

# OPORTUNIDADES Y RETOS DE LOS MACRODATOS (Big Data)

En la toma de decisiones  
sanitarias

Jesús Barea Mendoza  
Joan Rovira Forns  
Luis Quecedo Gutiérrez  
Jordi Gol Montserrat  
Juan del Llano Señarís



# OPORTUNIDADES Y RETOS DE LOS MACRODATOS (Big Data)

En la toma de decisiones  
sanitarias

Jesús Barea Mendoza  
Joan Rovira Forns  
Luis Quecedo Gutiérrez  
Jordi Gol Montserrat  
Juan del Llano Señarís



Edición: Fundación Gaspar Casal

Coordinación: Gema Pi Corrales

© Fundación Gaspar Casal

ISBN: 978-84-09-07174-6

Depósito Legal: M-39034-2018

Maquetación, impresión y encuadernación:

Impresión A2

Crta. M-506 – Km. 25,800 – Vía de Servicio

28320 - Pinto (Madrid)

## NOTA DE LOS AUTORES

---

En este documento hemos optado por emplear el término *macrodatos* como traducción al español de *Big Data*, tal y como recomiendan algunas voces autorizadas en el terreno de la adaptación de préstamos lingüísticos:

– Fundéu:

*“Macrodatos e inteligencia de datos son alternativas en español a la voz inglesa big data, que se emplea en el sector de las tecnologías de la información y de la comunicación para aludir a un conjunto de datos que, por su volumen y variedad y por la velocidad a la que necesitan ser procesados, supera las capacidades de los sistemas informáticos habituales.*

*El término macrodatos es una alternativa válida pues aporta, como big, el significado de ‘grande’; es una solución breve y no tiene, como ocurriría con megadatos, un posible riesgo de confusión con el término mega, muy frecuente también en los mismos ámbitos.”*



# ÍNDICE

---

<b>Presentación</b> <i>por Salvador Peiró Moreno</i> .....	7
<b>Prefacio</b> <i>por John P.A. Ioannidis</i> .....	13
<b>Preámbulo</b> <i>por Enrique Bernal-Delgado</i> .....	17
<b>Prólogo</b> <i>por Ignacio Hernández Medrano</i> .....	19
<b>Executive Summary</b> .....	21
<b>1. Conceptualización</b> .....	23
<b>2. Revisión sistemática de la literatura</b> .....	35
<b>3. Informe cualitativo: análisis del Grupo de discusión</b> .....	53
<b>4. Hoja de ruta</b> .....	109
<b>Bibliografía</b> .....	123
<b>Autores</b> .....	129



## PRESENTACIÓN

---

### Big data: ruido, esperanza y realidades.

El término *big data* es una de esas expresiones polisémicas que se usan tanto en forma de hipérboles (exagerar las cualidades) como de metonimias (designar una cosa con el nombre de otra), incluyendo sinécdoques (tomar el todo por la parte, o la parte por el todo) y antonomasias (usar el nombre apelativo por el propio, o el propio por el apelativo). Tanto tropo retórico hace a veces difícil saber de que se está hablando sin situarlo en algún contexto. Y más difícil dibujar sus potenciales y sus limitaciones, sus utilidades y sus traspiés.

A veces el término *big data* se emplea para referirse a los datos en si mismos. A más o menos grandes volúmenes de datos –más o menos desestructurados– que por sus características requieren métodos de análisis específicos, diferentes a los métodos estadísticos convencionales. En estos casos, el discurso *big data* suele venir ataviado con mucha lentejuela de gigabytes, terabytes y petabytes, adornos en forma de V (*volume, variety, velocity, veracity, value*) y la promesa –más o menos inmediata– de que nuestros datos acabarán por revelar los arcanos de nuestro futuro individual y las soluciones para construirlo a la medida de nuestros *big dreams*.

Otras veces, *big data* se refiere a los métodos de análisis de datos (*advanced analytics: predictive analytics, data mining, big data analytics, machine learning, neural networks, ...*) que se emplean fundamentalmente para la identificación de patrones o el desarrollo de modelos predictivos y, también, a los desarrollos de inteligencia artificial (*machine learning, deep learning, natural language processing, cognitive computing, ...*), empleados para el reconocimiento de patrones, el apoyo al diagnóstico y a la toma de decisiones (por ejemplo, en el diagnóstico anatómo-patológico del cáncer, el análisis del fondo de ojo o de diversas pruebas de imagen). También se emplea el término *big data* cuando se trata de análisis “ómicos”, de datos de imagen o de textos en lenguaje natural. En general, este tipo de usos busca identificar patrones, desarrollar nuevos biomarcadores, construir modelos predictivos, apoyar el diagnóstico o la toma de decisiones o adaptar diferentes trayectorias de tratamiento a diferentes pacientes en función de los resultados clínicos históricos en pacientes con esos patrones. Aunque raramente usan aproximaciones causales (ni falta que hace en estas aplicaciones) prometen el desarrollo de una medicina “personalizada” que –tras considerar todos nuestros datos,

incluyendo los genéticos– permitirá a los clínicos (o a los ingenieros biomédicos) prescribirnos un tratamiento específico, diferencial para cada persona, que incrementará los éxitos y evitará los eventos adversos.

En el mundo de la gestión sanitaria no es infrecuente referirse con el término *big data* a la explotación de los datos estructurados de los sistemas de información sanitaria (incluyendo la historia clínica electrónica) para el desarrollo de indicadores de productividad, calidad o eficiencia. Estos sistemas de *business intelligence* pueden utilizar (en algún caso) algún modelo basado en el análisis avanzado, pero casi siempre utilizan análisis convencionales (incluso obviando algunos aspectos elementales del análisis estadístico como los intervalos de confianza) para construir cuadros de mando (con más o menos indicadores pero de construcción e interpretación sencilla) que permiten monitorizar el comportamiento de los centros y servicios sanitarios en sus dimensiones más relevantes y apoyar la toma de decisiones organizacionales. La esperanza aquí es que este *big data* transformará las organizaciones sanitarias “tontas” en organizaciones “inteligentes”.

En evaluación de tecnologías sanitarias también se utiliza el término *big data* para referirse a los estudios observacionales retrospectivos realizados con grandes bases de datos procedentes de historias clínicas y otros sistemas de información sanitaria. A veces se usa como sinónimo de *Real World Data* (RWD), aunque este concepto incluye los estudios con registros, los ensayos clínicos pragmáticos y otros diseños que no se consideran *big data*. En estos trabajos –que intentan determinar la efectividad, seguridad o la eficiencia de medicamentos, tecnologías o servicios sanitarios– el establecimiento de relaciones causales entre la tecnología evaluada y los resultados clínicos de los pacientes es el objetivo esencial, y el control de los sesgos de selección, de información y del *confounding* son los elementos centrales del diseño y el análisis estadístico. En este campo hay numerosas promesas: que la industria farmacéutica y electromédica podrá reducir sus costes de investigación (en tiempo y dinero) al sustituir los ensayos clínicos por estudios RWD, que las agencias de medicamentos podrán establecer y monitorizar la seguridad de los fármacos basándose en estos datos, que los sistemas de salud podrán determinar de forma continua el valor (y el precio que deberían pagar) de medicamentos, productos y dispositivos, determinando la efectividad de los tratamientos en su uso real, y otras similares. Los RWD también se emplean en estudios de calidad asistencial, en la evaluación del rendimiento y efectividad de las organizaciones sanitarias, con posibilidades y limitaciones similares a cuando se evalúan tecnologías.

Todas estas promesas vienen, por el momento, acompañadas de logros discretos (aunque algunos sean importantes, especialmente en el campo del reconocimiento

de patrones) y de lenitivos imprecando menos excitación, menos “grandes” discursos comerciales, y más esfuerzo en la validación y evaluación de los desarrollos *big data* con los mismos criterios que se emplean para cualquier otra tecnología. Paradójicamente, buena parte de las críticas comparten confusión con lo criticado y no diferencian entre unos y otros tipos de datos o entre unos y otros usos, reclamando que todos los estudios *big data* cumplan los requerimientos de los estudios causales, o una mayor “calidad” a los datos de vida real (que, por definición, no pueden tener la calidad de los obtenidos en estudios prospectivos con definiciones operativas de cada variable).

Lo cierto es que la mayor parte de los problemas y limitaciones del *big data* son específicos de cada utilización y los investigadores que usan estas fuentes deben emplear abordajes específicos para cada objetivo. Por ejemplo, los modelos predictivos (se haya utilizado o no el *big data* en su construcción) no requieren establecer relaciones causales sino predecir adecuadamente (aportar información pronóstica útil sobre la previamente disponible a la modelización). Se les puede exigir una apropiada capacidad predictiva, una buena calibración, cierta parsimonia y que mantengan estas propiedades en centros diferentes. Si en la validación responden a estas características serán útiles en la práctica.

Algo parecido sucede en los desarrollos de apoyo al diagnóstico o a la toma de decisiones. Importa el rendimiento diagnóstico del “test” desarrollado (sensibilidad, especificidad, valores predictivos, razones de verosimilitud, estadístico C), especialmente cuando se valida en bases de datos diferentes a la que se utilizaron para su construcción, pero no tanto otras cosas. Por ejemplo, cuando se emplean sistemas de inteligencia artificial para el diagnóstico de cáncer en imágenes de alta resolución procedentes muestras de tejidos, lo relevante es que el rendimiento diagnóstico no sea inferior al de la revisión manual por patólogos. Cuando se utilizan biomarcadores genéticos o de imagen desarrollados a partir de procesos de análisis avanzado o redes neurales, lo relevante es que mejoren la información previamente disponible para el tratamiento del paciente. Si cumplen esta función, cualquier otro problema es lateral.

Del mismo modo, cuando se emplean datos de vida real para evaluar la efectividad de los tratamientos hay que asumir que pacientes que recibieron medicamentos distintos es porque eran distintos, que los profesionales sanitarios sólo registraron aquello que les pareció relevante en cada momento (usando apreciaciones subjetivas diferentes para cada profesional) y, en resumen, que la calidad de los datos no es la de los estudios prospectivos. Pero esos datos tienen valor. En los análisis RWD los datos *missing* son informativos. Esta situación obliga a los investigadores

a desarrollar diseños (restricción a pacientes con determinadas características, a casos incidentes, control de la selección de tratamientos por el médico, etc.) intentando controlar los diferentes sesgos y aproximar un contrafactual aceptable (*target trial*). Son elementos exigibles y la validez del estudio no dependerá tanto de la calidad general de los datos como de cómo se haya lidiado con esos problemas.

Por supuesto una organización sin liderazgo ni estrategias y extremadamente resistente al cambio organizativo no se volverá “inteligente” por más *big data* que le echen encima. Sin embargo, el *big data* ayudará a las organizaciones inteligentes a lograr sus objetivos. Por supuesto, si usamos modelos predictivos atolondrados para hacer cribados poblacionales imprudentes producirémos más daño que beneficio (como ya hacemos con algunos cribados que no se han desarrollado a partir de análisis *big data*). Pero en algunas poblaciones seleccionadas los desarrollos basados en *big data* pueden suponer avances. Por supuesto, si incorporamos a la práctica clínica modelos predictivos o instrumentos de ayuda a la toma de decisiones no validados en nuestro contexto, muchos pacientes tendrán problemas. Pero esto no es óbice para la incorporación de desarrollos validados y calibrados para nuestras bases de datos. Por supuesto, si autorizamos tecnologías basándonos en estudios retrospectivos con datos de mundo real, nos llevaremos más de un susto. Pero esto no quiere decir que no empleemos el RWD para incrementar la información que aportan los ensayos o cuando no es fácil disponer de ensayos. Recordemos que las sociedades deben tomar decisiones con la información disponible, y que disponer de evidencia “suficiente” de riesgos y beneficios no implica necesariamente eliminar toda incertidumbre.

Mientras España es uno de los países que más ha invertido en informatizar sus sistemas sanitarios, generando ingentes volúmenes de datos sanitarios, es también uno de los que menos esfuerzo ha hecho por explotarlos, por transformar estos datos en información útil para pacientes y organizaciones. El *big data* y el *real world data*, en todos sus significados y significantes, tiene un enorme potencial para aportar valor a la atención de los pacientes, a la gestión de las organizaciones, a la toma de decisiones regulatorias y a la competitividad de nuestras empresas. También tiene un gran potencial para las equivocaciones, el despilfarro y los errores, incluyendo los que causan daño a los pacientes. Más allá del ruido y de las modas, hace tiempo que es tiempo de que el SNS invierta en inteligencia para transformar sus grandes volúmenes de datos en información para la toma de decisiones. Hace tiempo que es tiempo para que el SNS desarrolle estrategias que estimulen la investigación (de calidad) con sus datos y la incorporación de los desarrollos a la mejora de su calidad y eficiencia. Hace tiempo que deberíamos estar hablando de las realidades del *big data* antes que de sus potencialidades. De los aciertos (y

los errores) que cometemos en su uso, que de las posibilidades teóricas de sus utilidades futuras.

**Salvador Peiró Moreno**

Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunidad Valenciana (FISABIO), Valencia



## PREFACIO

---

### Big Data: Forward and Forewarning

Big Data constitute an evolving landscape. What is Big Data today may be Relatively Small Data in a few years. Tools for collecting, analyzing and making inferences from Big Data (whatever these are) are also evolving. Opportunities abound to improve literacy and numeracy in ways that were difficult to imagine even a few years ago. At the same time, the risk of drowning in useless or even harmful Big Data is increasing. The idea that more Big Data will be good for us is not necessarily correct. Big Data may be helpful, if they are accurate enough, if they are relevant, if they can be operationalized in ways that they can be employed with consistency and with sufficient reproducibility in diverse settings, and if we use rigorous analytical tools to assess them. Furthermore, careful inferential conclusions need to be made and this is rarely easy when confusing black boxes of Big Data are involved.

There is pressure to change everything in new waves of Big Data and its corollaries of Precision Medicine and Precision Health: change methods, statistics, even the way that we think about what are the health and disease problems that we need to solve. Some of the innovation spirit is welcome. However, not everything needs to change just for the sake of changing. We need to ask people, not data, to learn what matters to people. We need to cure people, not cure Big Data. Moreover, Big Data run the risk of requiring enormous amounts of investment in a vain effort to remain current on what in the latest infrastructure, the latest add-on, the latest peculiar data science popular analytical tool – without offering real value to real people.

This quest for Big Data innovation at all cost is unlikely to lead to useful outcome for either individual or population health. For many tools, we already have very good options and very good methods to apply, but we don't use them in ways that would be efficient. Most of the biomedical literature is wrong and useless, not because we do not have Big-enough Data – even though small samples have indeed been problematic in many areas of research. The effective sample size of a study is a combination of both its numbers and its quality. Quality is usually not thoroughly protected and lots of Big Data may be of worse quality (even horrible quality) compared with the smaller datasets that were assembled in the past.

Coordinated efforts are needed to ensure that we get the best value of all these new opportunities. We need to combine the best data, with the best infrastructure, the best analytical tools and fit-for-purpose projects. Fragmentation of efforts in this space is already causing enhanced noise and geometrically growing confusion. The current effort is a genuine attempt to put some order in this chaos. Not all recommendations will be easy to implement, and plans may need to be revised downstream. However, a plan is necessary to have in place both at a national and international level.

Societies that know how to use Big Data may have an advantage in the future. Societies that just use Big Data without careful pre-thinking will probably have a disadvantage, since they will drown in useless information, while they could have focused on select items that matter. For individual patients and people, this challenge is even more acute. Big Data can allow us to dream about a new Digital world of Precision Health, or a nightmare thereof. The literature of screening with its massive past failures offers only a glimpse of the disaster that can ensue, for example, if Big Data lead to massive useless screening on a daily, hourly, or continuous real-time basis. Individuals who become captives of their own information are unlikely to become the super-healthy super-humans of the future.

Big Data is here to stay. Let's make sure that we know how to deal with this beast and civilize it.

### Macrodatos: avanzando y advirtiendo

Los Macrodatos son un paisaje en estado de evolución. Lo que se consideran Macrodatos hoy, quizás se consideran datos relativamente pequeños en unos pocos años. Las herramientas para recopilar, analizar y hacer inferencias a partir de Macrodatos (sean cuales sean) también están evolucionando. Abundan oportunidades para mejorar la alfabetización y la aritmética en formas que hace unos años eran difíciles de imaginar. De la misma manera, el riesgo de ahogarse en Macrodatos inútiles o incluso dañinos también está aumentando. La idea de que cuantos más Macrodatos, mejor para nosotros, no es necesariamente correcta. Los Macrodatos pueden ser útiles si son lo suficientemente precisos, si son relevantes, si pueden operacionalizarse de manera que puedan emplearse con consistencia y reproducibilidad suficiente en diversos entornos, y si utilizamos herramientas analíticas rigurosas para evaluarlos. Además, se deben hacer conclusiones inferenciales cuidadosas, y esto rara vez es fácil cuando se involucran las confusas "cajas negras" de los Macrodatos.

Hay presión para cambiarlo todo con las nuevas oleadas de Macrodatos y sus corolarios de Medicina de precisión y Salud de precisión: métodos de cambio, estadísticas, incluso la forma en que pensamos sobre cuáles son los problemas de salud y enfermedad que debemos resolver. Algo de espíritu innovativo es bienvenido, sin embargo, no todo tiene que cambiar solo por cambiar. Necesitamos que sean las personas, no los datos, quienes nos digan qué les importa a las personas. Necesitamos curar gente, no curar Macrodatos. Además, los Macrodatos corren el riesgo de requerir enormes cantidades de inversión en un vano esfuerzo por mantenerse al día, ser dotados de la última infraestructura, el último complemento, o la herramienta de análisis popular más reciente, sin ofrecer un valor real a personas reales.

Es poco probable que esta búsqueda de innovación a partir de Macrodatos lleve, a toda costa, a resultados útiles para la salud individual o de la población. Para muchas herramientas, ya podemos aplicar muy buenas opciones y muy buenos métodos, pero no los usamos de una manera eficiente. La mayoría de la literatura biomédica es errónea e inútil, y no porque no tengamos datos suficientemente grandes – si bien las muestras pequeñas han sido problemáticas en muchas áreas de investigación. El tamaño de muestra efectivo para un estudio es una combinación de sus números y su calidad. Generalmente, la calidad no se protege del todo, por lo que muchos Macrodatos pueden ser de peor calidad (incluso de pésima calidad) en comparación con conjuntos de datos más pequeños recogidos en el pasado.

Se necesitan esfuerzos coordinados para garantizar la obtención del máximo valor de todas estas nuevas oportunidades. Necesitamos combinar los mejores datos, la mejor infraestructura, las mejores herramientas analíticas y aquellos proyectos adaptados a un fin propuesto. La fragmentación de los esfuerzos en este espacio ya está causando un ruido intenso y una confusión que crece geométricamente. Este esfuerzo es un intento genuino de poner algo de orden en este caos. No todas las recomendaciones serán fáciles de implementar, y es posible que los planes deban revisarse posteriormente. Sin embargo, es necesario contar con un plan tanto a nivel nacional como internacional.

Las sociedades que sepan cómo usar los Macrodatos podrán tener una ventaja en el futuro. Las sociedades que solo usen Macrodatos sin realizar un cuidadoso pre-pensamiento, probablemente tendrán una desventaja, ya que se ahogarán en información inútil, mientras que podrían haberse centrado en seleccionar los elementos importantes. Para pacientes individuales y personas, este desafío es aún más agudo. Los Macrodatos pueden permitirnos soñar con un Nuevo Mundo Di-

gital de Salud de Precisión, o que tengamos pesadillas con el mismo. La literatura de cribaje, con sus pasados errores masivos, ofrece un vistazo del desastre que puede ocurrir, por ejemplo, si los Macrodatos nos conducen a una detección masiva inútil, de forma diaria, por hora o continua en tiempo real. Es poco probable que las personas que se vuelven cautivas de su propia información se conviertan en los “supersaludables”, “superhumanos” del futuro.

Los Macrodatos están aquí para quedarse. Asegurémonos de saber cómo tratar con esta bestia y civilizarla.

**John P.A. Ioannidis, MD, DSc**  
Stanford University

## PREÁMBULO

---

Es ocioso señalar que se está produciendo un cambio en la forma de relacionarnos con el mundo que nos rodea, y más específicamente, en la forma de aprender y actuar. El mundo de la medicina y de los sistemas sanitarios no es ajeno a este movimiento aparentemente browniano que ocupa todo.

La digitalización de nuestra vida cotidiana (cotidianos son también los múltiples contactos de las personas con los sistemas de salud) están poniendo en tensión el paradigma de conocimiento médico, arquetipo fuertemente inspirado por el conocimiento adquirido a través de la experimentación controlada en ensayos clínicos (o sus meta-análisis).

Sin entrar en disquisiciones sobre el verdadero valor de lo que allí se trata, al fin y al cabo la agenda de investigación con ensayos clínicos está capturada por el interés de sus financiadores, lo cierto es que muy habitualmente la aportación de los ensayos clínicos añade poco al conocimiento heurístico de aquéllos que tienen que decidir; así, el tipo de paciente o población experimental se parece poco al paciente o población tratados, las condiciones organizativas y el contexto institucional son muy distintos a aquéllos en los que se tienen que tomar las decisiones, y la información producida no tiene la vocación de ser actualizada. La sobredosis de validez interna del paradigma experimental tiene estos peajes.

Aquí radica buena parte del valor de la utilización de los datos de vida real, término acuñado por contraposición a los datos experimentales, y que habitualmente se presenta como sinónimo de *Big Data*, práctica que sugeriría evitar. Sea como sea, la reutilización de datos de vida real, tratados mediante métodos y tecnologías para datos masivos, heterogéneos y dinámicos, va a cubrir las limitaciones señaladas anteriormente – se observarán pacientes reales, tratados en contextos organizativos reales y, con información actualizada y adaptada a las necesidades de conocimiento local. Sin duda, éstas son aportaciones que están ya influyendo en el paradigma de conocimiento clásico, retando las más de las veces, refutando en ocasiones.

Lógicamente, la fascinación por esta nueva aproximación epistemológica no debe llevarnos a confundir fines y medios. La utilización de datos de vida real tiene limitaciones que deben ser conocidas y tratadas; así: siempre deberemos preguntarnos

por la veracidad de los datos (datos incompletos, datos inexistentes, datos heterogéneos, datos fraudulentos); las limitaciones propias de un estudio observacional, en definitiva el tipo de estudio que proporcionan este tipo de datos (especialmente, los riesgos del fenómeno de confusión y los sesgos de selección); y, las maldiciones propias del uso de datos masivos (dimensionalidad, correlaciones espurias o sobreajuste).

Tan cautelosos debemos ser en la consideración de las limitaciones de los datos de vida real, como preocupados debemos estar por las cuestiones adyacentes. La promesa de la reutilización de este tipo de dato para informar las decisiones de profesionales sanitarios, gestores, pacientes o ciudadanos requiere el diseño, desarrollo e implementación de lo que en términos anglosajones conocemos como un *Learning Health System*; un sistema que se organiza para el uso sostenido de los datos disponibles y se desarrolla tecnológicamente para su análisis y reporte inteligente, de forma continua y cercana en el tiempo al momento de la decisión.

El esfuerzo para ese desempeño es lo que constituye el *nudo gordiano* del debate, puesto que requiere fuertes dosis de liderazgo (los agentes alrededor de los datos de salud y sanitarios desempeñan roles frecuentemente contrapuestos), la implementación de complejos mecanismos de seguridad que garanticen el cumplimiento de los preceptos éticos y legales (mecanismos de *opting-out*, garantías para la despersonalización de los datos sin socavar la relevancia de la capacidad de seguir a cada individuo a lo largo de su experiencia sanitaria), sofisticación tecnológica para la integración, almacenamiento y refresco masivo de cualquier fuente de datos útil a los objetivos del sistema (registro electrónico de dato estructurado y no estructurado, lenguaje natural, imagen, etc.), y capacidad para el análisis masivo de ciclo rápido incorporando las cautelas señaladas sobre los datos observacionales.

Con las dosis justas de fascinación, obtener los beneficios esperados de la reutilización sistemática de los datos de vida real, requiere un debate sereno sobre su alcance real y sus límites; pero principalmente sobre el para qué y el cómo. La obra que aquí se prologa, puede resultar útil a este objetivo.

**Enrique Bernal-Delgado**

Investigador en Servicios Sanitarios y Políticas  
Instituto Aragonés de Ciencias de la Salud

## PRÓLOGO

---

A pesar de que el concepto ya había sido acuñado tiempo antes de 2014, en aquel año yo todavía no había oído ni hablar del término “big data”. Tenía, sin embargo, la intuición que otras muchas personas comparten, referente a que la ingente cantidad de datos relacionados que acumulamos encierra una gran cantidad de conocimiento, cuya potencial extracción sugiere enormes beneficios.

En realidad, hacía años que ya se usaban técnicas de la entonces llamada “minería de datos”, precisamente con la intención de encontrar correlaciones allá donde la mente humana no puede hacerlo. Lo que estaba cambiando no era por tanto la computación, ni las matemáticas (que son el lenguaje de la relación entre las entidades del Universo); lo que de verdad estaba albergando un giro en la trama era la aparición de los datos conectados, gracias a la implementación masiva de la era de internet, que ponía las posibilidades de la ciencia en ritmo exponencial.

La transición, en definitiva, de las investigaciones orientadas por hipótesis, hacia las investigaciones orientadas por datos, han venido a expandir en horizontal lo que la Estadística venía construyendo durante décadas. “Si metes basura, sacarás basura” o “correlación no es causalidad” han sido los mantras, esperables, que durante un tiempo hemos estado oyendo, toda vez que la mente humana se veía desafiada por algo tan incomprensible como certero. La idea, en suma, de que pueden generarse unas “cajas negras”, alimentadas por datos, que dan resultados que son ciertos, si bien lo hacen por vías que sencillamente no alcanzamos a comprender con nuestro sistema clásico de intuición humana.

Afortunadamente, la idea que planteamos en 2015, acerca de que, bien pensado, el “big data” no podía en caso alguno ser una moda pasajera, ha ido materializándose en innumerables casos de éxito, capaces de generar algoritmos predictivos que anticipan sucesos antes de que ocurran, segmentan soluciones con rapidez o describen realidades donde antes sólo veíamos sábanas de números inabordables.

El caso de la Medicina es especialmente llamativo. Si se piensa, probablemente no hay ninguna actividad humana que genere mayor cantidad de datos que la sanidad; nada como un hospital a la hora de producir “big data”. Por esta razón hemos estado esperando, impacientes, la aparición de soluciones o de proyectos de investigación biomédica basados en esta tecnología, hasta que al fin, en los dos últimos

años hemos visto aparecer los proyectos europeos, las publicaciones de alto impacto,... todo ello para poner a la atención sanitaria en el camino hacia los niveles de personalización, precisión y predicción en los que siempre hemos querido estar.

Ahora que el panorama se despeja parece obvio que ésta será una de las herramientas claves para hacernos mejores atendiendo pacientes. Ya casi nadie duda que, igual que el fonendoscopio aumenta el oído humano y la resonancia magnética, la vista, así el procesado masivo de datos y la consiguiente inteligencia artificial que se genera, aumentan la capacidad de la mente humana para gestionar variables. Pero esto no siempre fue evidente; por esta razón elogio a la Fundación Gaspar Casal por haber apostado por este mensaje desde el principio, apuesta que se materializa de nuevo en la edición de la presente obra.

Los años venideros son esperanzadores en términos de resultados en salud, al tiempo que retadores al respecto de conceptos como ciberseguridad y equidad, los dos grandes riesgos que esta tecnología destapa al tiempo que avanza. Todos estos conceptos están recogidos con precisión y, sobretodo, con madurez en el texto que el lector tiene entre las manos. Esta madurez es necesaria, pues con el “big data” como con todo salto disruptivo, no todo lo tecnológicamente posible es deseable.

**Ignacio Hernández Medrano**  
Socio Fundador, Savana Médica

## EXECUTIVE SUMMARY

---

### The digital transition in health: Big Data's Opportunities and Challenges regarding decision making.

The main objective of this research project has been to reflect the current situation of Big Data, as a factor of change in the Health sector: its conceptual framework, its tools, and the evidence that has already been generated from it, together to what is expected from its extended use in the future.

For this, the study has been carried out in three phases:

- Phase 1: Conceptualization. A theoretical framework has been developed, which includes the description of the big data concept and the essential elements for its incorporation into health systems.
- Phase 2: Systematic review of the literature. The aim has been to locate and describe the evidence already generated by big data in the health field.
- Phase 3: Focus group. A directed debate with the participation of different stakeholders (researchers, data producers, big data analysts, physicians, health managers, lawyers and decision makers) has been performed.

In view of the results of Phase 1, which has allowed to know the state of the art in depth, thoroughly reviewing everything published, three actions are recommended:

1. To ensure the quality, validity and replicability of the data used. For this, it is necessary that health institutions make an effort to improve the management of clinical information.
2. To train new professional profiles, capable of working in this field. These professionals should combine skills from fields that until now have been developed independently such as epidemiology, bioinformatics and health sciences.
3. To determine precisely which place will data mining take place in the generation of knowledge. It does not seem that this type of methodology replaces completely the classical methods of inference and production of knowledge. We highlight the ability of these new methods to generate new prediction models, more accurate than traditional ones, and provide much more data linked to real life.

Based on the analysis of Phase 2, the following eight actions are recommended:

1. To develop a cohesive, linear, powerful discourse of the entire National Health System, that overcomes the atomization and dispersion of efforts of everything that is being done now; the aim is to avoid failures and loss of opportunities. The generation of interoperability of all the relevant information between Autonomous Communities and all Institutions is detected as the most urgent step.
2. The risk of disintermediation must be considered. It may be possible that most of the public, local initiatives, in which resources and efforts have been invested, lead nowhere, and that the salvageable activity of it ends up in the hands of third parties (basically from the private sector, which probably won't use it with objectives of global interest).
3. Information, as a public good, must have affordable access for all legitimate users, such as health researchers and health managers.
4. The path that is wanted to be followed must be defined, and therefore it is needed to identify which are the basic requirements that must be shared so in the future the health system can promote projects and establish useful platforms.
5. The bases that put a common interoperability scheme, that allows us to grow in the sense of being connected, do exist. We are facing something very complex and heterogeneous. Therefore, an agreement on all the actors is necessary, especially doctors, pharmacists, managers and patients.
6. It is necessary to intelligently handle incentives, so that the Autonomous Communities can contribute data to an integrated information system under the control of the National Ministry of Health.
7. The randomized clinical trial (RCT) must continue being the Gold Standard, as a source of evidence in health and for the authorization of a medicine. However, it is convenient to investigate and regulate how Big Data, like Real World Data and other sources of information, can complement RCTs in certain situations, both to suggest hypotheses to be investigated, and to extrapolate or complement the results of the RCTs.
8. The new General Data Protection Regulation (GDPR) will not constitute a substantial obstacle in the use of data from the health system for the purposes of clinical practice, research or management. However, it is important to promote and monitor its development and strict application, to avoid any risk of commodification of the data in a specific sensitive field such as health.

## 1. CONCEPTUALIZACIÓN

---

# 1. CONCEPTUALIZACIÓN

---

1. Definición .....	25
2. Fuentes de información en materia sanitaria .....	28
3. Herramientas de macrodatos .....	29
4. Tipos de análisis de datos .....	31
5. Aplicaciones informáticas de aplicación clínica para el trabajo con macrodatos .....	31
6. Cuatro preguntas clave sobre el futuro de los macrodatos en el sector salud .....	34

El objetivo principal de este proyecto de investigación es reflejar la situación actual de los macrodatos (Big Data), como factor de cambio en el sector Salud: su marco conceptual, sus herramientas y la evidencia que ya se ha generado a partir de este, junto a lo que se espera de su uso extendido en el futuro.

Para ello, el estudio se ha llevado a cabo en tres fases:

- Fase 1: Conceptualización. Elaborar un marco teórico en el que se desarrolle el concepto de macrodatos y los elementos esenciales para una incorporación en los sistemas sanitarios.
- Fase 2: Revisión sistemática de la literatura. Localizar y describir la evidencia generada mediante macrodatos en el ámbito sanitario.
- Fase 3: Grupo de discusión. Debate dirigido con un grupo de expertos y usuarios de macrodatos.

