

Resumen

Se revisan datos y métodos para evaluar las políticas sanitarias y de salud, de enorme trascendencia económica durante la pandemia, y la institucionalización de la evaluación en España. Durante la pandemia, el gasto sanitario es solo una pequeña parte del coste total de la COVID. La prevención es altamente coste-efectiva. Proponemos áreas para evaluación: digitalización en sanidad, gestión indirecta de la gestión sanitaria, tecnologías de alto coste, intervenciones económicas y fiscales sobre la salud de la población, modelo MUFACE, y las consecuencias a largo plazo de la COVID-19 sobre la salud y sobre la desigualdad.

Palabras clave: evaluación de políticas, sanidad, análisis coste-efectividad, evaluación económica.

Abstract

We review data and methods for evaluating health and health policies, of enormous economic significance during the pandemic, and the institutionalization of evaluation in Spain. During the pandemic, health expenditure is only a small part of the total cost of COVID. Prevention is highly cost-effective. We propose areas for evaluation: digitalization in healthcare, indirect management of healthcare management, high-cost technologies, economic and fiscal interventions on population health, MUFACE model, and the long-term consequences of COVID-19 on health and inequality.

Keywords: policy evaluation, healthcare, cost-effectiveness analysis, economic evaluation.

JEL classification: H51, I14, I18.

EVALUACIÓN DE POLÍTICAS. SALUD Y SANIDAD

Beatriz GONZÁLEZ LÓPEZ-VALCÁRCEL (*)

Universidad de Las Palmas de Gran Canaria

I. INTRODUCCIÓN

EN los últimos dos años ha habido enormes avances en el conocimiento humano, con la ciencia como protagonista. Una enfermedad infecciosa ha marcado el paso a la economía mundial y específicamente al sector sanitario, que se ha vuelto más visible y pieza esencial en la gestión de las nuevas crisis, a la vez sanitarias y económicas.

El mayor reto de los países durante estos dos años ha consistido en encontrar soluciones para compatibilizar la pandemia con la vida de las personas y mantener el tejido productivo y el crecimiento económico a medio plazo. Contar con una sanidad bien organizada, reactiva, flexible y con los recursos necesarios es un requisito. La sanidad ya tenía su lista de problemas y retos pendientes, pero la agenda varió sustancialmente con la pandemia. Nunca la sanidad y la economía habían sido tan interdependientes, nunca las políticas sanitarias habían tenido tanto de políticas económicas. La pregunta del millón es si las decisiones que se tomaron en su momento consiguieron minimizar el impacto de la pandemia en los distintos frentes.

Articular una respuesta rigurosa a esa pregunta no puede hacerse sin desarrollar previamente un proceso de evaluación completo y metodológicamente robusto. La evaluación de polí-

ticas en general, y desde luego también la gestión de la pandemia, requiere datos accesibles y de calidad y el uso de métodos reproducibles (Casado y Pino, 2021). La disponibilidad global de datos de grano fino, sobre todo epidemiológicos, pero también económicos, de movilidad humana o de la gestión pública de la pandemia es cada vez mayor. Paralelamente, ha habido un enorme desarrollo y presencia de la *ciencia de datos*, agnóstica y omnipresente, al albur de la mejora exponencial en disponibilidad y acceso a algunos tipos de datos. La inteligencia artificial (IA) se ha convertido en objetivo de los Gobiernos, tal vez más como un fin en sí mismo que como un medio para avanzar en el conocimiento y el bienestar de la humanidad.

En todo caso, y más allá del caso particular de la respuesta a la pandemia de COVID-19, la evaluación de políticas de salud es particularmente compleja. Aunque es habitual que se pierda de vista, la salud, objetivo último, es más que sanidad, y se genera fuera del sector sanitario mediante políticas intersectoriales, de vivienda y urbanismo, laborales y fiscales, entre otras. Además, la evaluación de las políticas de salud está menos incentivada que, por ejemplo, la evaluación de tecnologías sanitarias como son los medicamentos o los procedimientos quirúrgicos, por lo que suele concentrar menos recursos y esfuerzos.

En este artículo se reflexiona sobre los datos y métodos para evaluar las políticas sanitarias y de salud y sobre la organización de dicha evaluación en España. Tras esta introducción, la sección segunda de este artículo se centra en los *datos* necesarios para evaluar los avances que se han obtenido en los últimos años y los problemas que han surgido. En la tercera sección se abordan los cambios en los *métodos* de evaluación, en parte relacionados con las tendencias en la disponibilidad de datos. La sección cuarta aborda específicamente la modelización epidemiológica y económica de la pandemia, que aporta *inputs* imprescindibles para la evaluación de las políticas públicas relacionadas con la COVID-19. Relacionado con esa última sección, la quinta sección considera la evaluación de la *macrogestión* de la pandemia, especialmente en España. La sexta sección se pregunta si los *métodos clásicos de evaluación económica* en sanidad siguen siendo aplicables o deben ser reformulados y adaptados, y se analizan los cambios recientes en la arquitectura institucional de la evaluación de tecnologías sanitarias en nuestro país. También se presentan algunas aplicaciones específicas de evaluación económica relacionada con la COVID-19. La sección séptima abre el foco, presentando ejemplos relevantes de evaluación de políticas públicas de salud y sanidad que no están relacionadas con la pandemia, y se apuntan temas relevantes para una agenda priorizada de evaluación de políticas en el ámbito sanitario y de la salud. Finalmente, se concluye con el resumen de los principales mensajes y prescripciones.

1. Los datos necesarios para evaluar

La materia prima esencial para evaluar las políticas de salud y sanitarias son los datos. La evaluación de políticas en la actualidad se ve influida por tres grandes tendencias:

- Disponibilidad de *registros sanitarios* poblacionales en lugar de muestras. Esos datos se crean en formato digital y por tanto están o pueden estar accesibles *online*. Hay dos retos, hacerlos interoperables y combinar bases de datos individuales de distintos registros, preservando la confidencialidad y el anonimato. Las historias clínicas combinadas con bases de datos de dispensaciones farmacéuticas y de imagen tienen un enorme potencial para la investigación observacional con datos en el mundo real (*real world data*). Algunas comunidades autónomas (CC. AA.) van por delante y ya cuentan con registros poblacionales de datos de salud, con usos en investigación evaluativa (García-Sempere *et al.*, 2020).
- Disponibilidad de *datos longitudinales*, que permiten controlar por heterogeneidad no observable entre personas –pacientes– y organizaciones sanitarias. Es el caso de las historias clínicas, pero también de los datos de utilización, actividad y calidad de los centros sanitarios a nivel de mesogestión, como los que publica en abierto el Observatorio de Resultados del Servicio Madrileño de Salud (<https://www.comunidad.madrid/servicios/salud/observatorio-resultados-servicio-madrileno-salud> o AQUAS

en Cataluña (<https://aquas.gencat.cat/ca/inici>)

- Disponibilidad creciente de *big data* (vs. datos estructurados), caracterizados por las cinco «*uves*»: volumen, variedad, velocidad, veracidad y valor. Representan un reto para los modelos económicos y estadísticos tradicionales y son *inputs* esenciales para la *ciencia de datos*.

El contar con información digitalmente accesible facilita la evaluación y la investigación evaluativa. Todavía no podemos valorar hasta qué punto las políticas de transparencia activa mediante la incorporación de estándares de datos abiertos y de la Carta Internacional de Datos Abiertos (IODC) (Ubaldi, 2013) han avanzado durante los años de pandemia. El índice OURdata mide la disponibilidad, accesibilidad y reutilización de los datos del Gobierno. Según la OCDE, se trata de «un bien público que debería ser provisto con un propósito definido, proactivamente y con foco en la reutilización, en línea con las necesidades del usuario» (OCDE, 2020). El *Digital Government Index* de la OCDE incorpora las seis dimensiones del *Digital Government Policy Framework*: 1) digital desde el diseño; 2) sector público impulsado por los datos; 3) Gobierno como plataforma; 4) abierto por defecto; 5) enfocado al usuario; y 6) proactividad. En 2019, último año para el que hay datos, España ocupaba el séptimo puesto de los 33 países de la OCDE en el índice global. Aunque destaca entre los primeros puestos (puesto 4 de 33) en las dimensiones 1 (administración de diseño digital), 2 (un sector público impulsado por los datos) y 6 (proactivo), hay mucho mar-

gen de mejora en el enfoque hacia el usuario (puesto 12), y sobre todo en el establecimiento del principio de accesibilidad del público a los datos por defecto, en el que ocupamos el puesto 23, a la cola.

En España se han dado algunos pasos adelante, como el acuerdo de abril de 2021 entre el INE, la Agencia Estatal de Administración Tributaria, la Seguridad Social, el Servicio Público de Empleo Estatal y el Banco de España para facilitar datos anonimizados cruzados para investigación. También las líneas estratégicas del Programa anual 2021 del Plan Estadístico Nacional 2021-2024 (Real Decreto 150/2021) (Póo y Galiano, 2021) y la estadística experimental del INE (<https://www.ine.es/experimental/experimental.htm>) representan avances en la disponibilidad de datos *de grano fino* casi en tiempo real. La estadística experimental incluye desde la información semanal de movilidad humana, cotidiana y estacional, obtenida de las antenas de teléfonos móviles, que permite predecir mejor la dinámica de la pandemia, hasta la estimación del número de defunciones semanales durante el brote de COVID-19: (https://www.ine.es/experimental/defunciones/experimental_edes_proyecto.pdf).

