrado de extremo a extremo, así como autenticación, ya que todos los mensajes deben estar firmados. Este sistema de comunicación ha sido probado en un entorno real, en dispositivos con recursos muy limitados y situados en diferentes redes. Los resultados muestran que la solución propuesta escala de manera eficiente para su uso en el rastreo de contactos digital.

### Contribuciones de los autores

Los autores participaron igualmente en la elaboración del manuscrito y aprobaron la versión final presentada.

### Financiación

Este trabajo es parte del acuerdo entre la Comunidad de Madrid y la Universidad Carlos III de Madrid para la financiación de proyectos de investigación del SARS-CoV-2 y la enfermedad COVID-19, para el proyecto «Multi-source and multi-method prediction to support COVID-19 policy decision making», que ha sido financiado con los fondos REACT-EU del fondo de desarrollo regional Europeo a «way of making Europe».

### Declaración de disponibilidad de datos

El código usado para implementar la solución propuesta se encuentra disponible en <https://github.com/fcaravaca/DigitalContactTracing>

### Conflicto de interés

Los autores declaran que no hay conflicto de interés.

**Proponemos utilizar datos de geolocalización escalables y precisos ya existentes, que probablemente sirvan para construir una solución digital eficiente de localización de contactos**

## Referencias bibliográficas

- Aleta, A., Martín-Corral, D., Bakker, M. A., Piontti, A. P. y., Ajelli, M., Litvinova, M., . . . Moro, E. (2020) Quantifying the importance and location of SARS-CoV-2 transmission events in large metropolitan areas. *medRxiv*. Retrieved from <https://www.medrxiv.org/content/early/2020/12/17/2020.12.15.20248273> doi: 10.1101/2020.12.15.20248273
- Apple and Google. (2021). *Exposure Notifications: Using technology to help public health authorities fight COVID-19*. (Accessed on: February 26, 2021. <https://www.google.com/covid19/exposurenotifications/>)
- Data World Bank. (2020). *Total Population*. (Accessed on: December 27, 2020 <https://data.worldbank.org/indicator/SP.POP.TOTL>)
- Demographics of Mobile Device Ownership and Adoption in the United States*. (2020, Jun). Pew Research Center. (Accessed on: December 27, 2020 <https://www.pewresearch.org/internet/fact-sheet/mobile/>)
- Dimoco. (2020, 04). *Market Insights*. (Accessed on: December 27, 2020 <https://dimoco.eu/carrierbilling/coverage/>)
- EU. (2016). *Regulation (EU) 2016/679 of the European Parliament (General Data Protection Regulation)*. European Union. (Accessed on: April 22, 2020 <http://eur-lex.europa.eu/eli/reg/2016/679/oj>)
- Facebook. (2020). *Facebook Marketing API*. (Accessed on: December 27, 2020 <https://developers.facebook.com/docs/marketing-apis>)
- FDVT. (2021, 02). *Contact tracing apps: number of downloads sources across countries*. (Accessed on: February 28, 2021 [https://fdvt.org/files/COVID\\_APPS\\_SOURCES.pdf](https://fdvt.org/files/COVID_APPS_SOURCES.pdf))
- Ferretti, L., Wymant, C., Kendall, M., Zhao, L., Nurtay, A., Abeler-Dörner, L., . . . Fraser, C. (2020). Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science*, 368(6491). Retrieved from <https://science.sciencemag.org/content/368/6491/eabb6936> doi: 10.1126/science.abb6936
- Google. (2020, Mar). *Wi-Fi location: ranging with RTT*. (Accessed on: April 20, 2020 <https://developer.android.com/guide/topics/connectivity/wifi-rtt>)
- GPS.gov. (2017, Dec). *GPS Accuracy*. (Accessed on: April 20, 2020 <https://www.gps.gov/systems/gps/performance/accuracy/>)
- Hinch et al., R. (2020). Effective configurations of a digital contact tracing app: A report to NHSX. *GitHub*. (<https://github.com/BDI-pathogens/covid-19-instant-tracing/blob/master/Report>)
- Lednicky, J. A., Lauzardo, M., Fan, Z. H., Jutla, A., Tilly, T. B., Gangwar, M., . . . Wu, C.-Y. (2020). Viable SARS-CoV-2 in the air of a hospital room with COVID-19 patients. *medRxiv*. Retrieved from <https://www.medrxiv.org/content/early/2020/08/04/2020.08.03.20167395> doi: 10.1101/2020.08.03.20167395
- Mokbel, M., Abbar, S., & Stanojevic, R. (2020, October). Contact Tracing: Beyond the Apps. *SIGSPATIAL Special*, 12(2), 15–24. Retrieved from <https://doi.org/10.1145/3431843.3431846> doi: 10.1145/3431843.3431846
- Nakamoto, I., Wang, S., Guo, Y., & Zhuang, W. (2020). A QR Code–Based Contact Tracing Framework for Sustainable Containment of COVID-19: Evaluation of an Approach to Assist the Return to Normal Activity. *JMIR mHealth and uHealth*, 8(9), e22321.
- Prather, K. A., Marr, L. C., Schooley, R. T., McDiarmid, M. A., Wilson, M. E., & Milton, D. K. (2020). Airborne transmission of SARS-CoV-2. *Science*, 370(6514), 303–304. Retrieved from <https://science.sciencemag.org/content/370/6514/303.2> doi: 10.1126/science.abf0521
- Rahman, M. T., Khan, R. T., Khandaker, M. R., Sellathurai, M., & Salan, M. S. A. (2020). An automated contact tracing approach for controlling COVID-19 spread based on geolocation data from mobile cellular networks. *IEEE Access*, 8, 213554–213565.
- Reichert, L., Brack, S., & Scheuermann, B. (2020). Privacy-Preserving contact tracing of COVID-19 patients. *IACR Cryptol. ePrint Arch.*, 2020, 375.
- Rodríguez, P., Granã, S., Alvarez-Leo´n, E. E., Battaglini, M., Darias, F. J., Herna´n, M. A., . . . others (2021). A population-based controlled experiment assessing the epidemiological impact of digital contact tracing. *Nature communications*, 12(1), 1–6.

Salathé, M., Althaus, C. L., Anderegg, N., Antonioli, D., Ballouz, T., Bugnion, E., . . . von Wyl, V. (2020). Early Evidence of Effectiveness of Digital Contact Tracing for SARS-CoV-2 in Switzerland. *medRxiv*. Retrieved from <https://www.medrxiv.org/content/early/2020/10/04/2020.09.07.20189274> doi: 10.1101/2020.09.07.20189274

*Scientific Brief: SARS-CoV-2 and Potential Airborne Transmission*. (2020, Oct). Centers for Disease Control and Prevention (CDC). (Accessed on: December 27, 2020 <https://www.cdc.gov/coronavirus/2019-ncov/more/scientific-brief-sars-cov-2.html>)

StatCounter Global Stats. (2020). *Mobile Operating System Market Share Worldwide*. (Accessed on: December 27, 2020 <https://gs.statcounter.com/os-market-share/mobile/>)

StraitTimes. (2020, April). *Call for more people to use contact-tracing app*. (Accessed on: April 20, 2020 <https://www.straitstimes.com/singapore/call-for-more-people-to-use-contact-tracing-app>)

Swiss Federal Statistical Office. (2020). *Swiss Covid App Monitoring*. (Accessed on: December 27, 2020 <https://www.experimental.bfs.admin.ch/expstat/en/home/innovative-methods/swisscovid-app-monitoring.assetdetail.13407769.html>)

Wymant, C., Ferretti, L., Tsallis, D., Charalambides, M., Abeler-Doerner, L., Bonsall, D., . . . others (2021). The epidemiological impact of the NHS COVID-19 App. *Alan Turing Institute*.

## Modelización estadística para la estimación y predicción de la incidencia de COVID-19 en España

### Statistical modelling for estimating and forecasting COVID-19 incidence in Spain

David Moraña<sup>a</sup>, Alessandra Ybargüen<sup>a</sup>

<sup>a</sup> Departamento de Econometría, Estadística y Economía Aplicada - Riskcenter-IREA, Universitat de Barcelona, España

#### Resumen

**Introducción:** Basar procesos de toma de decisiones en datos que contienen errores e imprecisiones es inevitable en muchos de contextos por diferentes razones. La situación derivada de la pandemia mundial de COVID-19 es un claro ejemplo, donde los datos proporcionados por fuentes oficiales no siempre fueron fiables debido a problemas de recopilación de datos y a la alta proporción de casos asintomáticos.

**Objetivos:** Cuantificar la gravedad de la información errónea en una serie temporal y reconstruir la evolución más probable del proceso, así como una discusión sobre los métodos estadísticos más adecuados para obtener predicciones en este contexto. **Métodos:** Se propone el uso de un modelo autoregresivo con heterocedasticidad condicional y estimación de los parámetros mediante Bayesian synthetic likelihood. **Resultados:** Solo alrededor del 51% de los casos de COVID-19 en el período 23 de febrero de 2020 al 27 de febrero de 2022 se notificaron en España, observándose también diferencias relevantes en la intensidad del subregistro entre comunidades autónomas. **Conclusión:** La metodología propuesta proporciona a los tomadores de decisiones en salud pública una valiosa herramienta para mejorar la evaluación de la evolución de una enfermedad bajo diferentes escenarios, ya que permite generar predicciones realistas en este contexto.

Palabras clave: modelización estadística; series temporales; datos infraregistrados; enfermedades infecciosas; COVID-19

#### Abstract

**Introduction:** Basing decision-making processes on data containing errors and inaccuracies is unavoidable in many situations. The COVID-19 pandemic related data is a clear example, where the information provided by official sources was often unreliable due to data collection mechanisms and the amount of asymptomatic cases. **Objectives:** To estimate the amount of misreported data in a time series and reconstructing the most probable evolution of the process and provides a discussion on the more appropriate statistical methods able to yield reliable forecasts in this context. **Methods:** The usage of a model based on autoregressive conditional heteroskedastic time series is proposed, estimating the parameters by Bayesian synthetic likelihood. **Results:** Only around 51% of the cases of COVID-19 in the period from February 23rd, 2020 to February 27th, 2022 were observed in Spain, also detecting remarkable differences in the reporting issues between Autonomous communities. **Conclusion:** The presented method allows generating realistic predictions under different possible scenarios, and therefore it represents a valuable tool for policy makers in order to improve the evaluation of the evolution of a situation.

Keywords: statistical modelling; time series; underregistered data; infectious diseases; COVID-19.

## Introducción

La situación provocada por el virus SARS-CoV2, desde finales de 2019, ha puesto de manifiesto que es fundamental contar con datos de calidad en la cadena de toma de decisiones, tanto en epidemiología como en muchos otros campos. Existe una enorme preocupación mundial en torno a esta enfermedad, lo que llevó, en marzo de 2020, a la declaración de emergencia de salud pública por parte de la Organización Mundial de la Salud (OMS) (Sohrabi et al., 2020).

La modelización matemática permite, mediante la utilización de herramientas matemáticas y la información disponible sobre una determinada enfermedad, representar y predecir una situación epidémica, estimar situaciones futuras y evaluar actuaciones ya realizadas. Desde el inicio de la pandemia se hizo evidente no solo la necesidad de disponer de información epidemiológica fiable, sino también la importancia de utilizar estas técnicas como apoyo en la gestión de la crisis por parte de las autoridades sanitarias.

Hasta la fecha, se han realizado muchos esfuerzos metodológicos para lidiar con los datos de COVID-19 mal informados, siguiendo las ideas introducidas en la literatura desde finales de los noventa (Alfonso, Løvseth, Samant & Holm, 2015; Arendt et al., 2013; Bernard et al., 2014; Gibbons et al., 2014; Rosenman et al., 2006; Winkelmann, 1996). Estas propuestas van desde el uso de factores de multiplicación (Stocks et al., 2018) hasta modelos basados en cadenas de Markov (Azmon et al., 2014; Magal & Webb, 2018) o modelos espacio-temporales (Stoner, Economou & Drummond Marques da Silva, 2019). Recientemente se han propuesto varios enfoques basados en series de tiempo discretas (Fernández-Fontelo, Cabaña, Joe, Puig & Moraña, 2019; Fernández-Fontelo, Cabaña, Puig & Moraña, 2016; Fernández-Fontelo, Moraña, Cabaña, Arratia & Puig, 2020) y continuas, una característica de los datos de COVID-19 y típicamente presente en el modelado de enfermedades infecciosas. En este sentido, en Moraña, Fernández-Fontelo, Cabaña, Puig, Monfil, Brotons & Diaz (2021) se introduce un nuevo enfoque para datos longitudinales que no presentan correlaciones temporales y en Moraña, Fernández-Fontelo, Cabaña & Puig (2021) se presenta un modelo capaz de manejar estructuras temporales usando un enfoque diferente. Una limitación típica de este tipo de modelos es el esfuerzo computacional necesario para estimar correctamente los parámetros.

**La modelización matemática permite representar y predecir una situación epidémica, estimar situaciones futuras y evaluar actuaciones ya realizadas**

## Métodos

La estimación de la incidencia de COVID-19 que se utiliza para generar las predicciones descritas en este trabajo se basan en un modelo de series temporales autorregresivas con errores heterocedásticos (ARCH), cuyos detalles se pueden encontrar en Moraña, Fernández-Fontelo, Cabaña, Arratia & Puig (2023). Una vez estimada la incidencia de COVID-19, incluyendo casos no reportados, se estudian las predicciones generadas por diferentes métodos tanto a largo plazo (predicciones para todo un año) como a corto plazo (a una y dos semanas en el futuro). Los diferentes modelos predictivos considerados incluyen smoothing splines (Hastie & Tibshirani, 1990) para las predicciones a corto plazo, y una red neuronal (Hochreiter & Schmidhuber, 1997) para generar las predicciones a largo plazo.

