

ICB digital

mayo 2025

Boletín para los Comités de Ética de Investigación

contenido nº150

artículo breve

Utilización de la Inteligencia Artificial en Investigación Farmacoepidemiológica

Miguel Ángel Maciá Martínez

noticias

Sesión informativa sobre los ensayos clínicos en la UE: situando al paciente en el centro

próximo número

nº 151 - julio 2025

© SOCIEDAD ESPAÑOLA DE FARMACOLOGÍA CLÍNICA

administracion@se-fc.org

www.se-fc.org/icbdigital

Utilización de la Inteligencia Artificial en Investigación Farmacoepidemiológica

Miguel Ángel Maciá Martínez. Jefe de Área de Farmacoepidemiología, División de Farmacoepidemiología y Farmacovigilancia, Agencia Española de Medicamentos y Productos Sanitarios.

Este artículo está basado en la ponencia del mismo título del autor en la mesa redonda “Inteligencia artificial en medicina: tecnología, investigación, docencia y ética” en el XXXII Congreso de la Sociedad Española de Farmacología Clínica celebrado del 16 al 18 de octubre de 2024 en Santiago de Compostela.

INTRODUCCIÓN

La inteligencia artificial (IA) está transformando numerosos campos, y la investigación biomédica es una de sus áreas de impacto. En este artículo, se examinarán las contribuciones potenciales de la IA en la investigación farmacoepidemiológica, subrayando su función en la utilización de datos de la vida real (Real World Data, RWD) para convertirlos en evidencia de la vida real (Real World Evidence, RWE). Se abordará específicamente la experiencia en el campo de la generación de la evidencia científica que sustenta la regulación de los medicamentos en Europa¹ y la experiencia concreta de la Agencia Española de Medicamentos (AEMPS) con el Programa BIFAP (Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público).

LOS DATOS EN LA REGULACIÓN: ESTRATEGIA PARA EL USO DE BIG DATA Y OTRAS FUENTES

La red regulatoria europea de medicamentos² fundamenta sus decisiones en diversos tipos de datos, incluyendo fuentes de “big data”. Dentro de esta red se estableció un grupo de trabajo dedicado a orientar un mejor análisis y uso del “big data” en la regulación de medicamentos. El “HMA-EMA Joint Big Data Steering Group”³ define el término “big data” como conjuntos de datos extremadamente grandes, complejos, multidimensionales, no estructurados y heterogéneos. Estos datos se caracterizan por su rápida acumulación y la posibilidad de ser analizados computacionalmente para descubrir patrones, tendencias y asociaciones.

Una prioridad clave en la estrategia de la red de agencias de medicamentos europeas es invertir en análisis de datos e inteligencia artificial (IA). Para ello, a partir de la experiencia del “HMA-EMA Joint Big Data Steering Group” se ha constituido el “Joint HMA-EMA Network Data Steering Group”⁴, entre cuyos objetivos se encuentra aprovechar el uso de IA para mejorar la toma de decisiones. Los métodos avanzados impulsados por IA son propuestos como herramientas para generar evidencia que respalde decisiones sobre la seguridad, eficacia y calidad de los medicamentos.