2. Cambios en los métodos de evaluación

«Para quien solo tiene un martillo, todo son clavos». Frente al enorme avance en la disponibilidad de datos, la metodología para procesarlos con fines de evaluación de políticas públicas está relativamente estancada. Desde la economía, los enfoques estructural (Heckman, 2001) y la revolución de la credibilidad

(Angrist y Pischke, 2010), que le valieron el Nobel de Economía en 2000 a Heckman y en 2021 a Angrist, Imbens y Card, tienen ya décadas. La disponibilidad omnipresente de datos masivos resulta en una tendencia a explorar microcausalidades locales más que relaciones generales de causa-efecto, con riesgo de sobreutilización de IA, que está resultando más útil para unas funciones (por ejemplo, para diagnosticar COVID-19 adiestrando algoritmos para que exploren placas de tórax) que para otras (por ejemplo, para contestar preguntas específicas sobre el impacto de determinadas políticas en la utilización sanitaria o en la salud de la población).

El ensayo clínico aleatorizado y doble ciego es el *gold standard* en diseño de datos experimentales en biomedicina, y durante años desde la economía hemos buscado buenos *proxies* cuando no es posible hacer experimentos sociales, cuasiexperimentos o contar con experimentos naturales. Con todo, el número de (auténticos) experimentos sociales relacionados con la salud y con las políticas de salud es abundante, sobre todo en países en desarrollo. El Banco Mundial ha seguido una estrategia de evaluación de impacto de las políticas (Gertler *et al.*, 2016) en la que el diseño e implementación de experimentos juega un papel central. Incluso ha avanzado en el diseño de nuevos métodos para minimizar los sesgos provocados por su propio *staff*. Es lo que se ha dado en llamar *práctica reflexiva* (McKee y Stuckler, 2015), recomendable para otras organizaciones que han de evaluar políticas de salud.

Pero en economía, en ciencias sociales y en los análisis del

comportamiento humano relacionado con la salud, los pequeños ensayos, aunque estén bien diseñados, tienen limitaciones. Por ese motivo, se está propugnando como un avance sustancial la implementación de *megaensayos* para establecer relaciones causa-efecto de distintas intervenciones sobre el comportamiento humano (Royer, 2021). Por ejemplo, un experimento masivo (Milkman *et al.*, 2021) estimó y comparó el efecto de 54 intervenciones diferentes y simultáneas, de cuatro semanas de duración, en la misma población (los 61.293 usuarios de una red de gimnasios en EE. UU.) utilizando la misma definición y medida de resultado (frecuencia de asistencia al gimnasio). Un total de 30 científicos de 15 universidades trabajaron en pequeños equipos independientes en el diseño de las intervenciones. Este megastudio supera las limitaciones de los pequeños ensayos, cuyos resultados solo tienen validez local y son difícilmente comparables entre experimentos.

Pero el tamaño no lo es todo, siendo más importante evitar sesgos por falta de representatividad de las muestras que conseguir tamaños muestrales masivos. Un ejemplo contrastado de la *paradoja del big data* es la estimación del porcentaje de personas vacunadas de COVID en EE. UU. entre enero y mayo de 2021 a partir de grandes encuestas en Facebook (n=250.000 por semana) y del *Census Household Pulse* (n=75.000 cada dos semanas). Con los datos reales de vacunaciones se puso en evidencia que ambas encuestas sobreestimaban el porcentaje de vacunados, en 17 puntos la de Facebook y en 14 puntos la otra (Bradley *et al.*, 2021). Un buen diseño muestral con una muestra 500 veces menor habría

proporcionado estimadores con error cuadrático medio menor y sin sesgo.

II. LOS MODELOS: DE LA EPIDEMIOLOGÍA A LA EPIECONOMÍA

Desde marzo de 2020, ha habido una *pandemia de modelos* COVID (1). La mayor parte han ido desapareciendo, en un proceso darwiniano en el que sobreviven los más capaces de aportar. Algunos han sido adoptados por las instituciones internacionales, como es el caso del modelo de la Universidad Politécnica de Cataluña por el European Centre for Disease Prevention and Control (ECDC) (Català *et al.*, 2021).

Los modelos epidemiológicos más simples partían de los modelos SIR (susceptibles, infectados y recuperados); SEIR (susceptibles, expuestos, infectados y recuperados) del siglo pasado. Son modelos *compartimentales* en los que cada persona pertenece a un grupo de un conjunto exhaustivo y mutuamente excluyente: susceptibles, expuestos, infectados, recuperados. Han ido añadiendo *complejidad* progresivamente, tendiendo hacia modelos de grano fino que incorporan cada vez más información micro sobre movilidad e interacciones sociales, y sobre políticas públicas (restricciones a la movilidad y a la actividad económica) e intentan contestar preguntas sobre efectos de las políticas e intervenciones públicas. Las aplicaciones para la actual pandemia estiman los efectos de la movilidad y las restricciones de los Gobiernos, sobre todo en relación a viajes y movimientos transfronterizos (Chinazzi *et al.*, 2020; Linka *et al.*, 2020; Poletto, Scarpino y Volz, 2020). No es que estos modelos epidémicos espaciales sean nue-

vos (Balcan *et al.*, 2010), pero nunca antes había habido tantos datos para utilizarlos.

Una función importante de los modelos es predecir la presión sobre el uso de los recursos y servicios sanitarios y el déficit de recursos disponibles para otros problemas de salud. Complementariamente al tradicional *forecasting* o predicción de futuro, ha surgido el *nowcasting*, (literalmente «predicción del ahora»), predicción a ultracorto plazo, o más bien análisis de lo que está pasando en tiempo real. Se había aplicado en meteorología, física y otras ciencias, y ahora se han encontrado aplicaciones para tomar decisiones exprés relativas a la pandemia (Rossman y Segal, 2022) (Wu *et al.*, 2021).

Estos modelos utilizan datos masivos *compartidos en tiempo real*, de acceso abierto. Hay proveedores universales, como la John Hopkins y Dataworld. La comisión COVID-19 de *The Lancet* (<https://data.covid19commission.org/>) tiene una base de datos de acceso abierto que actualiza semanalmente, con 154 indicadores para 244 países, más de 10 millones de datos en total.

Otros *inputs* para los modelos epieconómicos son los datos de movilidad (a nivel regional, las visitas diarias a tiendas y ocio, supermercados, parques, farmacias, estaciones de transporte, lugares de trabajo y zonas residenciales) que aporta Google como *inputs* para los modelos, obtenidos del seguimiento que hace Google Maps de los desplazamientos (<https://www.google.com/covid19/mobility/>).

Tanto los datos como los modelos se han globalizado. Hace apenas unos años hubiera sido

impensable que una organización como el Institute of Health Metrics and Evaluation (IHME) de la universidad de Seattle (EE. UU.) ofreciera predicciones regionales sobre la evolución de la pandemia bajo escenarios alternativos de uso de mascarillas y vacunación, no solo para España (<https://covid19.healthdata.org/spain?view=cumulative-deaths&tab=trend>), sino también para cada una de sus CC. AA.

Las reacciones de los Gobiernos frente a la evolución de la pandemia, las medidas que han ido tomando para *aplanar la curva* o para crear *zonas verdes* libres de virus intercomunicadas entre sí, se plasman y cuantifican en índices estandarizados, comparables internacionalmente (Wong *et al.*, 2020). Uno de los más conocidos y utilizados, también por la OCDE, es el *Stringency Index* de Oxford (Hale *et al.*, 2021), que contiene 23 indicadores de respuesta de los Gobiernos: restricciones, instrumentos económicos y respuestas desde el sistema sanitario.

Los modelos son para predecir a corto plazo, pero también para evaluar las políticas de gestión de la pandemia. Utilizando el método científico, se trata de aprender de los datos para explicar qué ha pasado y por qué (y no repetirlo), detectar tempranamente brotes (anticipación) y aconsejar sobre decisiones políticas. Asimismo, los modelos permiten contestar preguntas específicas y encontrar asociaciones. Sin embargo, tienen problemas por la incertidumbre intrínseca de una enfermedad que está en continua evolución. Los modelos de simulación de políticas tienen mayor incertidumbre que los predictivos, en parte debido a que necesitan *inputs* sobre la adherencia de la población a las

restricciones (el propio rebaño es quien determina su inmunidad). No es de extrañar que incluso los modelos más simples, epidemiológicos predictivos, obtengan buenos resultados en el corto plazo (y en el hipercorto, con el *nowcasting*), pero sus errores de predicción a medio plazo (varias semanas o meses) sean muy altos (para un *ranking* de capacidad predictiva de los modelos, véase Friedman *et al.*, 2021). Para Europa, también hay disponible un sitio web que compara predicciones de distintos modelos (2) supervivientes de aquel proceso darwiniano al que se hacía referencia al principio de esta sección.

La modelización ha logrado grandes progresos, pero todavía falta integrar mejor los modelos epidemiológicos y económicos, así como incorporar la dimensión de las *desigualdades* interpersonales y los efectos sobre la desigualdad.

1. La necesidad de evaluar la macrogestión de la pandemia

A grandes rasgos, la gestión de la pandemia abarca: 1) las medidas para limitar los contagios; 2) la disposición y organización de recursos y servicios socio-sanitarios para prevenir y tratar la enfermedad; y 3) las medidas de carácter económico para minimizar los efectos de la pandemia sobre el tejido productivo y sobre las economías familiares.

