



Analizando los datos del paciente para mejorar el sector salud

PABLO REBOLLO¹, CONCHA ALMARZA²

¹Principal HEOR – RWE Solutions, Spain. QuintilesIMS. ²General Manager, Spain. QuintilesIMS.

El artículo persigue clarificar algunos conceptos clave relacionados con la utilización de datos de salud para avanzar en el conocimiento médico y optimizar el manejo clínico del paciente. Estos datos, llamados “Real World Data” están almacenados fundamentalmente en la historia clínica informatizada, de la que el paciente es dueño y el sistema sanitario es el usuario fundamental. Aplicando las modernas metodologías de análisis ya disponibles sobre estas grandes bases de datos se facilitaría la generación de otro nivel de conocimiento: la Real World Evidence.

KEY WORDS: Real World Evidence, Big Data, e-Salud.

En los últimos años cada vez son más las referencias que se hacen en los medios de comunicación a los términos Real-World Data, Real-World Evidence, Big Data, Predictive Analytics, Inteligencia Artificial... cuando se habla de los datos de la salud de pacientes, recogidos en los sistemas informáticos de los servicios de salud, y de su posible utilización para el avance del conocimiento en Medicina. Muchos actores del sector salud consideran inevitable la transición del puro almacenaje de datos, situación actual, a la utilización de los mismos para generar nuevas evidencias que ayuden a mejorar el proceso de los cuidados del paciente individual. De hecho, hay voces que empiezan a hablar de la oportunidad que se está perdiendo por no haber empezado ya a organizar la explotación científica de estos datos. Cualquier cambio puede verse comprometido por las resistencias que genera, y estas a menudo son resultado de la falta de mensajes claros sobre la cuestión. En este sentido, el presente artículo intenta explicar, de una manera sencilla para el lector, los conceptos básicos en esta no tan nueva área dentro de la medicina y la investigación clínica, con el objetivo último de facilitar la incorporación de más personas al necesario debate de fondo sobre si es necesario, viable y ético el análisis de datos de salud del paciente para mejorar la salud de todos.

REAL-WORLD DATA (RWD) Y REAL WORLD EVIDENCE (RWE)

El término Real-World Data (RWD) se aplica en general a todos aquellos datos de pacientes que son recogidos fuera del marco de los ensayos clínicos aleatorizados (ECA), frecuentemente sin finalidad investigadora. Son datos recogidos en la historia clínica informatizada, en los laboratorios de análi-

Autor para correspondencia:

Pablo Rebollo
QuintilesIMS
Calle Provença, 392 - 3ª y 4ª pl.
Barcelona 08025. España



sis clínicos, en los servicios de radiodiagnóstico... y siempre en el marco de la práctica clínica habitual. Cuando se analizan con las metodologías adecuadas, los RWD pueden proporcionar conocimiento sobre la forma en que los pacientes son diagnosticados y tratados en condiciones de “vida real”, lo que constituye la Real-World Evidence (RWE). La RWE no pretende sustituir la evidencia que se obtiene de los ensayos clínicos y de su agregación en meta-análisis, sino complementarla, para poder evaluar las intervenciones fuera de las condiciones controladas que se aplican en los ECA. Los ensayos clínicos proporcionan información de la eficacia y seguridad de una intervención médica (principalmente medicamentos), y los RWD y RWE proporcionan información suplementaria sobre la efectividad y la seguridad.

Hasta no hace mucho tiempo la única forma posible de obtener RWD sobre el uso de los medicamentos era la realización de estudios observacionales, básicamente de tipo transversal, de casos y controles o de cohortes. Con diferentes diseños metodológicos, estos estudios persiguen responder a preguntas sobre los nuevos medicamentos que no pueden ser respondidas con el análisis estadístico de los datos de los ECA. Los estudios observacionales son útiles para describir cómo es el proceso diagnóstico del paciente con una determinada patología, cuál es la incidencia y prevalencia de una patología, qué tratamientos se pautan según las características del paciente y de su enfermedad y cuáles son los resultados del uso de un determinado medicamento (efectividad), ya sean estos medidos en unidades biológicas (cifras de hemoglobina en sangre, cifra de tensión arterial, nivel de colesterol...) o en unidades de calidad de vida, satisfacción del paciente con el tratamiento u otras de las llamadas PRO (del inglés *Patient Reported Outcomes*). El análisis

Muchos actores del sector salud empiezan a hablar de la oportunidad que se está perdiendo por no haber empezado ya a organizar la explotación científica de los datos de salud almacenados, para el beneficio del paciente

de los datos recogidos en los estudios observacionales, en la mayoría de los casos se realiza aplicando métodos estadísticos que pueden ser considerados como clásicos. Los diferentes tipos de estudios observacionales continúan siendo una fuente de RWE, ello a pesar de lo complicado que resulta a menudo ponerlos en marcha en nuestro país, sobre todo los de cohortes prospectivas.