### 1. Definición

En este primer apartado nos parece importante definir qué son los macrodatos sus características, el potencial y sus limitaciones, dado que es tratado por muchos como un cajón de sastre donde cabe todo. Los macrodatos son el conjunto de nuevas tecnologías y arquitecturas diseñadas para la obtención de valor de grandes volúmenes y variedad de datos de una forma rápida, facilitando su captura, procesamiento y análisis<sup>1</sup>.

Tal vez ésta sea la que mejor represente el concepto de macrodatos, al poner de manifiesto tecnologías y datos para la obtención de valor, caracterizando dichos datos por su Volumen, Variedad y Velocidad de generación: las tres (que ahora han aumentado a cinco: Veracidad y Valor) V's que han caracterizado a los macrodatos<sup>2,3</sup>.

Se han aportado múltiples definiciones al concepto de macrodatos en el ámbito sanitario. Así, Raghupathi & Raghupathi<sup>4</sup>, se refieren a macrodatos en el ámbito de la salud como conjuntos de datos que por su tamaño y complejidad son imposibles de manejar con el software y hardware tradicional. Con esta idea se han propuesto otras muchas definiciones haciendo hincapié en sus características (las V) y peculiaridades de procesamiento y análisis.

Ghani et al<sup>5</sup> los define como grandes conjuntos de datos que durante la fase de diseño no fueron limitados en tamaño y ámbito. Estos autores señalan como ca-

racterística que estos datos no han sido recogidos de forma específica para responder a ninguna pregunta de investigación o clínica, siendo cualquier hipótesis producto de un análisis a posteriori. Así Ghani centra la definición en uno de los aspectos más controvertidos del análisis de los macrodatos. En este análisis no se tiene una hipótesis de partida, sino que se exploran las relaciones existentes en los datos (asociaciones y predicciones), para generar nuevos modelos: es un análisis inductivo, nos enfrentamos a los datos sin una hipótesis inicial. Finalmente, Baro et al<sup>6</sup> ofrecen una definición cuantitativa de macrodatos, refiriéndose como base de datos con un tamaño como  $\text{Log}(n \cdot p) \geq 7$ .

Por el contrario, en los análisis de los Datos de la Vida de Real (registros poblacionales y estudios observacionales), suele haber una hipótesis inicial, el método empleado es deductivo y se persiguen relaciones causales.

Otra diferencia es el tipo de dato empleado en el análisis: DVR trabaja sólo con grades bases de datos estructurados. En análisis de macrodatos, también se incorporan los datos semiestructurados o no estructurados, como el lenguaje natural, audio, video, mediciones de sensores, etc. (tabla 1).

Tabla 1: Tipos de datos

Datos estructurados	Semi-estructurados o no estructurados
<ul style="list-style-type: none"> <li>- Aseguramiento/tarjeta sanitaria.</li> <li>- Prescripción y dispensación.</li> <li>- Registros hospitalarios: CMBD, ingresos, urgencias, HaD, farmacia externa, ...</li> <li>- Sistemas automatizados de órdenes médicas, equipos de laboratorio, ...</li> <li>- Registros de incapacidad temporal, ...</li> <li>- Registros de mortalidad.</li> <li>- Otros registros específicos: vacunaciones, cribados, transporte sanitario, conciertos, diálisis, sensores de monitorización, de telemedicina, de programas, ...</li> <li>- Datos contactos: diagnósticos, visitas, y otros aspectos estructurados.</li> </ul>	<ul style="list-style-type: none"> <li>- Imagen radiológica y fotográfica en PACS (imagen molecular) , e informes asociados.</li> <li>- Historia clínica: anotaciones médicas, gráficos, grabaciones y otros aspectos no estructurados [lenguaje natural].</li> <li>- Registros de monitorización continua o semicontinua (marcapasos, DAI, ...) y telehealth.</li> <li>- Ómicas: genómica, epigenómica, transcriptómica, proteómica, metabolómica y microbiómica ...</li> <li>- Metadatos (cookies), buscadores web, ...</li> <li>- Datos de APPs de salud, Facebook, hábitos (deportivos, sedentarismo, tabaquismo, ...) y otros.</li> </ul>

Tabla adaptada de: S Peiró, comunicación personal

Tabla 2: Resumen de diferencias en las aproximaciones analíticas

Aproximación Macrodatos	Aproximación DVR
<ul style="list-style-type: none"> <li>- Análisis avanzado (advanced analytics), mediante una perspectiva inductiva.</li> <li>- Guiado por las asociaciones multivariadas entre los propios datos (muchas veces no estructurados, con escasa densidad de información y una elevada proporción de datos perdidos).</li> <li>- Intenta descubrir relaciones desconocidas entre los datos, establecer asociaciones, facilitar las inferencias, identificar patrones y construir modelos predictivos.</li> </ul>	<ul style="list-style-type: none"> <li>- Se utilizan métodos deductivos.</li> <li>- Guiados por hipótesis sobre datos (casi siempre estructurados) con elevada densidad de información para medir/describir diferentes procesos o resultados (endpoints).</li> <li>- Se establece la presencia de asociaciones y causalidad.</li> <li>- Detectar tendencias y construir modelos predictivos.</li> </ul>

Tabla adaptada de: S Peiró, comunicación personal

Estas son las 5 uves:

**Volumen**, por el incremento de datos que determina un cambio de escala, pasando de terabytes a petabytes y zetabytes de información. Mucha de esta información según el tipo de utilización, puede pasar a tener un ciclo de vida de su valor muy corto, llegando a la obsolescencia de forma muy rápida.

La **Velocidad** con que los datos son creados y capturados ha aumentado de forma considerablemente. Es, sobre todo, requerida para hacer frente a la obsolescencia de los datos debido a su rápida capacidad de generación. Se trabaja en rentabilizar y optimizar el uso adecuado de los mismos aumentando la precisión y calidad de los resultados. La velocidad de adquisición y análisis de datos inherente a esta tecnología ofrece como ventaja hipotética la capacidad de modificar y corregir en tiempo real modelos de predicción derivados de ellos.

**Variiedad**, basada en la diversidad de los tipos de datos y de las diferentes fuentes de obtención de los mismos. Así, los tipos de datos podrán ser estructurados, semiestructurados o desestructurados, y sus fuentes podrán provenir de un texto, archivos de imagen, datos web, tuits, audios, vídeos, etc. Esta riqueza aumenta el grado complejidad tanto en su almacenamiento como en su procesamiento y análisis.

**Veracidad**, característica asociada a la calidad de los datos. Es el grado de confianza que se establece sobre los datos a utilizar. En la actual fase de desarrollo y aplicación de los macrodatos en salud se hace necesario cruzar los datos obtenidos por medios tradicionales con los obtenidos mediante estas técnicas. En ocasiones no existe un *gold standard* para comparar estas conclusiones haciendo más difícil la interpretación de los resultados.

**Valor aportado**, entendido como la mejora producida en la atención sanitaria derivada del empleo y análisis de los macrodatos. En ese sentido, para hablar de valor hemos de determinar los resultados sobre los que impacta esta nueva forma de análisis, debiendo ser relevantes para el paciente (ej: detección de complicaciones poco frecuentes tras empleo de un nuevo fármaco, pautas de adherencia al tratamiento determinadas mediante aplicaciones móviles).

El valor aportado puede disminuir a medida que aumenta el volumen y la complejidad de los datos, debido a la dificultad de explotación de los mismos. Aumentar el valor de los datos es uno de los retos actuales.

## 2. Fuentes de información en materia sanitaria

Es imprescindible que la información que se obtiene sea de calidad. En este sentido, resulta imprescindible acercarse a las diferentes fuentes de información de macrodatos en salud.

A continuación, presentamos algunas de ellas:

- La web y las redes sociales: Datos de Facebook, Twitter, LinkedIn, Instagram, sitios web de salud y aplicaciones de teléfonos móviles.
- Lecturas de los sensores, medidores y otros dispositivos (principalmente sanitarios).
- Satisfacción, reclamaciones, calidad y costes de atención médica en formatos semiestructurados y no estructurados.
- Datos biométricos: huellas dactilares, datos genéticos, escáner de retina, rayos X y otras imágenes médicas, presión arterial, pulso, lecturas de oximetría de pulso y otros.
- Registros médicos electrónicos, historias clínicas digitales, notas de los profesionales sanitarios, correos electrónicos y documentos en papel.

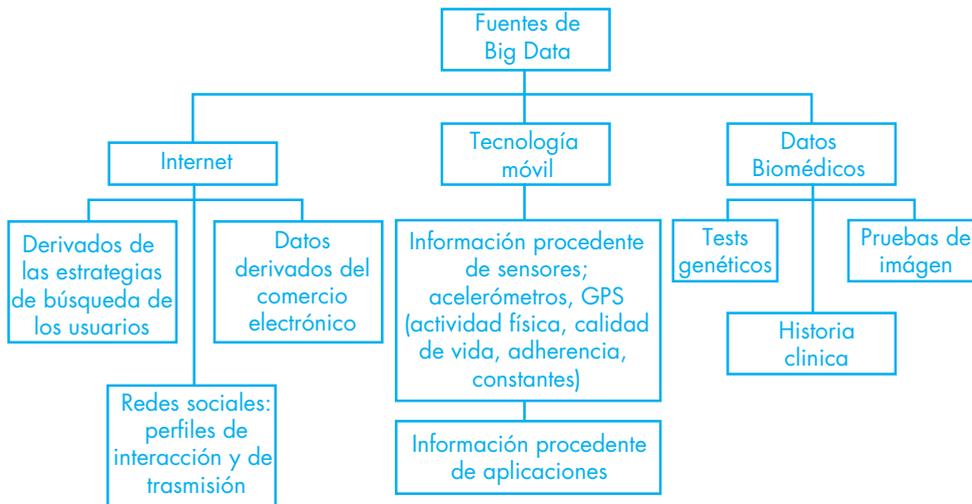


Tabla adaptada de Alonso S. J. *Med Syst* (2017)

### 3. Herramientas de macrodatos

A partir de los macrodatos se pueden desarrollar tareas intensivas de computación masiva mediante el análisis de grandes cantidades de datos en entornos distribuidos<sup>7</sup>.

Los datos pueden estructurarse en diferentes formas y tamaños. Se han desarrollado múltiples sistemas de gestión de esta información y una gran variedad de aplicaciones que ayudan en la programación sanitaria. Veamos algunos ejemplos<sup>8</sup>.

**Base de datos relacional** (Relational Database Management System - RDBMS). Gracias a su arquitectura, estándar durante décadas, permite trabajar con grandes volúmenes de datos. A medida que el tamaño de la información contenida crece, su rendimiento disminuye.

**Base NoSQL**, es decir, No sólo SQL (Structured Query Language o Lenguaje de Consulta Estructurada). Ofrece un entorno de trabajo con gran rendimiento siendo más ágil en el procesamiento para las bases de datos. Amazon o Google las aplican. El origen en la eficiencia de las bases NoSQL, en parte, tiene que ver con su configuración. Las bases de datos NoSQL no están estructuradas,

esto les permite trabajar a mayor velocidad que las relacionales. Las bases totalmente relacionales sólo aportan como ventaja la consistencia. La estructura no relacional de NoSQL se organiza con grupos de información distribuida. Esta organización permite tener varios nodos para el proceso, pudiendo incluso la información encontrarse alojada en diferentes servidores. Este tipo de bases presentan también escalabilidad horizontal, lo que nos permite distinguirlas de las RDBMS.

**Infraestructuras de computación en paralelo.** Una de las principales herramientas de aplicación en el campo de macrodatos son los métodos de análisis de los datos en red. La capacidad de ejecutar algoritmos de forma simultánea en computadoras conectadas en red es una condición necesaria para el análisis e implementación del análisis en macrodatos. Existen múltiples modelos de desarrollo que permiten este tipo de análisis, uno de ellos es MapReduce. A su vez MapReduce se compone de dos funciones de programación, la función Map y la función Reduce. La popularización del modelo MapReduce se debe en parte al desarrollo por parte de APACHE de un paquete software libre llamado Hadoop.

**Hadoop.** Es probablemente la plataforma de computación más conocida en el campo de los macrodatos. A diferencia de las bases de datos tradicionales, Hadoop almacena grandes cantidades de datos de forma no indexada. Aunque hace el proceso de consulta más complejo, permite el almacenaje y procesamiento de volúmenes de datos no asumibles por las bases de datos tradicionales. En realidad, Hadoop es el software que permite el almacenaje y computación de estos volúmenes de datos. Estos datos pueden almacenarse de forma distribuida en diversos servidores en red. Incluye una gran variedad de herramientas con utilidades concretas, algunas de las cuales son: tecnologías para el escalado eficiente del almacenamiento de datos (HDFS); herramientas para la gobernanza e integración de datos (Atlas, Falcon, Flume o Scoop); herramientas para la seguridad de datos (Knox o Ranger); o herramientas sobre gestión operativa del sistema (Ambari, Oozie o Zookeeper). De todas ellas, una de las más conocidas es MapReduce. MapReduce constituye un algoritmo de computación que permite realizar operaciones paralelas en múltiples clústeres. Se pueden así realizar una gran cantidad de operaciones en diferentes servidores permitiendo Hadoop el escalado horizontal. De esta forma, es posible llevar a cabo en pocos minutos operaciones que, sin este componente, hubiesen tardado horas. Hadoop es la herramienta que mejor almacena datos de los usuarios, lo que permite saberlo todo sobre ellos.

### 4. Tipos de análisis de datos

La mayoría de datos en bruto no ofrecen demasiado valor si no se han procesado. Importan la descriptiva, la predictiva y la prescriptiva<sup>9</sup>.

**Analítica descriptiva.** Agregación de las bases de datos en piezas de información más manejable. Resume lo que ha pasado.

**Analítica predictiva.** Utiliza algoritmos, minería de datos y técnicas de aprendizaje automático para estudiar datos recientes e históricos. Permite hacer predicciones sobre el futuro. Plantea escenarios probabilísticos.

**Analítica prescriptiva.** Recomienda itinerarios de acción y muestra las posibles consecuencias de cada decisión. Predice futuros posibles en función de las acciones y decisiones tomadas. Predice pues las consecuencias de las acciones sugeridas. Así, recomienda el mejor camino ante cualquier consecuencia preestablecida. Se alimenta de datos híbridos, estructurados (números, categorías) y no estructurados (vídeos, imágenes, sonidos, textos) que proceden de fuentes internas de la propia organización, o bien externas (redes sociales). Sobre estos datos se aplican modelos de Machine Learning; en nuestro sector, con las cautelas propias de confidencialidad y buenas prácticas. Todavía pocas organizaciones las utilizan.

### 5. Aplicaciones informáticas de aplicación clínica para el trabajo con macrodatos

De acuerdo con Luo et al<sup>10</sup>, podríamos clasificar este tipo de aplicaciones en cuatro grandes grupos según la función de la aplicación. Así, podemos distinguir aplicaciones destinadas al almacenaje de los datos, aplicaciones relacionadas con la recuperación y compartición de los datos, aplicaciones relacionadas con la seguridad y vinculadas al análisis de las mismas.

#### a) Almacenaje de datos

La forma en que se almacenan los datos resulta crítica ya que debe permitir un acceso eficiente a los sistemas de computación. Algunos de los ejemplos de almacenaje de grandes cantidades de datos clínicos se han producido con registros de ECG. De esta manera, la herramienta Cloudwave que emplea un modelo basado en Hadoop permite el acceso la integración de datos clínicos y registros de ECG en tiempo real. La comparación de los tiempos de análisis de

este software frente a los análisis convencionales determinó tiempos de computación para Cloudwave frente a los 20 minutos en los sistemas de computación y almacenaje al uso. Como se ha comentado en apartados previos, el empleo de NoSQL es el estándar en almacenamiento de bases de datos en forma no relacional en el contexto de datos masivos. En el ámbito de la imagen, tiene gran aceptación el empleo de PACS en el almacenaje y procesamiento en el contexto de la radiología clínica. Así, las iniciativas de macrodatos buscan integrar el empleo del PACS con sistemas de bases de datos NoSQL y MapReduce. El campo de la mamografía ha sido uno de los más estudiados.

### **b) Recuperación y acceso de los datos**

Los sistemas de recuperación y acceso a datos en un contexto de macrodatos deberían ser capaces de integrar diversas fuentes de datos (relacionales y no relacionales) permitiendo de forma simultánea la interacción e interoperabilidad de diferentes investigadores y sistemas de historia clínica electrónica. En los ámbitos de las pruebas de imagen es una necesidad la integración entre diferentes proveedores y clientes. Es común que para un mismo servicio (hospital, unidad asistencial) se requiera el acceso a varios servicios que proveen diferentes productos en red. Así, existen modelos de almacenamiento local de datos (federado) o modelos con almacenamiento centralizados (integral). En esa línea se ha desarrollado el sistema *Super PACS* el cual es capaz de dar soporte a múltiples clientes, sincronizar PACS y RIS y proveer interfaces múltiples.

### **c) Seguridad**

El trabajo con macrodatos en el ámbito de la salud debe conciliar la rapidez de acceso, análisis y recuperación junto con una alta garantía en criterios de seguridad. Esto, sin embargo, supone un reto para el ámbito de la bioinformática. Existen varias propuestas a este respecto. Un modelo es el propuesto por *MedCloud*. Este proyecto busca sumar las ventajas derivadas de los sistemas de computación basados en Hadoop con los requisitos de seguridad exigidos por la HIPAA (Health Insurance Portability and Accountability Act).

### **d) Análisis de los datos**

Decir que el empleo de macrodatos cambiará la forma en que hacemos medicina es ya una noticia del pasado. Sin embargo, los datos por sí mismos no son capaces de cambiar nada en el ámbito de la salud. Los verdaderos cambios

vendrán del empleo de nuevas formas de trabajar, integrar y analizar los datos. Los aspectos que se vislumbran de aplicación más inmediata en el ámbito de la salud son:

- Pronóstico: el empleo de nuevas técnicas de análisis de patrones (ej. machine learning) pueden de forma teórica mejorar la manera en que se realizan predicciones sobre el pronóstico de los pacientes. Los modelos tradicionales de predicción incorporan un número limitado de variables productoras (decenas) capaces de ser comprensibles por el ser humano. Los modelos obtenidos mediante nuevos tipos de análisis pueden construir modelos que incorporen miles de variables, así como diferentes combinaciones de interacciones. Dichos modelos podrían predecir con más exactitud grupos de pacientes asociados a un determinado pronóstico. Aunque se han propuesto modelos en este sentido (especialmente en el ámbito de la oncología) se requiere varios años para determinar la verdadera capacidad predictiva de estos nuevos métodos.
- Diagnóstico: algunos diagnósticos pueden resultar especialmente complejos de realizar ya que se basan en aislar pacientes con características y pronóstico similar (sepsis, artritis reumatoides, patrones de malignidad). Existe una importante cantidad de información no estructurada que puede resultar de gran utilidad pero que no se integra de forma adecuada. Además, en muchas ocasiones se realizan pruebas diagnósticas que no ofrecen una ventaja en la toma de decisiones. El empleo de los registros clínicos y su cruce con la evolución de los pacientes puede ofrecer nuevas herramientas de apoyo diagnóstico, determinado qué pruebas o diagnósticos son más adecuados para cada paciente.
- Diagnóstico por imagen y anatomía patológica: algunos ámbitos de la medicina se basan en la interpretación de patrones de imagen repetidos y asociados a una determinada evolución clínica (nódulos y cáncer, infiltrados y neumonía etc.). Con el desarrollo de sistemas capaces de almacenar y analizar grandes cantidades de imágenes (pixel a pixel) podrían generarse nuevos algoritmos diagnósticos capaces de superar la capacidad del ojo humano. Además, estos sistemas podrían incorporar datos clínicos o fisiológicos mejorando la exactitud diagnóstica.

6. Cuatro preguntas clave sobre el futuro de los macrodatos en el sector salud

Preguntas	¿Por qué es relevante?	Próximos pasos
¿Cómo puede mejorar el acceso a datos de seguimiento a largo plazo de los pacientes?	El almacenamiento de datos puede abarataarse, especialmente mediante computación en la nube. No obstante, el coste global de los datos biomédicos crece de forma importante, limitando el desarrollo de este campo.	Se requiere un análisis sobre el uso y manejo de los datos biomédicos. Es necesario el desarrollo de nuevas estrategias de almacenaje. Finalmente será muy necesaria una mayor inmersión.
¿Cómo podemos superar las barreras de acceso a los datos de ensayos clínicos y registros de salud?	Los datos anonimizados procedentes de registros y estudios clínicos ofrecen información relevante para investigación. Sin embargo, el uso de los datos queda limitado por cuestiones legales y de privacidad.	Los pacientes reclaman acceso a sus propios datos. Se requieren nuevas iniciativas que promuevan el acceso a los datos en un marco de seguridad y privacidad.
¿Cómo convertir la información procedente de macrodatos en información útil para la toma de decisiones a pie de cama?	El empleo de la medicina de precisión puede incrementar la utilidad y aplicabilidad de los registros médicos. Sin embargo, la toma de decisiones en este nuevo marco requerirá incorporar nuevas competencias en la formación de los clínicos.	Se requiere la integración de los nuevos datos en la toma de decisiones. Para ello, es necesario la incorporación de software de apoyo a la decisión, así como sistema de aprendizaje en los sistemas sanitarios (health-care learning systems).
¿Cómo mejorar el acceso a programas formativos relacionados con la ciencia de los datos y la bioinformática?	La escasez de programas formativos atractivos para el ámbito de los macrodatos ha limitado el desarrollo de profesionales que integren competencias en estadística, biología y computación.	Se requieren programas formativos y fondos destinados al desarrollo de estos perfiles profesionales. También será necesario evaluar la contribución y necesidades de estos nuevos perfiles profesionales.

Tabla adaptada de Bender E. Nature 2015

Este apartado se ha adaptado parcialmente de “Datos de vida Real y Macrodatos: su incorporación a la evaluación”. Pérez Huerta P; Edo Solsona M<sup>ª</sup>D; Poveda Andrés JL. Fundación Gaspar Casal, 2018.

## 2. REVISIÓN SISTEMÁTICA DE LA LITERATURA

---

## 2. REVISIÓN SISTEMÁTICA DE LA LITERATURA

1. Objetivos .....	37
2. Metodología .....	37
3. Resultados de la búsqueda .....	39
3.1. Estudios realizados en España .....	43
4. Discusión .....	43
5. Anexo I. Estrategias de búsqueda .....	46
6. Anexo II. Características de los estudios incluidos .....	47
7. Anexo III. Estrategia de búsqueda de los originales para España .....	52

### 1. Objetivos

El objetivo principal de la presente revisión de la literatura fue localizar y analizar las publicaciones existentes en las diferentes bases de datos internacionales que abordan las aportaciones del análisis de macrodatos, en el ámbito de los servicios sanitarios. La evaluación comprende sus diferentes perspectivas, clínica, docente, financiera o social.

Como segundo objetivo se analizaron aquellos trabajos que relacionan los resultados de la efectividad/eficacia de tecnologías sanitarias.

El tercer y último objetivo fue el de rastrear aquellos trabajos originales realizados en España.

### 2. Metodología

La revisión sistemática se ha desarrollado en 3 pasos:

2.1.1.1. Búsqueda bibliográfica

2.1.1.2. Selección de los artículos más relevantes, realizada en dos fases.

2.1.1.3. Lectura crítica de los artículos finalmente seleccionados.

Se emplearon criterios de inclusión amplios (alta sensibilidad) teniéndose en cuenta estudios que describen tanto la metodología de análisis como estudios clínicos aplicados que usaron macrodatos. El proceso se realizó en acuerdo a los estándares de la Colaboración Cochrane (Cochrane Handbook for Systematic Reviews of Interventions) y en acuerdo con la declaración PRISMA (Proposal to improve the publication of systematic reviews and meta-analyses).

#### **Criterios para considerar los estudios en la presente revisión.**

Se identificaron aquellas RS relevantes sobre macrodatos en Servicios Sanitarios:

a) Criterios de inclusión:

Población del estudio: se seleccionaron aquellos estudios cuya población comprende participantes humanos de cualquier edad y sexo.

Exposición: se localizaron aquellos estudios relacionados con el sector sanitario, indistintamente de que abordasen aspectos clínicos, financieros-administrativos, investigadores o éticos-sociales.

Se seleccionaron revisiones sistemáticas de la literatura, tanto de ensayos clínicos como de estudios observacionales, siendo indistintamente de tipo cualitativas o con meta-análisis.

### b) Criterios de exclusión:

Se excluyeron revisiones narrativas, ensayos clínicos, estudios observacionales, series de casos o registros;

Estudios no relacionados con el sector de servicios sanitarios, investigación sanitaria o la industria farmacéutica;

Asimismo, también se excluyeron trabajos no publicados en lengua inglesa o en español.

### Búsqueda bibliográfica.

Se consultaron las bases de datos electrónicas de Medline (OVIDSP), WOS (Web of Science) y Google Académico, sin una limitación de tiempo en la antigüedad de los estudios. Se utilizaron los siguientes descriptores tanto en lenguaje libre como controlado (Mesh), utilizando filtros metodológicos para seleccionar revisiones sistemáticas y/o meta-análisis como tipo de estudio. Se utilizaron los siguientes descriptores: big data, advanced analytics, data sets, data mining, machine learning, systematic review. Para analizar la concordancia de los resultados de los estudios de eficacia /efectividad con el análisis de big data, se utilizaron los descriptores: real world data, comparative effectiveness research, combinando mediante booleanos con los descriptores de big data.

### Selección de los estudios.

Tras el proceso de búsqueda de los estudios dos autores revisaron de forma independiente los resultados (Barea-Mendoza JA, Quecedo-Gutiérrez J). En una primera fase, se eliminaron los duplicados y se revisaron los estudios en formato de abstract y título en función de los criterios de elegibilidad preespecifica-

dos. En una segunda fase, los estudios fueron revisados nuevamente por dos autores a texto completo (Barea-Mendoza JA, Quecedo-Gutiérrez J). Los desacuerdos se resolvieron mediante discusión o con la participación de un tercer autor (Gol-Monserrat J). El resultado de esta fase determinó el número total de estudios incluidos. El resumen del proceso de búsqueda se publica mediante el esquema PRISMA. Tras la lectura detallada de los artículos seleccionados se evaluó la calidad de las evidencias, utilizando las recomendaciones Prisma para las revisiones sistemáticas de la literatura y SRQR para los estudios cualitativos.

### 3. Resultados de la búsqueda

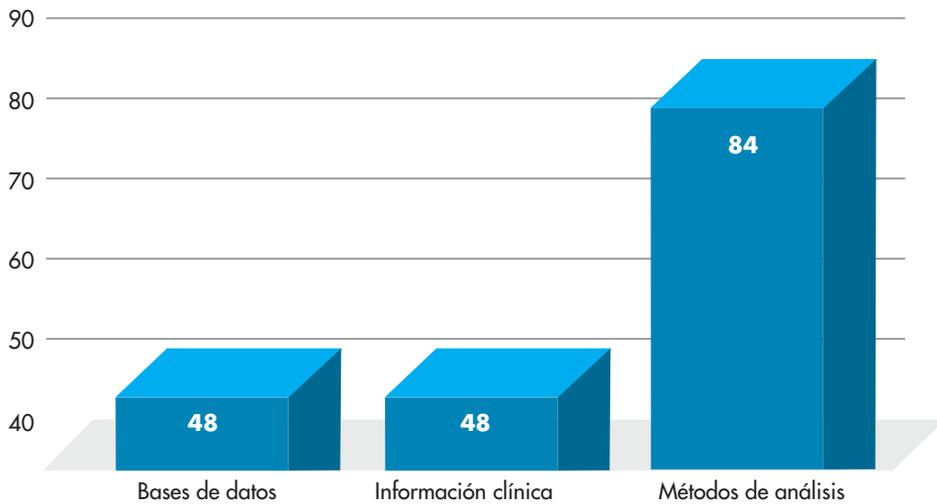
La estrategia inicial arrojó un total de 138 resultados tras la búsqueda en las bases de datos electrónicas Medline (70 artículos), WOS (46 artículos) y Google Académico (22 artículos). Tras la exclusión de duplicados se mantuvieron un total de 61 trabajos. En estos 61 artículos se analizó su contenido mediante la lectura y pertinencia de su contenido en el título y abstract quedando 37 estudios. En una segunda fase de análisis a texto completo, 12 fueron descartados, al detectarse incumplimientos en al menos una de las condiciones anteriormente citadas en los criterios de inclusión o exclusión, quedando finalmente seleccionados 25 estudios. Los 25 incluidos seleccionados fueron revisiones sistemáticas, todas ellas de estudios observacionales. Dada la gran heterogeneidad en el contenido de los estudios no fue posible realizar un análisis cuantitativo de los estudios incluidos.

#### Características de los estudios incluidos.

En la revisión narrativa se incluyeron un total de 25 meta-análisis de estudios observacionales. Los meta-análisis revisados incorporaron un total de 9.147 estudios, aunque el 79% de los mismos provienen de un único estudio (Machado-Díaz 2011).

El objetivo de estudio y el área de interés de los mismos fue variable. Así la temática tratada en los estudios fue: áreas clínicas (n=12), salud global (n=8) y bioinformática (n=5). Un 48% de los estudios incluyeron información sobre bases de datos que emplearon macrodatos, un 48% aportan datos sobre resultados en salud obtenidos mediante análisis de macrodatos y más del 80% aportaron datos sobre la metodología de análisis empleado (Figura 1).

Figura 1: Información incluida en los estudios revisados (% , n=25)



Las áreas de salud que estuvieron más presentes fueron la psiquiatría (Al-jawahiri 2016; Librenza-garcía 2017; Machado 2011), la cirugía (Balla 2017; Batista 2018; Celtikci 2018; Senders 2017; Sochacki 2018) y la enfermedad cardiovascular (Marinov 2011, Thakur 2016).

Desde un punto de vista metodológico se revisó la presencia de los criterios de inclusión en el texto, la descripción de las estrategias de búsqueda, así como la valoración de la calidad metodológica. Un 76% de las revisiones (n=19) establecieron una pregunta clara y describieron los criterios de inclusión. Por otro lado, el 56% (n=14) incorporaron las estrategias de búsqueda utilizadas en el estudio. Ninguna de las revisiones describió la calidad metodológica de los estudios incorporados de forma narrativa o mediante scores destinados a ello. En ese sentido, cabe destacar que un número importante de los estudios incorporan elementos del área de la bioinformática haciendo más difícil la valoración metodológica mediante las herramientas habituales.

#### Principales resultados procedentes de los estudios incluidos.

El resumen de las características y resultados de las revisiones incluidas se resume en la tabla adjunta (Anexo II).

Algunas de las revisiones incorporadas destacan por aportar datos relevantes que podrían ser de aplicación directa en la práctica clínica. Así el estudio publicado por Al-jawahiri y colaboradores recopila información relacionada con enfermedades del espectro autista. Se recopila información de más de 33 bases de datos con aspectos clínicos de diferentes ámbitos: bancos de imagen, fenotipos, genética o EEG entre otros. En la misma línea varias revisiones recogen datos relacionados con los resultados tras diferentes tipos de cirugía. Dos revisiones (Balla 2017 y Batista 2018) emplearon métodos de macrodatos para determinar los resultados quirúrgicos tras cirugía bariátrica y mínimamente invasiva en colon. Dichos estudios se basaron en el uso de grandes de bases de datos de carácter nacional en EE.UU. (National Inpatient Sample, Accreditation and Quality Improvement Program, American College of Surgeons National Surgical Quality Improvement (ACS-NSQIP). Los autores destacan el papel de estas bases de datos en la investigación de resultados en salud. También en el campo de la cirugía, Sochacki y cols. estudiaron la incidencia de complicaciones tras artroscopia mediante el análisis de grandes bases de datos (más de 30.000 pacientes).

