

Análisis en tiempo real de la situación sanitaria y epidemiológica por COVID-19 en el contexto internacional: dos fracasos y un éxito

Desarrollo de Francisco Rodríguez Cabrera, Elena Vanessa Martínez Sánchez y el equipo de datos del Centro de Coordinación de Alertas y Emergencias Sanitarias.

INTRODUCCIÓN

El contexto geográfico y la movilidad son dos de los factores más importantes en la distribución de las enfermedades infectocontagiosas respiratorias. En la pandemia gripal estacional, por ejemplo, es habitual que en España se comience en el norte del país y se distribuya hacia el sur en las siguientes semanas.

La vigilancia de esta situación sanitaria y epidemiológica del resto de países, con el menor retraso de notificación posible, supone un paso clave para la preparación y control de enfermedades como el COVID-19. En este sentido, la información agregada que los países comunicaban a su ciudadanía era frecuentemente más rápida que la recopilada en los registros individuales de los sistemas de información de vigilancia europeos (TESSy).

Por su elevada y rápida transmisibilidad, Ómicron supuso un reto de nuevo en la vigilancia de COVID-19. El gran crecimiento de los casos y la disminución de la letalidad, en parte por la variante y en mayor medida por la cobertura vacunal, hizo que disminuyera la credibilidad en los números de casos que aportaban los sistemas de vigilancia de muchos países. Por esta razón, la vigilancia de la situación sanitaria cobró aún mayor importancia para evaluar nuestro contexto. Acceder a la última información de confianza disponible sobre el resto de países permite tomar acciones más acordes, también, a la situación internacional, algo clave en un momento donde las limitaciones de movilidad por motivos sanitarios iban decreciendo. A partir de este momento se ideó un proceso nuevo para recopilar, validar, analizar y difundir un conjunto de indicadores COVID de otros países, comparándolos con los datos de España.

METODOLOGÍA

En primer lugar, se identificaron y evaluaron nuevamente los conjuntos de datos abiertos en COVID-19 disponibles en instituciones de mayor fiabilidad. Si bien prácticamente todas las fuentes recogían casos confirmados, tanto el desglose de estos como los indicadores de situación hospitalaria eran muy heterogéneos. Por ello, se utilizaron en primer lugar los portales de los países de mayor repercusión para España (tales como Portugal, Francia, Italia, Reino Unido o Alemania), el ECDC, la OMS (incluyendo OMS Europa) y, además, fuentes informales como *Our World in Data* cuando los datos no estaban disponibles en el resto de fuentes.

Se acudía en primer lugar a los portales de las autoridades sanitarias nacionales, dado que la labor del ECDC a menudo registraba la información con un rango de 1 a 2 días de retraso, un tiempo que podía ser valioso en determinados momentos de la pandemia. Esto hizo aumentar la complejidad:

- En Portugal, los datos en origen se disponían a través de un portal de ArcGIS, que dejó de funcionar durante 2022. A mediados y finales de ese año, únicamente el Instituto Nacional de Saúde Ricardo Jorge publicaba datos a través de un informe en portugués. Si se quería analizar Portugal, el código pasó a descargar el PDF, importar (escraper) el documento y buscar en función de determinadas palabras clave introducidas. Una situación similar sucedió con Francia en los meses posteriores. Tras la caída de publicación de estos informes, se pasó a los datos que recogía el ECDC.
- En Reino Unido, se descargan los datos a través de su [API oficial](#), si bien a menudo hacen referencia únicamente a Inglaterra.
- En Italia, se descargan los últimos datos a partir del GitHub del [Dipartimento della Protezione Civile](#).
- Para los países europeos, se accede a través de múltiples fuentes que mantenía en abierto el ECDC sobre incidencias ([1](#), [2](#), [3](#)), [cobertura vacunal](#), [variantes](#), [hospitalización](#) y [positividad de pruebas](#).
- Para todos los países, se accede a través de las bases de datos de la OMS ([1](#), [2](#)) y la agregación de [Our World in Data](#).

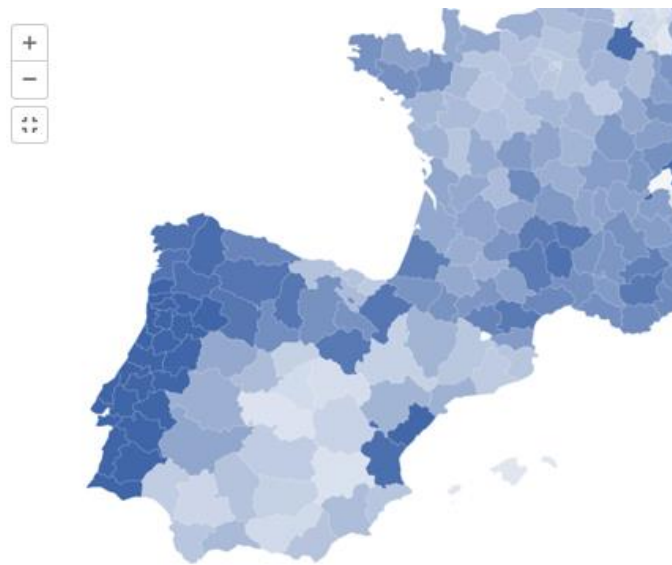
En origen, y dada la escasez de tiempo disponible, el código se produjo íntegramente con R, pero los mapas se realizaban ejecutando un do-file reutilizado de STATA desde el command prompt. código fue posteriormente traducido a R para la producción de los informes automatizados. En el análisis por regiones europeas, con el objetivo de solventar la heterogeneidad en los datos a nivel subnacional dentro de Europa, se reutilizó parcialmente el código del [COVID-19 European Regional Tracker](#) para producir mapas interactivos con Leaflet. La información sobre vacunaciones y variantes del ECDC o, en países fuera de la EU, provenían de agregaciones hechas por *Our World in Data*. En Reino Unido se descargaban también de su API.

