

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

Real COVID-19 incidence rate estimate in Spain

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

^a Departamento de Informática, Universidad Carlos III de Madrid, España

^b Centro Nacional de Epidemiología, Instituto de Salud Carlos III, España

^c CIBER en Epidemiología y Salud Pública (CIBERESP), España

^d CIBER en Enfermedades Infecciosas (CIBERINFEC), España

^e 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.

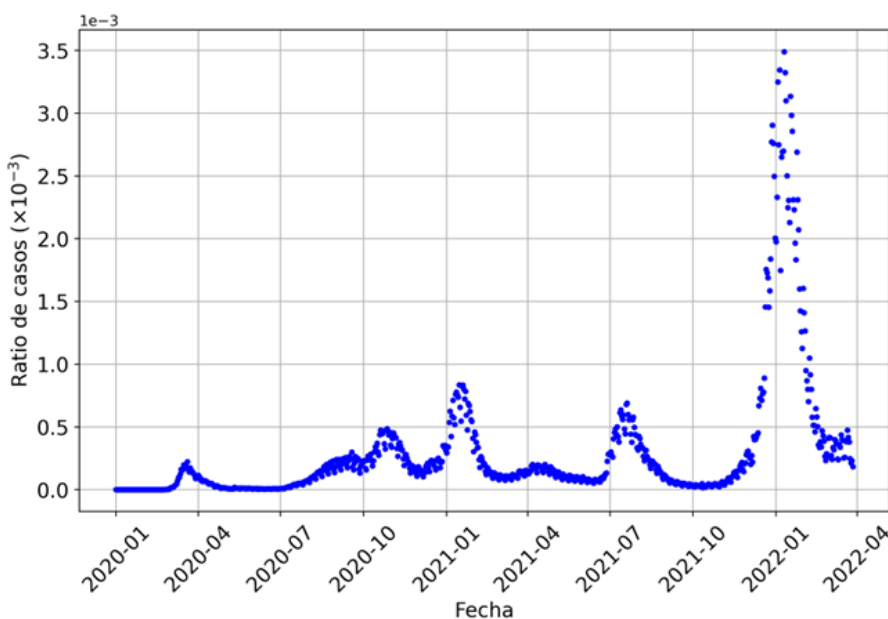


Figura 1. Ratio de casos notificados de COVID-19 del 1 de enero de 2020 al 28 de marzo de 2022 en España (fuente: SiViES).

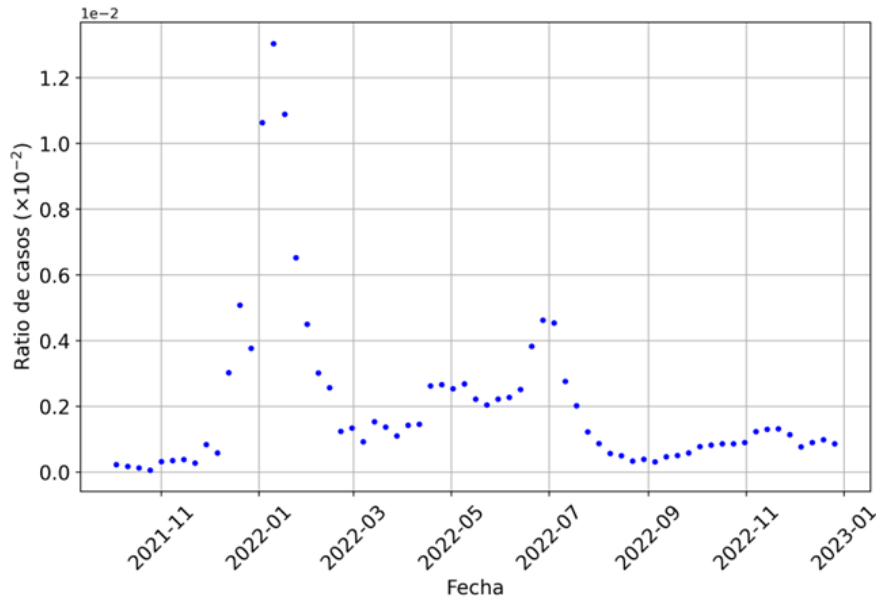


Figura 2. Ratio de casos semanales de la semana 40 de 2021 (31 de octubre de 2021) a la semana 31 de 2023 (30 de junio de 2023) (fuente: SiVIRA).

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.