## Resultados

Este apartado reporta los resultados más reseñables producidos por la metodología propuesta, en relación con la estimación de la incidencia real de COVID-19 en cada comunidad autónoma en el período 23 de febrero de 2020 al 27 de febrero de 2022 y a las predicciones generadas para la Comunidad de Madrid en todo el año 2022.

### *Estimación de la incidencia*

Los primeros casos de neumonía provocados por el betacoronavirus SARS-CoV-2 se detectaron en diciembre de 2019 en la ciudad de Wuhan (China) (Sohrabi et al., 2020), posteriormente diagnosticados como COVID-19. Teniendo en cuenta que muchos casos cursan sin desarrollar síntomas o solo con síntomas muy leves, es razonable suponer que la incidencia de esta enfermedad ha sido subregistrada. En Moraña et al., 2023 se analiza la incidencia semanal de COVID-19 registrada en España en el periodo del 23 de febrero de 2020 hasta el 27 de febrero del 2022, viendo que las fuentes oficiales reportaron 11.056.797 casos de COVID-19 en España, mientras que el modelo predice un

total de 21.639.627 casos (solo se reportaron el 51,10% de los casos reales). Este trabajo también reveló que, si bien la frecuencia de subregistro es extremadamente alta para todas las regiones, la intensidad de este subregistro no es uniforme entre las regiones consideradas: Aragón y Ceuta son las comunidades autónomas con mayor intensidad de subregistro mientras que Región de Murcia y Comunidad Valenciana son las regiones donde los valores estimados son los más cercanos al número de casos notificados.

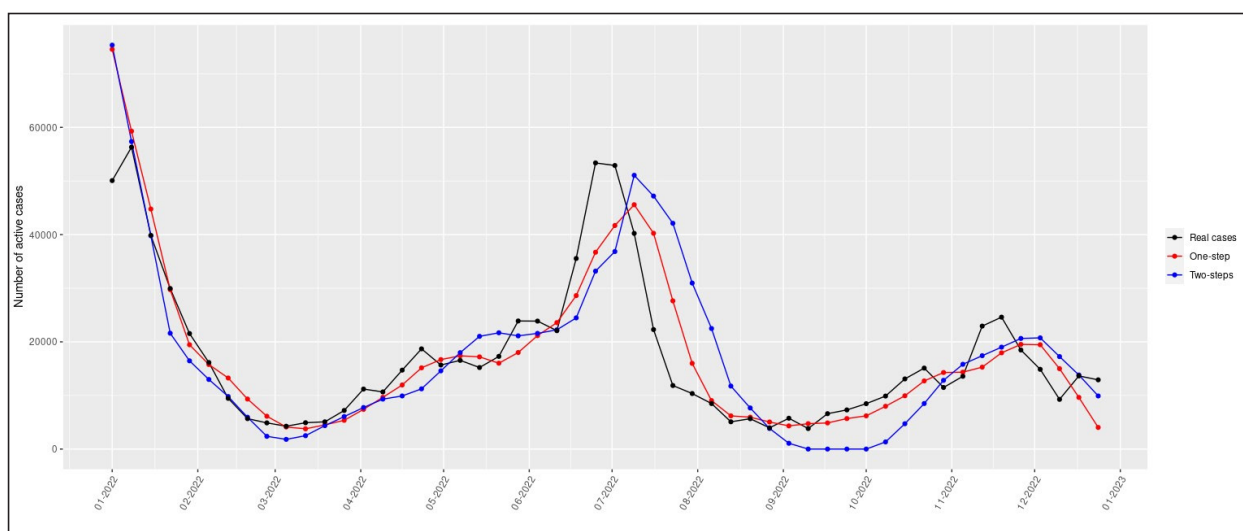
El modelo permite estudiar también el impacto de covariables que potencialmente estén relacionadas con la incidencia de la enfermedad que se está estudiando. En este caso se analiza el efecto del periodo de estado de alarma declarado entre el 15 de marzo y el 21 de junio de 2020 y el efecto de la vacunación, considerando como periodo post-vacunación el momento en el cual el 50% de la población española había recibido al menos una dosis de alguna vacuna contra el COVID-19, a partir de mayo de 2021. Aunque el principal impacto de los programas de vacunación se aprecia en los datos de mortalidad, puede apreciarse también un descenso significativo en el número de casos semanales en todas las comunidades autónomas excepto en Euskadi (Moriña et al., 2023).

### Generación de predicciones a corto plazo

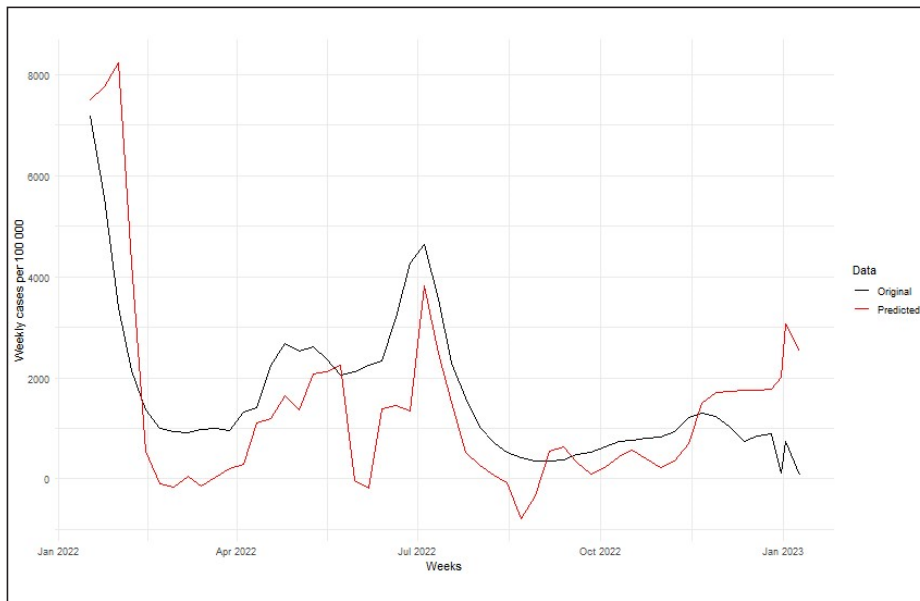
Una vez reconstruida la evolución más probable de la incidencia de COVID-19, se pueden obtener pronósticos a corto plazo a una y dos semanas en adelante. Estimando la incidencia semanal real en el período del 29 de octubre de 2021 al 31 de diciembre de 2021, se han producido pronósticos para cada semana de 2022, utilizando el último valor de incidencia conocido de 2021, a una y dos semanas en el futuro, basadas en la técnica de smoothing splines. Los valores pronosticados para las 52 semanas de 2022 en la Comunidad de Madrid se muestran en la Figura 1.

### Generación de predicciones a largo plazo

La Figura 2 muestra las predicciones a largo plazo para todo un año (53 semanas) en base a una red neuronal entrenada con los datos del período del 1 de enero de 2020 al 31 de diciembre de 2021, comparadas con los casos observados en 2022 y las primeras semanas de 2023.



**Figura 1.** Predicción del número de casos semanales activos según el método de *smoothing splines* de COVID-19 en la Comunidad de Madrid en 2022.



**Figura 2.** Predicción de la incidencia de COVID-19 por 100,000 individuos según el método de red neuronal de COVID-19 en España en 2022 e inicio de 2023.

## Conclusiones

Aunque es muy común en la investigación biomédica y epidemiológica utilizar datos de los registros oficiales, recientemente se ha hecho evidente una preocupación acerca de su fiabilidad, y se han realizado algunos esfuerzos para estandarizar los protocolos a fin de mejorar la precisión de los registros de información de salud (ver por ejemplo Harkener et al., 2019; Kodra et al., 2018). Sin embargo, como ha puesto en evidencia la pandemia de COVID-19, no siempre es posible implementar estas recomendaciones de manera adecuada.

Otro trabajo que analizó la carga acumulada de COVID-19 en España estimó que solo alrededor del 21% de los casos se notificaron en el período del 1 de enero de 2020 al 1 de junio de 2020 (Morña, Fernández-Fontelo, Cabaña, Arratia, Ávalos & Puig, 2021). Pero debemos tener en cuenta el fenómeno de la subnotificación se dió con mayor intensidad en las primeras etapas de la pandemia y, por lo tanto, se espera un menor subregistro general en el período más largo considerado en Morña et al. (2023). Adicionalmente, la metodología presentada permite un seguimiento en tiempo real y no solo el análisis de la incidencia acumulada en un periodo de tiempo. Disponer de datos fiables es clave para abordar cualquier proceso de toma de decisiones en salud pública de forma óptima, y para mejorar la precisión de los modelos dinámicos destinados a estimar la propagación de la enfermedad (Zhao et al., 2020) y predecir su comportamiento.

La metodología propuesta por Morña et al. (2023) puede tratar los datos mal informados de una manera muy natural y directa. Además, es capaz de reconstruir el proceso oculto más probable, proporcionando a los tomadores de decisiones de salud pública una herramienta valiosa para predecir la evolución de la enfermedad de interés bajo diferentes escenarios.

El análisis de los datos españoles de COVID-19 muestra que, en promedio, solo se notificó alrededor del 51% de los casos en el período del 23 de febrero de 2020 al 27 de febrero de 2022, y que existen diferencias importantes en la severidad del subregistro entre las distintas comunidades autónomas españolas. También es posible evaluar el impacto del programa de vacunación sobre la dinámica de la enfermedad, logrando una disminución significativa de la incidencia de COVID-19 en casi todas las comunidades autónomas después de que el 50% de la población tuviera al menos una dosis de la vacuna (aunque estos resultados probablemente serían notablemente diferentes si en el período de estudio se incluyeran variantes como la BA.4 o BA.5, con un escape de inmunidad mayor y que en el momento de escribir este trabajo son predominantes en muchos países), mientras que el modelo solo pudo detectar el impacto del confinamiento obligatorio en siete de las 19 comunidades autónomas analizadas.

Los resultados de este trabajo muestran también la dificultad de obtener predicciones a largo plazo para series temporales no estacionarias y con un comportamiento altamente volátil como la considerada, así como la importancia

**El análisis de los datos españoles de COVID-19 muestra que, en promedio, solo se notificó alrededor del 51% de los casos en el período del 23 de febrero de 2020 al 27 de febrero de 2022**



de combinar información proveniente de diversas fuentes para obtener predicciones a largo plazo fiables. La baja capacidad predictiva observada también hace altamente recomendable el uso de combinaciones de modelos predictivos o ensemble, cuya popularidad ha aumentado de forma notable recientemente, como consecuencia de la mejora de la capacidad computacional (Sherratt et al., 2023).

#### Contribuciones de los autores

Los autores participaron igualmente en la elaboración del manuscrito y aprobaron la versión final presentada.

#### Financiación

Esta investigación no recibió financiación.

#### Declaración de disposición de datos

Los datos presentados en este estudio pueden ser solicitados al autor de correspondencia.

#### Agradecimientos

Agradecemos al proyecto PredCov (Multi-source and multi-method prediction to support COVID-19 policy decision making) la oportunidad de presentar los resultados de nuestra investigación en este número especial.

#### Conflicto de interés

Los autores declaran que no hay conflicto de interés.

## Referencias bibliográficas

- Alfonso, J. H., Løvseth, E. K., Samant, Y. and Holm, J. (2015). Work-related skin diseases in Norway may be underreported: Data from 2000 to 2013. *Contact Dermatitis*, 72(6), 409-412. <https://doi.org/10.1111/cod.12355>
- Arendt, S., Rajagopal, L., Strohhahn, C., Stokes, N., Meyer, J., & Mandernach, S. (2013). Reporting of foodborne illness by U.S. consumers and healthcare professionals. *International journal of environmental research and public health*, 10(8), 3684-3714. <https://doi.org/10.3390/ijerph10083684>
- Azmon, A., Faes, C., & Hens, N. (2014). On the estimation of the reproduction number based on misreported epidemic data. *Statistics in medicine*, 33(7), 1176-1192. <https://doi.org/10.1002/sim.6015>
- Bernard, H., Werber, D., & Höhle, M. (2014). Estimating the under-reporting of norovirus illness in Germany utilizing enhanced awareness of diarrhoea during a large outbreak of Shiga toxin-producing E. coli O104: H4 in 2011—A time series analysis. *BMC Infectious Diseases*, 14(1), 116-116. <https://doi.org/10.1186/1471-2334-14-116>
- Fernández-Fontelo, A., Moriña, D., Cabaña, A., Arratia, A. and Puig, P. (2020). Estimating the real burden of disease under a pandemic situation: The SARS-CoV2 case. *PLoS ONE*, 15, e0242956. <https://doi.org/10.1371/journal.pone.0242956>
- Fernández-Fontelo, A., Cabaña, A., Joe, H., Puig, P. and Moriña, D. (2019). Untangling serially dependent underreported count data for gender-based violence. *Statistics in Medicine*, 38(22), 4404-4422. <https://doi.org/10.1002/sim.8306>
- Fernández-Fontelo, A., Cabaña, A., Puig, P. and Moriña, D. (2016). Under-reported data analysis with INAR-hidden Markov chains. *Statistics in Medicine*, 35(26), 4875-4890. <https://doi.org/10.1002/sim.7026>
- Gibbons, C. L., Mangen, M.-J. J., Plass, D., Havelaar, A. H., Brooke, R. J., Kramarz, P., Peterson, K. L., Stuurman, A. L., Cassini, A., Fèvre, E. M., Kretzschmar, M. E. E., & Burden of Communicable diseases in Europe (BCoDE) consortium. (2014). Measuring underreporting and under-ascertainment in infectious disease datasets: A comparison of methods. *BMC public health*, 14(1), 147. <https://doi.org/10.1186/1471-2458-14-147>
- Harkener, S., Stausberg, J., Hagel, C., & Siddiqui, R. (2019). Towards a Core Set of Indicators for Data Quality of Registries. *Studies in health technology and informatics*, 267, 39-45. <https://doi.org/10.3233/SHTI190803>
- Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized Additive Models*. CRC Press.
- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>