En su versión final, el código producía una tabla con los datos de indicadores de incidencias globales, incidencias por grupos de edad, proporción de variantes, hospitalizaciones convencionales, UCI, fallecidos y cobertura vacunal de los países solicitados. El conjunto de datos, filtrado con los indicadores seleccionados a priori, se cruzaba con la última información disponible sobre la situación en España a partir de la Red Nacional de Vigilancia Epidemiológica (SiViEs) y hospitalizaciones (CMC).

RESULTADOS

Mapas de la situación epidemiológica por regiones europeas

Eliminadas ya las restricciones de movilidad dentro del país, a excepción de algunas limitaciones sanitarias del espacio Schengen, las regiones limítrofes con España podrían suponer un mayor riesgo que aquellas de mayor lejanía geográfica y menores relaciones comerciales. Por esta razón, se diseñó un mapa con los datos de la situación regional de Francia, Portugal, Andorra y Gibraltar.

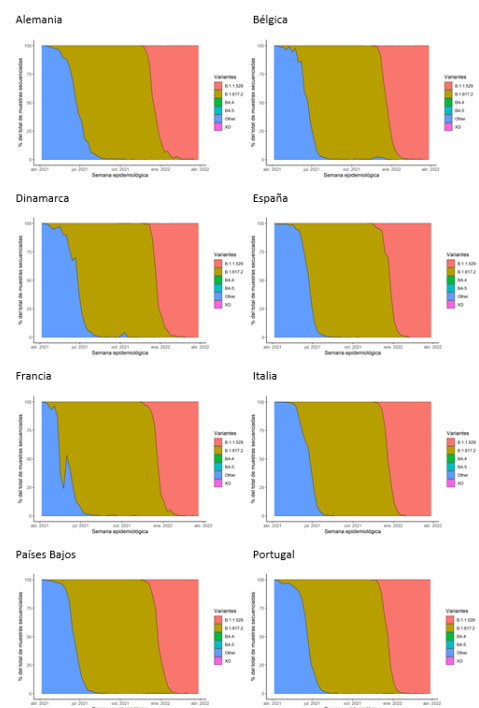


Este análisis desapareció a las pocas semanas de haberse producido tras la finalización de la publicación por parte de la fuente principal (ECDC y Comisión Europea).