Las decisiones de los tres tipos se han ido tomando con urgencia y en el ambiente de incertidumbre de cada momento. Conceptualmente, una evaluación *ex ante* ayudaría a decidir el curso de la acción, comparando los efectos de escenarios alternativos correspondientes a distintas

medidas. Las evaluaciones *ex post* permiten valorar *a posteriori* el acierto o la falta de acierto, las buenas y malas decisiones del pasado, para no reincidir cuando llegue la próxima oleada o pandemia.

En puridad, una evaluación de la (macro) gestión de la pandemia debería considerar las dos grandes alternativas de los Gobiernos, la *estrategia de supresión*, con zonas verdes en las que no hay casos de infección, que están intercomunicadas y aisladas de las zonas *rojas*, y la *estrategia de mitigación*, basada en restricciones tanto más serias cuanto peor es la incidencia y otros indicadores de los semáforos COVID (ocupación de camas, incidencia en grupos etarios específicos, etc.). La *green zone strategy* o estrategia de supresión se ha implementado en países-isla o con posibilidad de convertirse en territorios estancos (Nueva Zelanda y Australia, China y Estados-isla como Taiwán). Europa adoptó la estrategia de mitigación (*red zone strategy*). La mayor parte de la literatura económica aboga por esta última, basada en el cierre imprescindible, y destaca el enorme valor de los test individuales para mantener el tejido productivo en actividad. La literatura de salud pública, en cambio, aboga más por la estrategia de supresión (Oliu-Barton y Pradelski, 2021). Hoy por hoy no hay estudios rigurosos que permitan establecer la dominancia de una estrategia sobre la otra, y los escasos estudios (Zala *et al.*, 2020) se basan en proyecciones de modelos epidemiológicos que ya han demostrado ser erróneas.

Idealmente, serían necesarios modelos epieconómicos capaces de predecir las consecuencias

sobre la salud y sobre la economía de cursos de acción alternativos. Los *outcomes* son de salud (fundamentalmente, fallecimientos) y logros económicos. Como fuente de variación exógena, algunos modelos incorporan datos del pasado, sanitarios (incidencias, hospitalizaciones) y no sanitarios (intensidad de las restricciones entre otros) en distintas zonas o países, con el fin de anticipar los efectos, sobre la salud y económicos, de los posibles cursos de acción en la gestión macro de la pandemia. En España, las medidas más restrictivas fueron las de confinamiento domiciliario entre marzo y mayo/junio de 2020. Según una evaluación basada en modelos econométricos espaciales con datos provinciales (Orea y Álvarez, 2022), el confinamiento pudo haber reducido el número de casos en un 82,8 por 100. De haber confinado una semana antes, se habrían podido evitar un 11,6 por 100 adicional de casos y se habría conseguido un ahorro económico muy considerable.

La evaluación cuantitativa, aunque importante, no constituye el único método de evaluación útil, particularmente cuando lo que debe evaluarse es un proceso complejo de respuesta a una crisis de salud pública como la derivada de la epidemia de COVID-19, que ha tenido consecuencias en múltiples dimensiones. Ya a mediados de 2020, la OMS puso en marcha un proceso de evaluación independiente y exhaustivo con el fin de valorar la respuesta sanitaria internacional a la pandemia, del que pueden extraerse recomendaciones importantes para mejorar las capacidades de respuesta futuras (Sirleaf y Clark, 2021).

Sobre la base de la guía de la OMS para las denominadas «*after-action reviews*» (AAR) o «revisiones posactuación» (World Health Organization, 2019), el ECDC ha proporcionado un detallado protocolo de evaluación de las decisiones de respuesta a la pandemia. De acuerdo con el mismo, las AAR buscan evaluar las acciones de respuesta a un evento preocupante para la salud pública mediante la observación objetiva, el análisis de problemas y/o mejores prácticas, y la identificación de áreas de mejora en las actividades de preparación y respuesta, que suma a la recopilación de la evidencia científica disponible un proceso de reflexión que involucra a todos los agentes relevantes (ECDC, 2021).

Algunos países han publicado ya informes de evaluación que, sin seguir estrictamente el protocolo de las AAR, constituyen ejemplos de evaluación cualitativa orientada a identificar las lecciones que deben aprenderse de la gestión de la pandemia. Es el caso de Inglaterra, que explicita en su informe su espíritu constructivo, alejado de la búsqueda de culpables, un elemento que ha de considerarse crucial para el éxito y credibilidad de la evaluación (House of Commons, 2021). En España se ha reclamado una evaluación global de la gestión de la pandemia desde las primeras oleadas (García-Basteiro *et al.*, 2020; Martín-Moreno *et al.*, 2020) por el gran impacto en mortalidad, particularmente en residencias. El Consejo Interterritorial del Sistema Nacional de Salud (SNS) ha hecho el encargo en septiembre de 2021, y en el momento de publicación de este artículo el informe se encuentra en proceso de elaboración.

2. El principio del coste-efectividad como telón de fondo. Evaluación de las decisiones de cobertura, financiación de prestaciones del SNS y de la transformación digital

En el ámbito de la política sanitaria, la evaluación por excelencia ha sido la que consiste en aplicar el análisis coste-efectividad a los tratamientos, tecnologías e intervenciones sanitarias. La evaluación económica de procedimientos se ha estandarizado metodológicamente en el mundo, e incluso se ha convertido en una profesión, la farmacoeconomía, por ser los medicamentos el objeto más habitual de análisis. En cuanto a la metodología empleada, se vela por el uso de protocolos que impidan influencias indebidas o interesadas en los resultados, por ejemplo, a través del registro previo de los proyectos de evaluación, y se van produciendo avances importantes en lo que respecta a distintos aspectos, como por ejemplo el alcance de los costes indirectos.

Más allá del marco estrecho de los tratamientos médicos individuales, en sanidad hay que evaluar también cambios organizativos, intervenciones preventivas a nivel comunitario y políticas que inducen efectos fuera del sector sanitario y de la salud. Hay métodos como el *SROI* (*social return of investment*) que se han diseñado para ello y son de obligada aplicación para intervenciones sociales o, por ejemplo, para justificar actividades de las ONG en algunos países (para una revisión de aplicaciones del *SROI* en sanidad, véase González López-Valcarcel, 2019). Estos métodos se aplican cada vez más para evaluar políticas públicas

multisectoriales y del tercer sector. Aunque se ha avanzado en estandarización metodológica, todavía hay un gran trecho por recorrer.

La institucionalización de la evaluación económica de tecnologías sanitarias en España

Los desarrollos teórico-metodológicos y una enorme cantidad de aplicaciones de análisis de coste-efectividad en sanidad, accesibles en bases de datos como el registro de la Universidad de Tufts (<https://cevr.tuftsmedicalcenter.org/databases/cea-registry>) o el *Centre for Reviews and Dissemination* de la Universidad de York (<https://www.york.ac.uk/crd/>), han sido de gran utilidad en entornos institucionales en los que una cuarta barrera (demostrar el coste-efectividad) es legalmente preceptiva para incorporar una nueva tecnología a la cartera de prestaciones públicas. La mayor parte de países de la UE-15 exigen esa cuarta barrera actualmente.

España es una excepción, pues todavía no se reconoce formalmente la cuarta barrera para que un nuevo medicamento o prestación entre en la cartera de servicios y sea financiado por el SNS. A lo largo del tiempo, la necesidad de evaluación económica y el reconocimiento del criterio de coste-efectividad para la financiación pública de prestaciones se han ido incorporando de forma oblicua en el marco normativo (cuadro n.º 1).

La institucionalización de la evaluación de tecnologías sanitarias en España ha avanzado en los últimos años, pero todavía no está completada. Desde 2006 se ha ido conformando una red de colaboración entre agencias

CUADRO N.º 1

EL COSTE-EFECTIVIDAD Y LA EVALUACIÓN ECONÓMICA COMO CRITERIOS EN ESPAÑA. MARCO REGULATORIO

NORMA	CONTENIDO
Real Decreto 63/1995 sobre ordenación de prestaciones sanitarias.	«Para incorporar nuevas técnicas, tecnologías y procedimientos, la administración sanitaria del Estado ha de valorar su seguridad, eficacia y eficiencia».
Real Decreto 1030/2006 por el que se establece la cartera de servicios comunes del Sistema Nacional de Salud y el procedimiento para su actualización.	Para actualizar la cartera de servicios comunes del SNS «se deberá utilizar el procedimiento de evaluación más adecuado en cada caso que permita conocer el coste, la eficacia, la eficiencia, la efectividad, la seguridad y la utilidad sanitaria».
Real Decreto-ley 16/2012, de medidas urgentes para garantizar la sostenibilidad del Sistema Nacional de Salud.	La inclusión de medicamentos en la financiación del Sistema Nacional de Salud tendrá en cuenta (entre otros) el criterio del «valor terapéutico y social del medicamento y beneficio clínico incremental del mismo teniendo en cuenta su relación coste-efectividad». Las nuevas técnicas, tecnologías o procedimientos serán sometidas a evaluación, con carácter preceptivo y previo a su utilización en el Sistema Nacional de Salud, por la RedETS. «Para la decisión de financiación de nuevos medicamentos, además del correspondiente análisis coste-efectividad y de impacto presupuestario, se tendrá en cuenta el componente de innovación, para avances terapéuticos indiscutibles (...) y su contribución a la sostenibilidad del Sistema Nacional de Salud». «La Comisión Interministerial de Precios de los Medicamentos tendrá en consideración los análisis coste-efectividad y de impacto presupuestario».
Orden SSI/1833/2013, por la que se crea y regula el Consejo de la RedETS.	«La Red Española de Agencias de Evaluación de Tecnologías Sanitarias y Prestaciones del Sistema Nacional de Salud tiene como misión generar, difundir y facilitar la implementación de información destinada a fundamentar la toma de decisiones en el Sistema Nacional de Salud, contribuyendo de esta forma al incremento de la calidad, equidad, eficiencia y cohesión en el mismo». El pleno aprobará «el Plan Anual de Evaluación de Tecnologías Sanitarias (...) que se someterá a informe del Pleno del Consejo Interterritorial del Sistema Nacional de Salud».
Real Decreto Legislativo 1/2015 por el que se aprueba el texto refundido de la Ley de garantías y uso racional de los medicamentos y productos sanitarios.	«La ficha técnica o resumen de las características del producto (...) A la ficha técnica se acompañará, preceptivamente, información actualizada del precio del medicamento y, cuando sea posible, la estimación del coste del tratamiento» (art.15).