- Kodra, Y., Weinbach, J., Posada-De-La-Paz, M., Coi, A., Lemonnier, S. L., van Enckevort, D., Roos, M., Jacobsen, A., Cornet, R., Ahmed, S. F., Bros-Facer, V., Popa, V., van Meel, M., Renault, D., von Gizycki, R., Santoro, M., Landais, P., Torreri, P., Carta, C., ... Taruscio, D. (2018). Recommendations for improving the quality of rare disease registries. *International Journal of Environmental Research and Public Health*, 15(8). <https://doi.org/10.3390/ijerph15081644>
- Magal, P., & Webb, G. (2018). The parameter identification problem for SIR epidemic models: Identifying unreported cases. *Journal of Mathematical Biology*, 77(6-7), 1629-1648. <https://doi.org/10.1007/s00285-017-1203-9>
- Moriña, D., Fernández-Fontelo, A., Cabaña, A., Arratia, A., & Puig, P. (2023). Estimated Covid-19 burden in Spain: ARCH underreported non-stationary time series. *BMC Medical Research Methodology*, 23, 75. <https://doi.org/10.1186/s12874-023-01894-9>
- Moriña, D., Fernández-Fontelo, A., Cabaña, A., Arratia, A., Ávalos, G., & Puig, P. (2021). Cumulated burden of Covid-19 in Spain from a Bayesian perspective. *European Journal of Public Health*, 31(4), 917-920. <https://doi.org/10.1093/eurpub/ckab118>
- Moriña, D., Fernández-Fontelo, A., Cabaña, A., & Puig, P. (2021). New statistical model for misreported data with application to current public health challenges. *Scientific Reports*, 11(1), 23321. <https://doi.org/10.1038/s41598-021-02620-5>
- Moriña, D., Fernández-Fontelo, A., Cabaña, A., Puig, P., Monfil, L., Brotons, M., & Diaz, M. (2021). Quantifying the under-reporting of uncorrelated longitudinal data: The genital warts example. *BMC Medical Research Methodology*, 21(1), 6-6. <https://doi.org/10.1186/s12874-020-01188-4>
- Rosenman, K. D., Kalush, A., Reilly, M. J., Gardiner, J. C., Reeves, M., & Luo, Z. (2006). How much work-related injury and illness is missed by the current national surveillance system? *Journal of occupational and environmental medicine / American College of Occupational and Environmental Medicine*, 48(4), 357-365. <https://doi.org/10.1097/01.jom.0000205864.81970.63>
- Sherratt, K., Gruson, H., Grah, R., Johnson, H., Niehus, R., Prasse, B., Sandmann, F., Deuschel, J., Wolfram, D., Abbott, S., Ullrich, A., Gibson, G., Ray, E. L., Reich, N. G., Sheldon, D., Wang, Y., Wattanachit, N., Wang, L., Trnka, J., ... Funk, S. (2023). Predictive performance of multi-model ensemble forecasts of Covid-19 across European nations. *eLife*, 12, e81916. <https://doi.org/10.7554/eLife.81916>
- Sohrabi, C., Alsafi, Z., O'Neill, N., Khan, M., Kerwan, A., Al-Jabir, A., Iosifidis, C., & Agha, R. (2020). World Health Organization declares Global Emergency: A review of the 2019 Novel Coronavirus (Covid-19). *International journal of surgery (London, England)*, 76, 71-76. <https://doi.org/10.1016/j.ijsu.2020.02.034>
- Stocks, T., Britton, T., & Höhle, M. (2018). Model selection and parameter estimation for dynamic epidemic models via iterated filtering: Application to rotavirus in Germany. *Biostatistics*. <https://doi.org/10.1093/biostatistics/kxy057>
- Stoner, O., Economou, T. and Drummond Marques da Silva, G. (2019). A Hierarchical Framework for Correcting Under-Reporting in Count Data. *Journal of the American Statistical Association*, 114(528), 1481-1492. <https://doi.org/10.1080/01621459.2019.1573732>
- Taylor, S. J., & Letham, B. (2018). Forecasting at Scale. *The American Statistician*, 72(1), 37-45. <https://doi.org/10.1080/00031305.2017.1380080>
- Winkelmann, R. (1996). Markov chain Monte Carlo analysis of underreported count data with an application to worker absenteeism. *Empirical Economics*, 21(4), 575-587. <https://doi.org/10.1007/BF01180702>
- Zhao, S., Musa, S. S., Lin, Q., Ran, J., Yang, G., Wang, W., Lou, Y., Yang, L., Gao, D., He, D., & Wang, M. H. (2020). Estimating the Unreported Number of Novel Coronavirus (2019-nCoV) Cases in China in the First Half of January 2020: A Data-Driven Modelling Analysis of the Early Outbreak. *Journal of Clinical Medicine*, 9(2), 388. <https://doi.org/10.3390/jcm9020388>

## El papel de los modelos matemáticos en la Estrategia de Vacunación frente a COVID-19 en España

### The role of mathematical models in the Vaccination Strategy against COVID-19 in Spain

Carmen Olmedo Lucerón<sup>a</sup>, Laura Sánchez-Cambronero Cejudo<sup>a</sup>, Sonia Fernández Conde<sup>a</sup>, Ana Fernández Dueñas<sup>a</sup>, Elena Cantero Gudino<sup>b</sup>, Aurora Limia Sánchez<sup>a</sup>

<sup>a</sup> Área de Vacunas, Ministerio de Sanidad, España

<sup>b</sup> Área de Vacunas, Ministerio de Sanidad, Asistencia técnica Tragsatec, España

#### Resumen

La utilización de modelos matemáticos de predicción en las enfermedades transmisibles prevenibles por vacunación ayuda a evaluar el impacto de un programa de vacunación ya implantado o futuro. Hasta el momento ha habido escasa cultura de modelización en este ámbito, cuyos resultados sirvieran para realizar ajustes en la planificación y el seguimiento de estos programas. La pandemia de COVID-19 ha supuesto un impulso al desarrollo de modelos matemáticos de vacunación. Gracias a la colaboración entre modeladores y Salud Pública se ha contribuido al ajuste de la Estrategia de Vacunación COVID-19 en España. Es necesario consolidar y seguir impulsando esta capacidad investigadora, acercar los modelos matemáticos a los profesionales e incorporar de forma ágil el uso de estas herramientas a la toma de decisiones.

Palabras clave: Programas de vacunación; modelos matemáticos; vacunación COVID-19; evaluación.

#### Abstract

Mathematical models for the prediction of communicable diseases preventable by vaccination helps to evaluate the impact of implemented or future vaccination programmes. So far, there has been little culture of modelling in this area, the results of which will serve to make adjustments in the planning and monitoring of these programs. The COVID-19 pandemic has provided a boost to the development of mathematical vaccination models, due to the collaboration between modelers and Public Health professionals contributing to the adjustment of the COVID-19 Vaccination Strategy in Spain. It is necessary to consolidate and continue promoting this research capacity, bring mathematical models closer to professionals and quickly incorporate the use of these tools into decision-making.

Key Words: Vaccination programs; mathematical models; COVID-19 vaccination; assessment.

## Antecedentes

### *La utilización de modelos matemáticos en enfermedades transmisibles y vacunación*

**P**odemos definir los modelos matemáticos como una herramienta que utiliza recursos matemáticos para simplificar la realidad. La utilización de herramientas matemáticas para el desarrollo de modelos de predicción nos permite: conocer mejor el comportamiento de una enfermedad y proyectar una situación futura; estimar el impacto de intervenciones sanitarias para prevenir y controlar las enfermedades transmisibles; o realizar estudios de coste-efectividad de las intervenciones, sirviendo en todo caso, para asesorar las políticas de salud pública. Además, estos modelos pueden incorporar la vacunación como principal elemento modificador de la misma, en el caso de las enfermedades inmunoprevenibles, por lo que sirven también para evaluar o estimar el impacto de un programa de vacunación ya implantado o futuro (Bjørnstad, Shea, Krzywinski, & Altman, 2020).

A nivel internacional, los modelos matemáticos han sido utilizados ya desde hace tiempo. Su uso se vio impulsado durante la pandemia de gripe de 2009 (Wu & Cowling, 2011), ya que fue la base para la preparación y respuesta a otras pandemias futuras. En este caso, los modelos ayudaron a conocer cuáles eran las medidas más eficaces de mitigación de una pandemia (valoraron la utilidad de intervenciones farmacológicas y no farmacológicas), contribuyeron a optimizar estrategias de tratamiento y vacunación y a evaluar la viabilidad de intervenciones logísticas, además de mostrar las debilidades del sistema de salud pública.

Con los conocimientos previos, los modeladores matemáticos construyeron diferentes modelos que simulaban la posible evolución de situaciones pandémicas futuras. Además, también se utilizaron en otras enfermedades más controladas, y han desempeñado una función muy importante en enfermedades en fase de erradicación, como puede ser el sarampión o la rubeola (Cutts et al, 2019).

Sin embargo, en España ha existido una escasa cultura de investigación que se centrara en desarrollar modelos matemáticos aplicados a enfermedades inmunoprevenibles y cuyos resultados sirvieran para realizar ajustes en la planificación y seguimiento de los programas de vacunación. Aunque se han utilizado tradicionalmente modelos para la realización de estudios económicos para evaluar la incorporación de nuevas vacunas en los calendarios de vacunación (Lorente, Varona, Antoñanzas & REjas, 2016), han existido pocos equipos de investigación que desarrollaran estas predicciones.

### *El antes y después de la COVID-19*

La pandemia de COVID-19 puso de manifiesto la necesidad de contar, no solo con información epidemiológica actualizada, sino también la importancia de que las autoridades sanitarias utilizaran diferentes simulaciones en la gestión de la crisis. Desde su inicio, empezaron a surgir numerosas iniciativas nacionales e internacionales que utilizaron las herramientas matemáticas para adelantarse a los posibles escenarios de evolución de la pandemia.

Desde que surgieron los primeros desarrollos de vacunas frente a la COVID-19, se observó que su utilización sería una herramienta fundamental de prevención. En España, se comenzó a elaborar la Estrategia de Vacunación frente a COVID-19 varios meses antes de disponer vacunas disponibles en el mercado (GTT Vacunación COVID, 2020). En esta Estrategia se establecieron diferentes prioridades de vacunación con el objetivo de proteger a la población más vulnerable, reducir la morbimortalidad, minimizar el impacto y facilitar la vuelta a la normalidad de la sociedad. Esta Estrategia fue elaborada por la Ponencia de Programas y Registro de Vacunaciones, que emite las recomendaciones de vacunación para todo el país, elevando sus propuestas para aprobación por la Comisión de Salud Pública y, después, por el Consejo Interterritorial del Sistema Nacional de Salud (Limia, Olmedo, Soler, Cantero, & Sánchez-Cambronero, 2020). Además, se contó con un Grupo de Trabajo Técnico multidisciplinar, con participación de expertos en investigación, vacunas, bioética, comunicación y sociología, entre otros campos de conocimiento. Este Grupo elaboró la primera propuesta y participó en las sucesivas actualizaciones de la misma

**Los modeladores matemáticos construyeron diferentes modelos que simulaban la posible evolución de situaciones pandémicas futuras**

(Grupo de Trabajo Técnico de Vacunación COVID-19 de la Ponencia de Programa y Registro de Vacunaciones, 2023). Además, la Estrategia también fue revisada por asociaciones de profesionales y de pacientes.

### *La cultura de creación de modelos matemáticos en vacunación*

#### La importancia de la colaboración entre Salud Pública y los modeladores

Desde la Secretaría Técnica de la Ponencia (Área de Vacunas del Ministerio de Sanidad), que coordinaba la elaboración de esta Estrategia, se valoró como un elemento fundamental el disponer de modelos matemáticos que incorporaran a las vacunas frente a COVID-19. El objetivo fue decidir la mejor estrategia poblacional y adelantar diferentes escenarios de vacunación según el comportamiento de la infección y las variables asociadas. Esta posibilidad se materializó al conocer que numerosos investigadores en España estaban desarrollando modelos de predicción de la evolución de la COVID-19.

El Ministerio de Sanidad contactó con el Instituto de Salud Carlos III, que había puesto en marcha el programa “FONDO – COVID19 para la ejecución de proyectos de investigación SARS-COV2 y la enfermedad COVID19” en el marco del Real Decreto-ley 8/2020, de 17 de marzo, para conocer los proyectos cuyo objetivo fuera el desarrollo de modelos matemáticos de predicción de la evolución de la pandemia. Tras este contacto, se identificaron tres proyectos que podían incluir la vacunación a los modelos matemáticos que ya estaban desarrollando y que podían ayudar a orientar la Estrategia de Vacunación. En noviembre de 2020, antes de que comenzara la vacunación, se creó un Grupo de Trabajo multidisciplinar, con participación de personas expertas en vacunas, en epidemiología, en investigación en vacunas y los propios equipos de modeladores (físicos y matemáticos de tres universidades diferentes), con el objetivo de modelizar diversos escenarios de vacunación que posibilitasen la toma de decisiones. Estos modelos podían predecir el impacto de distintas estrategias de vacunación en el número de infecciones, hospitalizaciones y fallecimientos (Trawicki, 2017).