En otro ámbito, Librenza-García y cols. revisa el empleo de diferentes técnicas de Machine Learning en el campo de la psiquiatría. Los autores de esta revisión identifican hasta 51 estudios que emplean este tipo de análisis para la estratificación del riesgo de diversos resultados clínicos (cambios de ánimo, suicidio o depresión) en el trastorno bipolar. También en el ámbito de la predicción de resultados destaca el trabajo publicado por Senders y cols. En él se compararon la capacidad de los modelos predictivos tradicionales frente a los obtenidos mediante macrodatos en el ámbito de la neurocirugía; especialmente en el ámbito de la epilepsia, la patología medular y tumores. Los autores describen un AUC de 0,84 (0,82-0,88) y una mejora en la exactitud con esta metodología de hasta el 15% (10-22%).

Diversos estudios de los incorporados describen de forma global el empleo de macrodatos en el ámbito de la salud (Alonso 2017; Jingwei 2016; Kruse 2016). Algunos autores ofrecen recomendaciones, así como los principales retos, oportunidades y barreras en torno a este nuevo campo (Tabla 1 y 2).

**Tabla 1. Principales retos derivados la incorporación de macrodatos en el área de la salud.**

Se requiere la aplicación de tecnologías procedentes de otras industrias. Los profesionales de la salud podrían no tener las competencias necesarias.

Los métodos analíticos relacionados con los macrodatos no están basados en una hipótesis de partida si no en los datos (punto de partida datos vs hipótesis).

Incorporación y uso de fuentes de datos ideadas inicialmente para usos no clínicos.

Problemas vinculados a la estructura de los datos: fragmentación, gran volumen, falta de transparencia y heterogeneidad.

Se requiere mantener la confidencialidad de los datos. Para ello será necesario incorporar procesos de encriptación, validación y control de acceso.

Se requiere el desarrollo de nuevos perfiles profesionales capacitados para analizar las bases de datos generadas.

Variabilidad en la legislación internacional referente a protección de datos vinculados a la salud.

**Tabla 2. Principales oportunidades derivadas la incorporación de macrodatos en el área de la salud.**

Nuevas asociaciones determinadas mediante macrodatos pueden ayudar a mejorar la capacidad de predicción.

Capacidad para predecir resultados en salud que resultan difíciles por otros medios: predicción de epidemias, monitorización de estados de salud, comportamiento.

La integración de diferentes fuentes de información puede aportar nuevas visiones sobre el cuidado del paciente (ej. calidad de vida, adherencia a tratamiento).

La incorporación de dispositivos móviles permite adquisición de datos de la vida real relacionados con la salud.

### 3.1. Estudios realizados en España

Para localizar aquellos trabajos realizados en nuestro entorno, se realizó una búsqueda complementaria de forma independiente. Se utilizaron los mismos descriptores incorporando una limitación en el lenguaje de los trabajos al idioma español o referencia “Spain” en el título o abstract.

Se incluyen aquellos estudios originales relacionados con resultados de salud, sin restricciones temporales o de edad de los participantes. La búsqueda se circunscribe a las bases de datos Medline (Ovid SP) y WOS.

Los trabajos inicialmente obtenidos fueron de 81 artículos seleccionados en WOS y 42 en Medline. Tras la lectura del abstract y la aplicación de los mismos criterios de inclusión y exclusión, ningún trabajo cumplió los requisitos para su lectura a texto completo y posible posterior aceptación. Los trabajos no seleccionados fueron rechazados fundamentalmente por no estar sus resultados relacionados con la salud (sino principalmente relacionados con ciencias de la computación) o porque sus bases de datos o la metodología analítica no se correspondían con las características descritas en el primer capítulo de este estudio sobre lo que son macrodatos (principalmente son estudios con Datos de la Vida Real) (Anexo 3).

## 4. Discusión

Describir cuál es la situación actual de los macrodatos en el ámbito de la salud es, sin duda, una necesidad imperiosa. Sin embargo, la eclosión de proyectos e iniciativas entorno a éste hace que cualquier acercamiento sólo pueda ser parcial (una foto fija). Con la presente revisión de la literatura se intenta presentar una visión global de dicha situación. Los resultados de la misma se podrían agrupar entorno a tres preguntas:

1. ¿Se están empleando los macrodatos en la investigación en salud?
2. ¿Cuáles son las características de las iniciativas actuales?
3. ¿Cuáles son las principales oportunidades y retos que se plantean a este respecto?

Sin duda, la respuesta a la primera pregunta es sí. La actual revisión ha puesto de manifiesto que existen multitud de iniciativas que están empleando nuevas formas investigación que por la magnitud de datos y características del análisis

pueden considerarse macrodatos. Son varios los ejemplos que se exponen en la sección de resultados destacan áreas como la cirugía, la psiquiatría e imagen (Balla 2017; Batista 2018; Librenza-García 2017). Parece que una condición importante para su desarrollo es la disponibilidad de grandes volúmenes de datos de los pacientes; es por ello que grandes registros y bases de datos de pruebas de imagen o genómica son un sustrato muy apropiado para este tipo de iniciativas.

A pesar de las citadas experiencias hemos de decir que el ámbito de aplicación de éstas es por el momento limitado. La mayoría de los trabajos se centran en el desarrollo de modelos de predicción de enfermedad, estimaciones de riesgo o descripción de fenotipos en grandes poblaciones de pacientes. Ninguno de los trabajos encontrados ofrece resultados sobre eficacia de intervenciones. Esta es, probablemente, una consecuencia del tipo de análisis propio de los macrodatos. Varios trabajos comparan la capacidad de predicción de modelos obtenidos mediante métodos tradicionales con los que ofrecen los macrodatos. Los resultados reportados en ese sentido son similares a los obtenidos mediante epidemiología tradicional y en algunos son mejores (Balla 2014; Bruin 2013; Senders 2017). Algunas de las revisiones incluidas incorporan una descripción detallada de los diferentes métodos de análisis empleados, así como las plataformas de almacenamiento (Wang 2014; Yassin 2018). Siendo de interés para el ámbito de la bioinformática, la descripción en profundidad de estos métodos queda fuera de la intención de la presente revisión.

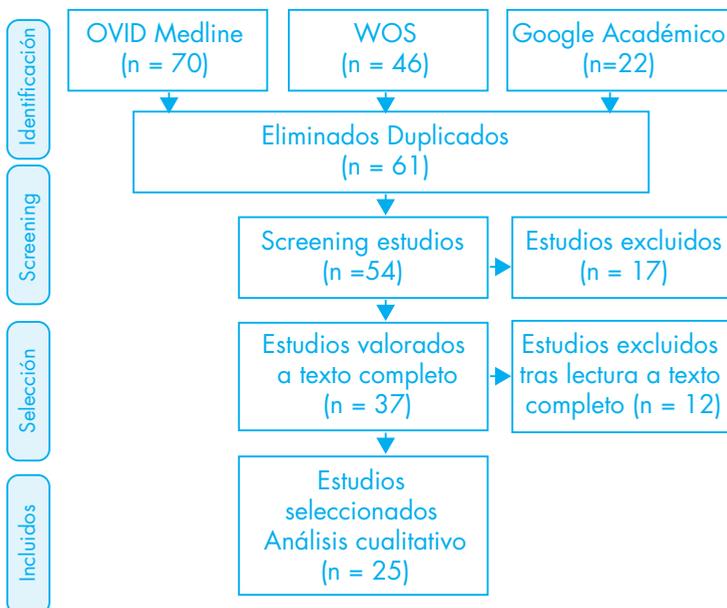
Los autores de los estudios incluidos describen las principales oportunidades y retos a los que las organizaciones sanitarias se enfrentan tras la incorporación de los macrodatos. Entre los retos destacan todos los que tienen que ver con asegurar la calidad, validez y replicabilidad de los datos empleados. Para ello, es necesario que las instituciones sanitarias hagan un esfuerzo en mejorar la gestión que hacen de la información clínica. En segundo lugar, se encuentra la necesidad de formar nuevos perfiles profesionales capaces de trabajar en este ámbito. Dichos profesionales deberán aunar competencias de campos que hasta ahora se han desarrollado de forma independiente como son la epidemiología, la bioinformática y las ciencias de la salud. Un elemento clave que deberá ser perfilado será determinar qué preciso lugar ocuparán los macrodatos en la generación del conocimiento. No parece que este tipo de metodología sustituya completamente a los métodos clásicos de inferencia y producción de conocimiento. Más bien cada deberá emplearse para responder aquellas preguntas a las que más se adecúen. Todos los autores coinciden en señalar una amplia

lista de oportunidades. Entre ellas destacamos la capacidad para generar nuevos modelos de predicción, más precisos que los tradicionales y aportar datos mucho más vinculados a la vida real. No obstante, muchas de las virtudes señaladas se le suponen, pero no han sido claramente demostradas.

La presente revisión presenta algunas limitaciones que deben ser expuestas. En primer lugar, dado que no pretende dar una respuesta exhaustiva se decidió incluir sólo revisiones sistemáticas. Así se han dejado fuera otros tipos de publicaciones como son revisiones narrativas, estudios observacionales aislados, puntos de vista, etc. En segundo lugar, como se expuso previamente, ninguna de las revisiones incluidas reportó la calidad metodológica de los estudios primarios. Por último, la gran heterogeneidad de las preguntas clínicas y ámbito de las revisiones incluidas no ha permitido obtener ningún resultado agregado, pudiendo ofrecer solamente una descripción cualitativa de los resultados.

Respecto a la búsqueda de originales en España, es llamativo no haber encontrado ningún trabajo que explote bases de datos más allá de aquellas que entran bajo el paraguas de los datos de la vida real. Algunos originales incorporan datos que por su volumen o método de análisis pueden considerarse macrodatos. Sin embargo, los objetivos de estos trabajos se dirigen a aspectos relacionados con las ciencias de la computación sin aportar información relevante sobre resultados en salud.

Figura 2. Diagrama PRISMA



Anexo I. Estrategias de búsqueda

# ▲ Searches	Results
1 ▶ big data.m_titl.	2402
2 ▶ advanced analytics.m_titl.	21
3 ▶ data sets.m_titl.	2050
4 ▶ machine learning.m_titl.	4086
5 ▶ data mining.m_titl.	2413
6 ▶ real world data.m_titl.	544
7 ▶ comparative effectiveness research.m_titl.	1442
8 ▶ big data.mp. [mp=tx, bt, ti, ab, ct, ot, nm, hw, kf, px, rx, ui, sy]	9242
9 ▶ systematic review.m_titl.	128252
10 ▶ 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8	19475
11 ▶ 9 and 10	70

# ▲ Searches	Results
1 ▶ big data.m_titl.	2402
2 ▶ advanced analytics.m_titl.	21
3 ▶ data sets.m_titl.	2050
4 ▶ machine learning.m_titl.	4086
5 ▶ data mining.m_titl.	2413
6 ▶ real world data.m_titl.	544
7 ▶ comparative effectiveness research.m_titl.	1442
8 ▶ big data.mp. [mp=tx, bt, ti, ab, ct, sh, ot, nm, hw, kf, px, rx, an, ui, ds, on, sy]	9242
9 ▶ 1 or 2 or 3 or 4 or 5 or 8	17530
10 ▶ 6 or 7	1983
11 ▶ 9 and 10	38

Anexo II. Características de los estudios incluidos

Autor; Año publicación	Revista	Objetivo y descripción	Características metodológicas	Resultados
Al-Jawahri; 2016	PeerJ — the Journal of Life and Environmental Sciences	Objetivo: Determinar la presencia de recursos basados en big data en el área del autismo. Estudios incluidos: 33 de bases de datos con información relevante. Área de interés: autismo.	Diseño: revisión sistemática que incluye descriptivos sobre de recursos y bases de datos tipo big data. Descripción de las estrategias de búsqueda; presente. Criterios de inclusión especificados: Presente. Valoración de sesgos de los estudios incluidos: ausente.	Se localizaron un total de 33 bases de datos que incorporan datos masivos en el área de interés. La naturaleza de los datos reportados es variable siendo los tipos más frecuentes: datos imagen, fenotípicos, muestras biológicas, eeg entre otros. No se presentan resultados de análisis sobre bases de datos de forma específica.
Alonso; 2017	Journal of Medical Systems	Objetivo: Determinar la presencia de recursos basados en big data en el área de salud así como técnicas de predicción más empleadas. Estudios incluidos: 32 de bases de datos con información relevante. Área de interés: sector salud	Diseño: revisión sistemática que incluye estudios descriptivos sobre de recursos y tipos de análisis para big data. Descripción de las estrategias de búsqueda; presente. Criterios de inclusión especificados: presente. Valoración de sesgos de los estudios incluidos: ausente.	Se localizaron un total de 32 publicaciones que cumplieron con los criterios de inclusión. Se describen las principales metodologías de análisis big data con ejemplos: k-means, árbol de decisión, redes neuronales bayesianas, naive bayes, part rules, Newton-Raphson. Algunos de los estudios incorporados comparan de validez en predicción mediante big data de algunas enfermedades crónicas: enfermedad cardiovascular (Palaniappan, & Avang, Sivagovry, et al.), enfermedad renal crónica (Kunwar et al.), enfermedad hepática (Alifisahin & Mantoro). Algunos de los beneficios descritos son: aumento en la capacidad de adquisición de datos, disminución en el consumo de recursos, desarrollo de modelos predictivos mejorados.
Ansanipour; 2016			Excluido. Es un abstract de un congreso.	
Ballier; 2017	Langenbeck's archives of surgery	Objetivo: determinar los resultados en salud tras la cirugía bariátrica mediante análisis big data. Estudios incluidos: 54 estudios observacionales. Área de interés: cirugía.	Diseño: revisión sistemática de estudios observacionales. Descripción de las estrategias de búsqueda; presente. Criterios de inclusión especificados: presente. Valoración de sesgos de los estudios incluidos: ausente.	Se describen un total de 54 estudios observacionales que describen los resultados tras la cirugía bariátrica. Todos los estudios se desarrollan a partir de 1 bases de datos americanas relacionadas con la cirugía como: National Inpatient Sample (NIS), Accreditation and Quality Improvement Program (MBSAQIP) y the American College of Surgeons National Surgical Quality Improvement (ACS-NSQIP). Se obtuvieron datos clínicos sobre resultados en las diferentes técnicas, morbilidad y mortalidad en cirugía bariátrica, readmisión al tercer día, resultados según patologías, formación en laparoscopia y resultados socioeconómicos. Además se postula el empleo de estas grandes bases de datos para el desarrollo de estudios sobre resultados en salud.
Battista; 2018	International Journal of Colorectal Disease	Objetivo: determinar los resultados en salud tras la cirugía colorectal mínimamente invasiva mediante análisis big data. Estudios incluidos: 56 estudios observacionales. Área de interés: cirugía.	Diseño: revisión sistemática de estudios observacionales. Descripción de las estrategias de búsqueda; presente. Criterios de inclusión especificados: presente. Valoración de sesgos de los estudios incluidos: ausente.	Se describen un total de 56 estudios observacionales que describen los resultados tras la cirugía bariátrica. Todos los estudios se desarrollan a partir de 1 bases de datos americanas relacionadas con la cirugía como: National Inpatient Sample (NIS), Accreditation and Quality Improvement Program (MBSAQIP) y the American College of Surgeons National Surgical Quality Improvement (ACS-NSQIP). Se obtuvieron datos clínicos sobre resultados en las diferentes técnicas, morbilidad y mortalidad en cirugía bariátrica, y resultados socioeconómicos.
Berger; 2014			Excluido. Presenta una descripción narrativa interesante.	
Bruin 2013			Excluido. No hay big data. Son registros electrónicos de resultados micro-biológicos.	

Autor; Año publicación	Revista	Objetivo y descripción	Características metodológicas	Resultados
Cettkici; 2018	<i>Turkish Neurosurgery</i>	Objetivo: determinar los estudios que emplean técnicas de machine learning (ML) en el ámbito de la neurociencia. Estudios incluidos: 51 estudios observacionales. Área de interés: neurociencia.	Diseño: revisión sistemática de estudios observacionales. Descripción de las estrategias de búsqueda: parcialmente. Criterios de inclusión especificados: presente. Valoración de sesgos de los estudios incluidos: ausente.	Se describen un total de 51 publicaciones que emplean las técnicas de ML en el ámbito de la neurociencia. Los temas incorporados fueron hidrocefalia (2), simulación cerebral (6), patología neurovascular (4), epilepsia (6), glioma (4), radiocirugía (7), patología espinal (7), trauma craneal (10), otros (4). No se describe información sobre las bases de datos originales, llamados de las mismas. Tampoco datos comparativos de otros métodos de análisis frente a ML.
Chute; 2013			Excluido. Es una revisión narrativa.	
Machado Díaz; 2011	<i>Asian Journal of Psychiatry</i>	Objetivo: empleo de técnicas de minería de datos en la búsqueda de publicaciones. Estudios incluidos: 7299. Área de interés: psiquiatría, estimulación magnética transcranial.	Diseño: revisión sistemática de estudios observacionales. Descripción de las estrategias de búsqueda: parcialmente. Criterios de inclusión especificados: presente. Valoración de sesgos de los estudios incluidos: ausente.	La estrategia de búsqueda inicial arrojó un total de 7299 resultados. Mediante estrategia de minería de datos se procedió a realizar un mapa conceptual de las principales materias. Así las áreas fueron depresión, parkinson, esquizofrenia y trastorno bipolar.
Fallah; 2017	<i>Healthcare Research Informatics</i>	Objetivo: determinar la utilidad de la minería de datos para la detección del efecto de apps móviles para pacientes. Estudios incluidos: 30. Área de interés: minería de datos, apps móviles para pacientes.	Diseño: revisión sistemática de estudios observacionales. Descripción de las estrategias de búsqueda: presente. Criterios de inclusión especificados: presente. Valoración de sesgos de los estudios incluidos: ausente.	Se incorporaron un total de 30 trabajos mediante técnicas de minería de datos. Los trabajos fueron categorizados según su uso por los pacientes en: monitorización, detección, cálculo, predicción, clasificación, recogida de datos y seguimiento. 20 trabajos recogieron la exactitud de la minería de datos para la función propuesta. 13 de los 20 reportaron una exactitud mayor del 90% y 7 por encima del 95%. Los métodos más empleados fueron SVM (support vector machine), clustering y árboles de decisión. Las situaciones en las que se aplicaron de forma mayoritaria eran cáncer, actividad física, nutrición y enfermedad cardiovascular.
Gianfredi; 2018	<i>Frontiers in Public Health</i>	Objetivo: determinar la utilidad de las nuevas fuentes de datos (Novel Data Stream, NDS) para la vigilancia de brotes epidémicos. Estudios incluidos: 47. Área de interés: big data, vigilancia epidemiológica.	Diseño: revisión sistemática de estudios observacionales. Descripción de las estrategias de búsqueda: presente. Criterios de inclusión especificados: presente. Valoración de sesgos de los estudios incluidos: ausente.	Se revisaron un total de 47 estudios que emplearon las técnicas NDS en la detección de brotes epidémicos. De los estudios revisados hacía referencia al Chikungunya, 6 al Dengue, 19 al Ébola, 2 a Malaria 1 Mýarayo Virus, 2 Virus del Nilo y 16 Zika. La fuentes de datos empleadas fueron muy variables. Las más importantes fueron Twitter, Google Trends, Youtube, CDC Facebook, Baidu entre otras.
Gloria Perez; 2015			Excluido. Es un artículo de opinión de Gaceta Sanitaria.	
Grams; 2011			Excluido. No hay big data.	
Gray; 2015			Excluido. Presenta una descripción narrativa interesante.	
Hamroui; 2017	<i>Wireless Communications and Mobile Computing</i>	Objetivo: determinar los diferentes mecanismos de seguridad empleados en big data para el ámbito sanitario. Estudios incluidos: 22. Área de interés: big data, seguridad informática.	Diseño: revisión sistemática de estudios observacionales. Descripción de las estrategias de búsqueda: parcial. Criterios de inclusión especificados: parcial. Valoración de sesgos de los estudios incluidos: ausente.	Tras la búsqueda se obtuvieron un total de 22 resultados. Se agruparon los diferentes mecanismos de seguridad así como los inconvenientes de los mismos. Finalmente los autores proponen dos modelos de seguridad combiando mecanismos encontrados.

Autor; Año publicación	Revista	Objetivo y descripción	Características metodológicas	Resultados
Hernani; 2015	Statistical Analysis and Review on Healthcare Big Data	Objetivo: valorar la incorporación de las estrategias de big data en el ámbito de la salud. Estudios incluidos: 321. Área de interés: big data, salud.	Excluido. Es una descripción narrativa.	Los autores realizan una revisión sistemática con una estrategia de alta sensibilidad que les permite recuperar una gran cantidad de publicaciones que relacionan big data y ámbito sanitario. Un 14,5% de los artículos fueron referidos a omics, un 14,1% eran referidos a especialidades médicas y un 73,4% abordan temas del ámbito de la tecnologías de la información en salud (registros electrónicos, bioinformática, web mining etc). Aunque los autores revisaron las publicaciones comprendidas entre 2002 y 2015 se observó un incremento exponencial de las publicaciones a partir de 2014. El 80% de las referencias datan de 2014 y 2015.
Kruse; 2016	Jmir medical Informatics	Objetivo: determinar las oportunidades y retos derivadas de la incorporación de big data en salud. Estudios incluidos: 28. Área de interés: big data, salud.	Diseño: revisión sistemática de estudios observacionales. Descripción de las estrategias de búsqueda: presente. Criterios de inclusión especificados: presente. Valoración de sesgos de los estudios incluidos: ausente.	Tras la revisión de los resultados los autores sumarizan las principales oportunidades y amenazas relacionadas con la introducción del big data. Se centran en elementos de componente más técnico. Entre los retos señalados por los autores que se incluyeron destacan la estructuración de los datos y su fragmentación, elementos relacionados con la seguridad y la estandarización de los datos. Se destacaron múltiples oportunidades, siendo las más repetidas la mejora de la calidad de la asistencia, detección precoz de la enfermedad así como datos relacionados con calidad de vida.
Librenza-garcía; 2017	Neuroscience and Biobehavioral Reviews	Objetivo: determinar el impacto de los análisis con machine learning en la investigación sobre trastorno bipolar. Estudios incluidos: 51. Área de interés: machine learning, psiquiatría.	Diseño: revisión sistemática de estudios observacionales. Descripción de las estrategias de búsqueda: presente. Criterios de inclusión especificados: presente. Valoración de sesgos de los estudios incluidos: ausente.	Los autores incluyen 51 estudios que valoran el papel de las técnicas machine learning en el manejo de pacientes con trastorno bipolar. 38 estudios estuvieron focalizados en el diagnóstico. Entre ellos 11 emplearon técnicas de neuroimagen, 13 imagen neurofuncional, 5 análisis genéticos, 3 test neuropsicológicos y 2 biomarcadores. 7 Estudios reportaron resultados clínicos como suicidio, depresión o cambios anímicos. Los autores concluyen que las técnicas de ML permiten desarrollar un tratamiento más personalizado ajustado al riesgo individual.
Luo; 2015	Biomedical Informatics Insights	Objetivo: revisar la bibliografía disponible sobre big data Estudios incluidos: 68. Área de interés: big data, salud.	Diseño: revisión sistemática de estudios observacionales. Descripción de las estrategias de búsqueda: parcial. Criterios de inclusión especificados: presente. Valoración de sesgos de los estudios incluidos: ausente.	Los autores clasificaron los resultados en 4 disciplinas: bioinformática, informática clínica, imagen y salud pública. La integración de diversas fuentes de información permite incorporar nuevos puntos de vista de la asistencia al paciente, la disponibilidad de datos en tiempo real permite nuevas formas de seguimiento y por último el archivo de datos y su análisis facilita el desarrollo de nuevas herramientas de toma de decisiones.
Marinov; 2011	Journal of Diabetes Science and Technology	Objetivo: revisar la aplicación de técnicas de minería de datos en el ámbito de la diabetes. Estudios incluidos: 17. Área de interés: minería de datos, diabetes.	Diseño: revisión sistemática de estudios observacionales. Descripción de las estrategias de búsqueda: parcial. Criterios de inclusión especificados: presente. Valoración de sesgos de los estudios incluidos: ausente.	Los autores incorporaron información de 17 artículos que emplearon técnicas de minería de datos. 4 artículos analizaron la interpretación de los niveles de glucemia, 4 sobre predicción y 3 sobre genética. Otros incluyeron objetivos diversos como la determinación de eventos adversos, guías clínicas, predicción de mortalidad, etc.
Metha; 2018	International Journal of Medical Informatics	Objetivo: determinar el alcance de bigdata en sector salud, estrategias para solventar los retos. Estudios incluidos: 58. Área de interés: big data, salud.	Diseño: revisión sistemática de estudios observacionales. Descripción de las estrategias de búsqueda: presente. Valoración de sesgos de los estudios incluidos: presente.	Los autores extraen información relevante de 51 artículos. Presentan información sobre el concepto y evolución de big data, fuentes de datos (big data) en salud, técnicas de análisis, ejemplos de aplicación clínica y retos. Algunas de las propuestas para afrontar los retos fueron el desarrollo de gestión adecuada, implementar culturas enfocadas a compartir información, incorporar medidas de seguridad, formar personal en el análisis de datos e incorporar sistemas de computación adecuados para en análisis big data.
Pashazadeh; 2017	Journal of Biomedical Informatics	Objetivo: determinar y describir los diferentes tipos de análisis presentes en el ámbito big data. Estudios incluidos: 29. Área de interés: big data, salud.	Diseño: revisión sistemática de estudios observacionales. Descripción de las estrategias de búsqueda: presente. Valoración de sesgos de los estudios incluidos: ausente.	Los autores desarrollaron la búsqueda que arrojó un total de 205 artículos. Tras la revisión 29 cumplieron los criterios de inclusión preseleccionados. Se clasificaron los artículos en función de la estrategia empleada para el manejo de los datos. En un 28% la estrategia fue basada en heurísticos, un 25% machine learning, 18% métodos híbridos, 18% basados en agentes y 11% métodos de computación en nubes.

Autor; Año publicación	Revista	Objetivo y descripción	Características metodológicas	Resultados
Piri; 2017		Objetivo: describir el conocimiento disponible sobre problemas de adherencia mediante análisis big data. Estudios incluidos: 22. Área de interés: big data, seguimiento clínico.	Diseño: revisión sistemática de estudios observacionales. Descripción de las estrategias de búsqueda: ausente. Criterios de inclusión de los estudios incluidos: ausente. Valoración de sesgos de los estudios incluidos: ausente.	De los 22 trabajos más frecuentes fue la salud pública (14) seguido de las especialidades médicas (7) y el grupo de las omías (1). La mayor parte de los estudios emplearon bases de datos administrativas o registros retrospectivos. Tres estudios emplearon datos adquiridos en tiempo real. Los autores constataron que el número de trabajos que comunican el empleo de estrategias de big data en el problema de la adherencia es limitado.
Santos; 2018	Advances in Intelligent Systems and Computing	Objetivo: describir de forma amplia la situación del big data en el ámbito de la salud. Estudios incluidos: 466. Área de interés: big data, salud.	Diseño: revisión sistemática de estudios observacionales. Descripción de las estrategias de búsqueda: parcial. Criterios de inclusión de los estudios incluidos: ausente. Valoración de sesgos de los estudios incluidos: ausente.	Se analizaron 466 artículos. Los países mayoritarios de publicación fueron China (117), EE.UU (108) e India (54). Los principales sectores fueron la salud (72) seguidos de la tecnología (29). Las herramientas de análisis presentadas de forma más frecuente fueron Apache Hadoop (204), MapReduce (164) y Hadoop Distributed File (77) entre otros. Los objetivos principales de las publicaciones fueron conceptuales (52%), arquitectura (22%), marco general (19%) y encuestas 19,7%.
Senders; 2017	World Neurosurgery	Objetivo: determinar la utilidad de machine learning en la predicción de resultados en neurocirugía. Estudios incluidos: 30. Área de interés: machine learning, neurocirugía.	Diseño: revisión sistemática de estudios observacionales. Descripción de las estrategias de búsqueda: presente. Criterios de inclusión de los estudios incluidos: ausente.	Se revisó la capacidad predictiva de los modelos desarrollados mediante ML frente a los tradicionales. Las patologías más frecuentes fueron epilepsia (9 estudios), tumores cerebrales (4), patología medular (5 estudios), hemorragia subaracnoidea (2) y parkinson (2). La mediana de exactitud de los modelos fue del 94% (87-95%) y el AUC de 0.84 (0.82-0.88). Cuando se compararon los modelos obtenidos mediante ML con la predicción habitual se observó una mejoría media en la exactitud del 15% (10-22%).
Sinnensberg; 2016	American Journal Public Health	Objetivo: revisar de forma sistemática el empleo de twitter en investigación sanitaria. Estudios incluidos: 137. Área de interés: big data, Twitter, salud.	Diseño: revisión sistemática de estudios observacionales. Descripción de las estrategias de búsqueda: presente. Criterios de inclusión de los estudios incluidos: ausente.	Se incluyeron un total de 137 trabajos que de forma conjunta habían analizado un total de 5 millones de tweets. Un 33% de los estudios se habían publicado de forma muy reciente (2015). La metodología empleada en los estudios fue análisis del contenido (65%), vigilancia epidemiológica (26%), seguimiento (14%) y reclutamiento (7%). Las áreas de mayor aplicación fueron la salud pública (23%) y las enfermedades infecciosas (20%).
Soehacki; 2018	The Journal of Arthroscopic and Related Surgery	Objetivo: comparar los resultados de la artroscopia reportados en estudios originales y en análisis big data. Estudios incluidos: 207. Área de interés: big data, traumatología.	Diseño: revisión sistemática de estudios observacionales. Descripción de las estrategias de búsqueda: parcial. Criterios de inclusión: presente. Valoración de sesgos de los estudios incluidos: ausente.	Se incluyeron un total de 207 estudios. De ellos 201 fueron trabajos originales (15.780 pacientes) y 6 derivaron de grandes bases de datos (20.825 pacientes). Los pacientes incluidos en las bases de datos fueron más mayores, con más frecuencia mujeres y presentaron mayor seguimiento (p < 0.001). Las complicaciones mayores, revisión y conversión a artroscopia fueron más frecuentes en los estudios que emplearon grandes bases de datos.
Thakur; 2016	-	Objetivo: determinar la utilidad del análisis big data en la enfermedad cardiovascular. Estudios incluidos: no determinado. Área de interés: big data, cardiovascular.	Diseño: revisión narrativa. Descripción de las estrategias de búsqueda: ausente. Criterios de inclusión: ausente. Valoración de sesgos de los estudios incluidos: ausente.	Los autores llevaron a cabo revisiones sistemáticas en 4 bases de datos en PDGene, AltGene, ScGene, CEA. La sensibilidad en la selección de artículos mediante minería de datos fueron 100% en las tres primeras bases de datos y 99% en el registro CEA. Las especificidades fueron respectivamente 90, 93, 90, 73%.
Wallace; 2012	Genetics in Medicine	Objetivo: determinar la utilidad de la minería de datos en el desarrollo de revisiones sistemáticas. Estudios incluidos: no determinado. Área de interés: big data, revisiones sistemáticas.	Diseño: revisión sistemática. Empleo de 4 bases de datos. Descripción de las estrategias de búsqueda: presente. Criterios de inclusión: ausente. Valoración de sesgos de los estudios incluidos: ausente.	Los autores llevaron a cabo revisiones sistemáticas en 4 bases de datos en PDGene, AltGene, ScGene, CEA. La sensibilidad en la selección de artículos mediante minería de datos fueron 100% en las tres primeras bases de datos y 99% en el registro CEA. Las especificidades fueron respectivamente 90, 93, 90, 73%.
Wang; 2017			Excluido. Es una revisión narrativa.	