Fuente: Elaboración propia.

de evaluación de tecnologías sanitarias, algunas de las cuales ya existían desde principios de los años noventa. La red se constituye formalmente por el Consejo Interterritorial del SNS en 2012 como Red Española de Agencias de Evaluación de Tecnologías Sanitarias y Prestaciones del Sistema Nacional de Salud (RedETS) (<https://redets.sanidad.gob.es/>). RedETS «pretende fundamentar la toma de decisiones sobre incorporación, condiciones de financiación/desinversión, y uso apropiado de las tecnologías sanitarias, con el fin de promover la equidad y sostenibilidad del SNS» (3). Actualmente está formada por ocho unidades, siete de comunidades autónomas (Andalucía, Aragón,

Canarias, Cataluña, Galicia, Madrid, País Vasco) y el Instituto de Salud Carlos III, que colaboran en red, trabajan por encargo y con financiación del Ministerio de Sanidad y de los Gobiernos regionales, y elaboran informes no preceptivos sobre determinadas tecnologías, destacando las no farmacológicas (dispositivos médicos, procedimientos, etc.). Aunque explícitamente no hay una especialización «oficial» de las agencias, en la práctica algunas trabajan más intensamente determinados temas. Además de la función *primaria* de evaluación de tecnologías sanitarias (1 y 2 en la lista siguiente), la mayor parte de unidades de la red asumen otras funciones:

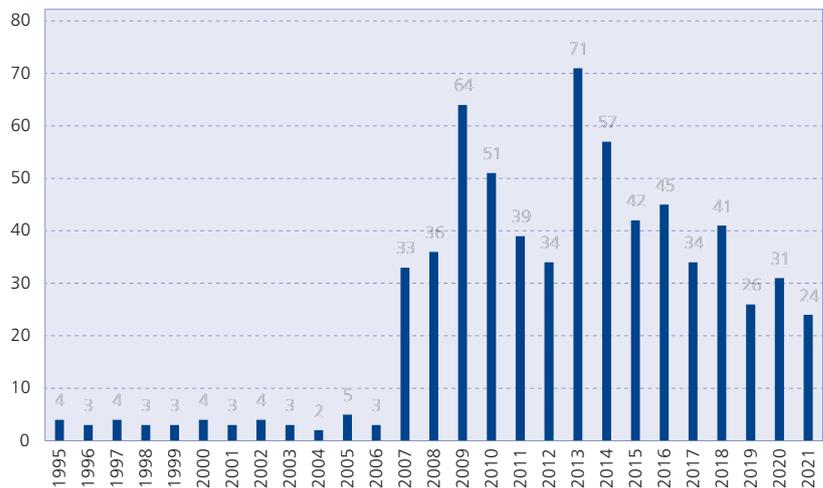
1. Informes de evaluación de tecnologías sanitarias (procedimientos, dispositivos, pruebas diagnósticas...).
2. Informes de evaluación de medicamentos (estudios coste-efectividad).
3. Guías de práctica clínica.
4. Proyectos de investigación (convocatorias competitivas nacionales e internacionales).
5. Compra pública innovadora.
6. Radar de detección de tecnologías emergentes a evaluar.
7. Guía de desinversión (detección de tecnologías obsoletas o que no aportan valor).

8. Metodología de la evaluación.
9. Formación y transferencia de conocimiento.

Algunas de estas agencias forman parte de unidades gubernamentales más amplias de investigación o generación y transferencia de conocimiento. Aunque se suele hablar de la «red de agencias», jurídicamente han ido evolucionando hacia otras formas jurídicas. En la mayor parte de los casos son unidades administrativas en el organigrama de las consejerías de sanidad (Andalucía, Madrid) o del servicio regional de salud (Canarias). AQuAS, por su parte, sí conserva el estatus de agencia independiente.

La RedETS aporta valor al SNS y aprovecha economías de alcance y de red. A su vez, las organizaciones de la RedETS participan en la Red Europea de Agencias de Evaluación (EUnetHTA). El nivel de recursos, actividad y producción de informes es muy desigual entre los miembros de la red. La más prolífica es, sin duda, la agencia catalana. En 2019, cuando AQuAS cumplía 25 años, había elaborado 1.130 informes de evaluación y publicado 750 artículos en revistas científicas (4). Durante las últimas dos décadas ha habido un enorme aumento tanto de la actividad de evaluación de tecnologías sanitarias en España como del número de miembros de la red. Considerando únicamente los informes de evaluación de tecnologías sanitarias en sentido estricto, hasta 2022 el conjunto de los ocho miembros actuales de la RedETS había publicado 669 informes de evaluación de tecnologías sanitarias (5) (gráfico 1). Hasta el año 2000, se publicaban una media de 3,5 informes anuales; en la

GRÁFICO 1
NÚMERO ANUAL DE INFORMES DE EVALUACIÓN DE TECNOLOGÍAS SANITARIAS EN ESPAÑA (RedETS) 1995 A 2021



Fuente: RedETS (2022). Listado de informes de evaluación de tecnologías sanitarias. Accesible en <https://redets.sanidad.gob.es/productos/buscarProductos.do?metodo=buscaTipos&tipoid=1>

década 2001-2010, 20,4; y en la última década 40,4 informes anuales. El 30 por 100 de los informes son posteriores a 2015.

Pese al indudable avance en el conocimiento, su aplicación sistemática a la toma de decisiones sigue postergada en nuestro país. En 2021, un grupo de académicos y profesionales de la sanidad y de la economía en España reivindicaba la constitución de una agencia independiente de evaluación de prácticas y políticas sanitarias, al estilo de la National Institute for Health and Care Excellence (NICE) británica, con capacidad legal para recomendar e inducir a los Gobiernos a cumplir las recomendaciones o, alternativamente, a explicar por qué no las cumplen. La iniciativa llegó a la comisión de sanidad del Congreso de los Diputados en noviembre de 2021 con una proposición no de ley (6), pero no pasó de ahí.

Hasta ahora, los pasos dados en este ámbito han sido claramente insuficientes. El Ministerio de Sanidad ha reconocido y asumido la metodología de evaluación de medicamentos del grupo GENESIS, una red de servicios de farmacia hospitalaria, como «oficial» para elaborar los informes de posicionamiento terapéutico (IPT), cuyo objetivo es analizar el valor añadido de los nuevos medicamentos. La ventaja de los informes GENESIS es su disponibilidad inmediata en tiempo real para tomar decisiones de compra, pero se trata de evaluaciones exprés que, a pesar de seguir una metodología común y consensuada, no siguen los estándares internacionales y están lejos de ser evaluaciones completas, lo que requiere de mucho tiempo y recursos. Los IPT sobre nuevos medicamentos se elaboran y publican desde 2013 por el Ministerio de Sanidad, pero hasta 2021 únicamente consideraban eficacia y efectividad,

sin incorporar información ni consideraciones de costes ni de coste-efectividad.

La preocupación específica por los medicamentos de alto coste también ha llevado a la creación de VALTERMED (Sistema de Información de Valor Terapéutico de Medicamentos), un sistema de información para determinar el valor terapéutico en la práctica clínica real de los medicamentos de alto impacto sanitario y económico en el SNS (<https://www.mscbs.gob.es/profesionales/farmacia/valtermed/home.htm>). Es una base de datos creada en noviembre de 2019, de acceso restringido. En diciembre de 2021, con dos años de existencia, había 13 protocolos fármaco-clínicos publicados en la web. Los informes podrían afectar a más de once mil pacientes (7).

Los mayores propulsores del gasto sanitario, particularmente los medicamentos oncológicos y otros de uso hospitalario, así como los tratamientos personalizados (CAR-T) constituyen, por su alto coste e impacto presupuestario, una prioridad en la evaluación. El disponer de criterios sociales de disposición a pagar por un año de vida ajustado por calidad, con posibilidad de diferenciación por grupos de pacientes o patologías, es una de las cuestiones abiertas. En el Reino Unido, que ha sido un pionero en este ámbito, la agencia de evaluación NICE está en proceso de redefinir umbrales por patologías.

En resumen, la arquitectura institucional de la evaluación de tecnologías sanitarias y medicamentos en España es compleja y poco resolutive, no hay cuarta barrera ni se ha declarado umbral de Ratio de Coste Efectividad

Incremental, tampoco existen criterios explícitos de priorización de grupos de pacientes o enfermedades, aunque se reconoce la necesidad de priorizar. Comparándonos con otros países europeos, «aún falta garantizar la independencia de las agencias de evaluación de tecnologías sanitarias, eliminar la influencia de la política económica e industrial de los comités de decisiones sanitarias, fijar precios basados en el valor terapéutico, mejorar la transparencia del proceso y de los resultados de las evaluaciones y, por último, aumentar la participación de los grupos de interés. Al igual que en otros países de su entorno, España tiene que clarificar el papel de la evaluación económica como criterio de decisión» (Epstein y Espín, 2020).