La metodología que utilizaban los modeladores de este Grupo de Trabajo era muy diferente: el modelo desarrollado por la Universidad Carlos III de Madrid estaba basado en agentes que reproducen características individuales, así como sus interacciones sociales utilizando el simulador EpiGraph (Singh et al, 2021); un segundo modelo desarrollado por la Universidad de Barcelona (Fernández Fontelo, Moriña, Cabaña, Arratia & Puig, 2020), utilizaba series temporales discretas y cadenas de Markov ocultas estimando los casos reales en cada momento y corrigiendo la infranotificación de personas que tenían enfermedad leve; mientras que el tercer modelo desarrollado por la UNED, en colaboración con la empresa Inverence, estaba basado en ecuaciones en diferencias finitas que le permitía de forma dinámica proyectar la evolución de la pandemia.

#### Las fuentes de información de Salud Pública

La colaboración de Salud Pública con los desarrolladores de modelos fue básica para poder obtener el mejor resultado posible. Pero los modelos matemáticos no están exentos de limitaciones, pudiendo dar como resultado estimaciones poco precisas si se basan en datos escasos o poco fiables. Igualmente, también dependen de las variables que se incluyen, ya que un elevado número de parámetros en el modelo general producirá una alta incertidumbre en las predicciones.

Por tanto, se consultaron fuentes rigurosas y la mejor evidencia disponible para conseguir predicciones con la menor incertidumbre posible. Desde el Ministerio de Sanidad, se facilitó la misma información para nutrir los tres modelos, que se iban actualizando periódicamente con los nuevos datos de la pandemia. Las fuentes de información utilizadas fueron las siguientes:

- Características epidemiológicas de la infección (incidencia, gravedad, transmisibilidad, aparición de nuevas variantes): Centro de Coordinación de Alertas y Emergencias Sanitarias (CCAES) del Ministerio de Sanidad. A partir del Sistema de investigación de Vigilancia Epidemiológica (SiViEs) de la Red Nacional de Vigilancia Epidemiológica (RENAVE) y las actualizaciones de la evidencia publicadas (Centro de Coordinación de Alertas y Emergencias Sanitarias, 2023).
- Eficacia de las vacunas: A partir de las fichas técnicas y los informes públicos de evaluación de las vacunas, publicados por la Agencia Europea del Medicamento (EMA).
- Efectividad: Se ha obtenido información de la efectividad de las diferentes vacunas y en diferentes grupos de población y momentos de la pandemia, además de la pérdida de protección con el tiempo. Estos datos se obtuvieron

a partir de los resultados de los estudios coordinados desde el Área de Vacunas del Ministerio de Sanidad y realizados por el Centro Nacional de Epidemiología (CNE) y la Agencia Española de Medicamentos y Productos Sanitarios (AEMPS) a partir de diferentes bases de datos (Ministerio de Sanidad, 2023).

- Seroprevalencia: La estimación de las personas que habían pasado ya la enfermedad se obtuvo del Estudio Nacional de Sero-epidemiología de la Infección por SARS-CoV-2 en España (ENE-COVID) (Ministerio de Sanidad, 2021).
- Coberturas de vacunación: Se obtuvo a partir del primer sistema de información para el seguimiento de la vacunación frente a la COVID-19 (REGVACU), que supuso la confluencia homogénea y normalizada de todos los datos de vacunación COVID-19 en España (Ministerio de Sanidad, 2022).
- Implantación de la Estrategia de vacunación (priorización de grupos de población, personas a vacunar en cada grupo, esquema de vacunación y escenarios epidemiológicos): A partir de la información del Área de Vacunas, de las actualizaciones de la Estrategia y los diferentes documentos propuestos desde la Ponencia de Vacunas y publicados por el Ministerio de Sanidad.
- Disponibilidad de Vacunas y velocidad de vacunación: Información obtenida del Área de Vacunas del Ministerio de Sanidad a partir de REGVACU.

### Modelos matemáticos de vacunación y el contexto de Europa

Además, se mantuvo una colaboración activa a todos los niveles dentro del marco de la Unión Europea. Se utilizó la información disponible desde el Centro Europeo de Control de la Enfermedad (ECDC). El ECDC creó grupos de trabajo internacionales y colaborativos de modelos matemáticos desde el inicio de la pandemia, combinando y comparando proyecciones de diferentes equipos de modelado en la Unión Europea, aumentando así la solidez de las conclusiones.

Se impulsó la participación de los equipos de modeladores españoles en estas plataformas que han elaborado predicciones en diferentes escenarios a medio y largo plazo, y que siguen colaborando de forma activa para ajustar los modelos a la nueva situación y fases de la pandemia (European Covid-19 Forecast Hub, 2022). Estos equipos incluyen, entre otros, dos modelos basados en agentes desarrollados por la Universidad de Varsovia y la Wrocław University of Science and Technology; un modelo compartimental desarrollado por el ECDC; un modelo extendido desarrollado por el equipo SIMID (Simulation Models of Infectious Diseases) de las universidades Hasselt University y University of Antwerp; y OpenCOVID, un método estocástico de simulación a nivel de individuo desarrollado por el Swiss Tropical and Public Health Institute.

### *El impacto de las predicciones en la toma de decisiones y el ajuste de la Estrategia*

#### Cuestiones que han contribuido a orientar

Debido al interés sin precedentes en disponer de la mayor información posible durante el desarrollo de la Estrategia de vacunación, con la información de partida se modelaron diferentes escenarios para responder a algunas de las siguientes preguntas:

- ¿Es mejor vacunar por edad cumplida o a determinados grupos poblacionales?
- ¿Es mejor separar o no los intervalos recomendados entre dosis?
- ¿Qué impacto tiene usar una determinada vacuna en un determinado grupo?
- ¿Se pueden utilizar pautas mixtas?
- ¿Es necesaria la vacunación de los grupos de edad más jóvenes?
- ¿Es necesario vacunar a la población infantil o adolescente?
- ¿Debe la estrategia de vacunación modificarse con las nuevas variantes?
- ¿Se puede dejar de utilizar la mascarilla y cuándo?
- ¿Es necesario añadir dosis de recuerdo a toda o algún grupo de la población?
- ¿Qué periodicidad deben de tener las dosis de recuerdo?
- ¿Es necesaria una vacunación estacional?

Los modelos, además de simular escenarios que priorizaban a las personas más vulnerables, también contemplaban otras posibles opciones como vacunar a aquellos grupos de población que tenían más impacto en la transmisión de la pandemia, o estrategias que consiguieran el máximo beneficio social. Por ello, los modelos fueron esenciales para priorizar los grupos de población a vacunar ante la baja disponibilidad de dosis inicial, o poder simular posibles escenarios futuros en el momento actual, como es la vacunación estacional.

## Impacto en la toma de decisiones

Uno de los principales resultados fue la predicción sobre cuál era la mejor estrategia posible en los primeros meses de la vacunación. En este caso, las conclusiones de los tres modelos utilizados fueron similares, lo cual da idea de su solidez. Según estos modelos, vacunar a las personas más vulnerables, seguido de una vacunación escalonada en función del riesgo, era la estrategia más acertada que reduciría la mortalidad más del 60% (Limia et al, 2022). En todo caso, hay que tener en cuenta que no siempre existía evidencia para todos los aspectos que eran considerados, lo cual pudo añadir incertidumbre a los resultados obtenidos.

También se simuló el escenario de vacunación infantil, en el cual se obtuvo una predicción que indicaba que, si se lograban altas coberturas de vacunación, se podían reducir hasta un 15% de las infecciones y hasta un 6% de las hospitalizaciones y fallecimientos (Expósito, Olmedo, Limia, Guzmán-Merino, & Carretero, 2022).

Además, se han presentado los resultados en escenarios a largo plazo, hasta 10 años. Estos escenarios han sido desarrollados en el marco de la colaboración europea para orientar las estrategias futuras, teniendo en cuenta diferentes escenarios de evolución de la pandemia y la pérdida de la inmunidad. Estas predicciones han estimado una disminución en el número de infecciones entre un 7% y un 17%, dependiendo de la diferente estrategia y escenario posible.

Todos los resultados de los modelos se presentaron periódicamente y fueron valorados por la Ponencia de Programa y Registro de Vacunaciones y el Grupo de Trabajo Técnico de Vacunación frente a COVID-19 en España en cada una de las actualizaciones de la Estrategia. Sin embargo, no siempre fue posible incorporar los resultados a las actualizaciones. Esto se debió a varias circunstancias: por una parte, a que la toma de decisiones estaba rodeada de una enorme incertidumbre en una situación en permanente cambio, lo cual dificultó su gestión. Por otro lado, se ha mencionado que existe una limitada cultura de desarrollar modelos matemáticos aplicados a los programas de vacunación en España. En este punto, hay que insistir en la importancia, no solo de crear capacidad investigadora del país para desarrollar modelos de vacunación, sino también formar a los profesionales del campo de la Salud Pública para que puedan comprender e interpretar adecuadamente los resultados de estos modelos. Por último, también hay que añadir que es necesario disponer de canales para poder hacer una evaluación rápida de las predicciones del modelo utilizando datos de vigilancia y protocolos claros, para incorporar más ágilmente los resultados del modelo en la toma de decisiones.

**Los modelos matemáticos son herramientas potentes de predicción ante diferentes escenarios que pueden ayudar en el proceso de toma de decisiones en salud pública**

## Conclusiones

Los modelos matemáticos son herramientas potentes de predicción ante diferentes escenarios que pueden ayudar en el proceso de toma de decisiones en salud pública, adecuar las estrategias de vacunación, y ayudar en la gestión de crisis sanitarias. Además, nunca antes se habían aplicado los modelos matemáticos en nuestro país de forma tan extensa y continuada en un programa de vacunación como en el caso de COVID-19.

La flexibilidad y adaptación dinámica de los modelos matemáticos han permitido incorporar nueva información de interés en relación con las características epidemiológicas de la infección y los cambios en la Estrategia de vacunación, y la adecuación ágil y rápida de los parámetros a la nueva evidencia, en contexto de permanente cambio, pudiendo realizar evaluaciones del impacto en las sucesivas actualizaciones.

Por tanto, además de haber contribuido a la toma de decisiones sobre la Estrategia de Vacunación, el modelo colaborativo de trabajo ha servido para impulsar la investigación de modelización en vacunas. Se han creado grupos multidisciplinares en red que amplían el horizonte en esta materia y cuyos desarrollos serán de aplicación en programas de vacunación actuales o futuros, contribuyendo, en última instancia, a tener una población más protegida frente infecciones prevenibles mediante la vacunación.

Finalmente, es necesario continuar reforzando esta capacidad investigadora lograda hasta el momento y desarrollar iniciativas para acercar estas herramientas a los profesionales que deben aplicarlas, además de establecer canales para incorporar la evaluación rápida de las predicciones y la incorporación de los resultados en el desarrollo de políticas de salud pública relacionadas con la vacunación.

## Agradecimientos

Al Grupo de Trabajo de Modelos Matemáticos dentro del Grupo de Trabajo Técnico de Vacunación frente a COVID-19 en España (Ministerio de Sanidad); Coordinación: Aurora Limia Sánchez (Área de Vacunas. MS); Miembros: Carmen Olmedo Lucerón (Área de Vacunas. MS), Javier Díez Domingo (FISABIO), David Expósito Singh (Universidad Carlos III de Madrid), José Luis Aznarte Mellado (UNED), José Almagro Pedreño (Inverence), David Moraña Soler (Universidad de Barcelona), M<sup>a</sup> José Sierra Moros (CCAES. MS), Amparo Larrauri Cámara (Centro Nacional de Epidemiología).

A la Ponencia de Programa y Registro de Vacunaciones.