Autor; Año publicación	Revista	Objetivo y descripción	Características metodológicas	Resultados
Wang; 2014	Jmir Medical Informatics	Objetivo: determinar las características actuales del análisis big data, incluyendo algoritmo más frecuentemente empleados. Estudios incluidos: no determinado. Área de interés: big data.	Diseño: revisión de la literatura. Descripción de las estrategias de búsqueda: ausente. Criterios de inclusión: ausente. Valoración de sesgos de los estudios incluidos: ausente.	Los autores centran sus conclusiones sobre la situación actual del big data. El origen de los datos en la investigación biomédica es muy diverso. Además existe una importante variabilidad en cuanto a tamaño de las bases de datos, dimensión, niveles de almacenamiento entre otros. Los principales plataformas para el almacenamiento son Apache Hadoop, NoSQL y Google BigTable. Los paquetes estadísticos empleados son SAS, Mplus, R, STATA entre otros. Los autores señalan algunas limitaciones como la interpretación de los resultados, sesgos de selección y manejo de las perdidas entre otros.
Yassin; 2018	Computer Methods and Programs in Biomedicine	Objetivo: determinar el papel de los métodos machine learning en los sistemas diagnósticos en cáncer de mama. Estudios incluidos: 154. Área de interés: machine learning, cáncer de mama.	Diseño: revisión de la literatura. Descripción de las estrategias de búsqueda: presente. Criterios de inclusión: presente. Valoración de sesgos de los estudios incluidos: ausente.	Los autores describen y comparan las diferentes técnicas de ML empleadas para las ayudas a la decisión en diagnóstico de cáncer de mama. Las técnicas más frecuentes fueron sistemas de vectores (SVM), redes neuronales (ANN) y K-NN. No se establece una comparación directa de las diferentes modalidades. Se recomienda estandarizar la producción de imágenes por parte de los servicios de radiodiagnóstico para mejorar los resultados.

### Anexo III. Estrategia de búsqueda de los originales para España

#### Medline OVID SP

1.	Big Data mp	2.602
2.	Advanced analytics m_Titl	15
3.	Data Sets m_Titl	1.303
4.	Machine learning m_Titl	2.104
5.	Data mining m_Titl	1.468
6.	1 or 2 or 3 or 4 or 5	7.286
7.	Limit 6 to Spanish	23
8.	Spain.mp, or SPAIN/	89.020
9.	6 and 8	25
10.	7 and 9	42

#### Web of Knowledge WOS

1.	Topic: (Big Data)	86.817
2.	Address: (Spain)	1.557.585
3.	Topic: (Healthcare)	268.712
4.	1 and 2 and 3	81

### 3. INFORME CUALITATIVO: ANÁLISIS DEL GRUPO DE DISCUSIÓN

---

### 3. INFORME CUALITATIVO: ANÁLISIS DEL GRUPO DE DISCUSIÓN

---

1. Objetivos y metodología .....	55
2. Transcripciones .....	57
3. Análisis del debate .....	99
4. Conclusiones .....	106

#### 1. Objetivos y metodología

##### *Objetivo de la sesión*

El objetivo de este Grupo de Discusión es complementar a las fases 1 y 2 mediante un debate abierto que aporte las experiencias y puntos de vista de una selección de panelistas procedentes de distintas disciplinas y permita esbozar el futuro plausible de los macrodatos.

##### *Metodología del Grupo de Discusión*

En primer lugar, los panelistas recibieron con antelación un dossier en el que se describían los antecedentes y la justificación del proyecto, sus objetivos, y las preguntas que se formularían en la sesión presencial.

En la sesión presencial del Panel de Expertos, moderada por Juan del Llano, se presentaron secuencialmente las preguntas y se permitió a los participantes intervenir de forma flexible contestando a las preguntas originales o reaccionando a las intervenciones de los otros miembros del panel.

Toda la sesión fue grabada y transcrita por Jordi Gol Montserrat. A continuación, Joan Rovira llevó a cabo una primera depuración del texto y solicitud de aclaraciones a los participantes, que tuvieron la oportunidad de revisar la transcripción final de sus intervenciones para garantizar que en la elaboración de la transcripción no se hubiese alterado el sentido de sus manifestaciones. Las afirmaciones y opiniones expresadas en el texto reflejan, por lo tanto, la posición de los participantes y no las de sus respectivas instituciones, ni las del equipo editorial de este informe, con la excepción de los comentarios finales.

##### *Participantes*

- Pablo García Mexia (Syntagma Consulting)
- Jaime del Barrio (Ernst & Young y Asociación Salud Digital)
- Jordi Colomer (Consultor)
- Mercedes Alfaro (MSSSI)
- César Velasco (Hospital Valle de Hebrón)
- Carolina González Criado (SERGAS)
- Francisco Estupiñán (IACS)

*Equipo organizador de la Fundación Gaspar Casal y presente en la sesión:*

- Moderador: **Juan del Llano**
- Análisis y redacción del informe: **Joan Rovira**
- Logística y grabación del debate: **Jordi Gol-Montserrat**

*Preguntas de la sesión*

1. ¿Qué iniciativas relacionadas con el empleo de Macrodatos (Big Data) en nuestro país conoce por experiencia directa o por la literatura? ¿Algún caso notable de éxito o fracaso?
2. ¿Conoce algún ejemplo de nuevo conocimiento generado mediante el análisis de Macrodatos que se haya aceptado e incorporado en la práctica real con un amplio consenso?
3. ¿Qué puede aportar a la efectividad-eficiencia el empleo de Macrodatos en el sector salud, tanto en la práctica clínica como en la gestión?
4. ¿Cuál es el sitio adecuado de los Macrodatos y del DVR? ¿Cómo se sitúan dentro del marco de la investigación y la toma de decisiones (efectividad, eficacia, seguridad, planificación, etc.)?
5. ¿Cree que los Macrodatos y el DVR pueden aportar evidencia de causalidad en cuestiones donde no sea viable realizar estudios experimentales?
6. ¿Puede indicar ejemplos donde la evidencia experimental (ECAs) sea contradictoria con la derivada de los Macrodatos o de los DVR?
7. Pregunta clave (clínicos/decisores/evaluadores): ¿Cómo respondería a situaciones en que este tipo de nueva evidencia (DVR y Macrodatos) fuese contradictoria con la “evidencia clásica/experimental”?
8. Interfaz entre regulación (RGPD) y acceso a datos. ¿Cuáles son los riesgos de la utilización de DVR y Macrodatos para la protección de la privacidad de los pacientes y ciudadanos?

## 2. Transcripciones

1. ¿Qué iniciativas relacionadas con el empleo de macrodatos en nuestro país conoce por experiencia directa o por la literatura? ¿Algún caso notable de éxito o fracaso?

**Francisco Estupiñán.** Nosotros estamos trabajando en una encomienda del Departamento de Sanidad del Gobierno de Aragón, para desarrollar macrodatos sanitarios. Ese es el alcance del proyecto: en primer lugar, queremos disponer de la infraestructura y de la arquitectura de datos; hemos estado haciendo pruebas y ya tenemos algún desarrollo basado en esa arquitectura. Todavía hay mucho trabajo por hacer para poder salir a producción, pero lo que esperamos es tener al menos un prototipo de la parte del portal de gestión clínica, uno de los desarrollos en los que se basa esta infraestructura y la plataforma tecnológica, que podamos someter a prueba a lo largo de este año. Ese sería un ejemplo.

**Juan del Llano.** Y para el 2020, ¿qué tenéis previsto?

**Francisco Estupiñán.** El proyecto es para dos años, tal y como lo tenemos planteado ahora mismo, y el objetivo es desarrollar, por un lado, la parte de gestión y luego desarrollar la parte de investigación, los dos pilares básicos de utilidad; aunque ya hay grupos de interés - estamos muy en contacto con todos los grupos de interés de la CA- que están poniendo encima de la mesa otras necesidades, que se explorarán en próximas fases.

**Joan Rovira.** Una precisión, si estáis desarrollando un proyecto, habréis tenido que adoptar una definición de macrodatos.

**Francisco Estupiñán.** Es la definición cualitativa de las Vs: es una cuestión de tamaño, de variación de las fuentes, del tipo de dato y de la velocidad de procesamiento... Dado que lo estamos desarrollando conjuntamente en el IACS, entre la Unidad de Biocomputación y el grupo de Investigación en Políticas y Servicios Sanitarios (ARiHSP), cuyo investigador principal es Enrique Bernal, una de nuestras preocupaciones de siempre, especialmente con la experiencia del trabajo que hemos desarrollado en el atlas de variaciones, es la veracidad y calidad de los datos. Por ello estamos siendo muy cautos con todas las amenazas conocidas en el tema de protección de datos, estamos siendo extremadamente escrupulosos en el tema de calidad, desarrollando y aplicando metodología

previa. En ese sentido vamos lentos en el desarrollo; y la definición la ponemos por alcance: tipos de datos, fuentes de información, volumen, todo el conjunto.

**Pablo G<sup>a</sup> Mexia.** En la línea que indicabas, me sorprende mucho - y esto no es una idea mía, sino de dos expertas británicas – que siempre se habla de las Vs famosas, las 27, 28 o 47 Vs; pero a lo mejor mi perspectiva legal me hace insistir un aspecto que estas profesoras resaltaban muy bien, el hecho de que los macrodatos lo que implican es la capacidad de interrelacionar esa información, la capacidad de cuajar, casar esos datos y además de generar unas conclusiones y unas consecuencias brutales desde todos los puntos de vista que podamos imaginar, por supuesto los legales, de un calado enorme, pero también, lo veremos después, los médicos, económicos, etc.

**Jordi Colomer.** Yo no conozco ninguna iniciativa. Conozco preocupaciones, inquietudes que vienen de hace mucho tiempo, con el CSIC y con sus expertos en inteligencia artificial. Luego posteriormente en los congresos empezamos a hablar de lo que son las redes neuronales, que es una forma de estudiar los datos masivos a través de la conectividad. Desde entonces, veo inquietudes en algún centro. Yo empecé en un centro a organizar los datos, a hacer un gobierno de datos, que es lo primero que hay que hacer, creo yo, y luego puse un jefe o directivo de datos, que tampoco existía. Pero al final, el día a día prosaico de los centros es otro, y cuando quieres hacer cosas pioneras como estas, te quedas más solo que la una. A partir de aquí, aparte de voluntades y experiencias preliminares, no conozco nada más. Te puedo comentar alguna iniciativa extranjera que he visto por ahí, pero poca cosa. Hace dos años estuve haciendo una búsqueda sobre redes neuronales en Berkeley y no había nada, y esto es como hablar de la prehistoria. Creo que aún estamos en una fase más conceptual que práctica.

**Carolina González-Criado.** Os cuento algunos aspectos de nuestra experiencia. Aprovechándose del proyecto InnovaSaúde financiado con fondos europeos, Galicia desarrolló a partir del año 2014 una plataforma que se llama Hexin, que explota datos clínicos, pero con una tecnología de macrodatos. Enlazando con lo que dijo Pablo, el nacimiento de esta idea y de esta iniciativa surgió porque nos dábamos cuenta de que siempre habíamos ido hacia una historia clínica electrónica única, que fue un gran acierto en su día, y nos dábamos cuenta de que teníamos un volumen de información valioso y grande, porque constantemente en el Sergas se generaba un volumen de consultas, episodios de hospitalización, informes, analíticas, datos de laboratorio, imágenes digitalizadas,

videos, etc. que era impresionante. Dentro de esta información encontrábamos que había a veces algún formato estructurado, que hacía muy fácil analizarla, pero una gran parte de la información se encontraba en lenguaje natural, en formatos no estructurados o semiestructurados. A veces era un gran volumen de información que veíamos que no lo íbamos a poder procesar de ninguna manera; además toda esa información era tanto de gran valor asistencial, como de gran valor en investigación clínica, porque era la realidad a la que se estaban enfrentando los clínicos en nuestras consultas, y nos podía dar lugar a inferir conocimiento. A partir de ahí, se empezó a desarrollar esta plataforma, que es capaz de nutrirse de nuestros distintos sistemas de análisis del dato estructural, pero también es capaz de interpretar el dato semiestructurado y el de lenguaje natural. Pero hay que diseñar reglas de explotación, es lo que yo le llamo “enseñándole a la máquina a pensar”. Depende de lo que tú quieras estudiar, investigar, etc., tienes que ir dándole unas pautas a la máquina. Por ejemplo, una de las primeras experiencias fue llegar a detectar nuestra verdadera prevalencia para enfermedades como el EPOC o la diabetes. Pero no nos conformábamos en leer del CMBD, analíticas o episodios de hospitalización, queríamos darle alguna pauta más. Porque imagínate que una persona con EPOC nunca haya estado ingresada, o no haya tenido contacto con primaria. Queríamos que, a partir del lenguaje natural de los cursos evolutivos, etc., se pudiese hacer inferencias, y no solo con lenguaje natural, sino también a partir de los medicamentos al paciente. Muchas veces el paciente no está bien codificado porque el día a día de la consulta asistencial a veces nos abruma, pero tú sabes por su medicación que es un paciente diabético. En tal caso, podríamos a partir de la medicación que recoge el paciente tanto a nivel de receta como de farmacia hospitalaria que es diabético. Leíamos de todas nuestras fuentes y combinábamos, pero lo importante era ayudar a la máquina a pensar en algoritmos de decisión. A la máquina le decimos, por ejemplo: tiene que darse esto y lo otro en lenguaje natural para poder etiquetarlo. Fuimos avanzando; También realizamos una iniciativa para detectar posibles enfermedades raras. Y ahora ya la tenemos bastante pulida y la queremos poner al servicio del clínico para su labor asistencial. Por ejemplo, un clínico en su consulta se plantea: yo este año quiero revisar a todos mis pacientes con EPOC que no estén con el tratamiento estándar. Primero hay que saber qué pacientes tiene en su cupo y, a partir de ahí, dirigir su labor asistencial. Si por ejemplo existiese una nota de seguridad con un medicamento, esta plataforma te puede ayudar a seleccionar directamente qué pacientes tienes en tu cupo o ámbito asistencial con el medicamento en cuestión; así toda esta información que se pueda combinar en formato estructurado, semiestructurado o no estructurado tendrá gran valor asistencial, y también, gran valor para la investigación.

**César Velasco.** Yo os puedo contar algunas experiencias. Es verdad que cuando has hecho la primera definición de vida real, muchas de las experiencias que estamos definiendo como macrodatos, dependiendo de cómo hagáis la definición de datos de la vida real (DVR), entrarían dentro de esa definición, porque al fin y al cabo son datos sanitarios que forman parte de lo que nosotros consideramos historia clínica electrónica; aunque sea un repositorio de datos clínicos, no vienen de *insights* de utilización de aplicaciones móviles, ni de *insights* de utilización de terminales móviles con geolocalización, o sea, que cruzan fuentes de datos distintas. Pero con la definición que hacía el MIT hace unos meses del *patient cloud*, donde los datos de pacientes vienen de la historia clínica electrónica, de las aplicaciones móviles tanto pasivas como activas, de los aparatos y tecnologías sanitarias, y del uso del IoT (Internet of things). Según vuestra definición, macrodatos incluiría al menos dos de estas cuatro patas; y experiencias de análisis de datos masivos que junten más de una de estas patas hay muy poquitas. Que yo haya vivido directamente, podría mencionar una de Naciones Unidas, que utilizó datos de genómica, de historia clínica de paciente y de la búsqueda de casos mediante epidemiología de campo durante el brote de ébola, y se fueron a las compañías telefónicas a pedirles datos de geolocalización de terminales móviles. Juntaron estas tres patas e hicieron este tipo de análisis que mezcla bases de datos de más de una naturaleza. Todas las demás de las que os puedo hablar son realmente componentes de datos que están siempre en lo que se considera historia clínica de pacientes. En cuanto a estos últimos, todos conocéis la experiencia que ha tenido AquAs: a través de la central de resultados sacó una plataforma para la consulta de datos masivos del sistema sanitario, que, aunque tuvo ciertos problemas, ahora podemos consultar. Recientemente se ha lanzado otra plataforma que se llama Padris, que es también una plataforma de acceso a los datos masivos de la sanidad de toda la CA de Cataluña, los datos están disponibles, aunque con muchas limitaciones. Dentro de lo que estamos haciendo más allá de esta plataforma de consulta, estamos utilizando muchísimo macrodatos para genómica, sobre todo en enfermedades raras (EERR), especialmente para la identificación de nuevas dianas terapéuticas, y en oncología. Además, estamos utilizando campo texto y lenguaje natural para la identificación de pacientes subsidiarios de reembolso mutual; eso se está haciendo en el Catsalut con un programa. Lo que se ha hecho es volcar todos los campos texto de la historia clínica compartida catalana, en TXT. A estos se les pasan unos algoritmos de lenguaje natural que seleccionan los pacientes que son subsidiarios de reembolso por una mutua o una federación deportiva, para aumentar el número casos detectados. Está en fase piloto bastante avanzada.

**Mercedes Alfaro.** ¿Perdona, son datos administrativos?

**César Velasco.** No, son datos clínicos. Son historias de pacientes, informes de alta, todo lo que está en la historia clínica compartida. Otros ámbitos donde estamos intentado entrar todavía más, dentro de lo que es el análisis de campo texto, lo estamos trabajando tanto con Savana como con otras dos empresas que trabajan con datos agregados. Por un lado, en un proyecto sobre enfermedades minoritarias estamos intentado no tener que utilizar la historia clínica habitual para marcar al paciente con enfermedad minoritaria, sino que se marquen de forma automática en el sistema mediante la definición de los criterios de laboratorio, de tratamiento y clínicos. De tal manera que reconoceremos a los pacientes y los tendremos marcados en el sistema, que actualmente no lo están. Al estar diagnosticados con la enfermedad minoritaria se pierden, es un problema que tenemos ahora mismo; como no están diagnosticados al alta, porque muchos de ellos no tienen una historia de alta, y el CMBD no cubre toda la casuística de enfermedades raras, queremos intentar entrar en los sistemas de información de EERR de una manera un poco más original y eficiente. Entre otras cosas interesantes y fuera del ámbito más convencional, estamos en un proyecto en alianza con los hospitales más significativos en el ámbito de innovación de Europa para generar cuadros de mando de gestión sanitaria que estén orientados a la predicción, pero sobre todo a la prospección de la toma de decisiones, a valorar indicadores clave para la decisión (*key decision indicators*), en vez de indicadores de comportamiento (*performance*), que estarían más orientados a lo que hemos hecho, es decir, a intentar aplicar analítica predictiva a las grandes bases de datos de los hospitales, a veces combinadas con datos externos, como puede ser la utilización del transporte para llegar al centro, la climatología de la ciudad en la que estamos y otros indicadores, tal como el consumo energético. (Estamos haciendo una revisión sistemática de la literatura y Delphi con una escuela de negocios para ver cuáles son los indicadores más adecuados a aplicar). Para hacer el cálculo predictivo del consumo energético de las instituciones, necesitamos datos que están fuera del entorno habitual de la historia clínica, y que entrarían en la definición de la que hablábamos. Por último, también estamos entrando un poquito en el campo del chatbox (*asistentes electrónicos virtuales, como "siri" de Apple, pero en versión teleoperadora*), la inteligencia artificial aplicada a la conversación humana. Hay una experiencia importante que se llama Babilon, que están llevando a cabo ahora mismo en Londres con médicos generalistas (*general practitioners o GPs*), con la que ya están dando servicio de GPs a la población de la ciudad de Londres a través de una aplicación de *chatbox*; nosotros estamos mirando si podemos empezar a

dar servicios de teleoperadora a través de este tipo de análisis de datos. Yo pensaba que era solo interpretación algorítmica de datos de voz, pero realmente sí tiene un componente de análisis, ya que interpreta datos porque capta las conversaciones previas, aprende de estas conversaciones, y de esta manera ofrece un mejor servicio de *chatbox*. Babilon es algo que merecería la pena incluir dentro del estudio. También trabajamos con ensayos clínicos de vida real, pero no se utilizan solo DVR, sino también macrodatos, porque emplean la historia clínica del paciente, pero también los procedentes de la sensórica y datos de aplicación móvil, para sustituir la clásica CRO (organización de investigación por contrato, ([https://es.wikipedia.org/wiki/Organizaci%C3%B3n\\_de\\_investigaci%C3%B3n\\_por\\_contrato](https://es.wikipedia.org/wiki/Organizaci%C3%B3n_de_investigaci%C3%B3n_por_contrato))). Sanofi está haciendo un estudio de DVR importante con sensores para diabetes, mediante glucómetros portátiles, que es bastante interesante. Hay algún otro laboratorio que también está empezando a hacer este tipo de ensayos de vida real. Esto nos evita todos los pasos del CRO clásico, como cumplimentar cuestionarios, y utiliza un poco más las bases de datos masivas.

**Jaime del Barrio.** En mi opinión no hay todavía ningún ejemplo de éxito o fracaso en nuestro país digno de ser destacado. Algunos casos, como comenta César, por ejemplo, Savana, tienen buenas expectativas, pero hay que aterrizar todavía más. Cuando lo intentas trasladar a algún sistema regional sanitario de nuestro país, tiene cosas mejorables, por decirlo de alguna manera. Conozco muchos proyectos que se están haciendo en todas y cada una de las CCAA, pero todavía son “pilotos”, iniciativas, preocupaciones, inquietudes, llamarlo como queráis. Esto me preocupa muchísimo, porque está involucrada mucha gente, dinero, ilusión, etc., y un porcentaje altísimo está abocado al fracaso en muy poco tiempo. El objetivo final de todo esto es hablar, compartir, evaluar, y en base a esto tomar medidas. La realidad es que somos 17 sistemas regionales de salud, cada uno de ellos con sistemas informáticos operativos diferentes, y todo lo que os podáis imaginar que de esta atomización se deriva. Incluso dentro de las CCAA, pasa lo mismo con los diferentes hospitales. Cada uno tiene sus propios sistemas operativos, sus bases de datos, cada uno cuenta y mide lo que cree oportuno, y la mayoría están reinventando las cosas, y a mí esto me preocupa por lo que se va a perder, dinero, trabajo, ilusión, y no vamos a avanzar mucho. Mi percepción ahora mismo, y perdona Carolina, es que probablemente la única CA en que yo siento que ha calado en toda la organización es Galicia. En Galicia no es cuestión de sistemas operativos, sino de cultura, de profesionales. Hay toda una organización que va a una. Y los que inicialmente eran reticentes ya se han subido a la ola. En resumidas cuentas, a mí me pre-

ocupa que mientras estamos aquí con “pilotos”, el mundo sigue girando y me temo que nos están chupando y vampirizando datos, vete tú a saber de qué manera, y dentro de poco otros nos van a dar las soluciones a muchas de estas cosas. Deberíamos ser más humildes, compartir más, hablar más sobre qué valor aporta realmente cada uno en esto. Y aquí hablo de lo público, de lo privado y de las empresas. Y termino: a nosotros, como consultora, nos están pidiendo cada vez más proyectos para validar este tipo de cosas; el 80% de nuestros clientes son de fuera de España y consisten en partenariados entre farmacéuticas, tecnológicas y universidades. Por ejemplo, César, sé que estáis haciendo un estudio de genética para enfermedades raras. Pero con datos de fuera, ¿no? Porque el tamaño de la base de datos que vas a tener en Cataluña es limitado.

**César Velasco:** Depende. Este tipo de estudios se hacen con biobancos de genómica y realmente lo que hacen es utilizar muchos datos de individuos sanos. Pero estoy completamente de acuerdo, hay alguna iniciativa europea de genómica.

**Jaime del Barrio.** Ya hay iniciativas en temas de predicción y prevención, como contabas antes, por ejemplo: para ver la evolución de la fiebre asiática; esto se ha hecho perfectamente. Se ha hecho también para la ubicación de dispositivos de emergencia en el terremoto de México. Nosotros hemos hecho un proyecto pequeño en base a perfiles de población para determinar dónde se pueden ubicar unidades de transfusión de sangre. Es un proyecto en el que no tienes dificultades de acceso a datos cumpliendo la normativa de protección, ya que puedes acceder a bases de datos públicas y en base a resultados decides nuevas ubicaciones de las unidades y con ello se ha aumentado la donación en un 30%. Esto son casos reales. El otro día estaba viendo temas del IBM Watson aplicados a medicina personalizada en oncología, y ya empieza a haber una plataforma en la que confluyen millones de datos, y no en *pdf* como hasta ahora, si no datos con los que se puede trabajar. Y actualmente están buscando hospitales con los que validar clínicamente lo encontrado hasta ahora.

En este sentido echo en falta un discurso cohesionado, lineal, potente, de todo el Sistema Nacional de Salud; me daría pena ver el fracaso, principalmente por atomización, de todo lo que se está haciendo ahora.

**César Velasco.** Comparto completamente lo que ha dicho Jaime.

**Juan del Llano.** Antes de darle la palabra a Mercedes para intentar ayudar a categorizar la respuesta desde fuera - pues es más fácil ver los toros y ser analista utilizando exclusivamente el sentido común – quiero comentar que, a mí, en políticas de salud digital, que es a lo que se refiere Jaime, me gustaría pactar un decálogo, unas grandes líneas. Sería muy sensato. Sobre todo, para no perder oportunidades haciendo millones de pilotos de los que muchos seguramente no van a llegar a ninguna parte. ¿Lo ves posible, desde el Ministerio?

**Mercedes Alfaro.** Estoy completamente de acuerdo con lo que comentaba Jaime, que sería una pérdida de oportunidades muy importante que se dispersasen los esfuerzos, y de hecho, ya se están dispersando. El problema es que nuestro sistema sanitario, incluso nuestra cultura de país es bastante individualista. Yo sé que eso no es una excusa, se puede cambiar y se puede promover un cambio. Pero os voy a contar dos ejemplos. Uno es la base de datos clínica de atención primaria (BDCAP), que ya funciona después de muchos años de esfuerzo. Una base de datos que se ha trabajado con las CCAA desde el minuto cero. Determinar qué nos interesa saber a todos, es algo que se ha sometido a consenso por activa, por pasiva y por reflexiva. Bien, ¿cuál es la situación real de la BDCAP? La BDCAP es una base de datos muestral que hace primero una selección de zonas básicas, como unidad primaria de muestreo; esas zonas básicas se quedan ya fijas en la muestra y, por lo tanto, la población que tu sigues a lo largo de los años, es una gran cohorte de 4 o 5 millones de personas, teniendo en cuenta, claro está, que habrá nuevas personas que irán a vivir a estas zonas y se incorporarán a la muestra, o, por el contrario, otras se irán a otros lugares de residencia y saldrán de la muestra. Pero es un seguimiento de una población concreta y desde ese punto de vista es muy interesante, porque además es una base de datos que se puede cruzar perfectamente con la tarjeta sanitaria para obtener variables sociales y económicas; o se puede cruzar con el CMBD para saber qué ha pasado con los pacientes que van al hospital. ¿Cuál es la realidad? Que está costando Dios y ayuda extraer los datos de las historias clínicas de una manera estructurada y útil para su procesamiento en la BDCAP. Hay CCAA que nos están pudiendo mandar el colesterol HDL y no nos pueden mandar el total, por ejemplo. El secreto posiblemente está en la manera de extraer los datos, en el extractor. Cada CA tiene una historia diferente en Atención Primaria, pero todas deberían poder proporcionar la información necesaria pues, aunque tiene cientos de variables, se ha ido casi a mínimos, si comparamos con la ingente información que tiene la historia clínica Y está costando mucho trabajo. Por eso yo sonrío siempre un poco cuando se habla de macrodatos, porque esto no son “macrodatos”, sino una base de datos muy grande,

que no se está sometiendo a un análisis estadístico que pueda considerarse como inteligencia artificial, En resumen, está costando tener unos datos de una calidad adecuada para obtener unas conclusiones fiables con una estadística tradicional. En el caso del CMBD, ahí la cosa está más pulida, tiene muchos años de desarrollo y consenso y ya se está trabajando con potentes motores de cálculo y ajuste de riesgo basados en redes bayesianas, que nos permite hacer ajustes por diferentes variables, demografía, riesgo de muerte, severidad, comorbilidad, etc. Yo pienso, en cualquier caso, que antes de avanzar, sería muy bueno definir hacia dónde queremos ir en el Sistema Nacional de Salud (SNS). Hay otros actores con los que hay que trabajar, pero, para empezar, hay que saber hacia dónde queremos ir nosotros y cuáles son los requerimientos básicos que todos tenemos que compartir para que en un futuro se puedan establecer plataformas, donde sea, pero que entre todos hayamos sentado las bases para poder participar de las plataformas que interesan a cada CA. Yo sí que creo que es posible, hasta el punto en que estamos trabajando ahora mismo con la SEIS, la Sociedad Española de Informática de la Salud, como punto de partida, a través de un convenio de colaboración que se firmó hace un año más o menos. La idea es intentar definir ese marco básico común, comentarlo con las CCAA, llevarlo a los comités, a la comisión de sistemas de información, etc. Y empezar a estudiarlo y debatirlo todos juntos.

**César Velasco.** Aquí hay un concepto que es la densidad digital, que realmente de lo que habla es que el mundo analógico que vivimos cada vez está más digitalizado. Generalmente lo que hacen los estudios de macrodatos, a nivel internacional, lo que está haciendo Oxford o los hospitales universitarios de Paris (APHP) es plantearse con qué datos contamos que se estén recogiendo ya, pero que no están integrados, sino en el *cloud* de paciente (base de datos online) que igual Medtronic tiene en su base de datos, o Toyota, o la compañía telefónica, mientras nosotros tenemos la nuestra; ¿Cómo podemos agruparlo todo en un sitio y hacer proyectos que den valor? Muchos proyectos de estos van orientados a la investigación científica, pero lo que están haciendo es captar los datos que se están recogiendo ya, y ver qué preguntas se pueden formular. Hay algún estudio que lo que hace, es decir: nos gustaría saber esto, y vamos a ver qué datos se están recogiendo que nos puedan ayudar a contestar lo que se supone que nos trae el mundo de los macrodatos. Esto te permite hacerte preguntas a ti mismo y ver qué datos tienes disponibles. Pero esto es un paso posterior.

2. ¿Conoce algún ejemplo de nuevo conocimiento generado mediante el análisis de macrodatos que se haya aceptado e incorporado en la práctica real con un amplio consenso?

Todos: No.

**Carolina González-Criado.** Siguiendo la pregunta, y complementando lo que dijo César, nosotros vamos por partes, por áreas terapéuticas, por cosas que queremos desarrollar. Yo creo que en macrodatos hay tres elementos clave: 1. Tienes que trabajar para enriquecer la calidad de esos datos; a lo mejor, con una primera extracción, haces un estudio de validación, de cómo reproduces una realidad, y luego hay que enriquecerla. 2. Tienes que tener muy claro lo que quieres medir, o sea, hacia dónde vas; necesitas una estrategia. 3. Otra cosa que yo creo que va a salir con la última pregunta es la gobernanza de los datos, los temas de seguridad, etc.

Nosotros hemos empezado a trabajar en muchos campos, y uno de ellos es la oncología, porque ahora mismo es una patología donde la innovación terapéutica se está incorporando a una velocidad muy grande y tenemos un proyecto que se llama “FIRES” (Farmacoepidemiología: Información más resultados en salud), donde nos preguntamos qué variables o datos utilizan nuestros oncólogos para prescribir y para hacer el seguimiento, y cuáles utilizan los farmacéuticos para la gestión de los tratamientos, pensamos en todo, y lo extrajimos y volcamos todo en un sistema de análisis, para ver qué encontrábamos. Y encontramos siete lenguajes distintos para identificar lo mismo, que se corresponden a las siete gerencias integradas que tenemos en el Servizo Galego de Saúde. Entonces concluimos que, si queremos tener un resultado más fiable, mezclándolo todo, tenemos que enriquecer esa calidad de la información mediante la homogeneización y normalización de la misma. Y otra cosa que da muy buenos resultados es comunicar lo que vamos a hacer, como lo vamos a hacer y para qué. Así nos reunimos con los oncólogos, y les mostramos la foto de lo que habíamos encontrado. Ellos fueron los primeros que se quedaron impresionados, por las muchas conclusiones que se podían sacar de ahí. Pero ellos también se dieron cuenta de que nos faltaba enriquecer la calidad de esos datos. Y ellos fueron los primeros que se subieron al carro de la normalización. Ellos mismos se preguntaron por qué no registrar de la misma manera un diagnóstico específico si siempre se va a realizar el registro del diagnóstico del paciente; si todos nos po-

nemos de acuerdo en cómo le vamos a llamar al cáncer de próstata y decidimos llamarlo CA de Próstata, dejemos de emplear otros términos como neoplasia maligna de próstata o NEO de próstata. Y también surgió la necesidad de homogeneizar cómo vamos a clasificar el estadio, la lateralidad, etc. Ellos mismos se dieron cuenta de que para enriquecer la calidad era importante trabajar en un lenguaje común, vieron la potencialidad que esto tenía y sin necesidad de aumentar esfuerzos. Lógicamente, no están a favor de duplicar registros, cosa que nosotros como servicio de salud también veíamos inviable. Pero sí estaban por normalizar el lenguaje a la hora de registrar. Eso es muy importante en todas estas tecnologías de macrodatos, para aumentar la calidad.