El uso de la evaluación económica (coste-efectividad) ligado a la pandemia: rastreo y vacunas

Hay dos tecnologías de particular interés relacionadas con la pandemia: el rastreo de contactos (que forma parte de una estrategia más general de test, rastreo y aislamiento [TRA]), y las vacunas.

En la medida en que la situación epidemiológica y el conocimiento sobre la enfermedad han ido cambiando, los análisis de coste-efectividad realizados para unas condiciones locales determinadas han ido perdiendo validez externa. A la altura de noviembre de 2020, estimábamos que la ratio beneficio/coste de la estrategia TRA en España podría estar en torno a 7/1: por cada euro gastado se recuperarían 7 en el medio plazo (Vallejo Torres y González López-Valcárcel, 2021). Por tanto, el TRA es una estrategia dominante en el contexto de una

determinada transmisibilidad e incidencia. Ese resultado depende de la incidencia acumulada y de la relación entre incidencia y uso de recursos sanitarios, y entre incidencia y fallecimientos. Dichas relaciones han cambiado sustancialmente con las variantes que han ido surgiendo, especialmente con ómicron. La clara dominancia de la estrategia TRA se debe, fundamentalmente, al hecho de que los principales costes de la COVID son costes indirectos, generados a través de los efectos en el tejido productivo y la economía.

Lo mismo ocurrió con las vacunas. Un estudio publicado en *Science* (Castillo *et al.*, 2021) por equipos de científicos de primera línea estimaba el coste-efectividad de expandir la capacidad de producción de vacunas en el mundo, acortando al máximo los tiempos hasta la inmunidad de grupo y la vuelta a la normalidad económica en los países de renta baja y media. Los resultados señalaban que el valor marginal de una dosis de vacuna en 2021 si se consiguiera expandir en mil millones de dosis la capacidad prevista de producción para ese año podría estar en torno a los 600-1.000 dólares. Un análisis estándar de coste-efectividad recomendaría comprar hasta ese precio, lo que resulta obviamente disparatado.

Otra estimación para la región metropolitana del norte de Cataluña (López *et al.*, 2022) basada en un análisis coste-beneficio concluía que la vacunación tiene una ratio beneficio/coste de 3,4 desde la perspectiva social (1,4 desde la perspectiva del sistema sanitario), con unos beneficios sociales de 116,67 euros por dosis de vacuna. La enorme disparidad entre estos resultados

y los del artículo de Castillo y colaboradores (Castillo *et al.*, 2021) radica en que este último considera también los costes y ahorros en el sistema económico que se producen si las vacunas ganan tiempo en la vuelta a la normalidad. El estudio para Cataluña no considera los ahorros derivados de evitar cierres por una vuelta adelantada a la actividad económica normal, que es la principal fuente de ahorro de costes en el estudio de Castillo. Como ya hemos señalado, los mayores costes inducidos por la COVID son sobre la economía.

Por otra parte, la incertidumbre ha marcado la validez de los análisis. La ratio de coste-efectividad incremental (RCEI) de las vacunas, a nivel mundial, se estimaba tan alta en la primavera de 2021 porque en aquel momento se confiaba o suponía que la vacuna inmunizaba durante un plazo largo, sin necesidad de dosis de refuerzo, y que la inmunidad de grupo se alcanzaría con un (indeterminado) alto porcentaje de vacunados. Unos meses más tarde se sabía que ninguno de esos dos supuestos era realista. En consecuencia, el coste-efectividad de las vacunas será muy inferior al estimado inicialmente.

3. Pero no todo es COVID. Evaluaciones de políticas públicas recientes en sanidad/salud y apuntes para una agenda priorizada

La evaluación en sanidad debe abarcar los tres niveles: macro (las políticas de salud y sanitarias), meso (la gestión y organización de centros y servicios sanitarios) y micro (la gestión clínica). Una parte importante

de la evaluación es la de cambios organizativos. España, por su descentralización sanitaria, cuenta con una enorme cantidad de experiencias que, si se evalúan bien, permitirán la mejora por comparación, el *benchmarking* y el escalado desde experiencias piloto hacia la generalización en todo el territorio. AQUAS aporta un valioso repositorio de experiencias de innovación en gestión y organización del que se puede aprender. Otra iniciativa interesante viene dada por los Premios Profesor Barea de Gestión Sanitaria, con los que desde 2002 la Fundación Signo viene reconociendo a centros y grupos de profesionales que llevan a cabo innovaciones valiosas (y evaluables) de gestión sanitaria (8).

En cuanto a políticas macro, la agenda de evaluación de intervenciones sanitarias de la Autoridad Independiente de Responsabilidad Fiscal española (AIReF) constituye un buen punto de partida. Refleja las prioridades del Gobierno, impelidas por las presiones del gasto sanitario. Los *spending review* (9) I y II han evaluado el gasto en medicamentos de prescripción, medicamentos de uso hospitalario y las inversiones sanitarias en bienes de equipo. Los resultados de dichas evaluaciones sugieren líneas de avance.

Sin pretensión de exhaustividad, se proponen aquí algunos ejemplos de evaluación en el área sanitaria, que apuntan hacia grandes áreas en las que se ha intervenido o es necesario hacerlo. Se han seleccionado casos de macrogestión, de mesogestión y de evaluación de los efectos sobre la salud de determinadas políticas económicas y fiscales.

La *digitalización* de la sanidad puede ser un legado posi-

tivo a largo plazo de la COVID-19. La pandemia ha acelerado un proceso en marcha, en algunos casos de forma brusca. España ha sido el país de la UE con mayor uso de consultas remotas en atención primaria (en marzo de 2021, más del 70 por 100 de la población había tenido alguna, frente al 38 por 100 del conjunto de la UE). Es el país de la Unión con mayor tasa de certificados COVID emitidos. La transformación digital de la sanidad impone nuevos retos para la evaluación económica de tecnologías sanitarias (González López-Valcárcel, Torres y Domínguez, 2021). La propuesta de la NICE, que diferencia entre tecnologías digitales según su individualización en pacientes concretos y el grado de intromisión en su vida, es un buen punto de partida para reformular estándares de evaluación económica. La UE señala la falta de evidencia sobre lo que funciona y lo que no funciona en la sanidad digital, y la necesidad de mirar más allá de la emergencia pandémica para evaluar cómo deberían reajustarse las tecnologías digitales, evitando las consecuencias adversas no intencionadas de su uso, y también cómo incentivar su uso (European Commission, DG SANTE, 2021). La evaluación sistemática de la digitalización en el sector sanitario y de las tecnologías digitales y el diseño y adaptación de la metodología de evaluación son tareas pendientes.

Evaluación de la gestión indirecta de la sanidad, en sus distintas formas y modalidades y de forma más general de las grandes políticas de concertación y colaboración público-privada en la gestión sanitaria. El llamado «modelo Alzira», de contratación externa de toda la gestión del

aseguramiento público y prestaciones de servicios con una entidad externa se puso en marcha en 1999 en la Comunidad Valenciana, primero solo para atención hospitalaria, ampliándose a la atención primaria en 2003. El modelo se extendió por toda la Comunidad, donde en torno al 20 por 100 de la población vivía en áreas de salud cubiertas por este tipo de acuerdos. En 2018, fecha de prescripción del contrato de quince años con la primera prestadora (Ribera Salud) comenzó el proceso de reversión para la comarca de Alzira. A pesar de que el modelo llevaba en funcionamiento casi veinte años y que llegó a convertirse en un escaparate de la modernidad en programas MBA de prestigio, como estudio de caso (10), no hay publicaciones científicas con estudios rigurosos de evaluación comparativa de la calidad y eficiencia de gestión entre los dos modelos. Un informe de 2017 encargado por la concesionaria (López-Casasnovas y del Llano, 2017) comparó los hospitales valencianos de la concesionaria, ajustando por complejidad y casuística, con hospitales catalanes, que formaron el comparador sintético. No se pudo hacer la comparación directa con los hospitales públicos de la Comunidad Valenciana por falta de información. Concluía que no hay diferencias significativas ni en indicadores clínicos ni económicos con los hospitales catalanes comparables. El propio informe advertía que la evaluación presenta limitaciones, fundamentalmente debidas a la falta de datos objetivos para poder establecer un comparador sintético creíble.

También se ha generado evidencia a través de análisis de frontera eficiente, que comparan

la productividad y calidad de hospitales españoles. La mayor parte de estudios sugieren que no es la forma jurídica ni la propiedad (hospital público o privado), sino la capacidad de gestionar recursos, la holgura de gestión, la que se asocia significativamente con mejores resultados (Comendeiro-Maaløe *et al.*, 2019). Una excepción es el estudio de Ortega y colaboradores (Ortega-Díaz *et al.*, 2020), que encuentra mayores probabilidades de estar en la frontera eficiente a los centros en régimen de colaboración público-privada en el período 2002-2012.