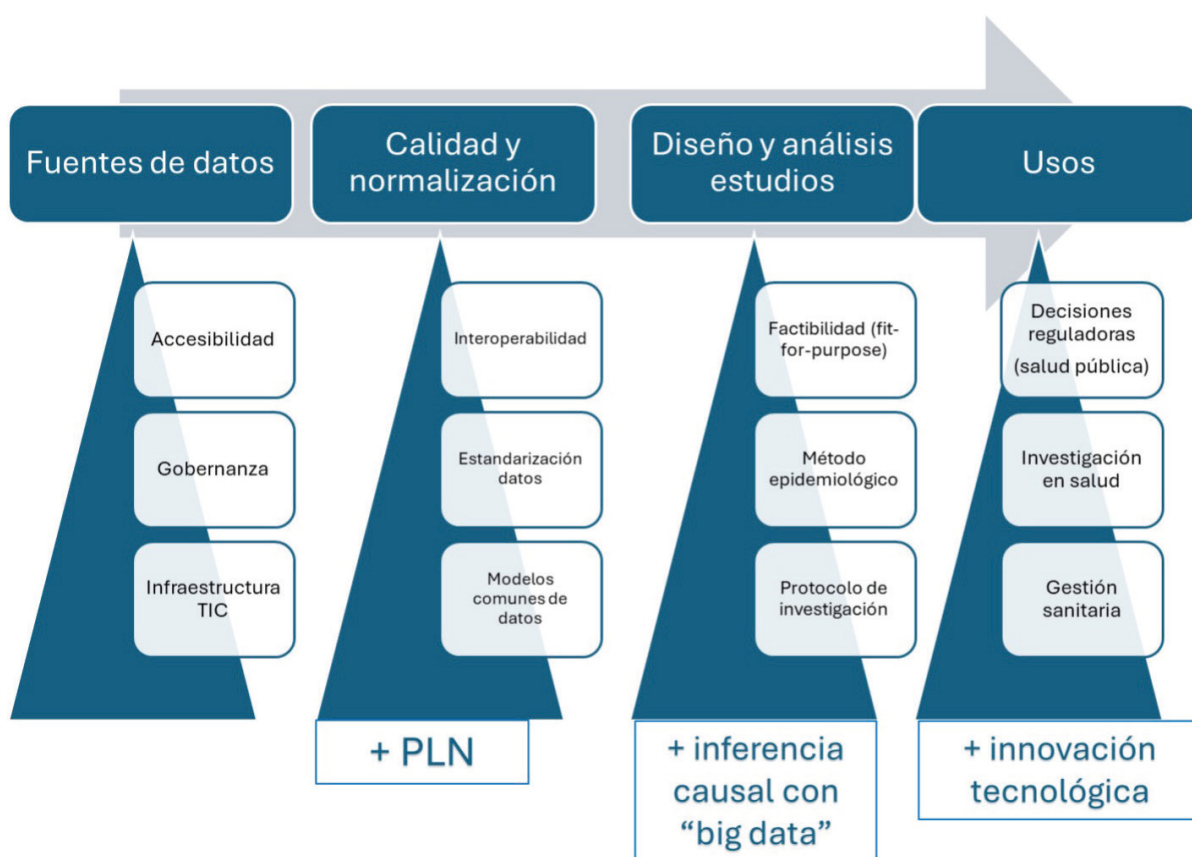
Las bases de datos de historias clínicas recopilan información de la práctica clínica de millones de pacientes y cumplen con la definición de “Big Data”. La generación de conocimiento científico sobre medicamentos mediante el uso secundario de datos de salud es una práctica consolidada en la evaluación de la seguridad de los medicamentos autorizados, constituyendo una herramienta fundamental en la farmacovigilancia. Recientemente, el potencial de los RWD se ha ampliado a diversas áreas de la investigación clínica, proporcionando evidencia científica cuando no es viable la realización de ensayos clínicos o bien complementando y creando sinergias con estos últimos⁵.

ASPECTOS CLAVE PARA CONVERTIR LOS DATOS DE VIDA REAL EN EVIDENCIA CIENTÍFICA

En el contexto europeo, el número de fuentes de RWD orientadas a la investigación biomédica está aumentando exponencialmente. El “EMA-HMA-Catalogue of Real World Data Sources”⁶ incluye más de 240 fuentes de información, entre las cuales hay más de 60 en España.

Sin embargo, solo un número limitado de estas fuentes está siendo utilizado en proyectos colaborativos que buscan aumentar la solidez de las decisiones sobre la calidad, efectividad y seguridad de los medicamentos y asegurar una relación beneficio-riesgo favorable. En general, se requiere una serie de elementos para hacer este viaje posible y generar conocimiento válido para la toma de decisiones (Figura 1). En cada uno de estos elementos, la innovación mediante la aplicación de recursos de IA tiene un potencial indudable para impulsar la generación de evidencia a partir de los datos.

Figura 1. De Real World Data a Real World Evidence



PLN=Procesamiento de Lenguaje Natural.

En primer lugar, las fuentes de datos deben ser fácilmente accesibles, proporcionando públicamente la información necesaria para que los investigadores puedan conocer las características de los datos y los procedimientos de acceso a los mismos. Esto debe realizarse mediante procedimientos y condiciones transparentes de acceso a los datos, asegurando al mismo tiempo el cumplimiento de la normativa de protección de datos. Catálogos de fuentes de datos como el ya mencionado⁶ de la EMA y la red de agencias de medicamentos europeas, y la publicación de la gobernanza de acceso a los datos, cumplen esta función⁷. También es imprescindible la existencia de infraestructuras tecnológicas con las capacidades necesarias para soportar el almacenamiento y el trasiego de los grandes volúmenes de datos de forma segura y eficaz.