Han entendido que es algo muy importante, gracias a que nos hemos sentado con ellos para explicarles qué vamos a hacer, cómo lo vamos a hacer, y qué resultados esperamos. Y que les vamos dando retroalimentación de lo que obtenemos, y ellos mismos se suben al carro.

**César Velasco.** Cosas que se hayan hecho aquí no conozco ninguna. Para terminar, unos comentarios que igual podían aportar. Que yo sepa, respecto a conocimientos que estemos aplicando que provengan de macrodatos, solo se me ocurren algunas dianas terapéuticas de terapias moleculares. Estas son globales, soluciones que dependen de proyectos internacionales de la industria farma o diagnóstica, con fines comerciales. Al final las iniciativas locales no van a llegar a ningún puerto, y un fenómeno que ha comentado Jaime y que es muy interesante, es la desintermediación. Van a venir compañías internacionales que ya están haciendo esto, nos van a ofrecer una solución, y acabaremos aceptando, si no nosotros, nuestros pacientes. Esto ha pasado con servicios metropolitanos de transporte, que estaban generando sus servicios de información de manera independiente, hasta que llegó Google Maps, ofreció el servicio, y ahora mismo los servicios metropolitanos de transporte pagan dinero a Google Maps para saber cómo se mueve la gente en nuestras ciudades. Nos va a pasar algo parecido en salud. Lo último que voy a comentar es que cuando planteabais sanitario o salud, si habláis de salud, yo sí que incluiría todo el concepto que no tiene que ver con servicios sanitarios; es decir, tanto en salud como en calidad de vida, en alimentación, en marketing minorista, se están haciendo muchas cosas con macrodatos que nosotros no estamos haciendo; pero a nivel de segmentación, en marketing digital de productos específicos, productos de gimnasio, productos de consumo, suplementos nutricionales, están utilizando macrodatos y ya están comprando datos de *profiling* o personalización de personas para hacer publicidad dirigida (segmentación de campañas de marketing

mediante Macrodatos). Y esto son macrodatos y se está haciendo ya, y tiene un impacto en salud, aunque los sistemas sanitarios no estemos ni acariciando el gran potencial que tienen estas cosas.

**Pablo G<sup>a</sup> Mexia.** En la onda de César. Efectivamente, a propósito de las prácticas de algunas entidades, sobre todo multinacionales, se está haciendo tanto, que se está empezando a hacer viciadamente. En concreto, ya hay Autoridades de competencia, sobre todo la alemana, la francesa también, las de la UE están un poco más atenazadas; pero claro están viendo que el algoritmo empieza a aprender que uno sube el precio, lo baja también, y jugando con la inteligencia artificial lo empieza a combinar; y al final claro, cada vez que intentáis buscar el hotel más barato, o el vuelo más barato, es imposible encontrarlo, porque el algoritmo listísimo se va entrometiendo. Bueno, que sepáis también que ya se está actuando legalmente, aunque todavía está muy focalizado. Y un apunte rapidísimo, también en la lucha contra las *fake news* se está empezando a utilizar algoritmos, a efectos de aprender.

**Juan del Llano.** Paco, según vosotros en el IACS, que tenéis mucha experiencia en *benchmarking* (tomar “comparadores”): ¿Podrían los macrodatos ayudar a que tomase más realidad la evaluación por comparación, y hasta cierto punto, que el ciudadano acabara decidiendo en función de información pública de médicos, rendimiento clínico, etc.? ¿Todo esto lo ves posible?

**Francisco Estupiñán.** Yo lo veo como una cuestión de alcance. Sería un alcance posible, pero muy lejano. Por medio hay un montón de problemas que hay que solucionar. Puede que haya una solución técnica, que sea posible aplicar, pero no se ha encontrado todavía.

Tenía un comentario sobre lo anterior. Hay mil cosas de los macrodatos, o de lo que se mueve alrededor de esta metodología, que, aunque ya están implantadas, nos pasan desapercibidas, como es el uso de una técnica de inferencia estadística que permite llegar a unas conclusiones diferentes a las que se llegaban antes, básicamente porque ahora puedes explicarlas, pues tienes más capacidad de computación y más datos. Cosas que nos pasan completamente desapercibidas, como la mejora de la definición de la imagen diagnóstica. Eso no sería posible si antes no se hubieran procesado millones de imágenes para mejorar la nitidez de la ecografía o la precisión de la resonancia magnética. O en oncología, la aplicación del tratamiento radioterapéutico con acelerador. La localización del tumor, la dosis de radiación, todo eso es al final es un caso de éxi-

to que nos pasa completamente desapercibido. Porque ya se ha asumido como una cosa normal en nuestra vida. Cuando lo llevamos a la gestión de personas, cuando lo llevamos a la gestión de sistemas de organizaciones lo tenemos un poco más complicado, porque la introducción de cuestiones que afectan a un montón de gente no es tan sencilla. La forma de introducir la información en el sistema es complicada.

Un tema general que ha estado mucho tiempo sobre la mesa es el de interoperabilidad. Se hablaba por aquí de la interoperabilidad técnica de sistemas, de la semántica. Estos son problemas a los que nos estamos enfrentando. En estos temas soy optimista. Creo que el Ministerio está haciendo una labor bastante interesante planteando a nivel nacional un esquema de interoperabilidad en el sector sanitario, que tiene sus historias, pero es necesario seguir creciendo. Entre otras cosas, porque nos enfrentamos a algo muy complejo y muy heterogéneo. Pero las bases para poner un esquema común de interoperabilidad que nos permita crecer en el sentido de estar conectados y de poder hacer las cosas que todos queremos hacer, están ahí. En un alcance más pequeño, el alcance técnico de sistemas de información, la primera derivada que nos hemos encontrado nosotros, en fase de desarrollo, la primera derivada que ya ha tenido un beneficio directo para el sistema sanitario es que nos hemos encontrado con problemas del sistema de información que no hubiéramos descubierto si no nos hubiéramos puesto a rascar. Dificultades semánticas entre los propios grupos de interés, que no hubiéramos encontrado si no nos hubiéramos sentado en algún momento todos en la misma mesa. Una de las cosas que estudiamos en el Atlas, es que en el CMBD existen patrones de codificación. Porque los codificadores van por cada hospital a las unidades de codificación, y han aprendido de un maestro. Y del mismo modo que existen variaciones en la práctica clínica que son directamente atribuibles a un estilo de práctica médica, existen variaciones en la codificación que son directamente atribuibles al lugar de donde se ha codificado ese registro. Este tipo de cosas no las ves hasta que no te pones a explorar.

**Jordi Colomer.** En realidad, no estamos hablando de macrodatos. En los ejemplos que se han explicado, en mi opinión, estamos tratando como máximo de CMBD, etc., y pienso que eso es importante, porque así la gente empieza a concienciarse de que tiene muchos datos y no se utilizan. Hasta ahora, ahí estaban, aburridos. El enfoque macrodatos es lo que nos ha empujado a decir, ¿qué hacemos con esto, radiografías, CMBDs, etc.? Pero esto, en sí mismo, para mucha gente no es macrodatos. Yo soy bastante talibán en este sentido con

aquello de coger datos a ver que sale. Siempre se tiene un objetivo. Y el gran mérito de los macrodatos es que empiezas con unos datos, a medias metes más, y te preguntas qué ha pasado con los que no he cogido y que pudieran ser importantes. Porque, aunque tengamos millones de datos en el mundo, los macrodatos está tratando solo el 2%. ¿Y cuántos datos no he tomado porque no he caído en ello? A los que trabajan con macrodatos generalmente no les preocupa la precisión del dato, que no es lo mismo que el gobierno de datos que habéis dicho. Prefieren, dentro de la incertidumbre, ver qué sale. Y luego, se está buscando la causalidad, pero macrodatos no identifican causalidad, sino correlación, que no tiene nada que ver. Esto es un cambio de mentalidad brutal.

**Mercedes Alfaro.** La causalidad hay que estudiarla después.

**Jordi Colomer.** Los macrodatos te orientan con las correlaciones por si más tarde puedes encontrar una causalidad. Pero su función no es esa. Ni la función tampoco es que las máquinas aprendan, esto también es un error. Lo que tienes que hacer es darle muchos más datos, con formas matemáticas, para que te calcule probabilidades que son correlaciones. Lo digo porque a veces se confunde macrodatos con inteligencia artificial, y no es lo mismo. Los macrodatos son otra historia. En mi opinión, aún nos queda mucho para llegar a los macrodatos. Pero sí que es importante que haya una política de la gente responsable, que se preocupe por esos datos. Esto no lo desmerezco. En medicina, en 200 años, tras mucha investigación, nos dimos cuenta de que lo importante era lavarse bien las manos. Ahora nos damos cuenta de la importancia del dato. Y en estos momentos es más barato guardar datos que limpiarlos. Porque es tan barato guardar datos, que muchas empresas tipo Amazon o Google van almacenando datos que vete a saber si algún día se utilizarán para algo. Porque saben que incluso los datos que no se utilizan inmediatamente, en último término generarán productividad. Esto es adelantarme 5 o 6 años. En definitiva, en el tema de los macrodatos aún queda un trecho muy grande por recorrer.

**Jaime del Barrio.** Estoy básicamente de acuerdo en todo lo que habéis dicho. Yo soy un optimista también, como Paco. Cuando Paco decía que estábamos consiguiendo hitos, miramos lejos. Y mirar lejos es lo que hace Jordi. Todo lo que estamos contando aquí, tomando como ejemplo la robótica quirúrgica, no dejan de ser si no los avances que tocan ahora, basados en nuevas tecnologías. Pero es verdad que al mismo tiempo estamos aprendiendo todos. Los sistemas, los profesionales, etc. Al hilo de lo que comentaba César, no perdamos de vista que estamos en una economía de datos. Y el sector salud es interesante, pro-

bablemente de los más interesantes para ir “a pescar datos”.

**Jordi Colomer.** Ojalá que los hospitales pusieran en sus activos el balance de los datos que tienen.

**Jaime del Barrio.** Por ejemplo. Y eso lo han entendido cantidad de empresas de otros sectores que vienen a Salud, repito “a pescar”. Google, Facebook, Microsoft, ... vienen aquí, por economía de datos.

**Jordi Colomer.** Y saben que tienes los datos.

**Jaime del Barrio.** Dos comentarios. 1. Aquí estamos desarrollando, pilotando. Eso sí: a ver quién me lo certifica, a ver quién me lo patenta, a ver quién me lo registra. Y miramos al sistema sanitario público. Mientras lo estamos haciendo desde aquí, por encima está la nube, y en la nube están todos los que están aportando datos, los pacientes, los ciudadanos. A mí se me abren las carnes cada vez que veo entidades del sistema de salud que se van a vender los datos a grandes empresas, y van a cobrar grandes cantidades para activar la patente. 2. Nos van a inundar de esto como nos han inundado de WhatsApp, gratis total. Y aquí estamos haciendo modelos de negocio para ver a quién le cobramos y cuánto le cobramos a una empresa farmacéutica, con la anuencia del sistema sanitario público. Este es el modelo que hay ahora. Cuando realmente, todas estas empresas que trabajan en economía de datos a lo que van es a datos, datos, datos, gratis, gratis, gratis, modelo WhatsApp. Y el mayor caladero de datos ahora mismo es el sector salud. Google le dedica el 33% de su inversión.

**Mercedes Alfaro.** Claro, es el sector más sensible, porque todos estamos dispuestos a arruinarnos por tener salud. El dato vale lo que vale el bien, y el “bien” salud vale mucho y, además, desgraciadamente, se considera como un bien de consumo más. El miedo de todo esto es hasta qué punto contribuimos a medicalizar innecesariamente la vida. Una vida medicalizada me parece rechazable desde todos los puntos de vista.

**Jaime del Barrio.** Mercedes, ahora mismo ya hay cantidad de planes y de proyectos que han medido la adherencia a los tratamientos y donde se ve que en algunas patologías el 50% de los pacientes ni toman lo que se les prescribe, ni cuándo se les prescribe. Para un uso adecuado de los medicamentos y la solución de todo esto, están las herramientas, sobre todo las informáticas. Pero nos tenemos que poner de acuerdo con nuestros compañeros los médicos y los farmacéuticos. Y termino: ¿Cómo conseguiríamos que todo el mundo aportase

datos al Sistema Nacional de Salud? Como ha hecho el Ministerio de Sanidad en otras épocas, dándole una zanahoria a cada uno. ¿Por qué dan datos las CCAA al Plan Nacional del Sida o al Plan Nacional de Drogas? Porque tenían un poquito de dinero. Todo el mundo daba datos a los planes nacionales en los que había un poquito de dinero. ¿Cómo se haría un plan integral de salud? Un poquito a cada uno, para que me den los datos, en tiempo y en forma. Es sólo el 10% del presupuesto, pero estimula.

**Pablo G<sup>a</sup> Mexia.** Pensando más en clave de la posible proyección de las cosas que digamos a políticos y a políticas, en todas partes cuecen habas. Recordad que para la OCDE hay dos temas digitales clave en salud: uno es precisamente los dispositivos (*devices*), y otro justamente la historia clínica, la necesidad de ir recabando información y centralizarla a escala nacional. Solamente 18 países de todos los que componen la OCDE, que son cerca de 40, tienen esto ya unificado. Y alguno está dando señales clarísimas, como Austria, Canadá o Suiza, pues siguen sin un historial clínico electrónico centralizado a escala nacional. Lo decía por romper una lanza a favor del optimismo.

**Mercedes Alfaro.** Optimistas hay que ser, pero, por ejemplo, en el tema de la historia clínica, España es un país descentralizado. No existe una historia única salvo algunas excepciones. Lo que se ha desarrollado aquí en buena parte son visores. Puedes atender al paciente porque ves su historia, pero no hay una integración de la información que tu produces con la que producen otros profesionales. Y, de hecho, por ejemplo, la historia clínica interoperable del Sistema Nacional de Salud consiste en once documentos clínicos que se supone que son clave, pero, de ellos, solo dos -que son el resumen de historia clínica de atención primaria y el resumen de historia clínica del paciente-, son de verdad interoperables, “integrables” con otra información. Son un destilado de la historia clínica de ese momento, que, junto a un extractor de datos, va a un modelo de datos común que puede ser consolidable en destino. Y que se está compartiendo. Lo demás son unos *pdf* muy bien estructurados, con unas codificaciones iguales y con un modelo de datos que establece adónde va cada dato común, pero al final son documentos, son *pdf*. Y España es de los países que más desarrollada tiene la interoperabilidad.

**Jaime del Barrio.** El ejemplo que estabas poniendo me sirve. Proyectos como Savana leen un *pdf*, un Word y un Excel. Pero luego el clínico o el paciente le hacen una pregunta, y Savana quizás no le puede responder, hasta este momento, a la pregunta.

**Jordi Colomer.** Es que está todo muy verde.

**Jaime del Barrio.** En cualquier caso, estoy de acuerdo en no flagelarnos tanto. Es bueno autoexigirse, pero sales de España y tampoco están tan avanzados.

**Joan Rovira.** Yo la idea que tenía de que macrodatos implicaba fundamentalmente meterse en la red, identificar bases de datos existentes, mirar las búsquedas que hace la gente y de esta información inferir, por ejemplo, cómo está evolucionando una epidemia, con una mayor rapidez que si he de esperar a que se procesen las declaraciones oficiales. Me parece muy bien que lo que se hace aquí algún día pueda llegar a considerarse macrodatos por el volumen de información acumulada, pero asociaba el enfoque macrodatos con la búsqueda y utilización en directo de datos que se generan espontáneamente por otras razones.

**Todos:** no.

**Jaime del Barrio.** Tú tienes un objetivo.

**Jordi Colomer.** Por ejemplo, en EE.UU. querían saber en qué viviendas podría haber drogadictos, etc. Entonces lo primero que hicieron fue ir al censo y surgen problemas de datos porque se enumeran diferentemente las calles, así que lo ponen todo ordenado, y empiezan a recoger datos. Y con esto pueden ver más o menos cuántas ambulancias han ido a ciertos sitios, a cuánta gente han atracado, y van cogiendo datos que saben que están relacionados con su pregunta, aquello que pueda explicar que aquel edificio es sospechoso de que haya narcotráfico o lo que sea.

**Joan Rovira.** Pero en este ejemplo que has dado se buscan datos existentes de otras instituciones, no se generan nuevos datos, mientras que muchas de las iniciativas descritas anteriormente implicaban generar nuevos datos.

**Mercedes Alfaro.** Para extraer datos.

**Jordi Colomer.** Para que veáis la complementariedad en este ejemplo que he explicado, en un momento en que el tema estaba bastante avanzado, notaban que fallaba en cuanto a encontrar edificios según aquellos parámetros que habían buscado. En cambio, cuando observaron además de la existencia de un vecino que se acompañaba de unas reformas en el edificio, se dieron cuenta que no habían introducido los datos de los permisos de mejoras en los edificios,

al dar por contado que eran edificios abandonados. Es decir, que se van enriqueciendo los análisis con el tiempo, buscando datos relacionados con el fin de afinar el objetivo inicial, que era identificar edificios que potencialmente albergaran narcotráfico, como en este caso. A pesar de tener un objetivo, siempre hay dudas sobre si hay indicadores relevantes que no se han incluido. Siempre hay que tener presente está a duda. Y por más digitalizado que pensemos que existe, estamos digitalizando el 2 o 3% de lo que lo está en el mundo y esta limitación se debe tener en cuenta.

**Carolina González-Criado.** Yo también creo que a los macrodatos les toca avanzar. Pero la plataforma gallega “Hexin” es macrodatos, porque también lee el lenguaje natural. Todavía no nos sirve para todo. Hay que ir diseñando por campos, porque le tenemos que enseñar a la máquina a pensar, pues el lenguaje natural todavía no piensa solo. Es verdad que los macrodatos te ayudan a predecir, pero el dato fiable quizá te lo da el ensayo clínico, o uno pragmático. A mí me impresionó que dijeran que Obama, en su segundo mandato, ganó las elecciones gracias a un estudio de macrodatos que hizo antes de la campaña electoral, un estudio sobre las preocupaciones de los americanos, por el que sabía tanto de sus votantes, como de los indecisos y de los que no pensaban votarle. Y estudiaba todo, qué compraban, adónde iban, las tarjetas de crédito, facebook, etc. Y al final ganó las elecciones.

**Mercedes Alfaro.** Los datos son una materia prima para muchas cosas, está claro.

**Jordi Colomer.** Pero la máquina no es que aprenda, sino que con el mismo algoritmo y con millones más de datos de los que tenía es capaz de sacar una probabilidad distinta.

**Mercedes Alfaro.** Para mí los ejemplos claros de macrodatos y que no están en el sector salud, son el Idealista y Fotocasa, que son los que llevan la política de precios de la vivienda. En España no hay una estadística de precios de venta de viviendas - que en cualquier caso siempre iría detrás de la realidad, porque iría a año vencido - pero podría servir para hacer un análisis retrospectivo y corregir algunos problemas y dar más transparencia. Quien tiene unas bases de datos tremendas sobre cuáles son las características de los compradores, en qué barrios miran, quién pide un préstamo, qué tipo de trabajo tienen, etc. son Idealista, Fotocasa y similares. Ellos están decidiendo lo que sube o baja el precio de la vivienda.

**Jaime del Barrio.** ¿Queríais datos reales? Tomad datos reales. Hace unos días hablaba con una persona del INE y me decía que estaban pendientes de que aprobasen los presupuestos generales del Estado de 2018. ¿Para qué? Para contratar a encuestadores, para empezar a hacer las encuestas ¡del año pasado!, cuando estén los resultados, estarán obsoletos.

**Mercedes Alfaro.** Quizás tenemos que cambiar el paradigma y comprar los datos a Fotocasa... (Es broma).

**Pablo G<sup>a</sup> Mexia.** Es que el sector privado muchas veces maneja mejores datos que el sector público.

**Jaime del Barrio.** Sobre la pregunta de qué son macrodatos, para mí macrodatos no es generación de datos, sino una plataforma que es capaz de leer todo lo que hay; y no hace falta ser tan finos en los indicadores, sino tener en cuenta las tres Vs: Uno, el valor, dar una respuesta a una pregunta que importe. Dos, el volumen, del orden de terabytes de datos, y tres, la velocidad. Esto es muy rápido.

**Jordi Colomer.** En algunas cuestiones, por ejemplo, perfiles de farmacia, según qué efectos secundarios de la farmacia, no puedes ir con correlaciones, sino que necesitas relaciones causa-efecto. Para muchas cosas, las correlaciones te dan la información para tomar decisiones, aunque las tomas sobre probabilidades. Para otras cuestiones, es necesario identificar una relación causa-efecto.

**Carolina González-Criado.** Es que el ensayo clínico aleatorizado (ECA) tiene que seguir siendo el *Gold Standard* para la autorización de un medicamento.

**Jaime del Barrio.** Esta discusión la tengo permanentemente con *startups*, empresas que trabajan con algoritmos de diagnóstico... No son conscientes de la regulación que hay detrás.

**3. ¿Qué puede aportar a la efectividad y a la eficiencia el empleo de macrodatos en el sector salud, tanto en la práctica clínica como en la gestión?**

**Jordi Colomer.** Mi opinión es que las correlaciones te pueden orientar a buscar una causalidad en unos aspectos concretos.

**Juan del Llano.** ¿En el corto plazo lo ves más útil para la práctica clínica o para la gestión?

**Jordi Colomer.** Para las dos actividades.

**Jaime del Barrio.** Para las dos.

**Juan del Llano.** ¿A la par?

**Jaime del Barrio.** No tienen por qué ir por separado, pueden ir en paralelo. Incluso hay temas que se mezclan. Nuestros compañeros médicos dicen: no tengo tiempo, no tengo tiempo. Pero no saben que el 40% de lo que hacen se podría automatizar ya, y podrían dedicar su tiempo a lo que realmente aporta valor: escuchar al paciente, tocarle la tripa, etc. Y en temas de gestión clínica, está claro que se pueden medir los resultados en salud, y que debemos medirlos y en base a ellos, planificar, organizar y asignar recursos. Esto tiene que ver con la financiación, el presupuesto y el gasto siempre limitado.

**Carolina González-Criado.** Te sirven de forma paralela tanto como para la investigación y la labor asistencial, como para la gestión. En investigación y labor asistencial es verdad que el ensayo clínico lo has hecho en un ámbito muy controlado, muy restringido, con una adherencia y unas patologías muy determinadas. Y cuando usas el medicamento en vida real, algunos no cumplen los criterios de inclusión, otros tienen comorbilidades y son de distintas edades, otros no se toman la pastilla, etc. y luego se va a ver como todo esto se manifiesta. También vas a tener estudios a largo plazo. Todo esto te sirve para complementar el ensayo clínico. Pero para la gestión, es importantísimo. Porque incluso todo esto te puede ayudar a replantearte decisiones tanto a nivel macro como meso. Por ejemplo, en Alemania, cuando sale un medicamento al mercado, el precio es libre, pero al cabo de un año, se vuelven a

sentar y evalúan los resultados. Y a nivel meso, te puede servir para analizar la variabilidad en la práctica clínica dentro de tu servicio de salud. Puedes comparar un hospital con otro; y si uno tiene mejores resultados, ver que se ha hecho distinto; lo cual te sirve para repartir recursos. Las estratificaciones de la población, ahora que nuestro gran problema va a ser la cronicidad, te sirven para dotar los recursos de otra manera. Incluso a nivel micro, como médico, te pueden servir para organizar tu consulta. Si cuando entra el paciente a la consulta, tienes una pirámide poblacional de estratificación de riesgo basada en macrodatos y datos de vida real, puedes decidir la prioridad. Tiene múltiples utilidades, tanto en el ámbito asistencial de investigación como en el ámbito de gestión.

**Francisco Estupiñán.** En términos generales, también. Digamos que nos movemos en una especie de escalera, donde hasta ahora hemos sido capaces de decidir. Lo hacíamos bastante bien, éramos capaces de analizar hasta cierto punto, en base a lo que teníamos disponible y éramos capaces de predecir, en ciertos casos, con un alcance muy limitado, porque para predecir tienes que fijar muy bien la pregunta de investigación. Y vamos hacia una mayor capacidad de predicción, para al final ser capaces de prescribir y ayudar a la toma de decisiones. Esto vale tanto para el profesional sanitario en la toma de decisiones clínicas, como para el gestor sanitario o planificador al asignar recursos. La su utilidad real del análisis basado en macrodatos se verá en los casos de uso que lleguemos a implementar, en el desarrollo de esos casos de uso, y en lo capaces que seamos de ponernos todos de acuerdo, tener una cultura común, para materializar los resultados que obtengamos a través de los análisis en mejoras para el Sistema de Salud.

**Mercedes Alfaro.** Estoy de acuerdo. La clave está en saber cuál es el objetivo, un punto que ha salido varias veces. Hacia dónde vamos y qué queremos. Si tenemos claro hacia dónde vamos, podrá ser tan útil para la gestión como para la práctica clínica. Si lo que estamos haciendo es esperar a ver qué encontramos, la utilidad será probablemente mucho más discutible. Yo quería decir una cosa de la estratificación. Para la gestión, y para la financiación, me parece una herramienta básica y muy cómoda; pero, por ejemplo, para un médico de primaria, el hecho de que tenga la necesidad de seguir a su paciente a golpe de “semáforo”, una de dos, o tiene mal ajustado el cupo de pacientes asignados, o no conoce a sus pacientes. Y lo digo con todo el cariño. Pero lo veo mucho más útil para financiar y planificar que para seguir a los pacientes.

**Carolina González-Criado.** Cada vez se ve más lo que tú dices, pero cuando un médico de un cupo sustituye a otro (inter-sustituciones), entonces la estratificación, cuando entra el paciente en la consulta, en la que, además, tienes más sobrecarga porque han repartido los pacientes del médico que falta, te ayuda ya que a simple vista te permite ver si es complicado, y su riesgo de morbilidad o empeorar.

**Mercedes Alfaro.** En ese caso tienes toda la razón, cuando no actúa el propio médico de familia puede ser útil. Pero si esto sucede a menudo, es como una contradicción en sí misma, el médico de familia significa continuidad.

**Jaime del Barrio.** Una aportación concreta en la práctica clínica y sobre todo en la investigación clínica, son los registros de pacientes. Este es un tema que seguimos históricamente demandando, necesitando, y que te lo va a facilitar muchísimo la gestión de los datos. De otra manera es imposible ahora mismo obtener resultados fiables y homologables en muchos proyectos de investigación y desarrollo. Con este tipo de herramientas, como plataformas de agregación de datos, etc., vamos a conseguir, no solamente registros de pacientes, sino también un reclutamiento importante para ensayos clínicos. O sea, aparte de decirle al médico o paciente lo que hay en las bases de datos disponibles, al mismo tiempo se abren nuevas líneas de trabajo, que teniendo en cuenta datos de la historia clínica electrónica, incluido su perfil genómico, permiten hacer avances en el diagnóstico, tratamiento y evolución de las enfermedades hasta hora insospechados, pero necesarios.

**Mercedes Alfaro.** Pero que voluntariamente han aceptado.

**Jaime del Barrio.** Sí claro. Pero Mercedes, la gente suele darlo todo: ahí tenemos como ejemplo: Google, Facebook...

**Pablo G<sup>a</sup> Mexia.** ¿No puede ayudar a auto-responsabilizar? ¿No puede ayudar a que el ciudadano tenga una información mucho mejor procesada, que esté en su poder, para de alguna manera, “luchar mejor contra sí mismo”?

**Jordi Colomer.** La gente que está en temas de seguridad da por hecho que todos damos los datos sin que nos importe. La solución no es tanto la autolimitación personal, sino regular bien al que utiliza los datos. Pedir a la gente que se autolimita en las redes sociales (RRSS) se da por perdido, en todo caso hay que regular bien cómo y para qué se utilizan los datos. Eso dice la literatura respecto

a la seguridad, porque la gente tiene tendencia a dar datos. Ahora con la nueva ley de protección de datos me he dado cuenta de qué terceros tenían mis datos sin que yo lo supiera o fuera consciente.

**Francisco Estupiñán.** Sobre lo que hablabais antes de aplicaciones prácticas, estamos trabajando en temas de sobre-medicalización, a petición de la CA de Aragón. Es hacer la medición del Compromiso por la Calidad de las Sociedades Científicas (Ministerio de Sanidad, Consumo y Bienestar Social), de las recomendaciones de “No hacer”. Hemos tenido un trabajo previo de desinversión, y ahora lo estamos usando para esto. Y el alcance máximo de esto es primero medir y asegurarnos de que lo que estamos midiendo es viable, y luego prescribir, poder informar al clínico sobre su práctica y oportunidades para dejar de prescribir fármacos o intervenciones, en situaciones concretas, sobre las que no hay evidencia o esta es discutida.

**Jordi Colomer.** La sociedad de geriatría dice que un paciente no puede estar tomando más de seis fármacos simultáneamente. Pues la media de un paciente de 80 años es de doce pastillas; no hacen falta los macrodatos para abordar este problema.

**Mercedes Alfaro.** ¡Menos mal que tienen poca adherencia...!

A mí me preocupa la medicalización que pueda generar este intento de tener muchos datos de todo, y que todos creamos que vamos a estar como más protegidos, porque se va a saber la velocidad del aire con el que pestañeamos, y eso va a ser predictivo de si al final vamos a tener una úlcera. No sé, me preocupa la medicalización, me parece otra forma de enfermedad. Por ejemplo, que una camiseta mida electrolitos en sudor... Yo os lo digo en serio, puedo entender que alguien que haga deporte de alta competición la necesite. Pero que cualquiera que salga a correr al Retiro lleve la camiseta que mide electrolitos en sudor, y luego vea que el sol está un poco bajo... la eficiencia, la efectividad, la eficacia, etc., va a venir a consultarnos y le vamos a dar una pastilla, no os engañéis. Creo que hay que buscar el equilibrio.

**Jaime del Barrio.** Y no hablamos de formación, educación, etc. Hoy a los alumnos se les sigue formando en el ámbito académico igual que hace 40 años, mientras que el entorno y las necesidades han cambiado, hoy en día, grandes corporaciones en USA tienen un 15% de sus profesionales sin titulación supe-

rior universitaria, en base a su talento, es un dato para tener en cuenta. Y todo lo que estas comentando va ligado a formación de profesionales.

**Mercedes Alfaro.** Exactamente.

4. ¿Cuál es el sitio adecuado de los macrodatos y del DVR? ¿Cómo se sitúan dentro del marco de la investigación y la toma de decisiones (efectividad, eficacia, seguridad, planificación, etc.)?

**Carolina González-Criado.** Se tienen que complementar con los ECAs.

**Jordi Colomer.** Tienen objetivos distintos.

**Jaime del Barrio.** Pero tienen que estar apoyándolos a todos. A mí lo que me preocupa es que ahora mismo, *startups* y empresas no tradicionales en el sector Salud, están pretendiendo liderar muchas iniciativas sin conocer que en Salud no hay atajos, es un momento para el trabajo transversal y multidisciplinar, en el que tienen cabida: ingenieros, informáticos, emprendedores, inversores, .... pero incuestionablemente profesionales sanitarios y autoridades regulatorias que han de validar en base a criterios establecidos, todos los avances.

**Jordi Colomer.** Para mí lo importante, es la relevancia de compartir entre distintas áreas de conocimiento. Si somos conscientes de esto, recorreremos mucho camino.

**Jaime del Barrio.** Claro, la autoridad competente tiene que ejercer como lo que es y marcar unas reglas de juego, claras y transparentes, además de posibilidades.