Evaluación de los efectos del cambio de copago de medicamentos de prescripción sobre la eficiencia, la equidad y la adherencia de los pacientes. En 2012 cambió el copago de medicamentos de prescripción en España. Los pensionistas, que tenían medicamentos gratuitos, empezaron a pagar el 10 por 100 o más del precio, con límites mensuales según renta. Los activos siguieron pagando, sin límites, el 40 por 100 del precio o subieron al 50 por 100 o 60 por 100 según renta, en función de tres tramos (base liquidable en IRPF hasta 18.000 euros anuales, entre 18.000 y 100.000 euros y más de 100.000 euros). Las evaluaciones publicadas, con distintos datos y métodos de análisis, apuntan a que la elasticidad-precio de la demanda de medicamentos es muy baja y en consecuencia el principal efecto ha sido el de redistribuir la carga financiera entre el Gobierno y las familias, si bien a corto plazo ha habido reducciones importantes del consumo de medicamentos, también de los esenciales como los que se han de tomar después de sufrir un infarto de miocardio (González López-Valcárcel *et al.*, 2017). Los efectos sobre la equidad han

sido mucho mayores que sobre la eficiencia y eliminar la diferenciación, artificial y trasnochada, entre «activos» y «pensionistas» se configura como una vía de avance técnicamente adecuada, pero políticamente costosa.

Evaluación de intervenciones económicas y fiscales sobre la salud de la población. Es el caso de políticas de salud pública basadas en medidas económicas y fiscales, como los impuestos a las bebidas azucaradas. El impuesto catalán de 2017 ha tenido impacto significativo de reducción de ventas (Vall Castelló y López Casasnovas, 2020) con efectos heterogéneos entre tipos de familias (Fichera *et al.*, 2021).

También se ha evaluado el Régimen Específico de Abastecimiento de Canarias (REA) desde la perspectiva de la salud. Las subvenciones a ciertos alimentos importados para garantizar el aprovisionamiento local de las islas han resultado en un daño colateral para la salud. Si se suprimieran las subvenciones al azúcar, mantequilla y queso se podrían evitar más de diez mil casos de obesidad en adultos en Canarias (Rodríguez-Mireles *et al.*, 2021).

Seguridad del paciente e intervenciones sanitarias de bajo valor. Una parte importante de intervenciones sanitarias son de bajo valor terapéutico, es decir, de baja efectividad. Detectar cuál es su prevalencia por centros y zonas de salud es un primer paso para avanzar hacia el paradigma de un sistema sanitario basado en el valor. En España, afortunadamente, se cuenta con información transparente y accesible de la frecuencia de uso de procedimientos de bajo valor (cesáreas

en embarazos de bajo riesgo, tonsilectomía en niños, liberación del túnel carpiano, etc.), gracias al *Atlas de variaciones en la práctica médica* (<https://www.atlasvpm.org/>), que también analiza las variaciones injustificadas en la calidad de los cuidados hospitalarios en España. La efectividad de muchos procedimientos y tratamientos médicos tiene gran variabilidad entre pacientes, de modo que una misma intervención puede ser muy efectiva en uno y muy poco en otro. Esas tecnologías «tipo II», en la terminología de Chandra y Skinner (Chandra y Skinner, 2012), son las que hay que protocolizar con más cuidado, y deben monitorizarse sus indicaciones para evitar sobreutilización que aporta escaso valor. No hay ningún riesgo, en cambio, de utilización inapropiada de un antiviral contra el VIH, que es efectivo en todos los pacientes.

En la evaluación en sanidad es importante cuantificar no solo el coste de lo que se hace, y su relación con el beneficio terapéutico, sino también los costes que se podrían evitar con una práctica correcta. Las hospitalizaciones potencialmente evitables aproximan de alguna manera la calidad del sistema de salud. Hay diferencias en Europa, como ha puesto en evidencia un estudio comparativo de seis países basado en microdatos homogéneos de pacientes y para varias enfermedades crónicas (Thygesen *et al.*, 2015). De forma paralela, es importante mejorar la medición de los resultados (salud): por ejemplo, es mejor hacerlo con la mortalidad prematura y sanitariamente evitable (MIPSE) que con la esperanza de vida. La MIPSE recoge las muertes producidas por causas que el sistema sanitario es capaz de evitar (posponer).

Trabajar en la seguridad del paciente y limitar la medicina defensiva (pedir pruebas o realizar procedimientos innecesarios para evitar posibles demandas) con incentivos y regulación adecuados son dos grandes retos de los sistemas sanitarios en los países desarrollados, que necesitan urgentemente evaluación *ex post* y *ex ante*. La OCDE estima que el daño a los pacientes que infringe el propio sistema sanitario representa en torno al 12,6 por 100 de todo el gasto sanitario, y más de la mitad (8,7 por 100) es prevenible (OCDE, 2020).

4. Evaluaciones pendientes necesarias para tomar decisiones. Apuntes para una agenda de priorización

En su informe 2021 sobre la salud en Europa, la Comisión Europea (European Commission, DG SANTE, 2021) señala tres prioridades: 1) conocer las consecuencias de la pandemia sobre la salud de las poblaciones, la COVID persistente y los efectos colaterales en la atención de otras enfermedades y su abordaje; 2) implementar una transformación digital de la sanidad con una hoja de ruta de evaluación para priorizar y redirigir; y 3) atender a la necesidad de más recursos humanos para la salud, más flexibles y con nuevos perfiles profesionales.

La evaluación de los efectos de la COVID sobre la salud en el medio plazo, como consecuencia de los retrasos en la atención, es un tema imperativo, que conecta directamente con el de la priorización y gestión de listas de espera. Entre los efectos a largo plazo de la COVID, los problemas de salud mental y las respuestas del sistema sanitario para

abordarlos requieren evaluación prospectiva.

Las desigualdades en el uso de servicios sanitarios, que pueden derivar en desigualdades en salud, han de ser evaluadas también. En este sentido, apuntamos la necesidad de estudiar cambios de diseño de los copagos que incorporen criterios de eficiencia, equidad y objetivos presupuestarios.

En cuanto a la fiscalidad del sector sanitario y la financiación de las prestaciones, convendría tener evaluaciones prospectivas de posibles cambios impositivos en el sector como, por ejemplo, la eliminación de las exenciones en IRPF/sociedades por los seguros de salud, y debería revisarse, como ya se ha apuntado, la arquitectura institucional de la evaluación económica de medicamentos y otras tecnologías para decidir cobertura y precio, garantizando la transparencia y la rendición de cuentas de las decisiones tomadas.

El llamado modelo MUFACE (Mutualidad General de Funcionarios Civiles del Estado) es una excepción española perdurable. Casi dos millones de funcionarios civiles del Estado, de justicia y de las fuerzas armadas tienen la posibilidad de elegir aseguradora entre las conveniadas, tanto privadas como el propio Instituto Nacional de la Seguridad Social. Ese modelo se ha puesto como ejemplo a seguir e incluso generalizar, desde posiciones de parte. Es una caja negra en la sanidad española, los datos que permitan comparar las prestaciones, acceso y salud lograda para los asegurados de ambas redes no están disponibles. El sesgo de autoselección lleva a que los *pool* de riesgos de ambos

colectivos no sean comparables, ni por tanto el coste unitario medio por persona. Una evaluación objetiva sería necesaria para fundamentar los debates y tomar decisiones.

Por último, un ensayo de políticas de gran interés para la evaluación es el área sanitaria única de Madrid, y convendría medir sus efectos sobre la eficiencia y la calidad de las prestaciones sanitarias públicas, así como los efectos cruzados sobre la utilización y financiación de los centros propios de la red pública madrileña. Con los datos disponibles del Observatorio Madrileño se podría hacer una primera aproximación.

III. CONCLUSIONES

1. En sanidad, las evaluaciones *ex post* de las intervenciones y políticas públicas son útiles, pero cada vez resulta más necesaria la evaluación prospectiva *ex ante* para facilitar la toma de decisiones fundamentada en datos.

2. Ha habido un enorme avance del conocimiento durante los años de pandemia 2020-2021, en parte gracias al uso de datos masivos compartidos, de acceso abierto y en tiempo real. La inteligencia artificial, que hace uso de esos datos, es complementaria, no sustitutiva, de la inteligencia humana y del conocimiento de los problemas cuyos datos se analizan.

3. Los métodos de evaluación de las políticas públicas se apoyan cada vez más en experimentos sociales y en la disponibilidad de registros y de datos cuasiexperimentales, así como en la posibilidad de explotar grandes masas de datos, datos de

vida real, estructurados o no, y de realizar megaensayos.

4. Los modelos epieconómicos son de enorme utilidad y pueden ayudar a la toma de decisiones, pero sirven para lo que sirven: explicar el pasado, predecir a corto plazo, retratar el presente (*nowcasting*) y avanzar hipótesis plausibles sobre el futuro (con enorme incertidumbre). Igual que la gestión de la pandemia, los modelos deben ser evaluados retrospectivamente, según su capacidad predictiva.

5. El mayor reto de los países durante estos dos años de pandemia ha consistido en encontrar soluciones para compatibilizar la pandemia con la vida de las personas y mantener el tejido productivo y el crecimiento económico a medio plazo. Nunca la sanidad y la economía habían sido tan interdependientes, nunca las políticas sanitarias habían tenido tanto de políticas económicas.

6. A grandes rasgos, la gestión de la pandemia abarca: 1) las medidas para limitar los contagios; 2) la disposición y organización de recursos y servicios sanitarios y sociosanitarios para prevenir y tratar la enfermedad; 3) las medidas de carácter económico para minimizar los efectos de la pandemia sobre el tejido productivo y sobre las economías familiares.

7. Hoy por hoy no hay estudios rigurosos que permitan establecer la dominancia de la estrategia de supresión sobre la de contención de la COVID, o viceversa.

8. Los métodos de investigación cualitativa deben reivindicar su utilidad, particularmente en la evaluación de la respuesta a

la pandemia. Las denominadas «after-action reviews» (AAR) o «revisiones posactuación» pueden servir para identificar las lecciones aprendidas y diseñar respuestas futuras más eficaces de forma constructiva, sin necesidad de identificar culpables.