También es esencial que las fuentes de datos dispongan la información clínica normalizada y estructurada en modelos comunes de datos. No hay que olvidar que los datos fueron registrados en las fuentes

originales para un uso primario fundamentalmente asistencial. La calidad del registro, el diseño de los sistemas de información y la forma de entrada de datos por los usuarios (profesionales sanitarios) refleja las necesidades de la asistencia sanitaria, no del uso, secundario en todos los sentidos, para investigación. Adolecerá también de todas las limitaciones del propio sistema sanitario. Al fin y al cabo son datos de vida real.

Su uso secundario para fines de investigación requiere un enorme esfuerzo de estandarización y de reorganización de la información en entornos que deben ser interoperables para facilitar el uso compartido por investigadores procedentes de distintos ámbitos.

El tercer pilar de la generación de evidencia científica a partir de RWD implica aplicar el método científico en la selección de fuentes de datos y el diseño del estudio para responder a las preguntas de investigación.

De este modo, las oportunidades para generar conocimiento a partir de los datos recopilados en la asistencia sanitaria pueden cubrir una amplia gama de objetivos, que podemos clasificar en las siguientes tres categorías:

1. La caracterización de poblaciones para generar hipótesis sobre los determinantes de la salud y la enfermedad, para comprender los resultados de salud de grupos específicos de población mediante análisis propios de la epidemiología descriptiva.
2. La estimación a nivel poblacional de los efectos (causales) de determinadas exposiciones (medicamentos, intervenciones, hábitos...) sobre los resultados de salud, mediante los métodos de inferencia causal propios de la epidemiología analítica.
3. La predicción a nivel de paciente de futuros resultados de salud a partir de características actuales, antecedentes y condicionantes del paciente, como ayuda a las decisiones clínicas, utilizando modelos predictivos.

El interés en el soporte para la toma de decisiones sobre el beneficio-riesgo de los medicamentos a nivel poblacional se enfoca principalmente en las dos primeras áreas, y constituye el ámbito central de la farmacoepidemiología. La tercera área contribuye específicamente al avance del campo de la medicina personalizada y la aplicación en este ámbito de técnicas avanzadas de IA está contribuyendo ya a avances importantes en el terreno de los modelos predictivos.

En este contexto, es fundamental diferenciar los métodos y roles que la ciencia de datos y la epidemiología desempeñan respectivamente en la resolución de preguntas de investigación predictivas frente a aquellas que implican inferencia causal⁸.

REALIDADES Y EXPECTATIVAS DE LA IA EN FARMACOEPIDEMIOLOGÍA A PARTIR DE LA EXPERIENCIA EN BIFAP

Nuestra experiencia en la AEMPS con el programa BIFAP⁹ incluye el uso de determinadas herramientas de IA y nos permite vislumbrar el futuro valor de la IA en la investigación farmacoepidemiológica y en la generación de conocimiento en general.

1. Acceso a las fuentes de datos y uso con fines de innovación tecnológica

La publicación del Reglamento¹⁰ de la Unión Europea (UE) sobre el Espacio Europeo de Datos de Salud (EEDS) supone un salto enorme que facultará, cuando sean aplicables sus disposiciones y se lleven a la práctica, un uso generalizado de fuentes de datos de salud para uso secundario compartidos en el seno de la UE. Establece fines específicos para el tratamiento de datos sanitarios electrónicos, que incluye explícitamente el de entrenamiento, la prueba y la evaluación de algoritmos de IA. La

utilidad de explotar grandes bases de datos de historias clínicas como BIFAP con fines de innovación tecnológica son enormes. El camino a seguir requiere abordar de manera rigurosa las cuestiones relacionadas con la garantía de la privacidad y la seguridad de los datos, tal como lo establece el propio Reglamento. Además, iniciativas como la acción conjunta TEHDAS2¹¹ contribuirán a definir las directrices y requisitos técnicos necesarios, incluyendo los aspectos de seguridad de los datos.

