

Utilización de datos de la vida real

Enrique Bernal Delgado



Los datos de vida real (a diferencia de los datos experimentales) son una mejor opción para informar la formulación de políticas

Los sistemas de salud están asistiendo a una transición en el paradigma del conocimiento. El enfoque clásico basado en la evidencia experimental (es decir, ensayos clínicos aleatorios bien realizados y metanálisis de ensayos clínicos) que ha demostrado ser esencial en la evaluación de la eficacia y la seguridad de las tecnologías sanitarias, está siendo cuestionado. Los estudios observacionales bien realizados basados en datos de la vida real podrían resultar más apropiados para informar las decisiones clínicas y políticas, ya que estos datos representan pacientes (o poblaciones) de la vida real que interactúan con el sistema de salud en condiciones de la vida real, exposición a múltiples intervenciones, en múltiples instancias de atención y que evolucionan con el paso del tiempo.

Intervención	Intervención
Intervenciones en el paciente	Regulación de la atención
Intervenciones de prevención (PI)	Intervención de necesidades no médicas
Modelos de pago innovadores	Diseño de un individuo
Modelos de procesos organizativos	Caracterización de actores o la atención
Identificación de actores relevantes	Modelos de PMSI, PMSU
Identificación de competencias relevantes	Modelos de uso de recursos tecnológicos en la organización y los procesos de atención
Impacto o la base de conocimientos	Eficiencia y eficacia en otros entornos

En efecto, si nos fijamos en las preguntas arquetípicas en gestión y política sanitaria (esas preguntas esenciales para la función de gobierno de los sistemas de salud) la mayoría de ellas no se pueden informar utilizando el paradigma experimental. Por ejemplo, preguntas sobre necesidades no satisfechas en determinados subgrupos de población, experiencias de los pacientes, utilización desigual de intervenciones efectivas, impacto de algoritmos para la toma de decisiones sobre la organización sanitaria, desempeño de los proveedores de salud o efectividad y eficiencia de una intervención en vida real son cuestiones para las que los ensayos clínicos, o sus meta-análisis, no podrán dar respuesta.



Cuando hablamos de datos de vida real nos referimos a datos que se coleccionan como consecuencia de la experiencia de salud y enfermedad de los individuos o de las poblaciones. Datos del mundo real que eventualmente podrían usarse para responder a las consultas de políticas antes mencionadas son, entre otros, registros electrónicos del cuidado sanitario, datos administrativos, registros de enfermedad, señales procedentes de dispositivos móviles, los textos de las conversaciones sobre la experiencia de usuarios, etc. Muchos de ellos proporcionan datos bastante estructurados (datos de reclamaciones, datos de facturación, datos de encuestas y registros), mientras que otros son semiestructurados (historiales médicos o de salud electrónicos) o no estructurados (datos de redes sociales o lenguaje natural en historias clínicas).



El uso extensivo y sistemático de los datos de vida real requiere la construcción y mantenimiento de sistemas de salud que aprenden de los datos que generan.

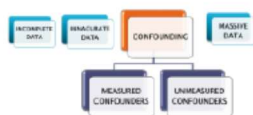
El esfuerzo consistirá en conciliar la informática, la ciencia de datos y las necesidades del sistema, en particular la mejora continua de calidad y la innovación, para que (al menos) el uso sistemático de aquellos datos que el sistema genera constantemente, respalde la toma de decisiones a nivel clínico, de gestión, para la regulación y para la política sanitaria. La reutilización de los datos de rutina en todo caso, debe desarrollarse y mantenerse en el contexto del diseño institucional.

Las características de un sistema de salud que aprende de los datos que genera serían: 1) los actores que toman decisiones deben ser propietarios del proceso; 2) cualquier tipo de datos de colección rutinaria son aprovechables; 3) la tecnología debe ayudarnos al aprovechamiento continuo de los datos, tan pronto como se generan; 4) la ciencia de datos debe ayudarnos a incorporar análisis de ciclo rápido, para que acercar el dato a la decisión informada; 5) todo el procesos debe estar imbricado (los resultados deben revertir) en los procesos de toma de decisiones; y 6) el esfuerzo debe ser sostenido en el tiempo.



El uso secundario de los datos que se recogen como consecuencia de la actividad asistencial (dato rutinario) requiere atender a cuestiones legales, organizativas, de interoperabilidad semántica y tecnológicas

Movilizar los datos para un uso distinto para el que fueron creados (uso secundario) y por tanto, facilitar el desarrollo de un sistema que aprende de los datos que genera, requiere un esfuerzo sostenido en distintas fases del ciclo de vida de los datos. Una vez el dato es recolectado por el *Data Controller* (así llamado en el Reglamento de protección de datos) y liberado para su uso (al menos) seudonimizado, el dato requerirá estandarización y armonización acorde a estándares de interoperabilidad semántica y publicación de los meta-datos de modo que el potencial usuario conozca el contenido de las fuentes de datos. Este trabajo, que debería ser realizado por instituciones especializadas en el procesamiento de los datos de vida real, continuará con funciones de enlace de fuentes e implementación de tecnologías y espacios para el procesamiento seguro de los datos.



El uso de datos del mundo real requiere una comprensión profunda de las limitaciones inherentes a la reutilización de datos de vida real y las cautelas que son intrínsecas a los estudios observacionales.

Los estudios observacionales basados en datos de vida real tienen que lidiar con algunas limitaciones: datos incompletos, inexactitudes de datos, confusión, sesgos de selección y, cuando se trata de datos masivos, algunas maldiciones matemáticas (dimensionalidad, heterogeneidad y sobreajuste).

Las limitaciones antes mencionadas pueden mitigarse, incluso superarse, tanto en el diseño de un estudio basado en datos del mundo real, como en los desarrollos analíticos. Como ilustración valgan las siguientes: a) las técnicas de imputación pueden mitigar la incompletitud; b) el análisis

visual puede proporcionar argumentos para la aceptación de la calidad de los datos y, *ex ante*, reducir las imprecisiones; c) el emparejamiento o el uso de variables instrumentales pueden ayudar a reducir los sesgos de confusión y selección; y, d) las matemáticas adecuadas reducirán los riesgos de correlaciones espurias al usar datos masivos.

En cualquier caso, dependiendo de la pregunta de interés, la evaluación y control de estas limitaciones es más o menos exigente. Los estudios de generación de hipótesis podrían ser menos exigentes en cuanto a la imprecisión de datos que la construcción de un modelo predictivo que defina qué pacientes críticos podrían beneficiarse más de un tratamiento.