9. La institucionalización de la evaluación de tecnologías sanitarias en España ha avanzado en los últimos años, pero todavía no está completada. La arquitectura institucional de la evaluación de tecnologías sanitarias y medicamentos es compleja y poco resolutive, y no existen criterios explícitos de priorización de grupos de pacientes o enfermedades, aunque se reconoce la necesidad de priorizar y se acepta el criterio de coste-efectividad.

10. Durante la pandemia quedó en suspenso el coste-efectividad como criterio de decisión, tanto por imperativo moral como por la apabullante evidencia de que prácticamente todo gasto sanitario en prevención y tratamientos COVID es dominante, teniendo en cuenta los costes (y ahorros de costes) indirectos por la vía del impacto en la economía.

11. La evaluación en sanidad debe abarcar los tres niveles: macro (las políticas de salud y sanitarias), meso (la gestión y organización de centros y servicios sanitarios) y micro (la gestión clínica). Una parte importante de la evaluación es la de cambios organizativos. España, por su descentralización sanitaria, cuenta con una enorme cantidad de experiencias que, si se evalúan bien, permitirán la mejora por comparación, el *benchmarking* y el escalado desde experiencias piloto hacia la generalización en todo el territorio.

12. En cuanto a políticas macro, la agenda de evaluación de intervenciones sanitarias de la AIReF constituye un buen punto de partida. Los resultados de dichas evaluaciones sugieren líneas de avance a considerar.

13. Hay áreas de especial relevancia para la evaluación: la digitalización de la sanidad, la gestión indirecta de la prestación sanitaria, las tecnologías de alto coste unitario (medicina personalizada entre otras). Hay que evaluar las intervenciones económicas y fiscales sobre la salud de la población, y analizar la actividad sanitaria con perspectiva de sistema basado en el valor. Los mayores propulsores del gasto sanitario, particularmente los medicamentos oncológicos y otros de uso hospitalario, así como los tratamientos personalizados (CAR-T) constituyen, por su alto coste e impacto presupuestario, una prioridad en la evaluación. El disponer de criterios sociales de disposición a pagar por un año de vida ajustado por calidad, con posibilidad de diferenciación por grupos de pacientes o patologías, es una de las cuestiones abiertas.

14. Finalmente, el modelo MUFACE requiere una evaluación urgente. En la agenda inmediata deberían entrar también las políticas sanitarias y de salud para hacer frente a las consecuencias a largo plazo de la COVID sobre la salud (también mental) de la población y la organización de los recursos necesarios para abordar el problema, así como su financiación. Las desigualdades en el uso de servicios sanitarios, que pueden derivar en desigualdades en salud, han de ser evaluadas sin demora.

NOTAS

(*) Agradezco a ROSA URBANOS los comentarios y aportaciones al texto.

(1) Ver GONZÁLEZ LÓPEZ-VALCÁRCEL *et al.* (2020). <https://theconversation.com/covid-19-pandemia-de-modelos-matematicos-136212>

(2) <https://covid19forecasthub.eu/>

(3) Presentación de la Agencia de Evaluación de Tecnologías Sanitarias del ISCIII. <https://www.isciii.es/QuienesSomos/CentrosPropios/AETS/Paginas/default.aspx>

(4) https://aguas.gencat.cat/ca/sobre_aguas/presentacio/commemoracio-25-anys-agencia

(5) Una base de datos en Excel con el listado actualizado de los informes está disponible en <https://redets.sanidad.gob.es/productos/buscarProductos.do?metodo=buscaTipos&tipod=1>

(6) <https://www.diariofarmacom/2021/11/21/el-pdecat-lleva-al-congreso-el-debate-sobre-impulsar-el-hispanice>

(7) <https://www.redaccionmedica.com/secciones/industria/los-protocolos-de-remdesivir-y-dupilumab-copan-las-visitas-de-valter-med-4905>

(8) <https://www.fundacionsigno.com/premios.php>

(9) <https://www.airef.es/es/spending-review/>

(10) <https://www.economista.es/valencia-na/noticias/8673190/10/17/El-modelo-de-gestion-sanitaria-de-Ribera-Salud-caso-de-estudio-en-Harvard.html>

BIBLIOGRAFÍA

ANGRIST, J. D. y PISCHKE, J.-S. (2010). The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics. *Journal of Economic Perspectives*, 24(2), pp. 3-30.

BALCAN, D., GONÇALVES, B., HU, H., RAMASCO, J. J., COLIZZA, V. y VESPIGNANI, A. (2010). Modeling the spatial spread of infectious diseases: The Global Epidemic and Mobility Computational Model. *Journal of Computational Science*, 1(3), pp. 132-145.

BRADLEY, V. C., KURIWAKI, S., ISAKOV, M., SEJIDINOVIC, D., MENG, X.-L. y FLAXMAN, S. (2021). Unrepresentative big surveys significantly overestimated US vaccine uptake. *Nature*, 600(7890), pp. 695-700.

CASADO, J. M. y DEL PINO, E. (2021). Evolución, situación actual y retos de la evaluación de políticas públicas en las Administraciones españolas (2000-2021). *Cuadernos Económicos de ICE*, 102, pp. 13-38.

CASTILLO, J. C., AHUJA, A., ATHEY, S., BAKER, A., BUDISH, E., CHIPTY, T., GLENNERSTER, R., KOMINERS, S. D., KREMER, M., LARSON, G., LEE, J., PRENDERGAST, C., SNYDER, C. M., TABARROK, A., TAN, B. J. y WIĘCEK, W. (2021). Market design to accelerate COVID-19 vaccine supply. *Science*, 371(6534), pp. 1107-1109.

CATALÀ, M., PINO, D., MARCHENA, M., PALACIOS, P., URDIALES, T., CARDONA, P.-J., ALONSO, S., LÓPEZ-CODINA, D., PRATS, C. y ÀLVAREZ-LACALLE, E. (2021). Robust estimation of diagnostic rate and real incidence of COVID-19 for European policymakers. *PLOS ONE*, 16(1), e0243701.

CHANDRA, A. y SKINNER, J. (2012). Technology Growth and Expenditure Growth in Health Care. *Journal of Economic Literature*, 50(3), pp. 645-680.

CHINAZZI, M., DAVIS, J. T., AJELLI, M., GIOANNINI, C., LITVINOVA, M., MERLER, S., PIONTTI, A. P., MU, K., ROSSI, L., SUN, K., VIBOUD, C., XIONG, X., YU, H., HALLORAN, M. E., LONGINI JR, I. M. y VESPIGNANI, A. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*, 368(6489), pp. 395-400.

COMENDEIRO-MAALØE, M., RIDAO-LÓPEZ, M., GORGEMANS, S. y BERNAL-DELGADO, E. (2019). A comparative performance analysis of a renowned public private partnership for health care provision in Spain between 2003 and 2015. *Health Policy*, 123(4), pp. 412-418.

ECDC (2021, septiembre 2022). *Protocol for a focused after-action review on evidence-based decision-making for selected COVID-19 response measures*. European Centre for Disease Prevention and Control. <https://www.ecdc.europa.eu/en/publications-data/protocol-focused-after-action-review-evidence-based-decision-making-covid-19-response>

EPSTEIN, D. y ESPÍN, J. (2020). Evaluation of new medicines in Spain and comparison with other European