## Referencias bibliográficas

- Bjørnstad, O.N., Shea, K., Krzywinski, M., & Altman, N. (2020). The SEIRS model for infectious disease dynamics. *Nature Methods*, 17, 557–558.
- Centro de Control de Alertas y Emergencias Sanitarias. (2023). *Enfermedad por nuevo coronavirus, COVID-19. Información actualizada*. Disponible en: <https://www.sanidad.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov/home.htm>
- Cutts, F.T., Dansereau, E., Ferrari, M.J., Hanson, M., McCarthy, K.A., Metcalf, C.J.E., Takahashi, S., Tatem, A.J..... Winter, A.K. (2020). Using models to shape measles control and elimination strategies in low- and middle-income countries: A review of recent applications. *Vaccine*, 38(5), 979-992.
- European Covid-19 Forecast Hub. (2023). *Evaluation Report for EuroCOVIDhub-ensemble*. Disponible en: <https://covid19forecasthub.eu/reports.html>
- Expósito, D., Olmedo, C., Limia, A., Guzmán-Merino, M., & Carretero, J. (2022) Estimación del impacto de la vacunación frente a COVID-19 en la población infantil de 5-11 años. *Revista Española de Salud Pública*, 96, e202202021.
- Fernández-Fontelo, A., Moraña, D., Cabaña, A., Arratia, A., & Puig, P. (2020). Estimating the real burden of disease under a pandemic situation: The SARS-CoV2 case. *PLoS One*, 15(12), e0242956.
- Grupo de Trabajo Técnico de Vacunación COVID-19 de la Ponencia de Programa y Registro de Vacunaciones (GTT Vacunación COVID). (2023). *Estrategia de Vacunación frente a COVID-19 en España*. Consejo Interterritorial del SNS. Ministerio de Sanidad. Disponible en: [https://www.msbs.gob.es/profesionales/saludPublica/prevPromocion/vacunaciones/covid19/Actualizaciones\\_EstrategiaVacunacionCOVID-19.htm](https://www.msbs.gob.es/profesionales/saludPublica/prevPromocion/vacunaciones/covid19/Actualizaciones_EstrategiaVacunacionCOVID-19.htm)
- Limia, A., Olmedo, C., Soler, M., Cantero, E., & Sánchez-Cambronero, L. (2020). Ponencia de Programa y Registro de Vacunaciones y evolución del calendario de vacunación en España. *Revista Española de Salud Pública*, 94, e202003018.
- Limia, A., Olmedo, C., Díez, J., Expósito, D., Aznarte, J.L., Almagro, J, Moraña, D, Sierra MJ, & Larrauri A. (2022). Predicciones de tres modelos matemáticos en relación a la Estrategia de Vacunación frente a COVID-19 en España. Junio 2021. *Revista Española de Salud Pública*, 96, e202202019.
- Lorente, R., Varona, J.L., Antoñanzas, F., & Rejas, J. (2016) La vacunación anti-neumocócica con la vacuna conjugada 13-valente en población inmunocompetente de 65 años: análisis del impacto presupuestario en España aplicando un modelo de transmisión dinámica. *Revista Española de Salud Pública*, 90, e40001.
- Ministerio de Sanidad. (2020). *Datos de vacunación en España*. Disponible en: <https://www.sanidad.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov/situacionActual.htm>
- Ministerio de Sanidad. (2023). *Efectividad e impacto de las vacunas COVID*. Disponible en: [https://www.sanidad.gob.es/profesionales/saludPublica/prevPromocion/vacunaciones/covid19/Efectividad\\_vacunaCOVID-19.htm](https://www.sanidad.gob.es/profesionales/saludPublica/prevPromocion/vacunaciones/covid19/Efectividad_vacunaCOVID-19.htm)
- Ministerio de Sanidad. (2021). *Estudio Nacional de sero-Epidemiología de la Infección por SARS-CoV-2 en España (ENE-Covid)*. Disponible en: <https://www.sanidad.gob.es/ciudadanos/ene-covid/home.htm>



- Singh, D.E., Marinescu, M.C., Guzmán-Merino, M., Durán, C., Delgado-Sanz, C., Gómez-Barroso, D., Carretero, J. (2021). Simulation of COVID-19 Propagation Scenarios in the Madrid Metropolitan Area. *Frontiers in Public Health*, 9, e636023.
- Trawicki, M.B. (2017) Deterministic SEIRs epidemic model for modeling vital dynamics, vaccinations, and temporary immunity. *Mathematics*, 5, 7.
- Wu, J.T., Cowling, B.J. (2011). The use of mathematical models to inform influenza pandemic preparedness and response. *Experimental Biology and Medicine*, 236(8), 955-961.

## Investigación sobre comunicación en COVID-19 y principales hallazgos COVID-19 communication research and main findings

Daniel Catalan-Matamoros<sup>a</sup>, Andrea Langbecker<sup>a</sup>, Carlos Elías<sup>a</sup>

<sup>a</sup> UC3M MediaLab, Universidad Carlos III de Madrid, España

### Resumen

Se revisan las investigaciones realizadas por UC3M MediaLab durante el proyecto PredCov, sobre la comunicación durante la pandemia. Se llevó a cabo una encuesta nacional con 1,800 personas para examinar el comportamiento de consumo de información sobre la pandemia en España. Los resultados indicaron que ideología y edad son factores determinantes que influyen en la elección de distintas fuentes de información, así como en la percepción hacia las vacunas. Adicionalmente, la encuesta reveló que las principales fuentes de información consultadas durante la pandemia fueron los medios tradicionales y las autoridades sanitarias. En relación con la cobertura periodística, se observó un enfoque destacado en las vacunas contra la COVID-19, caracterizado por un tono mayormente positivo. Destacó la publicación de más reportajes en profundidad sobre vacunas en comparación con el periodo pre-pandemia. Asimismo, se examinó el discurso público en Twitter en relación con las vacunas, identificando un mayor número de contenidos asociados a una actitud positiva hacia la vacunación. Se extendió el análisis al discurso público en Twitter en español, francés, inglés y portugués, durante las crisis relacionadas con la vacuna de AstraZeneca y la variante Omicron. El proyecto cerró la brecha en investigación de la comunicación sobre vacunas en España.

Palabras Clave: Comunicación; vacunas; COVID-19; pandemia; investigación.

### Abstract

Research conducted by UC3M MediaLab during the PredCov project on communication during the pandemic is reviewed. A national survey was conducted with 1,800 people to examine the information consumption behavior about the pandemic in Spain. The results indicated that ideology and age are determining factors influencing the choice of different sources of information, as well as the perception towards vaccines. Additionally, the survey revealed that the main sources of information consulted during the pandemic were the traditional media and health authorities. With regard to media coverage, there was a strong focus on the COVID-19 vaccines, characterized by a mostly positive tone. More in-depth reports on vaccines were published compared to the pre-pandemic period. Also, public discourse on Twitter regarding vaccines was examined, identifying a greater number of contents associated with a positive attitude towards vaccination. The analysis was extended to public discourse on Twitter in Spanish, French, English and Portuguese, during crises related to the AstraZeneca vaccine and the Omicron variant. The project closed the gap in vaccine communication research in Spain.

Keywords: Communication; vaccines; COVID-19; pandemic; research.

## Introducción

La pandemia COVID-19 se ha convertido en la primera pandemia global en la era de las redes sociales, planteando numerosos retos, entre ellos, la difusión masiva de desinformación y noticias falsas a través de los medios digitales. Al comienzo de la pandemia, en febrero de 2020, la Organización Mundial de la Salud (OMS) declaró el estado de "infodemia", advirtiendo sobre la propagación de desinformación, la cual se estaba extendiendo paralelamente al virus (Hao & Basu, 2020). La OMS recomendó a los países abordar activamente este nuevo fenómeno que acompañaba a la pandemia.

Según un artículo publicado en MIT Technology Review el 12 de febrero de 2020, la COVID-19 pasará a la historia como la primera pandemia global impulsada por las redes sociales, generando problemas significativos en la comunicación. Uno de los más evidentes ha sido la propagación masiva de desinformación y noticias falsas a través de plataformas digitales como redes sociales, sitios web y servicios de mensajería instantánea

como WhatsApp (Elías & Catalan-Matamoros, 2020). A pesar de que teorías absurdas, como la transmisión de un virus biológico a través de redes de telefonía móvil o la supresión del sistema inmunitario por la radiación de estas redes, podrían ser motivo de risa para personas con educación básica y sentido común, durante la devastación global causada por el coronavirus, tales afirmaciones han logrado convencer a muchas personas. Esto ha llevado a acciones extremas, como la quema de cientos de torres de telefonía 5G en diversos países, desde Australia y Nueva Zelanda hasta el Reino Unido, Irlanda, Finlandia, Suecia, Bélgica, los Países Bajos e Italia (Cerulus, 2020; Lewis, 2020). Estos casos son solo ejemplos de la amplia difusión de información falsa durante la pandemia, la cual, lamentablemente, ha sido compartida por figuras políticas, celebridades e influencers, acelerando su propagación incluso más rápido que el propio virus (Catalan-Matamoros, 2020).

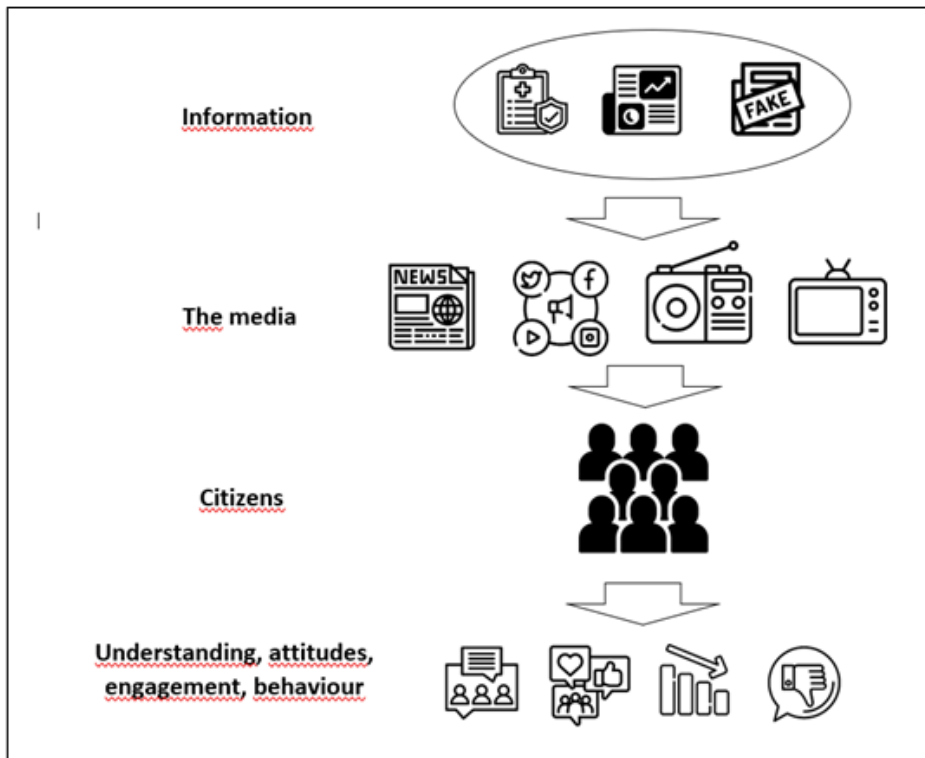
Este artículo revisa y presenta los resultados obtenidos en la línea 6 del Proyecto PredCov "Multi-source and multi-level prediction to support COVID-19 policy decision making", financiado por el programa REACT-EU bajo los fondos FEDER "a way of making Europe", a través del acuerdo entre la Comunidad de Madrid y la UC3M para la financiación de proyectos sobre SARS-CoV-2 y covid-19. Este proyecto consta de 8 líneas de investigación que implican a ingenieros, economistas, periodistas y científicos sociales de la UC3M. La línea de investigación número 6 se dedica al estudio de los medios de comunicación para analizar el discurso público, la percepción del riesgo y la identificación de noticias falsas en relación con la COVID-19. Esta línea de investigación se concentró en examinar el papel desempeñado por los medios de comunicación tradicionales (televisión, radio, prensa) y las plataformas de redes sociales (Twitter, YouTube, Facebook) en la configuración del discurso y el debate público acerca de la COVID-19, su vacunación y las medidas no farmacéuticas implementadas por las autoridades. Además, se exploró su influencia en la percepción pública del riesgo.

Para llevar a cabo esta investigación, el equipo de UC3M MediaLab analizó tanto las piezas informativas producidas por periodistas como el contenido generado por redes sociales y compartido por la población en relación con el COVID-19. Además, se llevaron a cabo estudios de diseño experimental para determinar la influencia de los medios de comunicación y las redes sociales en la población durante una crisis de salud pública. Para ello, se realizaron entrevistas en profundidad a personas con ideología antivacunas, así como una encuesta nacional en España para recopilar datos relevantes sobre la percepción de la población de todas las Comunidades Autónomas, incluyendo aspectos como el uso de los medios de comunicación para obtener información sobre COVID-19 y la vacunación, niveles de confianza en la información y las fuentes informativas, así como la percepción de noticias falsas y la percepción pública del riesgo.

El enfoque de esta línea de investigación se basa en el impacto que los medios producen sobre la población. De hecho, la influencia de los medios de comunicación y la difusión de información durante una pandemia incide en la percepción pública, con el objetivo de afrontar de manera más efectiva los desafíos asociados a la desinformación y a la percepción del riesgo en la sociedad. La Figura 1 representa el proceso mediante el cual la información de salud, compuesta por informes clínicos, datos de prevalencia y estudios científicos, se entremezcla con la desinformación antes de llegar a los ciudadanos a través de los medios de comunicación, internet y redes sociales. A partir de esta fusión de información, junto con otros factores como experiencias previas e ideología, los individuos elaboran

**El equipo de UC3M MediaLab analizó tanto las piezas informativas como el contenido generado por redes sociales en relación con el COVID-19**

su propia comprensión y perspectiva sobre la pandemia. Esta percepción influye directamente en la formación de actitudes, y, como consecuencia, en la participación y comportamiento relacionados, tales como las medidas preventivas para evitar el contagio, el uso de mascarillas, la disposición a vacunarse, entre otros aspectos.



**Figura 1.** Influencia de la información sobre salud en el comportamiento de la población. Fuente: Elaboración propia.

Para ello, esta línea de investigación se marcó tres objetivos principales:

- Analizar el discurso y el debate público sobre COVID-19 en las redes sociales, incluyendo la vacunación y el movimiento antivacuna.
- Contribuir con datos sobre medios de comunicación, redes sociales y desinformación relacionada con COVID-19 y vacunación.
- Evaluar la percepción pública del riesgo y la propagación de desinformación y noticias falsas sobre COVID-19 a través de encuestas poblacionales y estudios experimentales.