En parte por esto, en los últimos años viene cobrando relevancia la realización de estudios observacionales retrospectivos, utilizando fuentes de datos existentes: básicamente registros activos de pacientes y datos extraídos de historias clínicas de pacientes de determinadas zonas de salud, debidamente anonimizados y parametrizados. La utilización de estas bases de datos supone un primer paso para la explotación científica de los datos de la historia clínica, utilizando los análisis estadísticos propios del contraste de hi-

pótesis, como en el caso de los estudios observacionales con datos recogidos de manera primaria. De una forma relativamente rápida, estos estudios permiten contrastar hipótesis sobre grupos de pacientes, aunque con las limitaciones inherentes al diseño retrospectivo de los mismos. Aun siendo definidos como “estudios observacionales retrospectivos” y contando por tanto con un protocolo de investigación y un plan de análisis estadístico que son sometidos a la evaluación por parte de un Comité Ético de Investigación Clínica, apenas hay centros asistenciales en España que se muestren dispuestos a permitir la extracción de determinados datos de la historia clínica de los pacientes. Iniciativas como VISC+ que perseguían hacer accesibles a investigadores de cualquier institución, pública o privada, conjuntos finitos de datos anonimizados de pacientes para responder a preguntas de investigación con RWE, han quedado bloqueadas por

la alarma social generada en torno a la frase “se venden los datos de los pacientes”, que alguien con no mucha fortuna y peor intención, introdujo en los medios de comunicación social. El hecho claro es que el retraso en el acceso a los RWD de la historia clínica no está justificado en nuestro país, estando muy por detrás de otros países de nuestro entorno que, en lugar de dar la espalda a la realidad, han articulado maneras de utilizarlos de una manera segura, generando RWE que contribuye a mejorar la salud de la población.

BIG DATA Y PREDICTIVE ANALYTICS

A medida que han ido evolucionando los sistemas informáticos utilizados en la práctica clínica, ya sean estos de datos administrativos (como el CMBD), de historia clínica informatizada en atención primaria o especializada, o de receta farmacéutica electrónica, el volu-

men de datos almacenados se ha ido haciendo cada vez mayor, aumentando también la variedad de las fuentes y la velocidad de acceso a los mismos. Estas tres “v” (volumen, variedad y velocidad) son las características que comúnmente definen lo que se llama “Big Data”. En el ámbito de la salud se definen tres “v” adicionales¹: veracidad, variabilidad y valor. La veracidad de los datos recogidos se entiende que es crítica para poder utilizarlos en la toma de decisiones médicas. Se refiere a que los datos, los análisis y los resultados obtenidos están libres de error y son creíbles², aunque esto todavía no es una realidad. La variabilidad de los datos de salud es habitualmente elevada: por ejemplo una misma abreviatura recogida en historia clínica, puede tener varios significados. Por último, el valor del “Big Data” en salud deriva de la agregación de toda la información disponible en un único conjunto de datos que permita el análisis de los datos.





La utilización de “Big Data” en salud va por detrás del nivel de uso que se está haciendo en otros sectores, y esto es debido básicamente a obstáculos estructurales, sobre todo la baja inversión en sistemas de información, preocupaciones legales relacionadas con la privacidad y la seguridad, y complicaciones técnicas relacionadas con el intercambio de información¹. Sobre esta última cuestión, actualmente la mayoría de los RWD en salud no están estructurados y son difíciles de integrar; por ejemplo, la información contenida en una historia clínica puede incluir desde notas en plantillas personalizadas hasta archivos de sonido (MP3, WAV...), pasando por imágenes escaneadas en diferentes formatos, notas digitales escritas en una “tablet”, datos de laboratorio, fax, emails, y documentos de Microsoft u otros tipos de software. Además, estos datos pueden estar desperdigados en diferentes niveles asistenciales y por tanto en diferentes repositorios de información: consulta de atención primaria, consulta especializada del hospital, servicio de ingreso hospitalario, receta farmacéutica electrónica y registros de mortalidad. La integración de diferentes fuentes de datos según nivel asistencial y el desarrollo de ontologías basadas en computación que permitan estandarizar los datos escritos en la historia clínica informatizada, son dos áreas en pleno y rápido desarrollo que permitirán acelerar la utilización de “Big Data” en salud, generando lo que podríamos llamar “Big Real World Data”.