<p>countries. <i>Gaceta Sanitaria</i>, 34, pp. 133-140.</p> <p>EUROPEAN COMMISSION, DG SANTE (2021). <i>The State of Health in the EU: Companion Report 2021</i>.</p> <p>FICHERA, E., MORA, T., LÓPEZ-VALCÁRCEL, B. G. y ROCHE, D. (2021). How do consumers respond to «sin taxes»? New evidence from a tax on sugary drinks. <i>Social Science & Medicine</i>, 274, 113799.</p> <p>FRIEDMAN, J., LIU, P., TROEGER, C. E., CARTER, A., REINER, R. C., BARBER, R. M., COLLINS, J., LIM, S. S., PIGOTT, D. M., VOS, T., HAY, S. I., MURRAY, C. J. L. y GAKIDOU, E. (2021). Predictive performance of international COVID-19 mortality forecasting models. <i>Nature Communications</i>, 12(1), p. 2609.</p> <p>GARCÍA-BASTEIRO, A. L., LEGIDO-QUIGLEY, H., ÁLVAREZ-DARDET, C., ARENAS, A., BENGOA, R., BORRELL, C., VAL, M. D., FRANCO, M., GEA-SÁNCHEZ, M., GESTAL, J., GONZÁLEZ LÓPEZ-VALCÁRCEL, B. G., HERNÁNDEZ-AGUADO, I., MARCH, J. C., MARTÍN-MORENO, J., MENÉNDEZ, C., MINUÉ, S., MUNTANER, C., PORTA, M., PRIETO-ALHAMBRA, D. y LEGIDO-QUIGLEY, H. (2020). Evaluation of the COVID-19 response in Spain: Principles and requirements. <i>The Lancet Public Health</i>, 5(11), e575.</p> <p>GARCÍA-SEMPERE, A., ORRICO-SÁNCHEZ, A., MUÑOZ-QUILES, C., HURTADO, I., PEIRÓ, S., SANFÉLIX-GIMENO, G. y Díez-DOMINGO, J. (2020). Data Resource Profile: The Valencia Health System Integrated Database (VID). <i>International Journal of Epidemiology</i>, 49(3), 740-741e.</p> <p>GERTLER, P. J., MARTÍNEZ, S., PREMAM, P., RAWLINGS, L. B. y VERMEERSCH, C. M. J. (2016). <i>Impact Evaluation in Practice</i>. Second Edition. World Bank Publications.</p> <p>GONZÁLEZ LÓPEZ-VALCÁRCEL, B. (2019). Aplicaciones prácticas del método SROI. En: <i>El método SROI en la evaluación económica de intervenciones sanitarias</i>, pp. 69-81. Madrid. Weber. http://weber.org.es/wp-content/uploads/2019/06/Libro-SROI-digital-1.pdf</p> <p>GONZÁLEZ LÓPEZ-VALCÁRCEL, B. G., LIBRERO, J., GARCÍA-SEMPERE, A., PEÑA, L. M., BAUER, S., PUIG-JUNOY, J., OLIVA, J., PEIRÓ, S., y SANFÉLIX-GIMENO, G. (2017).</p>	<p>Effect of cost sharing on adherence to evidence-based medications in patients with acute coronary syndrome. <i>Heart</i>, 103(14), pp. 1082-1088.</p> <p>GONZÁLEZ LÓPEZ-VALCÁRCEL, B. G., TORRES, L. V. y DOMÍNGUEZ, J. P. (2021). La transformación digital del sector sanitario. <i>Economistas</i>, 174, pp. 75-84.</p> <p>HALE, T., ANGRIST, N., GOLDSZMIDT, R., KIRA, B., PETHERICK, A., PHILLIPS, T., WEBSTER, S., CAMERON-BLAKE, E., HALLAS, L., MAJUMDAR, S. y TATLOW, H. (2021). A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). <i>Nature Human Behaviour</i>, 5(4), pp. 529-538.</p> <p>HECKMAN, J. J. (2001). Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture. <i>Journal of Political Economy</i>, 109(4), pp. 673-748.</p> <p>HOUSE OF COMMONS (2021). <i>Coronavirus: Lessons learned to date – Health and Social Care, and Science and Technology Committees – House of Commons</i>. https://publications.parliament.uk/pa/cm5802/cmselect/cmstech/92/9203.htm</p> <p>LINKA, K., PEIRLINCK, M., SAHLI COSTABAL, F. y KUHL, E. (2020). Outbreak dynamics of COVID-19 in Europe and the effect of travel restrictions. <i>Computer Methods in Biomechanics and Biomedical Engineering</i>, 23(11), pp. 710-717.</p> <p>LÓPEZ, F., CATALÀ, M., PRATS, C., ESTRADA, O., OLIVA, I., PRAT, N., ISNARD, M., VALLÈS, R., VILAR, M., CLOTET, B., ARGIMON, J. M., ARAN, A. y ARA, J. (2022). A Cost&Benefit Analysis of COVID-19 Vaccination in Catalonia. <i>Vaccines</i>, 10(1), p. 59.</p> <p>LÓPEZ-CASASNOVAS, G. y DEL LLANO, J. (2017). <i>Colaboración público-privada en sanidad: El modelo Alzira</i>. Fundacion Gaspar Casal. https://www.riberasalud.com/wp-content/uploads/2019/03/Colaboracion-Publico-Privada-en-Sanidad-El-Modelo-Alzira.pdf</p> <p>MARTÍN-MORENO, J. M., ARENAS, A., BENGOA, R., BORRELL, C., FRANCO, M., GARCÍA-BASTEIRO, A. L., GESTAL, J., GONZÁLEZ</p>	<p>LÓPEZ-VALCÁRCEL, B., HERNÁNDEZ-AGUADO, I., LEGIDO-QUIGLEY, H., MARCH, J. C., MINUÉ, S., MUNTANER, C. y VIVES-CASES, C. (2020). Reflexiones sobre cómo evaluar y mejorar la respuesta a la pandemia de COVID-19. <i>Gaceta Sanitaria</i>, 36(1), pp. 32-36.</p> <p>McKEE, M. y STUCKLER, D. (2015). Reflective Practice: How the World Bank Explored Its Own Biases? <i>International Journal of Health Policy and Management</i>, 5(2), pp. 79-82.</p> <p>MILKMAN, K. L., GROMET, D., HO, H., KAY, J. S., LEE, T. W., PANDILOSKI, P., PARK, Y., RAI, A., BAZERMAN, M., BESHEARS, J., BONACORSI, L., CAMERER, C., CHANG, E., CHAPMAN, G., CIALDINI, R., DAI, H., ESKREIS-WINKLER, L., FISHBACH, A., GROSS, J. J. y DUCKWORTH, A. L. (2021). Megastudies improve the impact of applied behavioural science. <i>Nature</i>, 600(7889), pp. 478-483.</p> <p>OCDE (2020a). <i>The Economics of patient safety. From Analysis to Action</i>. https://www.oecd.org/health/health-systems/Economics-of-Patient-Safety-October-2020.pdf</p> <p>OCDE (2020b). <i>Open, Useful and Re-usable data (OURdata) Index: 2019</i>. https://ictlogy.net/bibliography/reports/projects.php?idp=4303</p> <p>OLIU-BARTON, M. y PRADELSKI, B. S. R. (2021). Green zoning: An effective policy tool to tackle the COVID-19 pandemic. <i>Health Policy</i>, 125(8), pp. 981-986.</p> <p>OREA, L. y ÁLVAREZ, I. C. (2022). How effective has the Spanish lockdown been to battle COVID-19? A spatial analysis of the coronavirus propagation across provinces. <i>Health Economics</i>, 31(1), pp. 154-173.</p> <p>ORTEGA-DÍAZ, M. I., OCAÑA-RIOLA, R., PÉREZ-ROMERO, C. y MARTÍN-MARTÍN, J. J. (2020). Multilevel Analysis of the Relationship between Ownership Structure and Technical Efficiency Frontier in the Spanish National Health System Hospitals. <i>International Journal of Environmental Research and Public Health</i>, 17(16), p. 5905.</p> <p>POLETTO, C., SCARPINO, S. V. y VOLZ, E. M. (2020). Applications of predictive modelling early in the COVID-19</p>
---	---	---

<p>epidemic. <i>The Lancet Digital Health</i>, 2(10), pp. e498-e499.</p> <p>PÓO, J. M. R. y GALIANO, A. S. (2021). Estadística oficial y evaluación de políticas públicas: Oportunidades y retos. <i>Cuadernos Económicos de ICE</i>, 102, pp. 51-67.</p> <p>RODRÍGUEZ-MIRELES, S., GONZÁLEZ LÓPEZ-VALCÁRCEL, B. y SERRA-MAJEM, L. (2021). When Industrial Policies Conflict With Population Health: Potential Impact of Removing Food Subsidies on Obesity Rates. <i>Value in Health</i>, 24(3), pp. 336-343.</p> <p>ROSSMAN, H. y SEGAL, E. (2022). Nowcasting the spread of SARS-CoV-2. <i>Nature Microbiology</i>, 7(1), pp. 16-17.</p> <p>ROYER, H. (2021). Benefits of megastudies for testing behavioural interventions. <i>Nature</i>, 600(7889), pp. 392-393.</p> <p>SIRLEAF, E. J. y CLARK, H. (2021). Report of the Independent Panel for Pandemic Preparedness and Response: Making</p>	<p>COVID-19 the last pandemic. <i>The Lancet</i>, 398(10295), pp. 101-103.</p> <p>THYGESEN, L. C., CHRISTIANSEN, T., GARCÍA-ARMESTO, S., ANGULO-PUEYO, E., MARTÍNEZ-LIZAGA, N., BERNAL-DELGADO, E. y ECHO CONSORTIUM. (2015). Potentially avoidable hospitalizations in five European countries in 2009 and time trends from 2002 to 2009 based on administrative data. <i>European Journal of Public Health</i>, 25(suppl1), pp. 35-43.</p> <p>UBALDI, B. (2013). <i>Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives</i>. OECD.</p> <p>VALL CASTELLÓ, J. y LÓPEZ CASASNOVAS, G. (2020). Impact of SSB taxes on sales. <i>Economics & Human Biology</i>, 36, 100821.</p> <p>VALLEJO TORRES, L. y GONZÁLEZ LÓPEZ-VALCÁRCEL, B. (2021). The costs of COVID-19 and the cost-effectiveness of testing. <i>Applied Economic Analysis</i>, 29(85), pp. 77-89.</p>	<p>WONG, M. C., HUANG, J., TEOH, J. y WONG, S. H. (2020). Evaluation on different non-pharmaceutical interventions during COVID-19 pandemic: An analysis of 139 countries. <i>The Journal of Infection</i>, 81(3), e70-e71.</p> <p>WORLD HEALTH ORGANIZATION (2019). <i>Guidance for after action review (AAR)</i> (WHO/WHE/CPI/2019.4). World Health Organization. https://apps.who.int/iris/handle/10665/311537</p> <p>WU, J. T., LEUNG, K., LAM, T. T. Y., NI, M. Y., WONG, C. K. H., PEIRIS, J. S. M. y LEUNG, G. M. (2021). Nowcasting epidemics of novel pathogens: Lessons from COVID-19. <i>Nature Medicine</i>, 27(3), pp. 388-395.</p> <p>ZALA, D., MOSWEU, I., CRITCHLOW, S., ROMEO, R. y McCRONE, P. (2020). Costing the COVID-19 Pandemic: An Exploratory Economic Evaluation of Hypothetical Suppression Policy in the United Kingdom. <i>Value in Health</i>, 23(11), pp. 1432-1437.</p>
---	--	---