#### *Infodemia y Redes Sociales: Un Reto Actual en Salud Pública*

Esta pandemia será recordada como la primera de la era digital, caracterizada por el desafío que la desinformación ha planteado a los servicios de salud pública, quienes no solo se enfrentaban al virus, sino también a la difusión de información falsa sobre el mismo. A lo largo de la pandemia, en un esfuerzo constante por contrarrestar la desinformación en medios digitales, plataformas de redes sociales prominentes como Facebook, Twitter, YouTube y WhatsApp han intensificado sus filtros de fact-checking. A pesar de estos esfuerzos, un estudio indica que una proporción significativa de contenido falso continuó circulando en estas plataformas (Brennen et al., 2020). Aunque estas plataformas han colaborado estrechamente con la Organización Mundial de la Salud y autoridades sanitarias para garantizar información precisa sobre la COVID-19, la persistencia de la infodemia planteó dudas sobre la eficacia de estas medidas. Además, el problema de la infodemia se revela como un fenómeno arraigado a lo largo del tiempo. Las teorías conspirativas actuales sobre la COVID-19, como el supuesto ataque al 5G, tienen sus raíces en narrativas previas que han sido revitalizadas. A pesar del rechazo científico, estas ideas persisten y han sido empleadas en pandemias anteriores, como SARS, H1N1, MERS, Ébola y Zika.

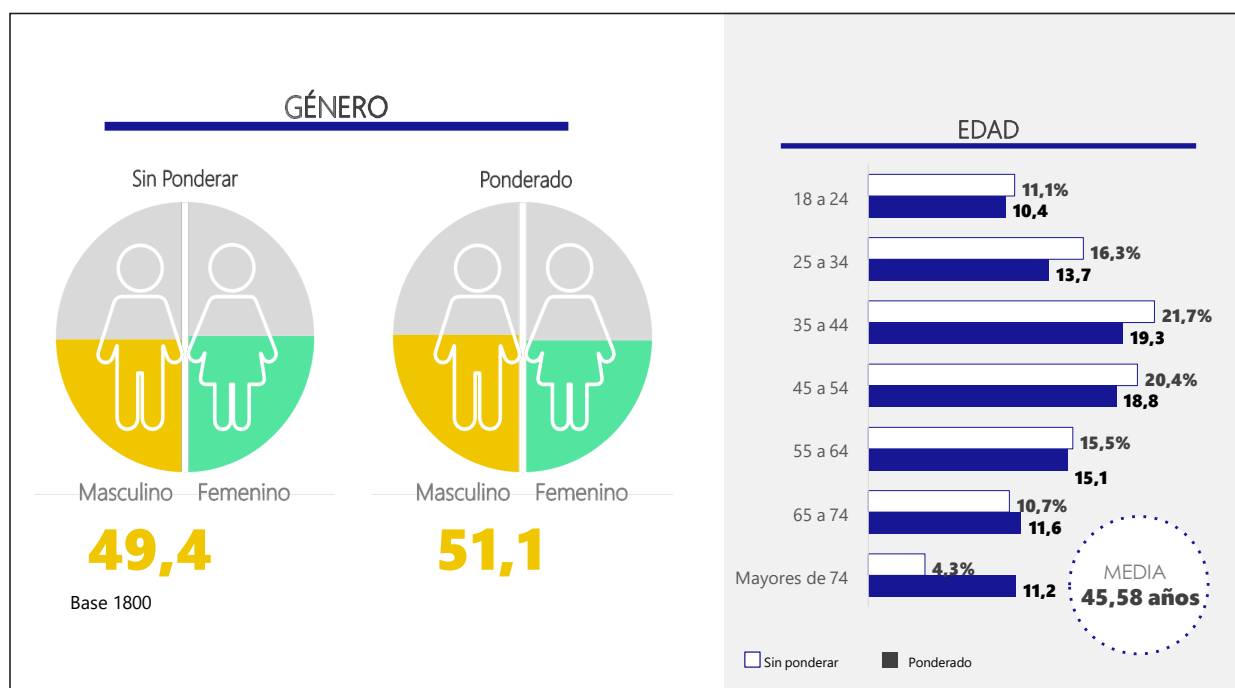
Este fenómeno no se limita a las redes sociales, ya que se origina en una compleja intersección de factores sociales, culturales y políticos. Desde valores personales hasta la baja alfabetización en salud y ciencia, la crisis

en las vocaciones STEM, el uso inadecuado de medios digitales y la baja inteligencia emocional. Estos elementos desempeñan un papel crucial en la propagación de la desinformación. La creencia previa de las personas y la resistencia a cambiar percepciones falsas presentan desafíos adicionales. Incluso los intentos explícitos de corregir falsas creencias con datos científicos pueden ser contraproducentes (Betsch et al., 2013).

Por lo tanto, abordar la infodemia va más allá de la verificación de datos y la corrección de información. Requiere un análisis profundo del contexto socio-cultural actual, marcado por el declive de expertos y el ascenso de políticos populistas sin perfil científico. La comprensión de la psicología social, incluyendo emociones, valores y creencias, es esencial para desarrollar estrategias de comunicación efectivas. En última instancia, combatir la infodemia del coronavirus implica no solo proporcionar datos científicos sólidos, sino también adaptar las estrategias de comunicación a las necesidades y la psicología de las audiencias. Este desafío multifacético exige una reflexión profunda sobre cómo las redes sociales amplifican la desinformación y cómo podemos redefinir el panorama informativo en un mundo impulsado por la conexión digital.

### *Consumo Informativo durante la Pandemia e Impacto de los Medios en la Opinión Pública*

Bajo el proyecto PredCov se elaboró una encuesta que fue diseñada por miembros del Grupo de Investigación Medialab de la UC3M y contó con el asesoramiento de investigadores del Ministerio de Sanidad y del Centro Nacional de Epidemiología. Esta encuesta estuvo financiada y forma parte del proyecto PredCov marcándose el objetivo de analizar el consumo de información sobre la COVID-19 en España y determinar la opinión pública sobre el papel de los periodistas, los medios de comunicación, la comunidad científica y las autoridades gubernamentales y de salud (Quian et al., 2023). La metodología incluyó la realización de una encuesta descriptiva a una muestra de 1,800 personas representativas de la población española, mayores de edad y residentes en las 17 comunidades autónomas en junio de 2022 (Figura 2).

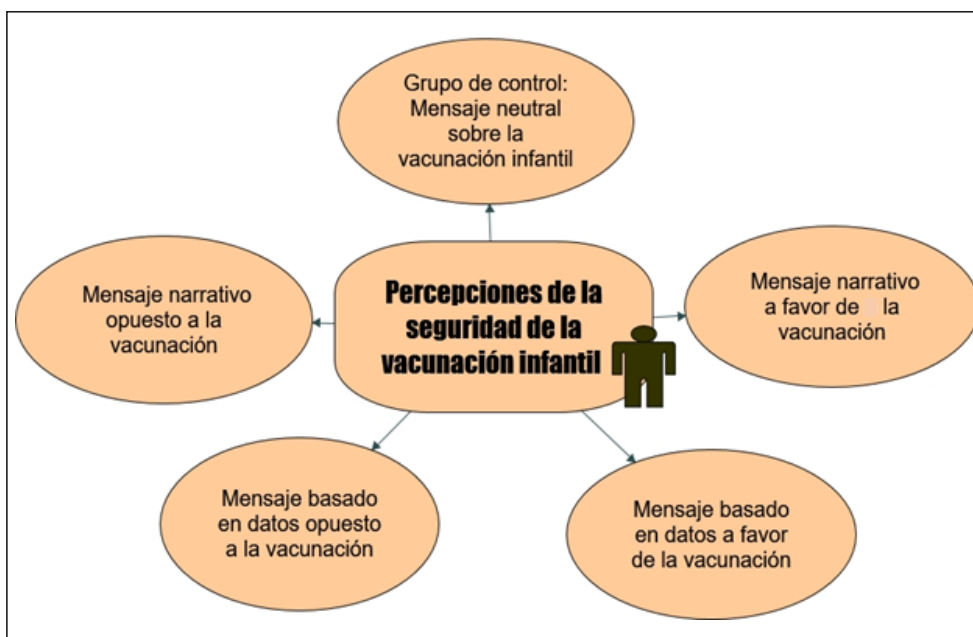


**Figura 2.** Caracterización de la muestra.

La edad, la inclinación política, la actitud hacia las vacunas y el nivel de educación fueron variables analizadas. Los resultados indicaron que la ideología y la edad son los factores que más condicionan el uso de diferentes tipos de fuentes de información. Las personas con ideología política de centro consumen más medios tradicionales que aquellos de izquierda o derecha, quienes son menos propensos a obtener información de medios tradicionales. Los de izquierda confían más en fuentes oficiales, como las autoridades sanitarias, en contraste con los centristas o de derecha. Los antivacunas prefieren fuentes alternativas. En relación al género, el uso de fuentes no difiere entre hombres y mujeres. Su comportamiento de consumo es similar, lo que indica que el género no es una variable que influya significativamente en el consumo de información, ni en la selección de fuentes ni en la percepción del papel

de la ciencia y el periodismo. En general, las principales fuentes de información consultadas durante la pandemia de COVID-19 fueron los medios tradicionales y las autoridades sanitarias (Catalan-Matamoros & Langbecker, 2023a). En el contexto de la pandemia, los jóvenes consumieron menos información y expresaron una mayor desconfianza en el periodismo y la ciencia.

Para conocer cómo influyen los medios de comunicación en la percepción y comportamiento de las personas, también desarrollamos un estudio de diseño experimental (Catalan-Matamoros, Prada & Langbecker, 2023). A través de la encuesta mencionada anteriormente se examinaron las fuentes de información en las que la población confía más en relación con los riesgos asociados a la vacuna de refuerzo para la Covid-19, así como para evaluar qué enfoques de mensajes son más efectivos para influir en la percepción de los riesgos de la vacuna infantil. Las variables dependientes fueron la percepción de los encuestados sobre (1) la eficacia de la vacuna de refuerzo Covid-19 y (2) la seguridad de la vacuna Covid-19 para los niños. Los participantes fueron expuestos de manera aleatoria a diversos mensajes relacionados con estas vacunas, utilizando diferentes fuentes de información y enfoques de mensajes. Además, algunos encuestados no recibieron ningún mensaje, formando así el grupo de control (Figura 3).



**Figura 3.** Experimento relacionado al enfoque de mensajes sobre la vacuna COVID-19 para niños.

Nuestros resultados indicaron que distintas fuentes y enfoques pueden afectar la percepción del riesgo asociado a las vacunas. La fuente "profesionales sanitarios" mostró un impacto positivo en la percepción del riesgo de la vacuna de refuerzo Covid-19 ( $p < 0,05$ ), y los mensajes a favor de las vacunas, ya sea en forma de narrativas (storytelling) o de exposiciones científicas, influyeron positivamente en la percepción de riesgo de la vacuna infantil por parte de los encuestados ( $p < 0,1$  y  $p < 0,05$ , respectivamente). Los hombres y los encuestados de mayor edad consideraron que las vacunas de refuerzo eran más eficaces que las mujeres y los más jóvenes. Aquellos con una ideología de derechas percibieron que las vacunas infantiles eran menos seguras. Estos hallazgos respaldan la necesidad de una comunicación estratégica en las campañas de vacunación llevadas a cabo por los servicios de salud, y avalan el impacto e influencia que los medios de comunicación pueden tener sobre la población durante la pandemia, en temas tan determinantes como la vacunación.

#### *Vacunas en el Centro del Debate: Analizando Actitudes, Información y Comportamientos*

La encuesta que se desarrolló bajo el proyecto PredCov, y que ha sido mencionada previamente, arrojó otros datos relevantes:

- El 92% de los encuestados se han vacunado contra la covid-19, mientras que únicamente el 64% se pondrían la dosis de refuerzo sin dudarlo, destacando los hombres mayores de 55 años.
- Las mujeres y jóvenes (18 - 34 años) se informaron sobre la covid-19 fundamentalmente a través de las redes sociales. Las personas con reticencia hacia las vacunas tienden a ser más activos en las redes sociales.

- Más de la mitad de los españoles consideran que la gestión de la pandemia es determinante para decidir su voto. Las personas de izquierdas se han vacunado más, y las de derechas tienden a tener mayor reticencia hacia la vacuna de la covid-19, considerando que la dosis de refuerzo es menos efectiva y que la vacuna es menos segura para los jóvenes.
- Los españoles muestran mayor credibilidad hacia los profesionales de la salud, y no tanto a fuentes gubernamentales. El público joven muestra una mayor credibilidad hacia los influencers.

Durante los últimos años, y especialmente desde que se desencadenó la pandemia de la covid-19, la actitud y la información recibida sobre las vacunas se ha convertido en un tema de debate habitual. Este fue precisamente el foco de la encuesta. Los medios de comunicación tradicionales (televisión y radio) fueron las fuentes de información más usadas por la población, aunque una parte de los encuestados duda de la calidad de la información recibida. Además, la encuesta incluyó dos experimentos para analizar cómo los mensajes sobre las vacunas influyen en la opinión pública en dos temas importantes: las dosis de refuerzo y la vacunación de niños (Catalan-Matamoros, Prada, & Langbecker, 2023). A mayor edad se encuentra menor reticencia hacia las vacunas, y que el nivel educativo puede influir en la comprensión y actitud hacia la vacunación, destacando el grupo de personas con estudios universitarios que muestran mayor reticencia hacia la vacunación. La encuesta también analizó el uso de la mascarilla como otra medida preventiva, encontrando que la población mayor de 55 años está más concienciada con su uso, y que Madrid es la comunidad autónoma donde la mascarilla más se usó en el transporte público.