Evolución de las variantes en el contexto europeo

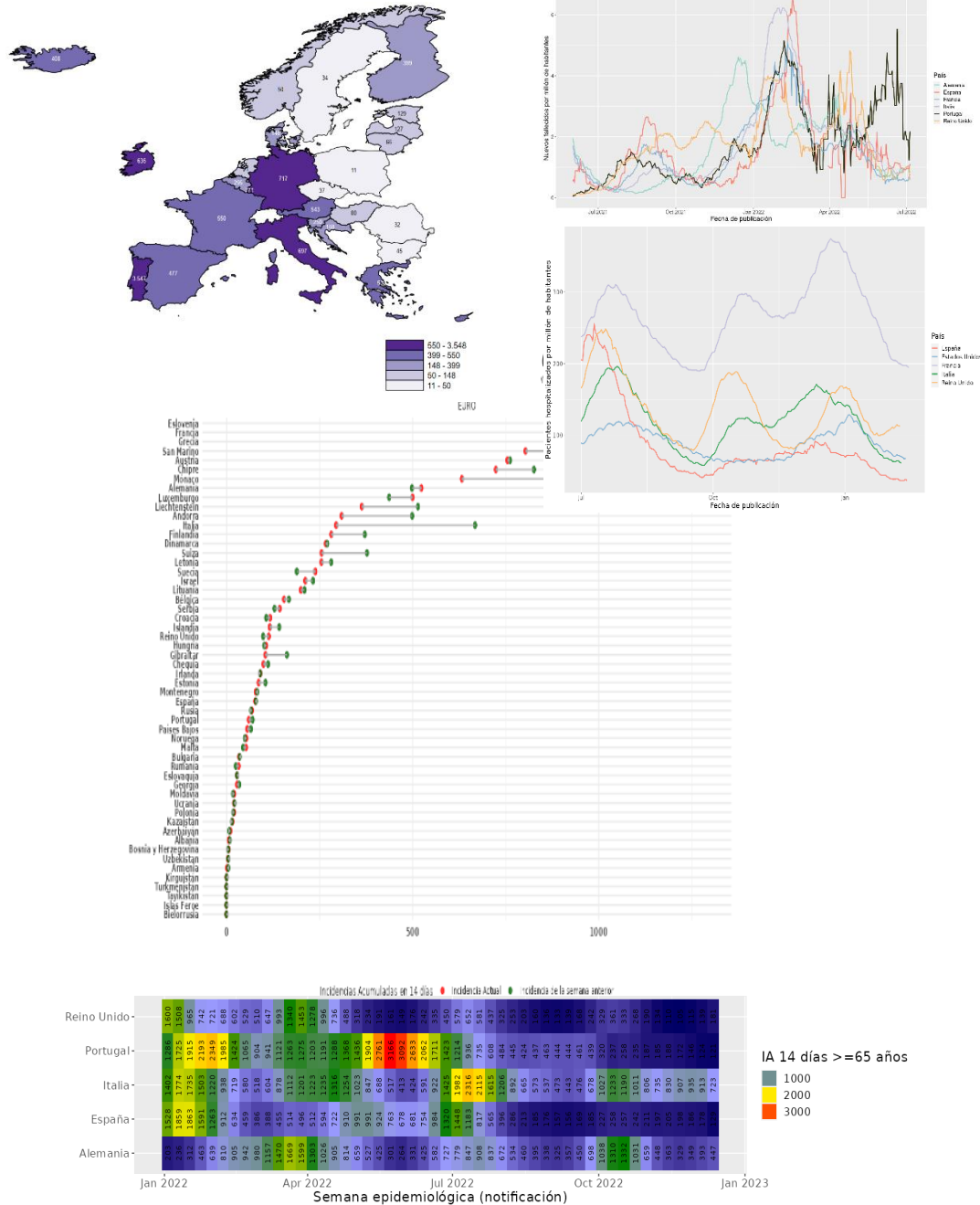
La rápida introducción de Ómicron en los países europeos reintrodujo la preocupación a nuevas variantes. Por esta razón, se incorporó una herramienta que agrupaba las variantes de especial interés e informaba de la introducción de algunas no obtenidas previamente en ninguno de los países. De las ya agrupadas, se obtenía la evolución en determinadas semanas epidemiológicas en los países europeos que se deseaba vigilar.

Este análisis tuvo escasas aplicaciones prácticas, por la mayor facilidad de uso y el desarrollo de los portales de GISAID.



Informes automatizados de situación epidemiológica y sanitaria, adaptable para cada país

En el código, se seleccionaban entre 2 y 8 países de comparación con España. A su vez, se seleccionaba una fecha de comienzo y, opcionalmente, una fecha de final del análisis. Este script finalizaba con un informe en formato Word con visualizaciones adaptadas a los objetivos (especialmente si se incluía o no el código del análisis de variantes). Entre otras consideraciones, incorporaba imputaciones de 1 a 6 valores intermedios entre valores válidos, en caso de que la información reportada por el país fuera semanal; las visualizaciones de fallecidos también incluían suavizaciones a través de una media móvil de 7 días, para evitar el efecto de infranotificación durante el fin de semana. Tras el cambio de vigilancia a finales de marzo de 2022, se filtró por grupos de edad. A continuación, se observan algunas visualizaciones de ese informe automatizado:



Se creó también un script para informes ad hoc únicos para un país, más completos, que analizaba la evolución de todos los indicadores disponibles de ese país. Estos informes estaban únicamente disponibles para países del contexto europeo.

Los scripts utilizados tanto los informes conjuntos como los informes ad hoc por país se siguen utilizando hoy día, en diferentes versiones dependiendo de la necesidad detectada.

CONCLUSIONES

A finales de 2021, el análisis de la situación epidemiológica fuera de España suponía a profesionales en la AGE un trabajo manual diario de entre media hora y una hora, limitándose hasta entonces al número de casos y la búsqueda de los portales web de países de mayor repercusión para España. Este trabajo implicaba normalmente a profesionales MIR, a los que se debía enseñar y, naturalmente, no siempre empezaban haciéndolo sin errores. La creación de informes duraba de 1 a 3-4 horas, en función de la complejidad del análisis requerido.

Con este proyecto, se consiguieron informes automatizados a partir de la ejecución de scripts de R que, en su versión más sencilla, tienen 2 minutos de duración en total. A pesar del intento de armonización llevada a cabo dentro de todo el proceso, las definiciones a una misma variable seguían pudiendo no concordar entre los países, especialmente en el caso de las hospitalizaciones y aquellos que fallecían por COVID-19 o de manera concurrente. No obstante, además de la rapidez, este proyecto ha permitido una comparación más flexible con los datos provenientes de nuestros sistemas de vigilancia nacionales, y poder tomar acciones pertinentes en nuestro contexto. Un trabajo que ha requerido de dos fracasos para poder obtener un éxito.