2. Calidad y normalización para asegurar la interoperabilidad y utilidad

Para garantizar la calidad y la normalización de los datos, es fundamental la interoperabilidad y la estandarización de datos mediante modelos comunes de datos (CDM) como OMOP¹². La incorporación de BIFAP¹³ al proyecto DARWIN-EU de la EMA, ha obligado a “traducir” el modelo de datos propio de BIFAP al CDM de OMOP.

DARWIN-EU es ejemplo de red federada europea, que optimiza la toma de decisiones regulatorias sobre los medicamentos mediante la generación de evidencia científica con RWD.

Un desafío para fuentes de datos como BIFAP es que gran parte de los datos de la historia clínica electrónica (HCE) son no estructurados o semiestructurados. Los datos no estructurados de la HCE en la investigación farmacoepidemiológica tienen un valor importante para:

- Confirmar la temporalidad para identificar el inicio de los síntomas, evitando así el sesgo de causalidad inversa, y para la selección de casos incidentes (no prevalentes) cuando así lo exige el protocolo del estudio.
- Identificar y capturar de la HCE los eventos clínicos no codificados debido a la falta de granularidad del diccionario, las características del sistema informático, o la decisión del profesional que realiza el registro.
- Establecer una definición diagnóstica compleja en casos de criterios clínicos múltiples, donde la información se anota únicamente en texto libre.
- Determinar la gravedad de la enfermedad cuando se utilizan criterios “blandos” de gravedad.

Además, incluso los datos de origen que se registran como codificados deben “mapearse” a las terminologías comunes al CDM de destino. La adaptación a versiones internacionales de los mismos diccionarios o de sus actualizaciones son retos adicionales para los equipos.

La incorporación de la IA en los procesos de normalización, codificación y transformación a estándares interoperables como OMOP puede mejorar la eficiencia de estos procesos, en particular con el uso componentes de procesamiento del lenguaje natural para el reconocimiento de entidades médicas en textos no estructurados y para poner a disposición del investigador solo textos “limpios” de cualquier identificador indirecto personal, como elemento para asegurar la privacidad.

3. Diseño y análisis de datos: utilidad y calidad

Las nuevas estrategias para aumentar la calidad de los estudios observacionales, como la emulación de ensayos clínicos (*target trial emulation*), necesitan disponer de muchas variables medidas en muchas personas¹⁴. En este contexto, las herramientas que se usan para “big data” se complementan con el método epidemiológico cuando se trata de realizar inferencia causal.

En todo caso, puesto que el objetivo es conocer los efectos favorables (efectividad) y desfavorables (seguridad) de un medicamento, aunque las herramientas de análisis de “big data”, son fundamentales para reconocer patrones y realizar predicciones en grandes volúmenes de datos, en la investigación causal farmacoepidemiológica no podemos prescindir del diseño adecuado, los análisis rigurosos y la interpretación cuidadosa de los resultados basados en el método epidemiológico.

El proyecto TARGET EU¹⁵ financiado por la EMA y para el que la AEMPS aportará los datos de BIFAP, busca mejorar el uso regulatorio de RWD mediante la aplicación de metodologías de emulación de ensayos clínicos y estimandos. Su objetivo es generar la evidencia que conecte los ensayos clínicos y los estudios observacionales, facilitando las decisiones regulatorias cuando los ensayos clínicos no son viables.

CONSIDERACIONES FINALES

La relevancia las redes de investigación colaborativa que utilizan RWD como fuente de datos y que tengan la capacidad de desarrollar proyectos de investigación integrados en un modelo común de datos, tiene un carácter estratégico en Europa. De hecho, el conocido como “Informe Draghi”¹⁶, dedicado a proponer estrategias para impulsar la competitividad de la Unión Europea en todos los terrenos, en relación con el ámbito farmacéutico, señala explícitamente la relevancia de *“Aprovechar la red DARWIN EU® para generar evidencia que impulse la innovación en el desarrollo de medicamentos, así como para la toma de decisiones de política sanitaria y clínicas apoyadas por el uso de IA. La experiencia existente debe orientarse hacia la generación de evidencia del “mundo real” mediante la realización de estudios observacionales, aprovechando el catálogo actual de fuentes de datos para expandir las actividades utilizando fuentes de datos adicionales en los Estados miembros que estén disponibles gracias al Espacio Europeo de Datos de Salud. La inteligencia artificial tiene un enorme potencial para acelerar la gestión y el análisis de datos de salud con este propósito.”*