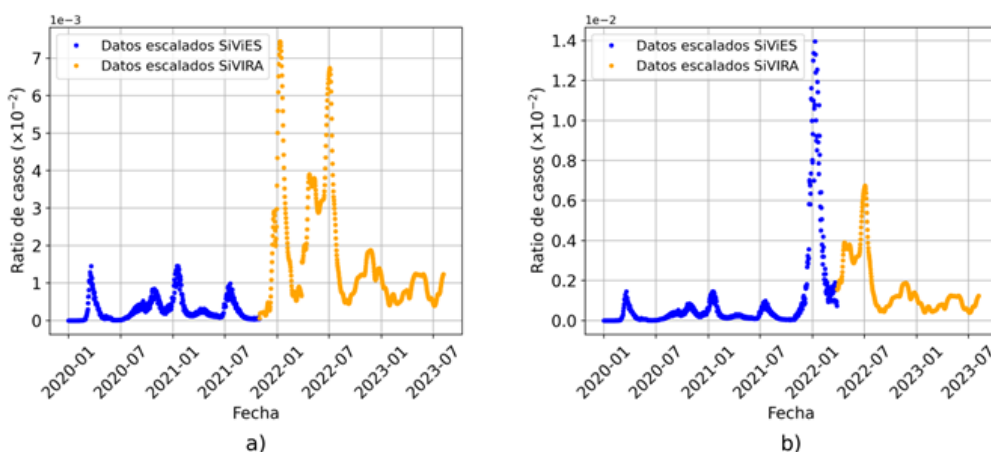


Figura 3. Tasas de incidencia estimadas (ratio a 1) para España. Se muestran dos posibles escenarios para representar la onda Omicron en España (diciembre de 2021 a marzo de 2022): (a) utilizando casos notificados por SiVIRA, (b) utilizando casos notificados.

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^a) 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 ¹
2021-07-01 – 2021-10-31	1.43 ¹
2021-11-01 – 2021-11-30	4.00 ¹
2021-12-01 – 2022-03-28	4.00 ¹
2022-03-29 – 2022-12-31	9.89

¹Valores 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.

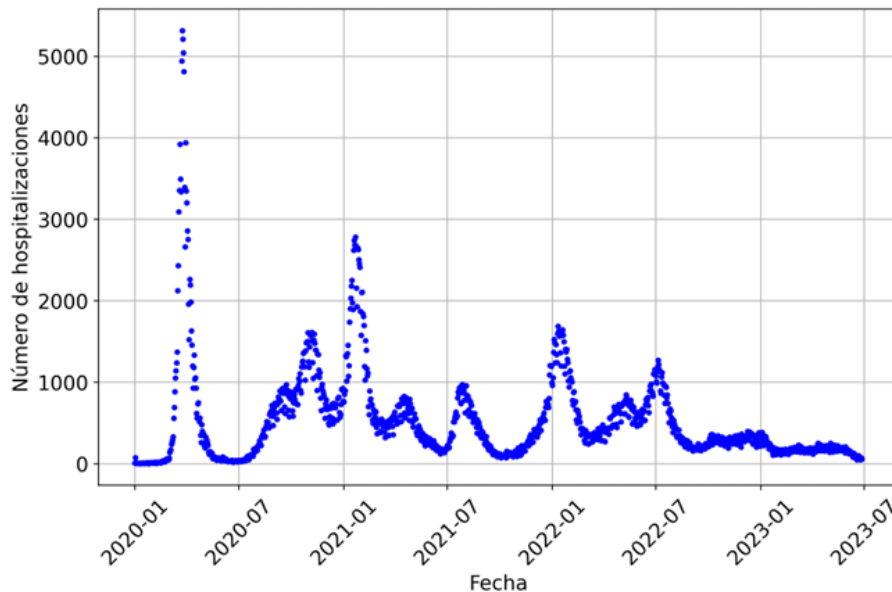


Figura 4. Número de hospitalizaciones por COVID-19 en España, del 1 de enero de 2020 al 29 de junio de 2023 (fuente: SiVIES).

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.

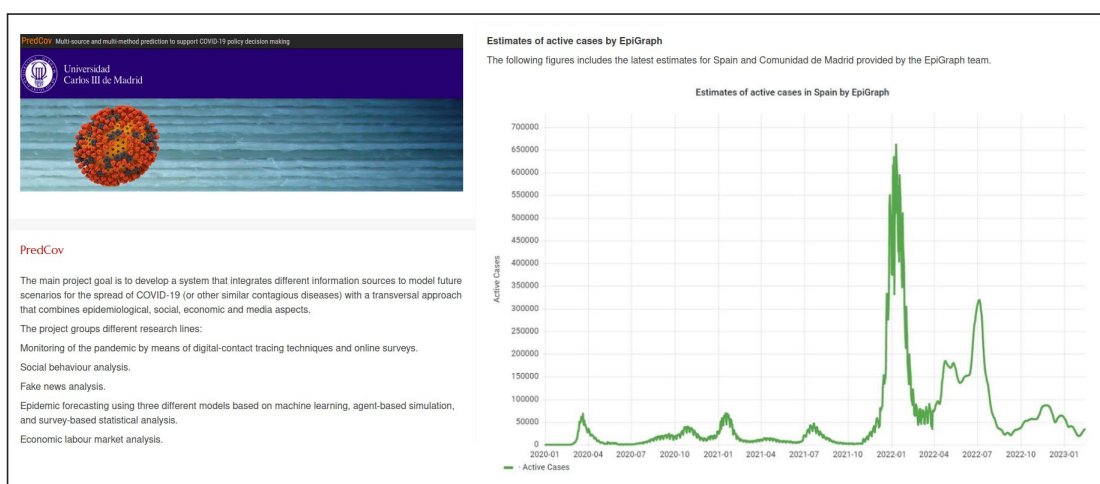


Figura 5. Página web del proyecto PredCov. A la izquierda se muestra la página principal, mientras que a la derecha se muestra la sección que tiene la estimación de la incidencia real de COVID-19.

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:

- ISCI3_RAW.png: datos originales de SiViES representados de forma gráfica.
- SiVIRA_RAW.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.

- TimeDependentSiVIRA_RAW.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+SiVIRA_SCALED.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