**Pablo G<sup>a</sup> Mexia.** Al hilo de Jordi, la Autoridad catalana de datos está marcando una pauta puntera en España y en Europa, justamente en esta línea, porque ha urgido a que, las empresas, en general, pero también las del sector sanitario, computen como activos sus datos como si fueran un intangible más, con vistas a su propia valoración, o a ser fusionadas con otras. En esta línea el derecho está perfectamente alineado con lo que comentas.

**Jaime del Barrio.** ¿De quién depende eso?

**Pablo G<sup>a</sup> Mexia.** De la Generalitat de Catalunya, es una Autoridad independiente dentro de la Administración autonómica.

**Jaime del Barrio.** Es que yo creo que estos son temas estratégicos que tienen que manejarse a un nivel muy alto.

**Pablo G<sup>a</sup> Mexia.** En este caso, es la Autoridad Catalana de la Competencia, que es un ente moderadamente independiente de las líneas que pueda marcar el gobierno catalán o por supuesto las autoridades de Bruselas.

**César Velasco.** Los macrodatos permiten avanzar en campos de la investigación en los que previamente era difícil incidir. La implementación de herramientas de macrodatos en investigación con grandes bases de datos genómicas ha llevado a una revolución en la identificación de causas genéticas, marcadores diagnósticos y posibles dianas terapéuticas. Los macrodatos se han situado en una posición crucial dentro del programa de la investigación biomédica en nuestro entorno. La capacidad que tenemos actualmente de analizar y dar valor a los datos del sistema sanitario es mínima en comparación con la que alcanzaremos gracias a los nuevos algoritmos de interpretación. Actualmente los centros sanitarios que cuentan con institutos investigación biomédica de excelencia están realizando acuerdos y convenios de colaboración con plataformas europeas de investigación con datos a gran escala. Han surgido diversas iniciativas tanto en el ámbito de los ensayos clínicos como en el ámbito de la asistencia sanitaria financiadas con fondos europeos para promocionar la utilización de los datos masivos sanitarios. A su vez varias convocatorias del programa H2020 europeo van orientadas a la producción de bases de datos masivos y a la mejora de las infraestructuras necesarias para generar investigación de valor.

En cuanto a los datos de vida real la situación es distinta, dado que hay menos iniciativas en la actualidad para promover la creación de datos de vida real integrados en bases de datos públicas y de calidad. Actualmente los profesionales sanitarios actúan como prescriptores de aplicaciones móviles o de aplicaciones de Internet de las cosas que generan datos sobre pacientes que tendrían gran valor a la hora de incorporarlos a la investigación científica, pero no están disponibles para el sistema sanitario. Sin embargo, los datos recogidos por gran mayoría de aplicaciones y soluciones de Internet de las cosas no están integrados en los sistemas y bases de datos comunes. Algunas instalaciones sanitarias, nuevas unidades de cuidados intensivos ya cuentan con dispositivos de medida de variables fisiológicas en tiempo real. Estas tecnologías son capaces de

almacenar los datos recogidos por los sensores incorporados al día a día del paciente ingresado. La información producida se puede almacenar y analizar posteriormente. La gran cantidad de datos que generan este tipo de unidades de cuidados intensivos todavía no se ha explotado de manera eficiente para producir datos de vida real con valor añadido para la práctica clínica o la investigación en nuestro entorno.

#### 5. ¿Cree que los macrodatos y DVR pueden aportar evidencia de causalidad en situaciones que no sea viable realizar estudios experimentales?

**Mercedes Alfaro.** Los macrodatos dan “proxis”, pero la causalidad hay que ir a buscarla y analizarla por otros medios. Es lo que comentaba antes Carolina, que, si no tienes el dato de un diagnóstico, pero ves que hay una insulina, puedes llegar a la conclusión de que se trata de un diabético. Hay casos en los que sí, en que posiblemente te puedan acercar a la causalidad, si la historia es muy biunívoca. En otros casos, no.

**Carolina González-Criado.** Yo también lo creo. Lo tenemos todos claro: los macrodatos solo dan correlaciones. Ocurre lo mismo con DVR, porque los registros no te van a servir para comprobar hipótesis; todo lo más a generarlas para un futuro estudio. A no ser que se trate de un estudio pragmático, observacional, completamente bien diseñado; a lo mejor ahí, sí encuentras causalidad, porque hay distintas fuentes de práctica clínica real. Habría que ver qué fuentes, claro.

**Mercedes Alfaro.** Eso, efectivamente, puede ocurrir.

**Francisco Estupiñán.** Yo creo que hay desafíos, pero que es posible encontrar causalidad en DVR. Con las técnicas que tenemos, yo creo que sí. Con DVR es posible hacer diseños cuasi experimentales, con grupo control, con un seguimiento retrospectivo-prospectivo, tratando de aproximar variables instrumentales características de estudios observacionales.

**Jaime del Barrio.** Ahí está el truco, lo has dicho tú y lo ha dicho Carolina. Estudio observacional; si no, se queda en hipótesis.

**Francisco Estupiñán.** Lo que te permite concluir esta historia, en mi opinión, es que donde antes tú tenías con mucho esfuerzo una perspectiva limitada de la

enfermedad de una persona, ya que ibas captando una serie de observaciones de momentos determinados del proceso de enfermedad, ahora puedes llegar a tener una visión muy amplia del continuo salud-enfermedad del paciente, siguiéndolo durante todo el tiempo que sea necesario, a través de todos los dispositivos que puedan acceder a información, y complementándolo con información que no tiene por qué estar relacionada con la salud. Esa imagen que te haces es mucho mejor que la que tenías antes.

**Jaime del Barrio.** ¿Y esos análisis pasan por los comités éticos?

**Francisco Estupiñán.** Sí, existe la obligación de supervisión por parte de los comités éticos para estos estudios, igual que para el resto.

**Carolina González-Criado.** De todas formas, ahí la posibilidad de sesgo es más grande. Habría que tener cuidado con el tipo de ensayo. De hecho, creo que ahora ya se está intentando validar herramientas para la lectura crítica de ese tipo de ensayos, porque la posibilidad de sesgo puede ser importante. Ya no solo para validación interna-externa, sino en el diseño. Lo que hablábamos, no se trata de ir a “pescar”, sino de tener publicado con anterioridad a qué preguntas vas a responder y qué variables vas a medir.

**Mercedes Alfaro.** Claro, no vale ir a ver qué encuentro y ya veré luego qué hago con ello.

**Juan del Llano.** Y los DVR están empezando a aventurar que quizás sí pueden ayudar a encontrar relaciones causa-efecto, si están bien diseñados metodológicamente.

**Joan Rovira.** Yo diría que, incluso si solo pretendes buscar correlaciones, con macrodatos puedes tener sesgos importantes. Por ejemplo, los individuos más preocupados con la privacidad estarán probablemente subrepresentados en las fuentes de datos en que puedan evitar dar su información personal.

**Jordi Colomer.** En cualquier caso, la correlación es muy importante cuando lo que estás buscando no es el porqué de lo que pasa, sino qué está pasando. Cuando es algo descriptivo. Tú vas buscando datos que crees que pueden ser útiles para aquel objetivo, siempre con la duda de cuánto estás dejando de introducir en el algoritmo. Las limitaciones están aceptadas desde el principio. Al final, este algoritmo, que es una caja negra, también depende de valores. Según

como hagas el algoritmo, puedes ser racista o xenófobo. En un futuro tal vez los algoritmos tengan que ser públicos y todo el mundo pueda explorar exactamente cómo se ha elaborado el algoritmo. Esta es una de las complejidades que veremos en los próximos años en materia de macrodatos. Por ejemplo, una anécdota. Un algoritmo analizaba qué sucedía en el hospital en cuanto a la EPOC. Se miraba pacientes con EPOC sin asma y con EPOC y asma, que es más grave. El algoritmo concluyó que se morían más los pacientes que tenían EPOC sin asma. La posible interpretación es que el médico, cuando veía un EPOC sin asma, le daba menos importancia y se focalizaba en los que presentaban EPOC con asma, con el resultado de que al final la mortalidad era más elevada para el EPOC sin asma que con asma. Encontrar el por qué, tuvo que ser resultado de la intuición del ser humano, que a partir de una correlación buscó la causa.

**Francisco Estupiñán.** Parece que vamos a descubrir la epidemiología, la confusión, las correlaciones espurias, y que no es fácil interpretar resultados y se tiene que tener mucha cautela. Eso es una realidad y es también la razón por la que el entusiasmo por estas cosas de la tecnología que producen tanta excitación, en general al final se desinflan, porque es cierto que no presuponen el conocimiento sobre su tratamiento. Por ejemplo, la arquitectura y la estructura que nosotros estamos desarrollando ahora mismo, está todo ello configurado sobre aplicaciones en código abierto. ¿Qué supone esto? Pues que ha habido mucha gente preocupada por hacer una implementación informática de las técnicas y metodologías necesarias para estos análisis de datos. Pero en muchos casos el uso que se hace de esta tecnología se ha hecho sin tener ni idea de cómo interpretar lo que sale del análisis, la matemática que hay detrás, y la informática que es necesaria para programar. Un ejemplo, aunque a mí me pongan el mejor TAC que hay disponible en el mercado, yo no me atrevo a diagnosticar, pues por muy buena que sea la herramienta, necesito tener conocimiento de cómo usarla y ser capaz de interpretar sus resultados. En ese sentido, para mí - sigo siendo optimista - las iniciativas piloto que estamos desarrollando las CCAA ponen de manifiesto dos puntos: uno, que tenemos muchas ganas de hacer cosas, y dos, que necesitamos un periodo de capacitación para convertirnos en usuarios habilitados, hasta cierto de punto, para usar esos instrumentos.

**César Velasco.** Recientemente se han comenzado proyectos que utilizan grandes bases de datos e inteligencia artificial para reproducir la fisiología humana en avatares humanos. En estos avatares humanos se podrán testar nuevas terapias de manera virtual sin necesidad de incluir voluntarios en un ensayo clínico. En este sentido los macrodatos y los datos de vida real servirán para aportar

evidencias de causalidad en entornos donde la realización de ensayos clínicos en humanos sea difícil. Queda un largo recorrido para poder desarrollar este tipo de tecnologías y validarlas científicamente, pero diversos proyectos internacionales están en camino. Del mismo modo resultará más sencillo realizar estudios observacionales post autorización de medicamentos gracias a la utilización de las bases de datos de vida real que ahora mismo ya son accesibles en los sistemas sanitarios de la Unión Europea.

Gracias al análisis de grandes datos seremos capaces de conocer cuáles son los efectos y la evolución de tratamientos convencionales en la población sin necesidad de realizar estudios de cohortes, aumentando la eficiencia y repercutiendo positivamente en la sostenibilidad del sistema sanitario. También seremos capaces de utilizar los datos generados por las tecnologías médicas y diagnósticas, estos datos actualmente están almacenados en las bases de datos de la industria que provee de tecnologías al sector. En un futuro cercano se comenzarán a cruzar estas bases de datos con los datos asistenciales de pacientes para estudiar la relación entre las pruebas diagnósticas y la evolución de las patologías.

6. ¿Puede indicar ejemplos donde la evidencia experimental, tal como, la procedente de ensayos clínicos, sea contradictoria con la derivada de macrodatos o de DVR?

7. Pregunta clave (clínicos/decisores/evaluadores): ¿Cómo respondería a situaciones en que este tipo de nueva evidencia (DVR y macrodatos) fuese contradictoria con la “evidencia experimental clásica”?

*(formuladas conjuntamente)*

**Francisco Estupiñán.** Yo tengo una respuesta muy poco “sexy”. Si el alcance de la pregunta es similar y la metodología que se ha usado en el diseño es el que corresponde y las respuestas son dispares, habría que plantearse la interpretación del resultado. El caso es que eso se da en muy poquitos casos. Y una de las cosas que encuentras a menudo es que, o bien el alcance no es el mismo, o la pregunta es distinta. Es el problema que tienes cuando haces un metaanálisis: al final tienes que calcular una medida de heterogeneidad.

**Mercedes Alfaro.** Es que no hay otra.

**Juan del Llano.** ¿Y conocéis algún ejemplo contradictorio? ¿Que un ensayo clínico diga una cosa y un estudio con DVR otra?

**Carolina González-Criado.** Yo lo estuve pensando estos días y no recuerdo ningún caso. Es más, muchas veces lo que se trata con los DVR es complementar el ensayo clínico, por ejemplo, extrapolar los resultados a una población que no se ha podido incluir por una comorbilidad, o porque por algún motivo no se le puede aplicar el protocolo, o no se adapta a la Guía clínica. En este caso, los DVR simplemente complementan.

**Jordi Colomer.** En muchos ensayos clínicos se han dado cuenta de que no han funcionado cuando se han aplicado, porque la población del ensayo tenía 40 años y se ha aplicado a pacientes reales de 80, como el sesgo que introducen algunos ensayos clínicos por un motivo de género. La literatura está llena de casos de estos. Pero no hace falta DVR ni macrodatos para evitar estos errores, solo sentido común.

**Mercedes Alfaro.** De todas maneras, aunque pueda parecer contradictorio, cuando lees sobre macrodatos, o empiezas a pensar en macrodatos, concluyes que al final se puede convertir en una especie de filtro, o sea justo lo contrario de lo que se sugería antes. Y en el mismo sentido que decía Jordi, si los macrodatos te dan una correlación, dices: aquí hay algo. Y ahí es donde empiezas a pensar en el por qué, y cuando te vas al ensayo clínico.

**Jordi Colomer.** Te ayuda a pensar.

**Mercedes Alfaro.** Claro, igual te ahorras los errores de elegir mal la población, el estadio de la enfermedad, la época del año, etc., en el diseño del ensayo clínico. Porque los macrodatos ya te han dado pistas.

**Jordi Colomer.** Por ejemplo, un DVR hecho ya, es el Cáncer Lin Q que utilizan en la Asociación Americana del cáncer (ASCO). Hace clústeres, de forma que si tengo un enfermo de cáncer de tales características y en esta base de datos hay 50 millones de individuos, seguro que mi paciente se puede incluir en algún clúster. Y con esto, que no son macrodatos, eres capaz de pensar si aquello que habías aplicado es lo más adecuado, si en sentido contrario te lo replanteas porque en función de aquellos datos puedes cambiar de parecer.

Entonces el problema está en adoptar la postura de “yo no me he equivocado, es la máquina”. El profesional se puede orientar a la postura fácil y concluir: “es la máquina la que me ha dicho lo que tengo que hacer”. Y esto no es verdad, porque la responsabilidad pasa por tu toma de decisiones. Pero esto es otro problema.

**Jaime del Barrio.** Respecto a la pregunta 6, yo creo que no hay ningún ejemplo. Y mi fuente es la EMA, la FDA, y la AEMPS. Recuerdo cómo y cuándo apareció la farmacogenómica; hay momentos en la historia de la regulación en que se introducen nuevos enfoques. Yo espero que, si nos volvéis a llamar dentro de dos años, las entidades reguladoras habrán incorporado una serie de expertos para la inclusión en los ensayos clínicos de todo esto que estamos comentando. Por ejemplo, la FDA ya tiene perfectamente definido lo que es evidencia científica derivada de las apps de salud; ya hay más de 600 trabajos publicados, metaanálisis, etc. Ahí no estoy de acuerdo con Joan, no somos pesimistas, somos optimistas. No vamos a decir que esto son fuegos artificiales y que los macrodatos ya están aquí, pero soy optimista; esto es lo que toca y creo que estará en las agencias reguladoras en un par de años. Igual que ahora hablamos del binomio inseparable molécula – biomarcador, impensable en otro tiempo.

**Juan del Llano.** ¿Qué os parece la aseveración de Mercedes de los macrodatos como cribado para generar hipótesis y preguntas de investigación?

**Mercedes Alfaro.** Entre otras funciones.

**Francisco Estupiñán.** Generar hipótesis y especificaciones de diseño de estudios.

**Jaime del Barrio.** Si ayuda al diseño, los reguladores lo tendrán que incorporar.

**Mercedes Alfaro.** Efectivamente.

**Juan Del Llano.** Lo acabáis de clavar.

**Carolina González-Criado.** De hecho, la EMA tiene puesto en marcha un piloto, *Adaptive Pathways*, que para determinadas autorizaciones de tipo condicionado prevé utilizar estudios de DVR para complementar la información que le falta para la autorización.

**Jordi Colomer.** Por lo general, si se había autorizado un medicamento, a pesar de estudios basados con una población pequeña, se trasladaba sin más a la población general, con la existencia de con una mayor población aparecieran efectos adversos. Una vez aceptado el medicamento, no es tan fácil que la agencia reguladora se retracte a menos de efectos muy perjudiciales. Y ahora se está valorando que bueno, que, con un registro previo, dentro de cuatro años veremos si seguimos adelante.

**Juan del Llano.** ¿Eso se sostiene jurídicamente, Pablo, ¿el dar una autorización provisional? O la empresa a posteriori te puede decir: oiga, a dónde va, me ha hecho perder el tiempo.

El asunto es que, hasta ahora, las autorizaciones de comercialización de fármacos se daban o no se denegaban. Si se daban, era con muchísimo fundamento. Lo que hablamos es de procedimientos rápidos (*fast track*), que suelen aplicarse a medicamentos para enfermedades raras y también para algún cáncer extraño; los están metiendo en vías (tracks) especiales en los que el resultado final es una autorización condicionada de dos años, por ejemplo. Mi pregunta es si eso, regulatoriamente, puede o no dejar alguna puerta abierta por la que luego puedan venir reclamaciones.

**Pablo G<sup>a</sup> Mexia.** La autorización dura tanto o tan poco como el regulador determina. ¿Os suena lo de las *sandboxes*? Autorizar provisionalmente, para, como dice la denominación, “jugar en la arena”.

De hecho, toda lo que “te estoy dando” es una autorización provisional para esta indicación y durante este periodo concreto “no te voy a sancionar”. Esto, por ejemplo, en Fintech se está haciendo muchísimo. En Europa con más intensidad, pero en España también se quiere hacer. Es perfectamente posible.

**Jordi Colomer.** Creo que hay algunos registros que se han pasado de frenada: se han lanzado al mercado con ensayos clínicos de 40-50 pacientes y luego han visto que, al pasar a la población general, con datos masivos si quieres, han aparecido efectos adversos evidentes.

**Francisco Estupiñán.** Esa es una utilidad clara de los datos masivos (farmacovigilancia), el descubrimiento de señales pequeñas que pasaban desapercibidas (eventos adversos infrecuentes), porque tenías que rellenar y enviar una hoja amarilla.

**Juan Del Llano.** ¿Mas comentarios sobre estas dos preguntas?

**Francisco Estupiñán.** Existe un programa de la FDA que está obligando redefinir el paradigma de farmacovigilancia y la evaluación de la efectividad de los fármacos post-comercialización llamado Sentinel. Y hay una iniciativa que me parece muy interesante, entre otras razones, porque tiene una dotación presupuestaria potente: el National Health Service (NHS) ha hecho un convenio con la Universidad de Newcastle para montar un Observatorio de Innovación (*Innovation Observatory*), que consiste básicamente en desarrollar un análisis prospectivo (horizon scanning) muy avanzado. El presupuesto total del proyecto es de 50 millones de libras. Están haciendo unas cosas espectaculares. Trabajan sobre 3 ejes. Uno que es la captación y análisis del lenguaje natural para análisis de diversas fuentes de información, desde registros de ensayos clínicos, hasta patentes y fusiones de empresas, a través de documentos útiles, notas de prensa, etc. para hacer el análisis prospectivo, prever qué es lo que está llegando y a qué velocidad se acerca al mercado; cuándo se espera que esté desarrollado, y así estar preparados para poder evaluar y para tomar decisiones informadas sobre su financiación. Otro eje en el que están trabajando es la integración de la perspectiva de los pacientes y de otros grupos de interés del sistema sanitario en la consideración de resultados y cuestiones relevantes para la evaluación del impacto de las intervenciones sanitaria. Se trata de obtener información de las RRSS, a través de foros, de retroalimentación directa con los *chatboxes*. Por último, trabajan en la integración de todo este conocimiento generado y su presentación a médicos y decisores de una forma operativa como herramientas de ayuda a la toma de decisiones informada.

**Juan del Llano.** Parece muy sensato y posiblemente tenga retorno.

**Francisco Estupiñán.** Lo más interesante es que están poniendo en abierto las herramientas que están desarrollando, y eso es una ganancia para todos.

**Jaime del Barrio.** El retorno no solo es en datos, sino en resultados en salud, economía, inversión, nuevas profesiones, nuevos desarrollos, nuevas empresas...

**Francisco Estupiñán.** Hay otras iniciativas muy interesantes. Aparte de Newcastle, en Inglaterra también tenemos Swansea, que está desarrollando un montón de librerías de distintos lenguajes de programación que luego se pueden utilizar para replicar los casos de estudio en los que trabajan; hay otra gente

que están desarrollando, en el eje de investigación, interfaces para el análisis de datos que incorporen todas las medidas de seguridad, a nivel técnico, para consulta de los clínicos y la realización de estudios avanzados.

**Juan del Llano.** ¿Aquí en España hay universidades que trabajen en estas cosas?

**Francisco Estupiñán.** Hay iniciativas muy interesantes. La Politécnica de Valencia está metida en todo lo que es calidad de los datos, da gusto verlo. En la Universidad de Barcelona hay un académico que está trabajando el tema de la seguridad de los datos y anonimización a un nivel técnico muy avanzado. Los campus de Zaragoza, Lérida y La Rioja están dando aproximaciones basadas en inteligencia artificial a temas básicos de gestión (stocks, regulación energética, amortización de las máquinas, mantenimiento de equipos, etc.).

**Jaime del Barrio.** En Gijón, concretamente en la Fundación Centro Tecnológico de la Información y la Comunicación (CTIC) también hay unas iniciativas muy interesantes. Y en el País Vasco también. Una pregunta Paco, en todas estas iniciativas internacionales que conoces, ¿hay algún socio español?

**Francisco Estupiñán.** Hay iniciativas internacionales en las que tenemos participación. Está funcionando una Acción Conjunta (Joint-Action) europea en información sanitaria en la que participa España a través del ISCIII. Uno de los objetivos de esa Acción Conjunta es desarrollar una infraestructura europea de investigación basada en datos sanitarios. Y después hay algún proyecto de convocatoria interregional del programa H2020, en concreto con Francia y Portugal.

**Mercedes Alfaro.** Sí, el IACS es muy vital y cada vez habrá más iniciativas. Lo que yo creo que pasa en España es que falta una cultura generalizada, es decir, cuando se habla de cambio del modelo productivo, no sé si entendemos y estamos de acuerdo en de qué estamos hablando. Porque luego miras a la administración y, en una escala muy pequeña, “cambio de modelo productivo” implica que no necesitamos tanto auxiliares administrativos que archiven, como expertos en semántica, por ejemplo. Le cuentas esto a tu secretario o al Ministerio de Administraciones Públicas, y no te suelen comprender Sin pretender que la administración lleve la iniciativa, la aplicación diaria de ese cambio de modelo productivo, que ya no estamos trabajando como antes, que trabajamos de otra manera, que producimos salud de otra manera, que producimos estadísticas de otra manera, no se acompasa con el capital humano; se necesitan unos

años para cambiar esa mentalidad. Hay que tenerlo en cuenta para desarrollar macrodatos.

**Juan del Llano.** Seguro que la brecha que apunta Mercedes es paralizante, en el sentido de que, si las ofertas públicas de empleo siguen con esos epígrafes, pues apaga y vámonos.

**Francisco Estupiñán.** Esa es una brecha importante.

**Juan del Llano.** A mí me parece tremenda. Pero quién frena eso, porque estáis hablando de e-administración desde el año 2000. ¿El statu quo?

**Jaime del Barrio.** No, la ley del Estatuto de los Trabajadores.

**Mercedes Alfaro.** No existe el cuerpo o escala de semánticos, o de bioestadísticos.

**Jaime del Barrio.** Ni se puede crear. Cuando dicen que el 50% de las profesiones actuales van a desaparecer en el año 2020, que es pasado mañana, miramos para otro lado. Me da igual que sea un 50 o un 20% pero sí que estamos seguros de que será un porcentaje significativo y sin embargo no estamos haciendo nada o muy poco a nivel individual o colectivo para reposicionarnos o ayudar a que otros lo hagan. Parece como que esto va con otros y nos quedaremos fuera si no lo corregimos.

**Francisco Estupiñán.** Y también hay que pensar que es muy complicado dar con los perfiles que se necesitan, porque son perfiles mixtos.

**Mercedes Alfaro.** Todos los desarrollos tecnológicos que puede haber están muy bien, y todos los acuerdos y consensos están muy bien, pero hay una cosa básica y es que se necesitan los perfiles profesionales adecuados para sacar todo esto adelante. No se puede sacar nada si no se va evolucionando conforme evoluciona la vida.

**Pablo G<sup>a</sup> Mexia.** Se acaba de perder una oportunidad de oro: crear un ministerio digital. Eso es una señal muy clara desde el nivel político correspondiente.

**Juan del Llano.** Lo que estáis diciendo es un claro limitante, es decir, la formación, los perfiles, y que la administración no tenga herramientas para incorporarlos.

**Francisco Estupiñán.** La aproximación que se persigue es de tipo multidisciplinar, pero el problema de los equipos es que tienes unos problemas de comunicación entre los distintos perfiles que son muy importantes. Hay que ir hacia esos perfiles mixtos, y yo creo que hay algunos planteamientos que ya lo hacen. Ahora están viniendo a nuestras puertas graduados en bioingeniería, que mezclan en su abordaje el tema biológico y médico con un punto de vista de ingeniería industrial o mecánica, pero que ya se van moviendo en la ingeniería informática. Esos perfiles cada vez van a ser más necesarios, con una flexibilidad suficiente para adaptarse a nichos específicos.

**Jaime del Barrio.** Todo esto hay que decirlo y reclamarlo, pero hay cosas que ya se pueden hacer. Por ejemplo, en el sistema nacional de salud, está el modelo MIR. Se sigue enseñando lo mismo que hace 40 años, las mismas especialidades. Los residentes van a ser como sus jefes. Todos imprimiendo artículos, ¿Cómo que imprimís artículos? ¿No los leéis en digital? Claro, sus jefes, son analógicos. Y los MIR, para seguir teniendo futuro, tienen que ser como sus jefes, analógicos en sus pensamientos y en sus obras. Se precisa ¡ya! un profundo cambio cultural.

**Francisco Estupiñán.** Hay una derivada de la que no hemos hablado, que es el impacto de los datos en formación médica. Una de las cosas que te plantean los DVR es que de repente dispones de una base de casos que no tenías antes. Y que puedes crear sistemas que sean interactivos con el médico/a que se está formando. Es algo muy interesante.

**Juan del Llano.** Creo que habéis acertado muchísimo, porque estáis hablando de un cambio disruptivo donde el asunto capital humano es crucial.

**César Velasco.** Una situación en que este tipo de nueva evidencia (DVR y macrodatos) fuese contradictoria con la evidencia experimental clásica probablemente daría lugar a un gran debate. Sin embargo, el valor de los datos de vida real podría ser incluso mayor que el de datos utilizados en condiciones experimentales clásicas. Esto se ha demostrado en otros campos en los que se han obtenido mejores resultados a través de nuevas metodologías en comparación con la aplicación de metodologías clásicas. Se deberían generar marcos conceptuales y consenso científico para estar preparados para esta eventualidad. Se comienza a discutir en círculos científicos y la clásica pirámide de evidencia científica GRADE está preparada para soportar las evidencias que se producirán gracias a las grandes bases de datos. También se comienza a comentar

que la inteligencia artificial y los algoritmos deben estar en la parte más alta de la pirámide GRADE de evidencia científica, Esto situaría a los resultados obtenidos mediante algoritmos que explotan grandes bases de datos por encima del metaanálisis y de los ensayos clínicos aleatorizados, ya que en este modelo de producción de la evidencia científica la intervención humana es menor. La capacidad que tendrán los proyectos que utilicen bases de datos masivas para encontrar asociaciones clínicas entre diversos factores será indiscutible. Además, podrán introducir en los algoritmos nuevos conceptos como movilidad, actividad física, o actividad diaria registrados en otros sistemas (bancario, telefónico, u otros).

#### 8. Interfaz entre regulación (Reglamento General de Protección de Datos) y acceso a datos. ¿Cuáles son los riesgos de la utilización de DVR y macrodatos para la protección de la privacidad de los pacientes y ciudadanos?

**Pablo G<sup>a</sup> Mexia.** Con el ánimo de enfocar la nueva normativa del Reglamento general de protección de datos de la UE desde el ángulo de la salud, la primera cosa, importantísima, como sabéis, es que estamos hablando de datos sensibles, por supuesto. Macrodatos y DVR, las dos variables concretas a las que se ciñe el Reglamento nuevo, entrarían en la categoría de datos sensibles, y por lo tanto éste es el campo de juego. Dice el Reglamento que no se pueden tratar. Fijaos en esto, porque aquí se ha hecho un enfoque un poquito más restrictivo en cuanto a la redacción. El Reglamento ha querido subrayar que, junto con otros datos sensibles, los datos genéticos y de salud no se pueden tratar salvo que se cuente con una serie de variables, a modo de válvulas de escape. Las cinco más importantes a efectos nuestros, serían: Uno, el consentimiento. Dos, que sea necesario en el ámbito laboral o de la Seguridad Social. Tres, que te vaya la vida en ello, es decir, que tengas que prevenir o curar una enfermedad, o gestionar esos aspectos. Cuatro, que haya un interés público en materia de salud, epidemias, calidad y seguridad de la sanidad, medicamentos, etc. Y quinto y último, por supuesto, la investigación científica. Éstas serían las cinco posibilidades de tratar este tipo de datos. Lógicamente el mayor riesgo es que éstos caigan en manos indebidas, y aquí también había anotado que este riesgo se acentúa por el hecho de que todo esto lo puedes interrelacionar. Puede que acabe perfectamente escaneado sin consentimiento.

Sobre los remedios, tengo aquí apuntados los tres más importantes que prevé la legislación. Algunos son nuevos, porque el derecho de oposición, es decir, que no consentas ya a que se siga tratando información tuya, habiendo consentido previamente, ya existía, pero se ha introducido uno nuevo, que es el derecho a oponerte a que se tracen perfiles tuyos en línea. Esto no existía antes como derecho y ahora sí. Puedes decir, yo no quiero que usted trace mi perfil en Internet. Porque resulta que es una compañía que elabora información financiera, y que al final va a acabar deduciendo mi saldo en la cuenta y todo lo que podáis imaginar. Y luego están las decisiones automáticas individualizadas. Que no haya un algoritmo que pueda tomar decisiones por ti sin que intervenga una persona, y que al final te acabe generando consecuencias, fundamentalmente financieras. Por ejemplo, que te suban la póliza de seguros a raíz de que se cruce tu información. Ése es un remedio doble que acabamos de ver muy claro. Otro, por supuesto, la seguridad, es determinante. Sin seguridad, no hay privacidad y viceversa. Constituyen un binomio. Finalmente, me interesa mucho subrayaros un remedio más, y lo haré con esta denominación, la de “información algorítmica”, que también está prevista en dos artículos del reglamento. Se trata de trasladar al titular de la información cómo se maneja la inteligencia en tu organización. Que se entienda en la mayor medida posible, y con qué consecuencias. Que te digan qué te puede pasar si no te opones a ceder tu información. Todo esto en materia de salud es determinante, porque son datos sensibles. Por otro lado, cuanto más puedas anonimizar, mejor, sería casi un cuarto remedio, ya que cuando un dato es anónimo, por así decir, “te libras” de la normativa. Ya no sería un dato personal, y por lo tanto ésta no se le aplica.