En el análisis de “Big Data” en salud hay dos niveles que pueden ayudarnos a responder preguntas. El primer nivel lo constituyen análisis descriptivos que ayudan al planteamiento básico del problema de salud a estudio respondiendo a las preguntas de “quién, qué, cuándo y dónde”. El segundo nivel corresponde a los análisis avanzados de tipo epidemiológico y

de tipo predictivo. Los análisis predictivos sobre “Big Data” no se pueden realizar con las técnicas clásicas de estadística, basadas en el contraste de hipótesis y por tanto conducidas por la formulación de las mismas, que depende del, a menudo lento, progreso científico. La obtención de RWE a partir de “Big Data” en salud mediante análisis predictivos incluye una variedad de técnicas estadísticas, como modelizaciones, máquinas de aprendizaje (“machine learning”), minería de datos, “random forest”, “neural networks” y “support vector machines” que analizan la realidad actual y pa-



sada para hacer predicciones sobre eventos futuros y desconocidos. Estas técnicas son flexibles, potentes, ideales para producir evidencia robusta en situaciones complejas. Descrito de una manera simple, mediante los análisis predictivos se crean algoritmos (perfiles de predicción) a partir de los pacientes observados en el pasado, para a partir de ellos obtener predicciones instantáneas para pacientes actuales³.

La utilización de “Big Real World Data” tiene realmente un potencial transformador en el sector salud, incluyendo la reducción de errores, la identificación de poblaciones de elevado riesgo, el soporte de la medicina basada en la evidencia y la mejora de los procesos asistenciales¹. El aumento de la exactitud de los diagnósticos, el apoyo a la medicina preventiva y salud pública, la identificación de necesidades no cubiertas en grupos de pacientes, la mejora de resultados de la actividad médica en el paciente individual, han sido también descritos como áreas en las que la RWE a partir de “Big Data” puede introducir mejoras significativas³.

¿DE QUIÉN SON LOS DATOS DE SALUD? ¿ESTÁ EL CIUDADANO DISPUESTO A CEDERLOS PARA LA MEJORA DEL SECTOR SALUD?

Parece claro que los datos de salud de una persona, son de esa persona, quien los cede a las instituciones para facilitar la asistencia sanitaria. Las instituciones sanitarias los almacena dissociando los datos identificativos y manteniéndolos en servidores seguros, cumpliendo en todo momento con la Ley Orgánica de Protección de Datos personales (LOPD). Es posible el acceso a datos agregados de salud en los servicios asistenciales de nuestro país, para investigación o docencia, de acuerdo a la Ley 14/2007 de Investigación Biomé-

dica y siempre aplicando un proceso previo de anonimización. Sin embargo, el actual marco legislativo no parece estar preparado para el acceso y uso de “Big Real World Data”, por lo que habría que solicitar un consentimiento del titular de los datos para todo fin distinto de la prestación de asistencia⁴. Sobre esta cuestión el debate todavía está activo y sería relevante conocer hasta qué punto los ciudadanos usuarios del sistema de salud no estarían más que contentos de colaborar en el progreso de la medicina, cediendo sus datos de salud, para poder tratar de manera más eficiente a otros pacientes en el futuro.

Según una encuesta realizada por el instituto Vodafone centrada en lo que los ciudadanos de países europeos piensan sobre el uso de “Big Data” en diferentes ámbitos de la vida cotidiana⁵, cuyo informe ha sido publicado en enero de 2016, el 86% de los ciudadanos españoles entrevistados se muestran de acuerdo con la recogida y análisis de información de salud de gran cantidad de personas, realizada por instituciones sanitarias para permitir la mejora en la detección y tratamiento de enfermedades. Esta cifra es la más elevada entre los países participantes en la encuesta y contrasta con el 42% de los ciudadanos alemanes que se muestran de acuerdo. Es más, el 69% de los encuestados en España darían, a estas instituciones sanitarias, acceso a sus datos de salud siempre que fuesen unidos a los de otros, anonimizados y usados solo para fines de investigación. Parece que no es tanto el recelo de los ciudadanos de nuestro país a compartir sus datos de salud, siempre que se cumpla con lo dispuesto en la ley, que obliga a proteger adecuadamente sus datos identificativos, y está claro que con las tecnologías adecuadas, todo se puede hacer bien reduciendo al máximo el riesgo de identificación de los pacientes.

- 1 Thorpe JH Gray EA. Big data in ambulatory care. Breaking down legal barriers to Support effective use. J Ambulatory Care Manage 2015; 38: 29-38.
- 2 Raghupathi W, Raghupathi V. Big analytics in healthcare: promise and potential. Health Information Science and Systems 2014; 2:3.
- 3 Winters-Miner LA. Seven ways predictive analytics can improve healthcare. <https://www.elsevier.com/connect/seven-ways-predictive-analytics-can-improve-healthcare>
- 4 Carnicero J. y Rojas D. (Coordinadores). La explotación de datos de salud: Retos, oportunidades y límites. Pamplona: Sociedad Española de Informática de la Salud, 2016. <http://www.seis.es>
- 5 Vodafone Institute. “Big Data. A European survey on the opportunities and risks of data analytics”.