REFERENCIAS

1. Ioakeim-Skoufa I, Atkins K, Hernández-Rodríguez MÁ. Optimizing real-world evidence studies for regulatory decision-making and impact assessment in pharmacovigilance. *Br J Clin Pharmacol*. 2025;91(4):1092-1095. doi:10.1111/bcp.16393
2. European Medicines Agency: The European medicines regulatory network <https://www.ema.europa.eu/en/about-us/how-we-work/european-medicines-regulatory-network>
3. Heads of Medicines Agencies: HMA/EMA Joint Big Data Steering Group <https://www.hma.eu/about-hma/working-groups/hma/ema-joint-big-data-steering-group.html>
4. Network Data Steering Group (NDSG) | European Medicines Agency (EMA) <https://www.ema.europa.eu/en/about-us/how-we-work/data-regulation-big-data-other-sources/network-data-steering-group-ndsg>
5. Arlett P, Kjaer J, Broich K, Cooke E. Real-World Evidence in EU Medicines Regulation: Enabling Use and Establishing Value. *Clin Pharmacol Ther*. 2022;111(1):21-23. doi:10.1002/cpt.2479
6. HMA-EMA Catalogues of real-world data sources and studies <https://catalogues.ema.europa.eu/>
7. BIFAP (Base de datos para la Investigación Farmacoepidemiológica en el Ámbito Público). Gobernanza de acceso a los datos. <https://bifap.aemps.es/data-governance>
8. Hernán MA, Hsu J, Healy B. A second chance to get causal inference right: a classification of data science tasks. *Chance*. 2019;32(1):42-49. doi:10.1080/09332480.2019.1579578
9. Maciá-Martínez MA, Gil M, Huerta C, et al. Base de Datos para la Investigación Farmacoepidemiológica en Atención Primaria (BIFAP): A data resource for pharmacoepidemiology in Spain. *Pharmacoepidemiol Drug Saf*. 2020;29(10):1236-1245. doi:10.1002/pds.5006

10. Reglamento (UE) 2025/327 del Parlamento Europeo y del Consejo, de 11 de febrero de 2025, relativo al Espacio Europeo de Datos de Salud, y por el que se modifican la Directiva 2011/24/UE y el Reglamento (UE) 2024/2847. «DOUE» núm. 327, de 5 de marzo de 2025
11. Second Joint Action Towards the European Health Data Space – TEHDAS2. <https://tehdas.eu/>
12. Observational Health Data Sciences and Informatics (OHDSI) Data Standardization –. The Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM). <https://www.ohdsi.org/data-standardization/>
13. AEMPS. Nota Informativa: 07 de marzo de 2024. Referencia: MUH, 05/2024. [BIFAP, la mayor base de datos española de registros médicos para investigación independiente, se integra en la red de big data de la Agencia Europea de Medicamentos](#)
14. Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am J Epidemiol.* 2016;183(8):758-764. doi:10.1093/aje/kwv254
15. Comparative effectiveness and safety studies using the target trial emulation and estimand frameworks (TARGET-EU). <https://catalogues.ema.europa.eu/node/4440/>
16. European Commission. The Draghi report on EU competitiveness. [The future of European competitiveness: Report by Mario Draghi.](#)

Sesión informativa sobre los ensayos clínicos en la UE: situando al paciente en el centro

El pasado 19 de mayo, la Agencia Española de Medicamentos y Productos Sanitarios (AEMPS) en colaboración con la Agencia Europea de Medicamentos (EMA, por sus siglas en inglés) celebró una sesión informativa online con el objetivo de dar a conocer las diferentes iniciativas que se están llevando a cabo tanto en España como en la Unión europea para resaltar el papel que tienen los pacientes en la investigación clínica.

La grabación del evento está disponible en la web oficial de la AEMPS, a través del siguiente enlace: <https://www.aemps.gob.es/la-aemps/sesion-informativa-sobre-los-ensayos-clinicos-en-la-ue-situando-al-paciente-en-el-centro/>



Entidades patrocinadoras y colaboradoras de ICB digital



Chiesi



Lilly



Menarini



Novartis



Roche