**Jordi Colomer.** Los macrodatos se alimentan de la historia clínica, pero también del internet de las cosas, del sudor de la camiseta, y geolocalización, de las RRSS; lo único que puedes controlar es la historia clínica, porque todo lo demás son datos que tú estás dando y que alguien utilizará algún día. Por eso se piensa que el problema no es tanto que la gente no diga que no se utilicen sus datos, sino que el que los va a utilizar diga para qué va a hacerlo. Como el caso de aquél padre que se dio cuenta de que su hija estaba embarazada, porque ella utilizaba su ordenador y toda la propaganda que recibía era de anticonceptivos. Hay datos, por ejemplo, de las epidemias, que Google detecta sin ningún dato clínico, simplemente con datos de transportes. Por eso digo que los datos sanitarios son regulables, y se regularán, pero los macrodatos se alimentan de cantidad de datos que tú estás dando voluntariamente, por ejemplo, con la camiseta que compras que mide el sodio en el sudor. Venden estos datos, y ganan más vendiendo los datos que vendiendo las camisetas.

**Joan Rovira.** Me gustaría insistir en este tema con una pregunta concreta: según el nuevo reglamento, ¿es legal negarse a que el sistema de salud registre y guarde cualquier dato tuyo? ¿Puedes ir al médico y decirle que no quieres que se quede con tu información?

**Mercedes Alfaro.** Yo lo que he entendido del reglamento es que cuando alguien es asistido por un médico, la historia clínica no es exclusivamente del paciente. Creo que la historia clínica también es una herramienta de los profesionales, sino no, no la cumplimentaría nadie tal como se hace. Una cosa es que, de esa historia clínica, cuando sale un informe clínico consolidado, esa información pertenezca al paciente o cuando hay un caso de litigio esté disponible. Pero si yo, como médico, la utilizo para apuntar datos del paciente de los que no me podría acordar, tal como “creo que no toma la medicación”, “me parece que sigue bebiendo”, “tiene problemas con el hijo”, o apunto hipótesis que luego no confirmo, es distinto. Y ya sé que las apreciaciones subjetivas, incluso cuando las pide un juez, se pueden quitar, pero por eso es una herramienta también del médico. Un paciente tiene que asumir que cuando va a recibir asistencia hay una herramienta que se llama historia clínica que sirve para atenderle con una mínima seguridad para su bienestar (del paciente), y para eso el médico no lo puede llevar todo en la cabeza; otra cosa es que no se pueda usar para ningún otro fin, ni para la investigación, ni para nada, excepto para ser asistido, si no hay unas normas claras. Yo entiendo que es así.

**Pablo G<sup>a</sup> Mexia.** Habría varios canales. Por un lado, tú inicialmente has consentido. Nos estamos planteando el supuesto de que dejas de consentir.

**Joan Rovira.** A mí me parecería razonable que a partir de ahora se pudiera decir, por ejemplo, “dejo mis datos, pero solo para ser usados para mi asistencia”. ¿Qué habría que hacer? ¿Pedirlo y exigir que te den un certificado conforme a que será así?

**Pablo G<sup>a</sup> Mexia.** Fijaos que yo os he mencionado otros supuestos que van más allá del consentimiento, que son los que facultan a la Administración a seguir usando esos datos: interés público para prevenir epidemias o incluso para la gestión sanitaria, investigación científica... Habría que hacer un análisis más detenido de hasta qué punto es necesario identificarte y a lo mejor sería conveniente hacer un esfuerzo de anonimización por parte de la Administración. Esto nos sacaría de la ley, ya no sería necesario tomar esa cautela. Hay cinco acomodos, y uno de ellos, solamente uno de ellos, es el consentimiento, que te

da pie para que incluso con esa denominación personal, la Administración siga tratando esa información con fines de investigación científica.

**Joan Rovira.** A mí nadie me pide normalmente permiso para utilizar la información de mi historia clínica, aunque estoy de acuerdo para no tener que explicar mi historia clínica desde el principio cada vez que voy a una consulta. Mi pregunta era si a partir de este reglamento, igual que lo han hecho todas las empresas que operan a través de internet, el médico también tiene la obligación de informar y pedir el consentimiento.

**Pablo G<sup>a</sup> Mexia.** Eso sí.

**Mercedes Alfaro.** Sí. Además, existe una figura que es el consentimiento amplio, porque igual tú lo que tienes es una lesión en el tobillo y has ido por ello al médico; y resulta que están haciendo una investigación de tobillo y tú piensas que te interesa muchísimo que se investigue este tema, porque te está dando mucha guerra. El consentimiento amplio significa que permites que se usen tus datos para que se investigue en todo lo relacionado con el aparato locomotor, pero no en oncología u otras áreas. Se está barajando un consentimiento amplio, por grandes áreas de conocimiento. Porque es cierto que, si tuviesen que pedir consentimiento para cada proyecto, la investigación se pararía.

**Pablo G<sup>a</sup> Mexia.** Sí, esto hay que asociarlo al principio básico del tratamiento, que es el de la finalidad. Se tiene que saber por parte del paciente qué va a hacer el médico con sus datos ahora, e incluso si va a hacer otras cosas después, si lo va a hacer con terceros, y qué fines va a perseguir con esa información. Enlazaría una cosa con la otra.

**Mercedes Alfaro.** Creo que, si todos los fines fueran benéficos, nadie en su sano juicio se negaría a dar sus datos para la investigación. El problema es que todos sabemos que se usan tanto para fines benéficos, como para fines espurios. Por ejemplo, conozco dos casos que, a partir de una consulta con un seguro privado en una especialidad, se encontraron con una subida en la póliza. Es la excepción, pero puede suceder simplemente por ir a una consulta. El tema es muy complicado, porque los datos es la materia prima más importante para que funcione una sociedad. En España somos muy de normativizar, nos encanta que nos digan “prohibido”, hasta aquí sí, por aquí no, etc., estamos todos muy contentos con eso. Esto es lo contrario al espíritu del nuevo Reglamento de Protección de Datos de la UE que es una filosofía FQM, es generar una cultura

de protección de los datos; usted tiene que tener hecho un análisis de riesgo, de impacto, cada vez que vaya a ceder los datos que le han confiado. Tiene que tener un sistema en su empresa que le garantice que todas las personas que acceden a sus datos sepan qué normas tienen que cumplir

**Pablo G<sup>a</sup> Mexia.** El consentimiento tiene limitaciones, puedes tratar la información con bases de legitimación que sean distintas del consentimiento, una de ellas es el interés público.

**Joan Rovira.** O sea que, aunque yo me hubiese negado a ello ¿si el sistema considera que hay interés público podría compartir mis datos?

**Pablo G<sup>a</sup> Mexia.** Estaría legitimado. Eso sí, te ha de informar con detalle, y con garantías reforzadas.

**Mercedes Alfaro.** Es muy, muy difícil el equilibrio.

**Jordi Colomer.** Yo estoy de acuerdo con vosotros, esto solo es el principio. La relación médico paciente es de confianza, a diferencia de EE. UU., donde es contractual. Nos puede gustar o no, pero vengo de pelearme durante años para que le den la información al paciente. No digo que sea solo suya, pero debe tener un acceso fácil.

**Juan del Llano.** Paco, ¿vosotros, que hacéis investigación aplicada de gran calidad, veis este reglamento como un freno?

**Francisco Estupiñán.** No. Es un tema muy complejo, y las amenazas están presentes siempre. Pero no lo vemos como un freno. Lo vimos como un freno en su concepción y tramitación en el Parlamento Europeo; allí sí que hubo momentos en que lo pasamos mal, con el artículo 29. No había consenso y veíamos peligrar la investigación en general. Pero el producto final me parece un instrumento muy interesante, que pone las bases para hacer lo que tienes que hacer.

**Mercedes Alfaro.** Quiere crear una cultura y creo que lo conseguirá.

**Pablo G<sup>a</sup> Mexia.** Pero tenemos un problema. El 25 de mayo de 2018 tendría que haber estado ultimada la ley que desarrolla un Reglamento que en principio no tendría que ser desarrollado, que es el RGPD. Porque resulta que las directivas comunitarias sí admiten la posible implementación por parte de los estados

miembros y los reglamentos, en cambio, se supone que son comunes, no admiten en teoría desarrollo, y resulta que este Reglamento deja una serie de vacíos, de agujeros, con lo cual parece que estamos en esta situación. Por cierto, España no es el único país que está llegando tarde. Salvo Austria, Alemania, y los países nórdicos (3 o 4 más), todos los demás estamos retrasados. Algunos, vamos notoriamente tarde. Nosotros no íbamos a estar en ese grupo hasta que ocurrió lo de la moción de censura. Ahora mismo el Reglamento está en el Congreso en fase de ponencia, pero de ponencia tiene que pasar a dictamen de comisión, de ahí al pleno, luego al Senado, ambas cámaras con mayorías distintas, de vuelta al Congreso; y con el verano de por medio, ahora mismo si llegamos a aprobarlo antes de final de año será de milagro. El problema es que la ley antigua no está adaptada al Reglamento nuevo.

**Francisco Estupiñán.** Y tenemos por medio la ley general de sanidad, la de salud pública, la regulación de los biobancos, la regulación de los ensayos clínicos... tenemos diferentes normativas que afectan a los datos y que van tocando el tema desde distintas aproximaciones, y que configuran un marco bastante complejo.

**Pablo G<sup>a</sup> Mexia.** Muy interesante tu observación. No os preocupéis, pues es verdad que hay como 400 enmiendas; es decir, que el texto final puede cambiar, pero en principio, en materia de salud, no dice casi nada. Vuestro marco clave es el Reglamento europeo.

**Mercedes Alfaro.** Es que el reglamento y la ley se plantearon así y de verdad que me parece muy coherente. Pretende no ser sectorial, sino muy transversal y no entrar en especificaciones. Al final ha habido un montón de presiones, sobre todo desde la parte de investigación, y es posible que al último borrador que nos ha llegado, se le añada una parte adicional que llevaría temas de salud. En mi opinión todo lo trascendental está incluido en el reglamento, por lo que desde mi punto de vista no haría falta especificarlo al detalle en la ley nacional, pero viene a tranquilizar a los sectores que están reclamando verlo ahí escrito. No lo sé, pero como está en tramitación...

**Pablo G<sup>a</sup> Mexia.** Yo también tengo entendido que hay un acuerdo entre los dos grandes.

**Mercedes Alfaro.** Esencialmente no innova nada, simplemente tranquiliza a determinados sectores.

**César Velasco.** La utilización de datos masivos no debería suponer un riesgo en cuanto a la protección de los datos personales de pacientes y ciudadanos. Eso sí, la gestión de las bases de datos de pacientes debería ser siempre propiedad de la administración sanitaria para evitar segundos usos. Es necesario además hacer un seguimiento de la implantación del GDPR en los diferentes países de Europa, que lo están aplicando con diferencias que pueden suponer grandes barreras para compartir datos de manera transfronteriza en un futuro próximo.

#### 3. Análisis del debate

En cuanto a la definición de “macrodatos” los panelistas asumen la referencia estándar a las tres (o más) Vs que caracterizan la información en este enfoque: Uno, el valor, dar una respuesta a una pregunta que importe. Dos, el volumen, del orden de terabytes de datos, y tres, la velocidad. En algunos casos se pone el acento en aspectos tales como la conectividad de la información. O que macrodatos no se refiere tanto a la generación de datos, sino a una plataforma que es capaz de leer todo lo que hay. En general, parece que los expertos no están especialmente preocupados por una definición precisa de macrodatos o por que se pueda solapar con otros conceptos de información, como los datos de la vida real (DVR), sino por la claridad de los objetivos para los que se utilizan y por la calidad de la información.

*Pregunta 1. ¿Qué iniciativas relacionadas con el empleo de macrodatos en nuestro país conoce por experiencia directa o por la literatura? ¿Algún caso notable de éxito o fracaso?*

Los panelistas mencionan diversas experiencias, aunque casi todas ellas están en una fase preliminar:

- Una de ellas ha sido desarrollada conjuntamente en el IACS, entre la Unidad de Biocomputación y el grupo de Investigación en Políticas y Servicios Sanitarios (AriHSP). Tienen una dilatada experiencia en el análisis de variaciones de la práctica clínica. El objetivo en una primera fase es desarrollar, por un lado, la parte de gestión y luego desarrollar la parte de investigación, los dos pilares básicos de utilidad; otros grupos de interés están planteando otras necesidades, que se explorarán en próximas fases.

- En Galicia se desarrolló a partir del año 2014 la plataforma Hexin, que explota datos clínicos con tecnología de macrodatos. La plataforma pretende aprovechar el volumen de información valioso y grande que genera el Sergas: volumen de consultas, episodios de hospitalización, informes, analíticas, datos de laboratorio, imágenes digitalizadas, etc. La plataforma puede nutrirse de distintos sistemas de análisis del dato estructural, pero también es capaz de interpretar el dato semiestructurado y el de lenguaje natural. El proyecto “FIRES” (Farmacoepidemiología: Información más resultados en salud) analiza qué variables o datos utilizan nuestros oncólogos para prescribir y para hacer el seguimiento, y cuáles utilizan los farmacéuticos para la gestión de los tratamientos.
- En Cataluña AquAs, a través de la central de resultados, estableció una plataforma para la consulta de datos masivos del sistema sanitario. Recientemente se ha lanzado Padris, una plataforma de acceso a los datos masivos de la sanidad de toda la CA de Cataluña. Por otra parte, se están utilizando macrodatos para genómica, especialmente para la identificación de nuevas dianas terapéuticas sobre todo en enfermedades raras (EERR) y en oncología. Asimismo, en el Catsalut se está utilizando campo texto y lenguaje natural para la identificación de pacientes subsidiarios de reembolso mutuo. Se colabora en un proyecto con los hospitales más significativos en el ámbito de innovación de Europa para generar cuadros de mando de gestión sanitaria que estén orientados a la predicción y a la prospección de la toma de decisiones.
- En cuanto a conocimientos que se estén aplicando que provengan de macrodatos, se mencionan algunas dianas terapéuticas de terapias moleculares. Estas son soluciones globales que dependen de proyectos internacionales de la industria farmacéutica o diagnóstica, con fines comerciales.
- Fuera de España se menciona una experiencia de Naciones Unidas, que utilizó datos de genómica, de historia clínica de paciente y de la búsqueda de casos mediante epidemiología de campo durante el brote de ébola, a partir de datos de geolocalización de terminales móviles de las compañías telefónicas. Hay iniciativas en temas de predicción y prevención, por ejemplo, para ver la evolución de la fiebre asiática y para la ubicación de dispositivos de emergencia en el terremoto de México.

Existe consenso en que no hay todavía ningún ejemplo de éxito o fracaso en nuestro país digno de ser destacado, pues lo que existe son “pilotos” e iniciativas, de los que un porcentaje altísimo está abocado al fracaso en muy poco tiempo.

*Pregunta 2. ¿Conoce algún ejemplo de nuevo conocimiento generado mediante el análisis de macrodatos que se haya aceptado e incorporado en la práctica real con un amplio consenso?*

La respuesta unánime del Panel a esta pregunta es que no.

Hay múltiples aplicaciones de los macrodatos, o de lo que se mueve alrededor de esta metodología, que, aunque ya están implantadas, nos pasan desapercibidas, por ejemplo, la mejora de la definición de la imagen diagnóstica, que no hubiese sido posible si antes no se hubieran procesado millones de imágenes para mejorar la nitidez de la ecografía o la precisión de la resonancia magnética.

En el tema de la historia clínica, España es un país descentralizado. No existe una historia única salvo algunas excepciones. No hay una integración de la información que se produce por parte de los distintos profesionales que la haga interoperable, “integrable” con otra información.

El Ministerio está haciendo una labor bastante interesante planteando a nivel nacional un esquema de interoperabilidad en el sector sanitario, que tiene sus dificultades, debido en parte a que los 17 sistemas autonómicos de salud tienen sistemas informáticos operativos diferentes e incluso dentro de las CCAA, pasa lo mismo entre hospitales.

*Pregunta 3. ¿Qué puede aportar a la efectividad y a la eficiencia el empleo de macrodatos en el sector salud, tanto en la práctica clínica como en la gestión?*

Los macrodatos permiten una mayor capacidad de predicción, para al final ser capaces de prescribir y ayudar a la toma de decisiones. Esto vale tanto para el profesional sanitario en la toma de decisiones clínicas y la labor asistencial, como para la investigación y también para el gestor sanitario o planificador al asignar recursos.

Preocupa también que este intento de tener muchos datos de todo haga que todos creamos que vamos a estar como más protegidos y genere una mayor medicalización de la vida

*Pregunta 4. ¿Cuál es el sitio adecuado de los macrodatos y del DVR? ¿Cómo se sitúan dentro del marco de la investigación y la toma de decisiones (efectividad, eficacia, seguridad, planificación, etc.)?*

Los macrodatos permiten avanzar en campos de la investigación en los que previamente era difícil incidir. La implementación de herramientas de macrodatos en investigación con grandes bases de datos genómicas ha llevado a una revolución en la identificación de causas genéticas, marcadores diagnósticos y posibles dianas terapéuticas. Los macrodatos se han situado en una posición crucial dentro del programa de la investigación biomédica en nuestro entorno.

Las empresas, en general, pero también las del sector sanitario, computan como activos sus datos como un intangible más, con vistas a su propia valoración

*Pregunta 5. ¿Cree que los macrodatos y DVR pueden aportar evidencia de causalidad en situaciones que no sea viable realizar estudios experimentales?*

De entrada, se acepta que, los macrodatos, igual que los DVR, solo dan correlaciones. No sirven para comprobar hipótesis, todo lo más a generarlas para un futuro estudio. En este sentido, los macrodatos pueden complementar, pero no sustituir los ensayos clínicos.

En muchos casos las correlaciones te dan la información para tomar decisiones. Pero en algunas cuestiones, por ejemplo, perfiles o efectos secundarios de fármacos, no basta con correlaciones, sino que hay que identificar relaciones causa-efecto. Es verdad que los macrodatos ayudan a predecir, pero el dato más fiable a efectos de causalidad lo proporciona sin duda posiblemente el experimento controlado, un ensayo clínico, o uno pragmático.

Pero en esta cuestión existen algunas divergencias de opinión, aunque no radicales, sino de matiz.

Se acepta que hay casos en los que los macrodatos pueden acercarse a la causa-

lidad, si se trata de un estudio pragmático, observacional, completamente bien diseñado para el fin buscado. O de diseños cuasi experimentales, con grupo control, con un seguimiento retrospectivo-prospectivo, tratando de aproximar variables instrumentales características de estudios observacionales.

La FDA ya tiene definido lo que es evidencia científica derivada de las apps de salud; y la EMA tiene puesto en marcha un piloto, *Adaptive Pathways*, que para determinadas autorizaciones de tipo condicionado prevé utilizar estudios de DVR para complementar la información necesaria para la autorización.

Los macrodatos se puede convertir en una especie de filtro, o sea justo lo contrario de lo que se sugería antes: si los macrodatos muestran una correlación, pueden sugerir la conveniencia de llevar a cabo un ensayo clínico. Así, macrodatos se podrían aplicar como un cribado para generar hipótesis y preguntas de investigación

En cualquier caso, la correlación es muy importante cuando el análisis tiene un objetivo descriptivo, cuando lo que se busca no es el porqué de lo que pasa, sino qué está pasando. Se mencionan proyectos que utilizan grandes bases de datos e inteligencia artificial para reproducir la fisiología humana en avatares humanos. En estos avatares humanos se podrán testar nuevas terapias de manera virtual sin necesidad de incluir voluntarios en un ensayo clínico. En este sentido los macrodatos y los datos de vida real servirán para aportar evidencias de causalidad en entornos donde la realización de ensayos clínicos en humanos sea difícil.

*Pregunta 6. ¿Puede indicar ejemplos donde la evidencia experimental, tal como, la procedente de ensayos clínicos, sea contradictoria con la derivada de macrodatos o de DVR?*

Los panelistas no logran identificar ningún ejemplo en que se dé este supuesto. Lo habitual es que mediante macrodatos o de DVR se intente complementar un ensayo clínico, por ejemplo, extrapolar los resultados a una población distinta a la del ensayo.

*Pregunta 7. ¿Cómo respondería a situaciones en que este tipo de nueva evidencia (DVR y macrodatos) fuese contradictoria con la “evidencia experimental clásica”?*

Una situación en que este tipo de nueva evidencia (DVR y macrodatos) fuese contradictoria con la evidencia experimental clásica probablemente daría lugar a un gran debate. Sin embargo, el valor de los datos de vida real puede ser incluso mayor que el de datos utilizados en condiciones experimentales clásicas; la capacidad que tendrán los proyectos que utilicen bases de datos masivas para encontrar asociaciones clínicas entre diversos factores será indiscutible. Además, podrán introducir en los algoritmos nuevos conceptos como movilidad, actividad física, o actividad diaria registrados en otros sistemas (bancario, telefónico, u otros).

Un problema más preocupante en la introducción de nuevos medicamentos es que en algunos casos se han lanzado al mercado con ensayos clínicos de 40-50 pacientes y luego con la ayuda de macrodatos han visto que, al pasar a la población general, han aparecido efectos adversos evidentes.

Esa es una utilidad clara de los datos masivos (farmacovigilancia): el descubrimiento de señales pequeñas que pasaban desapercibidas (eventos adversos infrecuentes).

Un claro limitante a la generalización de macrodatos es la ausencia de la formación y los perfiles necesarios y que la administración no tenga herramientas para incorporarlos. Por ejemplo, en modelo MIR se sigue enseñando lo mismo que hace 40 años, las mismas especialidades. No existe el cuerpo o escala de semánticos, o de bioestadísticos. La aproximación que se persigue es de tipo multidisciplinar, pero el problema de los equipos es que existen unos problemas de comunicación entre los distintos perfiles que son muy importantes. Hay que ir hacia perfiles mixtos.

**Pregunta 8.** *Interfaz entre regulación (Reglamento General de Protección de Datos) y acceso a datos. ¿Cuáles son los riesgos de la utilización de DVR y macrodatos para la protección de la privacidad de los pacientes y ciudadanos?*

Los panelistas son conscientes de los problemas de seguridad que plantean los macrodatos. Los expertos en seguridad suelen dar por hecho que todos damos los datos sin que parezca que nos importe mucho la seguridad. Pedir a la gente que se autolimita en las redes sociales (RRSS) se da por perdido, en todo caso hay que regular bien cómo y para qué se utilizan los datos. La solución no debería buscarse tanto en la autolimitación personal, sino en regular adecuadamente al que utiliza los datos.

Los macrodatos y los DVR, las dos variables concretas a las que se ciñe el Reglamento nuevo, entrarían en la categoría de datos sensibles, que en principio no se pueden tratar. El Reglamento ha querido subrayar que, junto con otros datos sensibles, los datos genéticos y de salud no se pueden tratar salvo que se den una serie de condiciones. Las cinco más importantes a efectos de este tema serían: Uno, el consentimiento. Dos, que sea necesario en el ámbito laboral o de la Seguridad Social. Tres, que es un asunto de supervivencia, es decir, prevenir o curar una enfermedad, o gestionar esos aspectos. Cuatro, que haya un interés público en materia de salud, epidemias, calidad y seguridad de la sanidad, medicamentos, etc. Y quinto y último, por supuesto, la investigación científica.

Los remedios más importantes que prevé la legislación incluyen el derecho de oposición, es decir, que no se consientas ya a que se siga tratando información tuya, habiendo consentido previamente, pero se ha introducido uno nuevo, que es el derecho a oponerte a que se tracen perfiles tuyos en línea (Internet). Luego están las decisiones automáticas individualizadas, es decir, que no haya un algoritmo que pueda tomar decisiones sin que intervenga una persona, y que al final te acabe generando consecuencias. También están la seguridad y privacidad, que constituyen un binomio. Finalmente, el reglamento establece que se debe trasladar al titular de la información cómo se maneja la inteligencia en la organización y con qué consecuencias. Todo esto en materia de salud es determinante, porque son datos sensibles.

Hay que tener en cuenta que de todos los macrodatos relevantes en el tema de la salud prácticamente lo único que se puedes controlar desde el sistema de salud es la historia clínica. Una gran parte de los datos relevantes los suministra el sujeto más o menos consciente y voluntariamente por otras vías. Por eso se acepta que el problema no es tanto que la gente acepte formalmente que se utilicen no o no sus datos, sino que se controle a quien vaya a utilizarlos y que este explicita el uso que les va a dar.

En el ámbito de la salud puede ser conveniente aplicar el denominado consentimiento amplio, que significa que alguien permite que se usen sus datos para que se investigue en todo lo relacionado, por ejemplo, con el aparato locomotor, pero no con oncología u otras áreas. Se está barajando un consentimiento amplio, por grandes áreas de conocimiento, pues si los investigadores tuviesen que pedir consentimiento para cada proyecto al que se aplicasen los datos registrados de un paciente, la investigación mediante macrodatos se haría inviable.

#### 4. Conclusiones

La utilización de macrodatos en España está en una fase incipiente. Existen bastantes iniciativas, pero se echa en falta en España un discurso cohesionado, lineal, potente, de todo el Sistema Nacional de Salud, que supere la atomización y la dispersión de esfuerzos de todo lo que se está haciendo ahora, para evitar que termine en fracasos y pérdida de oportunidades. La interoperabilidad de toda la información relevante entre CCAA y a nivel de instituciones es posiblemente el paso más urgente.

Se debería tener en cuenta el riesgo de desintermediación: van a venir a España compañías internacionales que ya están en esta actividad y que van a ofrecer soluciones que acabarán por aceptar, tanto las instituciones como los pacientes. Algo similar a lo que ha pasado con servicios metropolitanos de transporte. Se corre el riesgo que al final la mayoría de las iniciativas locales en que se han invertido recursos y esfuerzos no lleguen a ningún puerto y que la actividad acabe en manos de unas pocas empresas multinacionales.

En cuanto a la información, como bien público que es, es importante asegurar el acceso asequible de todos los usuarios legítimos a los datos relevantes para la investigación y gestión en salud.

Para empezar, hay que saber hacia dónde se quiere ir y cuáles son los requerimientos básicos que se debe compartir para que en un futuro se puedan promover proyectos y establecer plataformas útiles.

Las bases para poner un esquema común de interoperabilidad que nos permita crecer en el sentido de estar conectados están ahí. Pero es necesario seguir en esta línea porque nos enfrentamos a algo muy complejo y muy heterogéneo. Es necesario poner de acuerdo los médicos y los farmacéuticos.

Para conseguir que todo el mundo aporte datos al SNS el Ministerio de Sanidad se debería dar algunos incentivos; así es como las CCAA dan datos al Plan Nacional del Sida o al Plan Nacional de Drogas.

El ensayo clínico aleatorizado (ECA) tiene que seguir siendo el Gold Standard como fuente de evidencia en salud y para la autorización de un medicamento. Sin embargo, es conveniente investigar y regular en qué forma y medida los macrodatos, igual que los DVR y otras fuentes de información, pueden complementar los ECA en determinadas situaciones.

Cada vez más se señala la incapacidad de los ECA para aportar información relevante en toma de decisiones en algunas situaciones y contextos. Ciertos

ámbitos clínicos carecen de la necesaria “equipoise” (incertidumbre clínica) para realizar un ECA. Muchos ejemplos provienen del ámbito quirúrgico donde el manejo conservador podría tener indicación, pero la estrategia quirúrgica se realiza desde tiempos muy remotos. También son situaciones limitadas para el ECA aquellas situaciones de riesgo vital donde el desarrollo del ECA presenta inconvenientes éticos, metodológicos y operativos. En todos ellos el empleo de estrategias basadas en macrodatos probablemente complementen la información actual.

Otro aspecto a tener en cuenta es la integración de macrodatos en fases previas a los diseños de ECA. Actualmente la eficacia estadística de los ECA ha disminuido en muchos ámbitos clínicos. Esto es debido en parte, a la disminución progresiva de la mortalidad en muchas patologías que determina riesgos basales bajos y falsos negativos en los diseños. También, la inclusión en los estudios de participantes cada vez más complejos (pluripatológicos, polimedcados etc.) junto con la necesidad emplear desenlaces a largo plazo relevantes van en la misma línea. Así el empleo de estrategias con macrodatos puede permitir extraer información suficiente para determinar y perfilar los criterios de inclusión los ECA, seleccionar subpoblaciones o informar sobre la viabilidad del ECA.

Aunque los miembros del Panel no perciben que el nuevo Reglamento General de Protección de Datos vaya a constituir un obstáculo substancial para el uso de datos del sistema de salud a efectos de práctica clínica, investigación o gestión, es importante promover y monitorizar su reglamentación específico y su aplicación estricta en un tema tan sensible como es el de la salud.



## 4. HOJA DE RUTA

---



## 4. HOJA DE RUTA

Hoy es todavía difícil vislumbrar todas las oportunidades que podrán derivarse de la implantación de estas nuevas formas de entender y manejar los datos en los sistemas sanitarios<sup>1</sup>. El hecho de interconectar diferentes bases de datos, tanto sanitarias como sociales, la contribución de la eSalud y el IoT (Internet of Things o Internet de las Cosas), puede suponer una auténtica revolución en la sanidad con un potencial prometedor.

Lo que sí hay es un cierto consenso en la literatura estudiada acerca de la aplicabilidad de la utilización de los datos masivos en el entorno sanitario para:

- Obtener información sobre las intervenciones sanitarias que se producen en la práctica asistencial.
- Estimar la efectividad y seguridad de una intervención sanitaria a largo plazo frente a la práctica habitual.
- Obtener y comunicar los resultados de pruebas clínicas con mayor rapidez.
- Conocer mejor el impacto en la calidad de vida, la satisfacción de los pacientes o sobre la capacidad y funcionalidad de las distintas intervenciones sanitarias.
- Ayudar a calcular los costes de los servicios de salud a tiempo real y de manera más completa.
- Mejorar el impacto de intervenciones sanitarias, especialmente preventivas.

Y, aquí van, tras la reflexión a la que nos ha llevado este estudio, algunas de las oportunidades de la utilización de los macrodatos en sanidad:

### **Mejora de la solvencia y la perdurabilidad en el tiempo de los servicios sanitarios**

Dada la transición demográfica en curso y el nuevo patrón epidemiológico, estas nuevas tecnologías pueden ayudar a mejorar la eficiencia<sup>3</sup>:

- Diagnósticos más rápidos y precisos al poder comparar las variables clínicas del paciente con grandes bases de datos poblacionales. Derivado de ello, se requerirán un menor número de pruebas de imagen o laboratorio y por tanto un ahorro en el coste por diagnóstico. Se espera que sea especialmente destacable en el caso del diagnóstico de enfermedades raras.
- La utilización de los sistemas de teleasistencia sanitaria a través de apps móviles o el IoT puede derivar en una optimización de los tiempos de consulta y evitar desplazamientos innecesarios. También, la telemonitorización de constantes a distancia puede contribuir a evitar episodios agudos en urgencias e ingresos innecesarios.
- Prescripción de tratamientos que han demostrado ser más eficaces y seguros en pacientes con las mismas características. Permite seleccionar el fármaco óptimo y evitar iatrogenia y despilfarro.
- Recopilar y explotar datos demográficos, de estilos de vida (conducta) y sociales de la población para mejorar y optimizar las estrategias de prevención secundaria y cribado de determinadas enfermedades.

Otra pregunta pertinente sería, ¿disminuirán, el empleo de los macrodatos, los costes en los sistemas sanitarios tras su implantación y uso extenso? Ciertamente, no lo sabemos. Tras el análisis de las necesarias inversiones iniciales tanto en adquirir e implantar estas nuevas tecnologías, como en la contratación de analistas de datos, parece probable y todo apunta a que una vez estén implantadas y desarrolladas, puedan generar ahorros, vía el uso adecuado, es decir, evitando sobreutilización de recursos. Sin duda, un factor clave para alcanzar la eficiencia en estos proyectos es que sea el Sistema Nacional de Salud quien ocupe una posición de liderazgo. Si la gestión y análisis de la información queda en manos de externos, es muy probable que el control sobre la información y la dependencia económica merme muchas de las potencialidades propuestas.

### **Puede orientar y complementar la investigación clínica**

La medicina basada en la evidencia (MBE), con el ensayo clínico aleatorizado como *gold standard*, debe ser valorada críticamente y aunque supera en bondades, tiene una serie limitaciones:

- Tiempo: los ensayos clínicos deben tener una duración adecuada en función de la intervención y la enfermedad. En ocasiones este tiempo se queda corto para medir efectos positivos o negativos a largo plazo.