### *Vacunas en la Prensa durante la Pandemia: Un Análisis de la Cobertura Periodística en España*

La vacunación es la medida preventiva más eficaz para combatir una pandemia. Las personas que tienen dudas acerca de la vacunación se caracterizan por no tener claridad en cuanto a si desean recibir la vacuna o si, por el contrario, se oponen completamente a ser vacunadas, a pesar de tener acceso a las vacunas (MacDonald, 2015). En consecuencia, los niveles de incertidumbre respecto a la vacunación están estrechamente vinculados con la aceptación o el rechazo de vacunarse (Larson et al., 2014). Las dudas en torno a la vacunación conforman un fenómeno complejo en el que intervienen diversos factores, como la confianza en la eficacia de las vacunas (Thorpe et al., 2022) y la percepción del riesgo, aunque esta percepción del riesgo no necesariamente refleje la realidad del riesgo, al ser subjetiva. De todos modos, esta percepción puede tener un impacto en las decisiones relacionadas con la vacunación (Malecki et al., 2021). En la percepción, la información recibida por los individuos es clave. Y es aquí, donde los medios de comunicación juegan un papel muy importante en el comportamiento y toma de decisiones de las personas durante la pandemia.

Las epidemias, al ser novedosas y actuales, atraen la atención mediática y forman parte de los "ciclos mediáticos epidémicos", adaptándose a los intereses periodísticos. Estas noticias relacionadas con enfermedades que afectan a diversos grupos sociales pueden generar temores. En crisis sanitarias, la comunicación desempeña un papel crucial en la formación de interpretaciones colectivas, aunque el tema sobre la vacunación suele carecer de interés para la prensa, según investigaciones previas en España y Portugal (Catalan-Matamoros & Peñafiel-Saiz, 2020; Langbecker & Catalan-Matamoros, 2021). En la actualidad, abundan estudios sobre vacunas que abordan aspectos como desinformación, intención de vacunación, efectos adversos, rapidez en el desarrollo y ensayos clínicos, contenido emocional y polarizado, movimientos antivacunas en redes sociales y desconfianza hacia la vacunación. Referencias como Salmon et al. (2015) y Larson et al. (2011) exploran la indecisión ante las vacunas y sus causas y consecuencias, mientras que Paul et al. (2021) y Freeman et al. (2023) examinan las actitudes hacia las vacunas COVID-19 y sus implicaciones para la comunicación.

A pesar de la relevancia del tema, pocos estudios analizan la cobertura periodística sobre vacunas en España, ya que la mayoría se enfoca en medios anglosajones. Los pocos estudios encontrados investigan la prensa española sobre la vacuna contra el virus del papiloma humano, la viruela y la vacunación en general. Nuestro estudio (Catalan-Matamoros & Langbecker, 2023b) buscó cerrar esta brecha al investigar las representaciones sociales sobre las vacunas mediante el análisis de la cobertura periodística en la prensa española durante diez años. Se comparó el período previo a la pandemia del coronavirus (2012-2019) con el período pandémico (2020-2021), marcados por contextos sanitarios, sociales y epidemiológicos distintos. Se examinó la cobertura de vacunas en los

**Pocos estudios analizan la cobertura periodística sobre vacunas en España, ya que la mayoría se enfoca en medios anglosajones**

diarios españoles El País y El Mundo, a través del método de análisis de contenido. En el periodo pre-pandémico, la cobertura de vacunas se centraba principalmente en aspectos científicos, destacando el desarrollo de nuevas vacunas e investigaciones biomédicas. No obstante, en el periodo pandémico, la atención se dirigió principalmente a las vacunas contra el COVID-19, haciendo hincapié en la prevención y campañas de vacunación. Durante la pandemia, los artículos periodísticos ofrecieron más información en términos de extensión y profundidad en los temas tratados. Además, tanto en el periodo pre-pandémico como en el pandémico, predominó un tono positivo hacia la vacunación y un enfoque centrado en el “interés humano”. Esto muestra cómo el periodismo se convirtió en una estrategia esencial para mantener a la población bien informada acerca de la situación de salud y las medidas preventivas.

Por otro lado, se llevó a cabo un estudio con el objetivo de evaluar cómo la pandemia había afectado el uso de fuentes entre los periodistas de medios de comunicación tradicionales (Langbecker & Catalan-Matamoros, 2023a). Dado que la crisis del coronavirus ha afectado no solo la salud, sino también diversos ámbitos, incluido el de la comunicación, este estudio se propuso examinar cómo la pandemia de COVID-19 ha influido en las fuentes de información utilizadas en la cobertura de vacunas en los periódicos españoles El País y El Mundo. Este estudio analiza las fuentes de información de contenidos sobre vacunas en los medios de prensa española El País y El Mundo, comparando las características de estas fuentes entre los periodos pre-pandémico (2012-2019) y pandémico (2020-2021). Fueron identificadas y categorizadas 2817 fuentes, concentrándose el 80% en el periodo pandémico. Las fuentes de carácter científico predominaron durante el periodo pre-pandémico, como las organizaciones científicas gubernamentales, las asociaciones profesionales y los científicos universitarios. Durante la pandemia, destacaron las fuentes no científicas, siendo las más frecuentes las de organizaciones gubernamentales, medios de comunicación y grupos de ciudadanos.

#### *Twitter y Vacunas: Un Análisis del Discurso Público en Tiempos de Pandemia*

Durante esta pandemia, muchas personas recurrieron a plataformas de redes sociales para obtener información o expresar sus opiniones sobre acontecimientos significativos. Numerosos estudios han analizado el discurso público acerca de las vacunas contra el COVID-19 en Twitter. Tomeny et al. (2017) investigaron las creencias antivacunas en tweets publicados desde 2009 hasta 2015 en Estados Unidos, encontrando que la mitad de los tweets contenían estas creencias. Durante la pandemia COVID-19, los estudios se enfocaron principalmente en análisis de sentimientos sobre las vacunas utilizando técnicas de aprendizaje automático. Alam et al. (2021) identificaron un sentimiento neutral en la mayoría de los tweets, pero este sentimiento no fue consistente a lo largo del periodo investigado. Además, Ali et al. (2021) también encontraron un sentimiento neutral en tweets relacionados con las vacunas contra el COVID-19 en Estados Unidos, demostrando que este sentimiento cambió según el estado del país. Hu et al. (2021) identificaron un aumento en tweets positivos y una disminución en tweets negativos en el mismo país. Ansari & Khan (2021) analizaron tweets publicados en varios países, encontrando un sentimiento negativo en la mayoría debido a la falta de confianza y al miedo a las vacunas. Bonnevie et al. (2021) sugieren que aquellos que se oponen a las vacunas contra el COVID-19 en esta red social fomentan la desconfianza hacia las autoridades sanitarias.

Además, solo unos pocos estudios han comparado específicamente diferentes periodos del debate público sobre este tema en Twitter, como los periodos pre-pandémico y pandémico. Por ejemplo, Engel-Rebitzer et al. (2021) identificaron que los legisladores estadounidenses participaron menos en la discusión pública sobre la vacunación antes de la llegada del SARS-CoV-2. Sin embargo, cambiaron su comportamiento con la pandemia, lo que podría influir en la aceptación de la vacuna contra el COVID-19 entre sus seguidores. Estos hallazgos sugieren matices en diferentes contextos epidemiológicos y sociales, que pueden influir en los discursos que circulan en las redes sociales y, por lo tanto, en el debate público sobre las vacunas. A su vez, estos posibles cambios discursivos también implican modificaciones en la renuencia de las personas a vacunarse, lo cual dependería, de manera importante, del contexto (Larson et al., 2014).

A través del proyecto PredCov, analizamos cómo impactó la pandemia en el discurso público sobre vacunas en Twitter (Langbecker, Catalan-Matamoros, & Prada, 2023). Este estudio examinó el discurso público sobre las vacunas en Twitter durante la pandemia de COVID-19 y lo comparó con los discursos sobre vacunas antes de la pandemia. Se llevaron a cabo codificaciones manuales de discurso, perfiles de usuarios, contenido, tono, vacunas



y fuentes de información. Se identificó una diversidad de discursos relacionados con las vacunas. El discurso que más destacó durante la pandemia estuvo vinculado a las medidas de gestión pandémica. Sin embargo, tanto los discursos negacionistas como los antinegacionistas fueron menos frecuentes en comparación con el periodo anterior. Además, se observó menos polarización y más discurso neutral durante la pandemia. Ambos periodos estuvieron marcados explícitamente por tweets con contenido relacionado con opiniones y experiencias personales sobre las vacunas. Algunos de ellos se caracterizaron, en general, por la informalidad en la comunicación de ideas por parte de los usuarios. De manera interesante, los usuarios del ámbito médico-científico no participaron más en Twitter durante la pandemia. Hubo menos publicaciones que incluían hallazgos científicos y más tweets sobre recursos para pacientes. Por otro lado, los medios de comunicación y los periodistas estuvieron muy activos en este periodo al difundir información y expresar opiniones sobre las vacunas, reflejando su papel esencial en crisis de salud. Sin embargo, la presencia de canales de televisión, como Fox News, podría indicar ideas conservadoras sobre las vacunas. De esta manera, los discursos expresaron significados diferentes en relación con el tema y disputaron su legitimidad en este mercado simbólico y virtual.

Otra investigación que se desarrolló bajo el proyecto PredCOV (Prada, Catalan-Matamoros, & Langbecker, 2023) consistió en describir el diálogo público y el lenguaje utilizado en torno a las vacunas durante los programas de vacunación contra el COVID-19. Para ello, llevamos a cabo un análisis de contenido manual, de carácter cualitativo, de una muestra de publicaciones en Twitter escritas en inglés que incluían la palabra vacuna y sus variantes. Clasificamos 7 variables relacionadas con el contenido de los mensajes y categorizamos el tipo de usuario que los publicó, así como el número de retweets. Posteriormente, se exploraron más detalladamente los patrones de asociación entre estas variables. Entre los hallazgos más relevantes, tenemos que entre los tweets con una perspectiva desfavorable hacia las vacunas, el 33% presentaba discursos negacionistas, el 29% discursos de protesta o desafío, el 13% abordaba las medidas de gestión de la pandemia y otro 13% ofrecía un enfoque científico. Los resultados de investigaciones, datos de vacunación e información práctica se vincularon más con una actitud positiva hacia las vacunas, mientras que las noticias se relacionaron con un tono neutral. Las cuentas de medios y periodistas fueron las más retuiteadas, seguidas por las cuentas gubernamentales y las organizaciones científicas asociadas al gobierno. En promedio, los tweets con mensajes preventivos obtuvieron más retweets. Los discursos más vinculados con información objetiva fueron los preventivos, institucionales y médico-científicos, así como aquellos relacionados con las diversas medidas para abordar la pandemia. Por el contrario, los tweets más subjetivos incluyeron discursos negacionistas, antinegacionistas y de protesta. A pesar de la existencia de tweets que expresan una oposición directa a las vacunas, una parte sustancial del contenido negativo adopta la forma de discursos de protesta y críticas hacia las acciones gubernamentales y las medidas para enfrentar la pandemia. De esta manera, los discursos desfavorables durante la pandemia involucraron casos significativos de duda hacia las vacunas, los cuales no solo se alimentaron de la desconfianza en la ciencia, sino que también estuvieron notablemente conectados con la insatisfacción hacia la gestión pública de la pandemia. Finalmente, en base a este estudio se puede determinar que: a) el contenido negativo sobre vacunas no solo presenta discursos negacionistas, b) el contenido positivo sobre vacunas no se centró únicamente en la prevención, c) las publicaciones de profesionales de la salud no recibieron retweets en Twitter, y d) las noticias de medios y mensajes preventivos obtuvieron más retweets.

Finalmente, elaboramos un análisis (Langbecker & Catalan-Matamoros, 2023b) para examinar el comportamiento de los usuarios de redes sociales en relación con la vacunación contra el COVID-19, utilizando la encuesta nacional que abarcó a 1800 individuos representativos de la población española. Se observó que los usuarios de WhatsApp, Facebook e Instagram mostraron una mayor tasa de vacunación, mientras que TikTok y YouTube albergaron el mayor porcentaje de individuos no vacunados. Respecto a la dosis de refuerzo, la mayoría de los usuarios expresó una actitud positiva, especialmente aquellos en la plataforma de Facebook, mientras que los usuarios de YouTube exhibieron la mayor resistencia hacia la vacunación. Además, los usuarios que pasaron más tiempo en las redes sociales se vacunaron menos con las dosis de recuerdo, además presentaron un porcentaje mayor de no vacunados. Este estudio también ofreció datos concretos para elaborar estrategias de comunicación como segmentación y escucha social que deben realizarse durante una crisis.

**Las cuentas de medios y periodistas fueron las más retuiteadas, seguidas por las cuentas gubernamentales y las organizaciones científicas asociadas al gobierno**

### *Las crisis de comunicación durante la pandemia*

Las crisis sociales a gran escala, como la pandemia de COVID-19, generan un aumento en el discurso público sobre las políticas y estrategias de respuesta más correctas en las democracias contemporáneas (Rauchfleisch et al., 2021). Durante la pandemia de COVID-19, las redes sociales han sido el escenario de diversos tipos de discursos. Los medios digitales magnifican estas situaciones al permitir que los ciudadanos expresen sus opiniones libremente y participen en el discurso público, mientras que los medios de comunicación, periodistas o autoridades pierden su posición única como guardianes e intermediarios de la información (Wallace, 2018). Los datos de Twitter, por ejemplo, abren nuevas vías para estudiar estas dinámicas. Los datos de trazas digitales de las redes sociales brindan oportunidades valiosas para analizar los patrones de comunicación durante las crisis.

A medida que la comunicación mediada ocurre cada vez más en plataformas digitales, los datos de esas plataformas se convierten en una herramienta útil para analizar objetivamente el discurso público (Wicke & Bolognesi, 2021). Sin embargo, la investigación integral sobre cómo las crisis influyen en el discurso público en diferentes comunidades aún es limitada (Rauchfleisch et al., 2021). Por lo tanto, en el proyecto PredCov investigamos el impacto del COVID-19 en el discurso público en el ámbito de Twitter en cuatro comunidades según el idioma (inglés, francés, español y portugués), centrándonos en dos crisis específicas que ocurrieron durante la pandemia: (a) la vacuna AstraZeneca contra el COVID-19 y (b) la variante Ómicron (Catalan-Matamoros, Prieto-Sanchez, & Langbecker, 2023). Nuestro objetivo fue analizar e identificar el discurso público en situaciones de crisis sanitaria en diferentes comunidades internacionales. Utilizando Tweepy y palabras clave relacionadas, recopilamos 3,748,302 publicaciones de las comunidades de Twitter en inglés, francés, portugués y español relacionadas con dos crisis específicas durante la pandemia: (a) la vacuna AstraZeneca contra el COVID-19 y (b) la variante Ómicron.

En relación con AstraZeneca, el principal enfoque del discurso público fue el "coágulo de sangre". Mediante clasificaciones cuantitativas y algoritmos de procesamiento del lenguaje natural, se obtienen resultados para cada idioma. El discurso en inglés y francés se centró más en la "muerte", siendo la comunidad francesa la que generó el sentimiento más negativo. El discurso en portugués fue el único que hizo referencia directa a un político, el expresidente brasileño Bolsonaro. En la crisis de Ómicron, el discurso público se centró principalmente en el seguimiento de casos de infección y el número de muertes, reflejando una mayor cercanía al riesgo real. El discurso público durante las crisis sanitarias puede influir en diversos comportamientos. Mientras que el discurso público sobre AstraZeneca podría actuar como una barrera para las medidas preventivas al aumentar la duda hacia la vacunación, el discurso sobre Ómicron podría fomentar comportamientos más preventivos, como el uso de mascarillas. De esta manera, se amplía la comprensión de la comunicación en crisis al revelar el papel de las redes sociales en la construcción del discurso público desde un enfoque global, al analizar diversas comunidades según el idioma.

### Financiación

Estudio financiado a través del convenio entre la Comunidad de Madrid (Ministerio de Educación, Universidades y Ciencia) y la Universidad Carlos III de Madrid, con la concesión de € 4.859.000 euros para el desarrollo de actividades de investigación sobre Covid-19 y Sars-CoV-2, con recursos React-EU del Fondo Europeo de Desarrollo Regional "Una forma de hacer Europa". Este estudio también ha recibido financiación de la Convocatoria 2022 de Proyectos de Generación de Conocimiento del Ministerio de Ciencia e Innovación (ref. PID2022-142755OB-I00) y del proyecto Comsalud "Pseudociencia, teorías conspirativas, fake news y alfabetismo mediático en la comunicación en salud".

## Referencias bibliográficas

- Alam, K. N., Khan, M. S., Dhruva, A. R., Khan, M. M., Al-Amri, J. F., Masud, M., & Rawashdeh, M. (2021). Deep Learning-Based Sentiment Analysis of COVID-19 Vaccination Responses from Twitter Data. *Computational and Mathematical Methods in Medicine*, 2021, 1-15. <https://doi.org/10.1155/2021/4321131>
- Ali, G. G. Md. N., Rahman, Md. M., Hossain, Md. A., Rahman, Md. S., Paul, K. C., Thill, J.-C., & Samuel, J. (2021). Public Perceptions of COVID-19 Vaccines: Policy Implications from US Spatiotemporal Sentiment Analytics. *Healthcare*, 9(9), 1110. <https://doi.org/10.3390/healthcare9091110>

- Ansari, M. T. J., & Khan, N. A. (2021). Worldwide COVID-19 Vaccines Sentiment Analysis Through Twitter Content. *Electronic Journal of General Medicine*, 18(6), em329. <https://doi.org/10.29333/ejgm/11316>
- Betsch, C., Renkewitz, F., & Haase, N. (2013). Effect of Narrative Reports about Vaccine Adverse Events and Bias-Awareness Disclaimers on Vaccine Decisions: A Simulation of an Online Patient Social Network. *Medical Decision Making*, 33(1), 14-25. <https://doi.org/10.1177/0272989X12452342>
- Bonnevie, E., Gallegos-Jeffrey, A., Goldbarg, J., Byrd, B., & Smyser, J. (2021). Quantifying the rise of vaccine opposition on Twitter during the COVID-19 pandemic. *Journal of Communication in Healthcare*, 14(1), 12-19. <https://doi.org/10.1080/17538068.2020.1858222>
- Brennen, J. S., Simon, F., Howard, P., & Nielsen, R. K. (2020, abril 7). *Types, sources and claims of Covid-19 misinformation*. Reuters Institute. <https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation>
- Catalan-Matamoros, D. (2020). La comunicación sobre la pandemia del COVID-19 en la era digital: Manipulación informativa, fake news y redes sociales. *Revista Española de Comunicación en Salud*, Sup. 1, 5. <https://doi.org/10.20318/recs.2020.5531>
- Catalan-Matamoros, D., & Langbecker, A. (2023a). How Does the Public Receive Information about Vaccines during the COVID-19 Pandemic? A Nationwide Cross-Sectional Study in Spain. *Societies*, 13(3), 62. <https://doi.org/10.3390/soc13030062>
- Catalan-Matamoros, D., & Langbecker, A. (2023b). Cómo ha cambiado la comunicación durante la pandemia covid-19: La prensa española ante las vacunas. *Estudios sobre el Mensaje Periodístico*, 29(3), 555-566. <https://doi.org/10.5209/esmp.84698>
- Catalan-Matamoros, D., Prada, E., & Langbecker, A. (2023). Fact or fiction: An experiment on how information sources and message framing influence vaccine risk perception. *Profesional de la información*, e320510. <https://doi.org/10.3145/epi.2023.sep.10>
- Catalan-Matamoros, D., Prieto-Sanchez, I., & Langbecker, A. (2023). Crisis Communication during COVID-19: English, French, Portuguese, and Spanish Discourse of AstraZeneca Vaccine and Omicron Variant on Social Media. *Vaccines*, 11(6), 1100. <https://doi.org/10.3390/vaccines11061100>
- Catalan-Matamoros, D., & Peñafiel-Saiz, C. (2020). The print media in times of anti-vaccine lobby: A content analysis of national newspaper reporting in Spain. *Western Journal of Communication*, in press.
- Cerulus, L. (2020, abril 29). *How anti-5G anger sparked a wave of arson attacks*. Politico. <https://www.politico.eu/article/coronavirus-5g-arson-attacks-online-theories/>
- Elías, C., & Catalan-Matamoros, D. (2020). Coronavirus in Spain: Fear of 'Official' Fake News Boosts WhatsApp and Alternative Sources. *Media and Communication*, 8(2), 462. <https://doi.org/10.17645/mac.v8i2.3217>
- Engel-Rebitzer, E., Stokes, D. C., Buttenheim, A., Purtle, J., & Meisel, Z. F. (2021). Changes in legislator vaccine-engagement on Twitter before and after the arrival of the COVID-19 pandemic. *Human Vaccines & Immunotherapeutics*, 17(9), 2868-2872. <https://doi.org/10.1080/21645515.2021.1911216>
- Freeman, D., Lambe, S., Yu, L.-M., Freeman, J., Chadwick, A., Vaccari, C., Waite, F., Rosebrock, L., Petit, A., Vanderslott, S., Lewandowsky, S., Larkin, M., Innocenti, S., McShane, H., Pollard, A. J., & Loe, B. S. (2023). Injection fears and COVID-19 vaccine hesitancy. *Psychological Medicine*, 53(4), 1185-1195. <https://doi.org/10.1017/S0033291721002609>
- Hao, K., & Basu, T. (2020, febrero 12). *The coronavirus is the first true social-media "infodemic"*. <https://www.technologyreview.com/2020/02/12/844851/the-coronavirus-is-the-first-true-social-media-infodemic/>
- Hu, T., Wang, S., Luo, W., Zhang, M., Huang, X., Yan, Y., Liu, R., Ly, K., Kacker, V., She, B., & Li, Z. (2021). Revealing Public Opinion Towards COVID-19 Vaccines With Twitter Data in the United States: Spatiotemporal Perspective. *Journal of Medical Internet Research*, 23(9), e30854. <https://doi.org/10.2196/30854>

- Langbecker, A., & Catalan-Matamoros, D. (2023a). Las fuentes de información sobre vacunas en El País y El Mundo: El efecto covid-19. *Palabra Clave*, 26(3), 1-30. <https://doi.org/10.5294/pacla.2023.26.3.9>
- Langbecker, A., & Catalan-Matamoros, D. (2023b). Social media and covid-19 vaccination: analysis of user behaviour in Spain. *Perspectivas de la comunicación* 16(2), 1-25. doi: 10.56754/0718-4867.2023.3306.
- Langbecker, A., Catalan-Matamoros, D., & Prada, E. (2023). How has COVID-19 impacted the public discourse around vaccines? A comparative analysis of Twitter. *Communication & Society*, Under Review.
- Langbecker, A., & Catalan-Matamoros, D. (2021). In the era of disbeliefs and uncertainties: News coverage of vaccines in Portuguese newspapers. *Saúde e Sociedade*, 30(3), e200929. <https://doi.org/10.1590/s0104-12902021200929>
- Larson, H. J., Cooper, L. Z., Eskola, J., Katz, S. L., & Ratzan, S. (2011). Addressing the vaccine confidence gap. *The Lancet*, 378(9790), 526-535. [https://doi.org/10.1016/S0140-6736\(11\)60678-8](https://doi.org/10.1016/S0140-6736(11)60678-8)
- Larson, H. J., Jarrett, C., Eckersberger, E., Smith, D. M. D., & Paterson, P. (2014). Understanding vaccine hesitancy around vaccines and vaccination from a global perspective: A systematic review of published literature, 2007–2012. *Vaccine*, 32(19), 2150-2159. <https://doi.org/10.1016/j.vaccine.2014.01.081>
- Lewis, K. (2020, marzo 26). *5G is not accelerating the spread of the new coronavirus*. <https://fullfact.org/health/5G-not-accelerating-coronavirus/>
- MacDonald, N. E. (2015). Vaccine hesitancy: Definition, scope and determinants. *Vaccine*, 33(34), 4161-4164. <https://doi.org/10.1016/j.vaccine.2015.04.036>
- Malecki, K. M. C., Keating, J. A., & Safdar, N. (2021). Crisis Communication and Public Perception of COVID-19 Risk in the Era of Social Media. *Clinical Infectious Diseases*, 72(4), 697-702. <https://doi.org/10.1093/cid/ciaa758>
- Paul, E., Steptoe, A., & Fancourt, D. (2021). Attitudes towards vaccines and intention to vaccinate against COVID-19: Implications for public health communications. *The Lancet Regional Health - Europe*, 1, 100012. <https://doi.org/10.1016/j.lanepe.2020.100012>
- Prada, E., Langbecker, A., & Catalan-Matamoros, D. (2023). Public discourse and debate about vaccines in the midst of the covid-19 pandemic: A qualitative content analysis of Twitter. *Vaccine*, 41(20), 3196-3203. <https://doi.org/10.1016/j.vaccine.2023.03.068>
- Quian, A., Elías, C., & Soengas-Pérez, X. (2023). Consumption of information and citizen's perception of the sources consulted during the Covid-19 pandemic: A study of the situation based on opinion polls. *Profesional de la información*, e320413. <https://doi.org/10.3145/epi.2023.jul.13>
- Rauchfleisch, A., Vogler, D., & Eisenegger, M. (2021). Public Sphere in Crisis Mode: How the COVID-19 Pandemic Influenced Public Discourse and User Behaviour in the Swiss Twitter-sphere. *Javnost - The Public*, 28(2), 129-148. <https://doi.org/10.1080/13183222.2021.1923622>
- Salmon, D. A., Dudley, M. Z., Glanz, J. M., & Omer, S. B. (2015). Vaccine Hesitancy. *American Journal of Preventive Medicine*, 49(6), S391-S398. <https://doi.org/10.1016/j.amepre.2015.06.009>
- Thorpe, A., Fagerlin, A., Butler, J., Stevens, V., Drews, F. A., Shoemaker, H., Riddoch, M. S., & Scherer, L. D. (2022). Communicating about COVID-19 vaccine development and safety. *PLOS ONE*, 17(8), e0272426. <https://doi.org/10.1371/journal.pone.0272426>
- Tomeny, T. S., Vargo, C. J., & El-Toukhy, S. (2017). Geographic and demographic correlates of autism-related anti-vaccine beliefs on Twitter, 2009-15. *Social Science & Medicine*, 191, 168-175. <https://doi.org/10.1016/j.socscimed.2017.08.041>
- Wallace, J. (2018). Modelling Contemporary Gatekeeping: The rise of individuals, algorithms and platforms in digital news dissemination. *Digital Journalism*, 6(3), 274-293. <https://doi.org/10.1080/21670811.2017.1343648>
- Wicke, P., & Bolognesi, M. M. (2021). Covid-19 Discourse on Twitter: How the Topics, Sentiments, Subjectivity, and Figurative Frames Changed Over Time. *Frontiers in Communication*, 6, 651997. <https://doi.org/10.3389/fcomm.2021.651997>