- **Coste:** realizar un ensayo clínico supone un impacto económico muy relevante que solo se pueden permitir grandes corporaciones.
- **Poblaciones especiales:** para poder extraer conclusiones de los ensayos clínicos se deben evitar al máximo las variables de confusión. Es por ello que se selecciona muy bien a los pacientes que pueden participar. Sin embargo, la población a la que va destinada el fármaco no es la ideal, por lo que puede no estar representada de manera adecuada en los ensayos clínicos.
- **Variables a estudio:** en un ensayo clínico se monitorizan una serie de variables que se prevé que puedan afectar al resultado de la intervención. Sin embargo, se obvian otras variables que aún no se han estudiado pero que acompañan al paciente e influyen en los resultados.
- **Monitorización:** los pacientes incluidos en los ensayos clínicos son monitorizados de manera estrecha durante todo el tiempo a estudio. Así, se minimizan los problemas derivados de la intervención y se asegura un cumplimiento del protocolo estrictamente por parte de los profesionales sanitarios y los pacientes. Presentan mejor adherencia que en la vida real.

Es posible que sea la mejor estrategia para incorporar con garantías las nuevas tecnologías a la práctica asistencial. Cumple con los cinco principios que rigen la bioética sanitaria: beneficencia, no maleficencia, autonomía, justicia y eficiencia entendida como no despilfarrar.

Los macrodatos en el futuro muy cercano ofrecerán un complemento muy interesante. Así, la nueva tecnología se aplicará a población real, sobre la que se realizará una monitorización telemática continua de más variables que las empleadas en los ECA, con mayor rapidez y a un coste mucho menor y con la misma seguridad.

### **Caminaremos hacia una medicina más personalizada**

Los desarrollos en genómica unidos al potencial de los macrodatos van a permitir seleccionar la terapia óptima para cada tipo de paciente en función de sus características intrínsecas. Además, contaremos con datos del entorno social y ambiental junto la monitorización de los estilos de vida que permitirán una mayor personalización de la práctica de la medicina.

Será posible predecir qué medicamentos son más efectivos para un determinado perfil de paciente y personalizar el tratamiento para cada individuo.

### **Mejoraremos en prevención y diagnóstico temprano de enfermedades**

La monitorización continua de los usuarios mediante el Internet de las cosas, las aplicaciones móviles o las redes sociales es ya una realidad. Las variables de salud que recogen estos dispositivos analizados mediante la tecnología de macrodatos con las historias clínicas electrónicas detectarán a través de los análisis predictivos, enfermedades que aún no se han desarrollado y permitirán a los sistemas sanitarios actuar a tiempo.

### **Será más fácil orientar el modelo sanitario hacia los pacientes crónicos**

Los pacientes con enfermedades crónicas obtendrán un impacto más significativo de las intervenciones sanitarias por el manejo inteligente de estas nuevas tecnologías.

Las razones son:

- Posibilidad de monitorizar a pacientes domiciliarios mediante dispositivos médicos conectados remotamente. Permitirá prevenir episodios agudos y por lo tanto visitas a urgencias y hospitalizaciones innecesarias.
- Mayor implicación en la investigación clínica. Los pacientes crónicos se caracterizan por su comorbilidad y polifarmacia, situación que les suele dejar fuera de los ensayos clínicos tradicionales. Con los nuevos enfoques ya comentados van a poder participar en el desarrollo de la innovación.
- La evolución lenta de las enfermedades crónicas las hace más adecuadas para realizar análisis predictivos y evaluar las consecuencias de las acciones empleadas.

Qué duda cabe que hay barreras que pueden limitar la implantación y utilización de los macrodatos en el ámbito de la sanidad. Las describimos a continuación:

### **Estructura de los sistemas de información sanitaria**

El empleo de macrodatos requiere un sistema organizativo que permita que la información fluya a través de toda la red (estructuras e instituciones). Por el contrario, a día de hoy, la mayor parte del SNS está formado por un conjunto de compartimentos estancos con la capacidad de generar un gran volumen de información, que queda almacenada sin que apenas sea utilizada o explotada con inteligencia y diligencia. Si bien es cierto que la propia tecnología intrínsecamente y el empeño de los responsables de los sistemas de información cada más cualificados y motivados, están generando avances muy relevantes. Un ejemplo serían los distintos canales digitales para los pacientes que propician un acceso más fácil y mejor a toda la información que el sistema tiene de ellos, de la que son legítimos propietarios.

Sin embargo, siguen existiendo ciertos problemas de interoperabilidad con prescripciones e historiales clínicos, más en algunas comunidades autónomas que otras.

Por último, hay margen de mejora en la conexión entre la atención primaria y la especializada, entre el sector social y el sanitario y, entre la sanidad privada y la pública.

### **Confidencialidad**

Los datos de salud, como no podía ser de otra manera, son considerados muy sensibles. Existe una estricta legislación para conservar el derecho a la privacidad y confidencialidad de los pacientes en la que no está contemplada la utilización de los datos clínicos de forma masiva aun garantizando su anonimización. Representa un obstáculo para extender todo el potencial de la investigación de servicios sanitarios. Es preciso contar con una regulación que permita, por un lado, respetar la privacidad, y por otro, poder aprovechar el potencial de la tecnología de macrodatos en la investigación, en la práctica clínica y en la gestión de los servicios sanitarios.

### **Soporte técnico**

Para desarrollar la tecnología de los macrodatos se deben afrontar primero algunos avances tecnológicos como la analítica en tiempo real, el procesado de

datos escalable y con baja latencia, las interfaces de usuario nuevas y enriquecidas, la interacción y enlace de datos, información y contenido<sup>12</sup>.

La realidad del Sistema Sanitario Español está lejos de haber alcanzado los estándares necesarios para la aplicación efectiva de los macrodatos. Aunque el avance es imparable. Transitamos a una digitalización completa del sector.

### Recursos humanos

La dotación de personal especializado en macrodatos, tal como se ha comentado profusamente por parte de los panelistas del Grupo de Discusión, es una de las limitaciones a la que menos importancia se le otorga<sup>13</sup>. Sin embargo, se trata de una de las más importantes a día de hoy.

Para poder desarrollar iniciativas con el empleo de macrodatos en cualquier ámbito se debería dotar a los recursos humanos que operen con suficiente capacidad y formación en su manejo específico. Para lograrlo sería necesario actuar en tres vertientes: formación interna, redefinir puestos de trabajo y la selección de nuevo personal especializado, hoy por hoy, imposible en el sector público.

Sin embargo, en la función pública los procedimientos de selección de personal están muy limitados, obligando a la contratación mediante el sistema de concurso oposición. Este hecho dificulta la selección de perfiles muy especializados, por lo que se hace aún más imprescindible la formación interna del personal ya contratado en estas nuevas tecnologías de macrodatos, mientras se despeje la posibilidad de que la oferta de empleo público pueda incorporar este nuevo tipo de perfiles (*data scientist*).

### Profesionales sanitarios

Uno de los riesgos más desconocidos, que también fueron discutidos en el Grupo de Discusión por los panelistas y que no han sido estudiados a fondo, es el relacionado con las importantes reticencias que pueden aparecer entre los profesionales sanitarios. La incorporación de los macrodatos a la práctica asistencial de la medicina podría requerir un mayor nivel de manejo por parte de los clínicos de estas nuevas tecnologías. Para poder extraer datos fiables, los profesionales deben incorporar a su quehacer diario, de manera sistemática y

con la mayor calidad posible, solvencia en el empleo de las mismas. Este hecho, en el caso de los reticentes o inexpertos puede suponer un auténtico desafío que entra de lleno en el terreno de la resistencia al cambio.

La extracción de conclusiones poblacionales y su incorporación a la práctica asistencial también puede tener problemas. Pasaremos de “a mis pacientes les va bien” a los resultados de cientos de miles de pacientes similares extraídos mediante tecnologías todavía no enseñadas ni pilotadas con la participación activa de los profesionales sanitarios. Perder autonomía generará reticencias que sólo se pueden vencer con formación continua bien planeada y mejor implantada.

### ACCIONES

Nos quedaría graduar, a la vista de lo todo lo anterior, qué ocurrirá, arrojando una cierta luz al desarrollo del potencial que tiene el empleo de los macrodatos en España.

La primera parte de este trabajo ha permitido conocer el estado del arte en profundidad, revisando a fondo todo lo publicado. A partir de esta **se recomiendan tres acciones:**

- 1. Asegurar la calidad, validez y replicabilidad de los datos empleados.** Para ello, es necesario que las instituciones sanitarias hagan un esfuerzo en mejorar la gestión que hacen de la información clínica.
- 2. Formar nuevos perfiles profesionales capaces de trabajar en este ámbito.** Dichos profesionales deberán aunar competencias de campos que hasta ahora se han desarrollado de forma independiente como son la epidemiología, la bioinformática y las ciencias de la salud.
- 3. Determinar qué preciso lugar ocupará la minería de macrodatos en la generación del conocimiento.** No parece que este tipo de metodología sustituya completamente a los métodos clásicos de inferencia y producción de conocimiento. Destacamos la capacidad para generar nuevos modelos de predicción, más precisos que los tradicionales y aportar datos mucho más vinculados a la vida real.

En la segunda parte de este informe se ha realizado un análisis del discurso sobre el desarrollo de un muy estimulante Grupo de Discusión. A la vista de los resultados y las conclusiones, **se recomiendan las siguientes ocho acciones:**

1. **Elaborar un discurso cohesionado, lineal, potente, de todo el Sistema Nacional de Salud, que supere la atomización y la dispersión de esfuerzos de todo lo que se está haciendo ahora, para evitar que termine en fracasos y pérdida de oportunidades. Se detecta la interoperabilidad de toda la información relevante entre CCAA y a nivel de instituciones como el paso más urgente.**
2. **Se debe tener en cuenta el riesgo de desintermediación.** Puede que la mayoría de las iniciativas locales del sistema público en que se han invertido recursos y esfuerzos no lleguen a puerto alguno y que la actividad salvable acabe en manos de terceros, fundamentalmente del sector privado, que no lo utilicen con objetivos de interés global.
3. **La información, como bien público que es, debe contar con un acceso asequible, de todos los usuarios legítimos, a los datos relevantes para la investigación y gestión en salud.**
4. **Hay que definir hacia dónde se quiere ir e identificar cuáles son los requerimientos básicos que se deben compartir para que en un futuro se puedan promover proyectos y establecer plataformas útiles.**
5. **Existen las bases para poner un esquema común de interoperabilidad que nos permita crecer en el sentido de estar conectados. Nos enfrentamos a algo muy complejo y muy heterogéneo. Por ello, es necesario poner de acuerdo a todos los actores, especialmente a los médicos, a los farmacéuticos, a los gestores y a los pacientes.**
6. **Hay que manejar incentivos con inteligencia para que las CCAA's aporten datos a un sistema de información integrado bajo el control del Ministerio de Sanidad, Consumo y Bienestar Social.**
7. **El ensayo clínico aleatorizado (ECA) tiene que seguir siendo el *Gold Standard* como fuente de evidencia en salud y para la autorización de un medicamento. Sin embargo, es conveniente investigar y regular en qué forma y medida los macrodatos, igual que los DVR y otras fuentes de in-**

formación, pueden complementar los ECA en determinadas situaciones, bien para sugerir hipótesis a investigar, como para extrapolar o complementar los resultados de los ECA.

8. El nuevo Reglamento General de Protección de Datos no constituirá un obstáculo substancial para el uso de datos del sistema de salud a efectos de práctica clínica, investigación o gestión. Sin embargo, es importante promover y monitorizar su desarrollo y su aplicación estricta que evite cualquier riesgo de mercantilización de los datos en un campo tan sensible como es el de la salud.

Por último, terminamos con deseos que esperemos no se separen mucho de las realidades antes expuestas:

### Ganará terreno la medicina predictiva

La medicina predictiva se basa en la capacidad de los macrodatos de realizar predicciones a partir de grandes volúmenes de información que ya recopilamos. Proceden de los dispositivos de monitorización de los pacientes a tiempo real. Con ella aplicamos herramientas de análisis y correlación junto a otros datos procedentes de la historia clínica electrónica (como analíticas, imágenes médicas, tratamientos prescritos...). Con toda esta información relacionada se detectarán patrones que alertarán de la posible aparición de enfermedades antes incluso de que aparezcan los primeros síntomas.

### El paciente será más experto

Durante los últimos años estamos asistiendo a profundos cambios sociales en los que el ciudadano reivindica un mayor poder de participación y expresa su voluntad de implicarse en la toma de decisiones que afectan a su vida. En el ámbito de la salud sucede lo mismo y el paciente se corresponsabiliza cada vez en mayor medida de la gestión de su enfermedad y su tratamiento, buscando activamente información, participando en foros y redes sociales y estableciendo relaciones tanto con otros pacientes como con profesionales sanitarios a través de internet.

En este contexto, ha tomado fuerza el concepto de paciente empoderado, activo o experto, término que hace referencia a aquella persona preparada, capacitada y comprometida con su salud y su tratamiento. Se trata de un paciente activo en lo que se refiere a su salud y que los canales de comunicación digitales para mantener y mejorar su estado de salud. Además, el término no se refiere solamente al propio paciente, sino también a las personas de su entorno.

Dado que está muy reconocido que la figura del paciente activo se relaciona con mejores resultados clínicos, tanto a nivel individual como desde el punto de vista del sistema de salud, en los últimos años se han puesto en marcha programas que pretenden formar a los pacientes en habilidades, técnicas y competencias que les permitan mejorar su nivel de autocuidado y participar activamente en la toma de decisiones que afectan a su salud.

Más allá del evidente beneficio que la implicación del paciente en su salud tiene a nivel individual, no hay que olvidar las consecuencias que también tiene a nivel del sistema sanitario en términos de menor y mejor utilización de los recursos disponibles, lo que sin duda puede contribuir decisivamente a hacer el sistema sostenible. Esto es especialmente relevante en el caso de los pacientes crónicos, teniendo en cuenta que se estima que alrededor del 80% del gasto sanitario corresponde a estos pacientes. Por tanto, conseguir su implicación en la gestión de su salud resulta crítico para la sostenibilidad del sistema. De hecho, se estima que el empoderamiento del paciente debe ser una de las partes fundamentales de cualquier reforma de la gestión de las enfermedades crónicas por su contribución a maximizar la eficiencia.

El desarrollo del paciente experto o activo está muy relacionado con el auge de los macrodatos. Un paciente formado e involucrado con su enfermedad y su tratamiento va a contribuir mucho mejor a recoger información y que ésta sea de mayor calidad que un paciente que no lo esté. La información que recoge este paciente solo lo puede hacer él mismo, por lo que resulta aún más imprescindible su contribución.

Ya existen herramientas digitales destinadas a la recogida de datos por parte de los propios pacientes que permiten la vinculación con bases de datos clínicas. El ejemplo más destacado es el de las aplicaciones móviles de tipo recordatorio: medicación, citas, ... Muchas de estas aplicaciones permiten conectar estos datos recogidos por los pacientes con los profesionales sanitarios a través de las carpetas de salud desarrolladas en muchas comunidades autónomas.

### Aprenderemos de la realidad virtual

La virtualidad permite realizar una ruptura en la relación entre el usuario y sus coordenadas espaciotemporales, es decir, el dónde y el cuándo se encuentra físicamente. La simulación del entorno físico puede incorporar patrones de movimiento y velocidad que alteran los parámetros originales de la relación tradicional del sujeto con la realidad física. Así, en la realidad virtual se utilizan imágenes y sonidos, sensores y superficies que recrean un entorno de forma artificial y que permite la interactividad. Se suele utilizar mediante gafas y sistemas donde no se puede apreciar la realidad existente.

Durante años, se ha considerado una tecnología exclusiva para la industria del entretenimiento. Sin embargo, en los últimos años se han encontrado muchas aplicaciones prácticas para la realidad virtual que van más allá<sup>14</sup>.

En este sentido, ya se está empezando a utilizar la realidad virtual en las ciencias de la salud. Algunas de las aplicaciones que se están desarrollando son:

- Educación teórica: formación de profesionales sanitarios en anatomía mediante la recreación del cuerpo humano.
- Formación práctica: permite a los profesionales realizar simulaciones de diferentes intervenciones antes de tener que realizarlas sobre los pacientes. Por ejemplo, realización de intervenciones quirúrgicas complicadas.
- Terapia de exposición: animaciones que recrean entornos controlados para tratar a pacientes con estrés postraumático, fobias o adicciones.
- Rehabilitación cognitiva: para pacientes que han sufrido algún tipo de daño cerebral.
- Concienciación: permite a la sociedad ver el mundo como lo perciben algunos colectivos con discapacidad haciendo una labor de sensibilización social.

La creación de estos entornos simulados de imágenes y sonidos requiere un volumen de datos y una velocidad de conexión extremadamente altos, por lo que se puede decir que el avance de la realidad virtual va ligada al progreso de los macrodatos. Nos será ni lo primero ni lo más importante para nuestro sistema sanitario eminentemente público. Pero es una realidad insoslayable.

### Digitalización in crescendo

Las diferentes aplicaciones de los macrodatos ya están transformando el mundo de la sanidad. Ofrecen a usuarios y profesionales una nueva gama de posibilidades que van desde la historia clínica electrónica hasta el desarrollo de aplicaciones móviles que permiten la monitorización constante y a tiempo real de los pacientes. De todas ellas las que más impacto están teniendo en los últimos años son las relacionadas con la eSalud, como las aplicaciones móviles o los dispositivos digitales. La que llegará a revolucionar los sistemas sanitarios, son los macrodatos que a través de la minería está desarrollando la inteligencia artificial a gran velocidad.

Se apunta a que la tecnología de la información en el entorno de la salud supondrá una mejora significativa, cuya magnitud aún no se puede predecir completamente, en la calidad de la atención a los usuarios, además de la prevención, diagnóstico y tratamiento temprano de enfermedades, unido a una disminución de los costes sanitarios.

Sin embargo, para que se logren estas metas, los macrodatos tienen que superar, como hemos visto, una serie de limitaciones. Los sistemas sanitarios deben continuar por el camino de la integración de los datos y su interoperabilidad. Hay que cualificar a los profesionales de todos los estamentos en estas nuevas tecnologías con planes y programas de formación continua ad hoc. Hay que establecer un marco normativo que consiga alinear los requisitos necesarios de confidencialidad y privacidad, con la necesaria evolución de la innovación en curso, que es ciertamente disruptiva y que cambiará estatus quo.

Pese a que ya existen algunas iniciativas que utilizan estas tecnologías en el sector de la sanidad, aún no se ha conseguido que arraiguen en la estructura de nuestros servicios sanitarios regionales. Están en fase incipiente y ciertamente poco compartidas. Uno de los mayores retos que tiene el empleo de la minería a los macrodatos es dejar atrás las fases piloto y establecerse en la práctica clínica asistencial como un elemento más que aporte valor. Paradójicamente, este cambio depende menos de los retos tecnológicos que implica, que de la propia voluntad de los gestores en ponerlo en práctica, ya que son conscientes de la gran dificultad que entraña cualquier cambio de tal magnitud, donde muchos de los “afectados” se resistirán.

Veremos.

## BIBLIOGRAFÍA

---



## BIBLIOGRAFÍA

---

1. Jyoti R (2017). *IDC's Worldwide Big Data and Analytics Software Taxonomy*. Disponible en: <https://www.idc.com/getdoc.jsp?containerId=US42555117>
2. Martínez Sesmero JM. "Big Data": aplicación y utilidad para el sistema sanitario. *Farm Hosp.* 2015;39(2):69-70.
3. Fundación Vodafone España y Red.es (2016). Big Data en salud digital. Informe de resultados. Disponible en: <http://www.fundacionvodafone.es/publicacion/informe-big-data-en-salud-digital>
4. W. Raghupathi, V. Raghupathi, Big data analytics in healthcare: promise and potential, *Heal. Inf. Sci. Syst.* 2 (2014) 3, <http://dx.doi.org/10.1186/2047-2501-2-3>.
5. K.R. Ghani, K. Zheng, J.T. Wei, C.P. Friedman, Harnessing big data for health care and research: are urologists ready? *Eur. Urol.* 66 (2014) 975–977, <http://dx.doi.org/10.1016/j.eururo.2014.07.032>.
6. Emilie Baro, Samuel Degoul, Regis Beuscart, E. Chazard, Toward a literature-driven definition of big data in healthcare, *BioMed. Res. Int.* (2015).
7. De Roos D. *Hadoop® For Dummies®*. John Wiley & Sons. 2014. Hoboken, New Jersey.
8. Power Data: Especialistas en Gestión de Datos. ¿Qué es un Sistema de Gestión de Datos Big Data? Blog. 10 enero, 2017. Disponible en: <https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/que-es-un-sistema-de-gestion-de-datos-de-big-data>
9. Grupo IGN (2017). Tipos de analítica Big Data. Disponible en: <https://ignsl.es/analitica-big-data>
10. Luo, J, Wu, M, Gopukumar, D, & Zhao, Y. (2016). Big data application in biomedical research and health care: a literature review. *Biomedical informatics insights*, 8, BII-S31559.
11. Del Llano Señarís JE, Del Castillo JA, Dilla Quintero T, García López JL. Datos de la vida real en el sistema sanitario español. Ed. Fundación Gaspar Casal, 2016. Disponible en: [http://www.fgcasal.org/publicaciones/Datos\\_de\\_la\\_vida\\_real\\_SSE\\_2016.pdf](http://www.fgcasal.org/publicaciones/Datos_de_la_vida_real_SSE_2016.pdf)
12. Instituto Tecnológico de Informática (ITI) (2015). Big Data: El valor de los Datos. Disponible en: <https://www.coiicv.org/actualidad/descargas/download/23-monografias/467-big-data-el-valor-de-los-datos>

13. Fundación Cotec para la Innovación (2017). Generación de talento. Big data en España. Disponible en: <http://cotec.es/media/BIG-DATA-FINAL-web.pdf>
14. Li L, Yu F, Shi D, Shi J, Tian Z, Yang J, Wang X, Jiang Q. Application of virtual reality technology in clinical medicine. *Am J Transl Res*. 2017 Sep 15;9(9):3867-3880.

Referencias de los artículos incluidos en la Revisión Sistemática de la literatura:

1. Al-jawahiri, R., & Milne, E. (2017). Resources available for autism research in the big data era: a systematic review. *PeerJ*.
2. Alonso, S., de la Torre Díez, I., Rodrigues, J., Hamrioui, S., & López-Coronado, M. (2017). A Systematic Review of Techniques and Sources of Big Data in the Healthcare Sector. *Journal of Medical Systems*.
3. Balla, A., Rodríguez, G. B., Corradetti, S., Balagué, C., Fernández-Ananín, S., & Targarona, E. M. (2017). Outcomes after bariatric surgery according to large databases: a systematic review. *Langenbeck's archives of surgery*, 402(6), 885-899.
4. Celtikci, E. (2018). A systematic review on machine learning in neurosurgery: The future of decision-making in patient care. *Turkish Neurosurgery*.
5. Dewangan, R., Thombre, D., & Patel, C. (2016). Big data technology in health and biomedical research: A literature review. *International Journal of Database Theory and Application*.
6. Dias, Á., Mansur, C., Myczkowski, M., & Marcolin, M. (2011). Whole field tendencies in transcranial magnetic stimulation: A systematic review with data and text mining. *Asian Journal of Psychiatry*.
7. Fallah, M., & Kalhori, S. (2017). Systematic review of data mining applications in patient-centered mobile-based information systems. *Healthcare Informatics Research*.
8. Gianfredi, V., Bragazzi, N., Nucci, D., Martini, M., Rosselli, R., Minelli, L., & Moretti, M. (2018). Harnessing Big Data for Communicable Tropical and Sub-Tropical Disorders: Implications From a Systematic Review of the Literature. *Frontiers in Public Health*.
9. Hamrioui, S., De La Torre Díez, I., Garcia-Zapirain, B., Saleem, K., & Rodrigues, J. (2017). A systematic review of security mechanisms for big data in health and new alternatives for hospitals. *Wireless Communications and Mobile Computing*.
10. Kruse, C., Goswamy, R., Raval, Y., & Marawi, S. (2016). Challenges and Opportunities of Big Data in Health Care: A Systematic Review. *JMIR Medical Informatics*.

11. Li, J., Ding, W., Cheng, H., Chen, P., Di, D., & Huang, W. (2016). *A Comprehensive Literature Review on Big Data in Healthcare*.
12. Librenza-Garcia, D., Kotzian, B., Yang, J., Mwangi, B., Cao, B., Pereira Lima, L., . . . Passos, I. (2017). The impact of machine learning techniques in the study of bipolar disorder: A systematic review. *Neuroscience and Biobehavioral Reviews*.
13. Marinov, M., Saleh, A., Mosa, M., Yoo, I., & Boren, S. (2011). *Data-Mining Technologies for Diabetes: A Systematic Review*.
14. Mehta, N., & Pandit, A. (2018). Concurrence of big data analytics and health-care: A systematic review. *International Journal of Medical Informatics*.
15. Pashazadeh, A., & Navimipour, N. (2018). Big data handling mechanisms in the healthcare applications: A comprehensive and systematic literature review. *Journal of Biomedical Informatics*.
16. Pirri, S., & Turchetti, G. (2017). The Application Of Big Data Analysis To Measure And Better Understand Medication Adherence: A Systematic Literature Review. *Value in Health*.
17. Rodríguez, G. B., Balla, A., Corradetti, S., Martinez, C., Hernández, P., Bollo, J., & Targarona, E. M. (2018). What have we learned in minimally invasive colorectal surgery from NSQIP and NIS large databases? A systematic review. *International journal of colorectal disease*, 1-19.
18. Santos, A., Teles, Í., Siqueira, O., & de Oliveira, A. (2018). Big data: A systematic review. *Advances in Intelligent Systems and Computing*.
19. Senders, J., Staples, P., Karhade, A., Zaki, M., Gormley, W., Broekman, M., . . . Arnaut, O. (2018). Machine Learning and Neurosurgical Outcome Prediction: A Systematic Review. *World Neurosurgery*.
20. Sinnenberg, L., Buttenheim, A., Padrez, K., Mancheno, C., Ungar, L., & Merchant, R. (2017). Twitter as a tool for health research: A systematic review. *American Journal of Public Health*.
21. Sochacki, K., Jack, R., Safran, M., Nho, S., & Harris, J. (2018). There Is a Significant Discrepancy Between “Big Data” Database and Original Research Publications on Hip Arthroscopy Outcomes: A Systematic Review. *Arthroscopy - Journal of Arthroscopic and Related Surgery*.
22. Thakur, S., & Ramzan, M. (2016). A systematic review on cardiovascular diseases using big-data by Hadoop. *Proceedings of the 2016 6th International Conference - Cloud System and Big Data Engineering, Confluence 2016*.
23. Wallace, B., Small, K., Brodley, C., Lau, J., Schmid, C., Bertram, L., . . . Trikalinos, T. (2012). Toward modernizing the systematic review pipeline in genetics: Efficient updating via data mining. *Genetics in Medicine*.

24. Wang, W., & Krishnan, E. (2014). Big data and clinicians: A review on the state of the science. *Journal of Medical Internet Research*.
25. Yassin, N., Omran, S., El Houby, E., & Allam, H. (2018). Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review. *Computer Methods and Programs in Biomedicine*.

## AUTORES

---



## AUTORES

---

### **Jesús A. Barea Mendoza**

Licenciado en Medicina “Cum Laude”. Universidad de Granada (2005-11). Facultativo Especialista Medicina Intensiva. Hospital 12 de Octubre (Madrid). Jefe de Residentes del Hospital 12 de Octubre (2017-18). Especialista en Metodología de la Investigación Científica. Universidad de Alcalá de Henares. Máster administración y dirección de servicios sanitarios. Universidad Pompeu Fabra - Fundación Gaspar Casal. (2016-2018). Autor activo del Grupo Cochrane Anaesthesia, Critical, and Emergency Care (ACE). Cochrane Collaboration.

*\* Jesús A. Barea declara no tener conflicto de interés vinculado al presente trabajo. No ha recibido ninguna compensación económica por su participación en este proyecto.*

### **Joan Rovira Forns**

Es Profesor Emérito del Departamento de Teoría Económica de la Facultad de Ciencias Económicas y Empresariales de la Universidad de Barcelona y Profesor Asociado de la Escuela Andaluza de Salud Pública. Colabora habitualmente como consultor de la Fundación Gaspar Casal en la realización de estudios relacionados con la Evaluación de Tecnologías en Salud (ETS) y las políticas de precios de los medicamentos. Socio fundador y director de investigación de SOIKOS SL, que llevó a cabo un gran número de estudios de evaluación económica de tecnologías para la industria y para organismos sanitarios y que desarrolló la Base de Datos de Costes Sanitarios. Fue miembro de la Junta Directiva de ISPOR, y Economista de la Salud Senior en el Departamento de Salud, Nutrición y Población del Banco Mundial, Washington DC.

### **Luis Quecedo Guitérrez**

Doctor en Medicina, Universidad Autónoma de Madrid. Especialista de Medicina Interna. Especialista en Anestesiología y Reanimación. Máster en Administración y Dirección de Servicios Sanitarios, ADOS e ICADE- Universidad Pontificia Comillas de Madrid. Máster en Economía de la Salud y Gestión Sanitaria, Universitat Pompeu Fabra y Fundación Gaspar Casal (FGC). European Healthcare Leadership Programme, INSEAD. Facultativo Especialista de Área, Servicio de Anestesia, Hospital de La Princesa de Madrid.

### **Juan E. del Llano Señarís**

Licenciado en Medicina y Cirugía (1981) y Doctor en Medicina (1990) por la Universidad Complutense de Madrid. Especialista en Medicina Preventiva y Salud Pública (MIR 1982-85, Hospital La Paz). M. Sc. Community Health, Usher Institute, University of Edinburgh (1985-86). European Healthcare Leadership Programme, INSEAD (1999-2000). Programa de Alta Dirección de Instituciones Sanitarias, IESE, Universidad de Navarra (2004). Advanced Health Leadership Forum, Universitat Pompeu Fabra, Barcelona y University of California, Berkeley (2005). Salzburg Global Seminar, ediciones 1994 y 2015.

Desde 1998 es director de la Fundación Gaspar Casal. Desde 1989 es Director Académico y profesor de Salud Pública y Gestión Clínica del Master de Administración y Dirección de Servicios Sanitarios, Fundación Gaspar Casal y Universitat Pompeu Fabra (anterior al 2000, ICADE-UPCO). Presidente de la Asociación Española de Evaluación de Tecnologías Sanitarias, desde 2003. Editor Asociado de Gestión Clínica y Sanitaria, desde 1999. Investigador Asociado del Centro de Investigación en Economía y Salud (CRES), Universitat Pompeu Fabra, desde 2000. Evaluador del Área de Biomedicina, Agencia Nacional de Evaluación y Prospectiva. Profesor Ayudante Doctor, Agencia Nacional de Evaluación, Calidad y Acreditación (PAD: 2012-7582).

Ha sido Asesor de la Subdirección General de Atención Hospitalaria del Instituto Nacional de la Salud (1987); Vicepresidente Ejecutivo de la Asamblea de Madrid de la Cruz Roja Española (1988); Socio Director de Análisis y Diseños Operativos en Salud, S.A. (1989-1996) y Gerente Senior de Nuevos Clientes de MSD (1997)

**Jordi Gol-Montserrat**

Graduado en Economía por la Universidad Complutense de Madrid, especialidad de Análisis Económico. Master's Programme in Decision Support and Risk Analysis, Stockholm University. Actualmente es Coordinador de Investigación en la Fundación Gaspar Casal. Ex Pasante en la iniciativa PROVAC -creada para fortalecer la capacidad técnica en los países para la toma de decisión basada en evidencia ante la introducción de una nueva vacuna de la Organización Panamericana de la Salud, Washington DC. Ex Pasante en la Consejería de Agricultura, Alimentación y Medio Ambiente de la Embajada de España en Estados Unidos, Washington DC.









G

A

S

P

A

R

□

C

A

S

FUNDACIÓN  
GAS  
PAR  
■ CA  
SAL



**Fundación Mylan**  
para la Salud

A

L