



Revisión

Análisis estadístico en investigación biomédica potenciado por software e inteligencia artificial: mejorar precisión y eficiencia en datos complejos

César Dilú Sorzano

Albera Salut, Girona. Grupo de Salud Digital de la SEMG.

INFORMACIÓN DEL ARTÍCULO

Historia del artículo:

Recibido el 29 de diciembre de 2025

Aceptado el 30 de abril de 2026

On-line el 16 de junio de 2026

Palabras clave:

Inteligencia artificial

Interpretación estadística de datos

Investigación biomédica

Aprendizaje automático

Modelos de lenguaje grande

Medicina familiar y comunitaria

R E S U M E N

La investigación biomédica enfrenta una transformación metodológica impulsada por la inteligencia artificial (IA). Este artículo analiza la evolución del análisis estadístico, desde paquetes tradicionales hasta los modelos de lenguaje grande, y propone un marco para su integración responsable en la medicina de familia.

Se realizó una revisión narrativa basada en una búsqueda estructurada (PubMed, arXiv, Google Scholar) de publicaciones recientes y herramientas emergentes. Los resultados indican que la IA simplifica el análisis de datos complejos, aunque persisten desafíos como los sesgos algorítmicos, la privacidad de datos y la necesidad de validación rigurosa.

La sinergia entre la estadística clásica y la IA es fundamental: la IA para la exploración y generación de hipótesis, y los métodos tradicionales para la validación e inferencia causal. Es primordial que los médicos de familia desarrollen una comprensión conceptual de estas tecnologías para asegurar una implementación ética que mejore su producción científica.

© 2026 Sociedad Española de Médicos Generales y de Familia.

Publicado por E-Medfarma 2020, S.L.

Statistical analysis in biomedical research powered by software and artificial intelligence: improving accuracy and efficiency in complex data

A B S T R A C T

Biomedical research is undergoing a methodological transformation driven by artificial intelligence (AI). This article analyzes the evolution of statistical analysis, from traditional software packages to large language models, and proposes a framework for their responsible integration into family medicine.

A narrative review was conducted based on a structured search (PubMed, arXiv, Google Scholar) of recent publications and emerging tools. The results indicate that AI simplifies the analysis of complex data, although challenges such as algorithmic bias, data privacy, and the need for rigorous validation persist.

Keywords:

Artificial intelligence

Statistical data interpretation

Biomedical research

Machine learning

Large language models

Family practice

The synergy between classical statistics and AI is fundamental: AI for exploration and hypothesis generation, and traditional methods for validation and causal inference. It is essential that family physicians develop a conceptual understanding of these technologies to ensure ethical implementation that enhances their scientific output.

© 2026 Sociedad Española de Médicos Generales y de Familia.
Published by E-Medfarma 2020, S.L.

Introducción

La investigación biomédica ha pasado de un paradigma de datos escasos a uno de abundancia, generando desafíos que los métodos estadísticos tradicionales no pueden abordar en su totalidad¹. Los enfoques clásicos, aunque rigurosos, presentan limitaciones en escalabilidad, manejo de datos multimodales y descubrimiento de patrones complejos no hipotetizados^{2,3}. En este contexto, la inteligencia artificial (IA) emerge como un complemento indispensable para extraer conocimiento de los vastos y complejos conjuntos de datos biomédicos actuales⁴.

Este artículo tiene como objetivo analizar la transformación del análisis estadístico impulsada por la IA, describir la evolución de las herramientas analíticas disponibles y proponer un marco para su integración responsable en la investigación en Medicina de Familia.

Material y métodos

Este trabajo se fundamenta en la conferencia “Análisis estadístico en investigación biomédica potenciado por software e inteligencia artificial” (jornada INVESTIGA-MF, Madrid, 23 de octubre de 2025), y ha sido ampliado y actualizado mediante una revisión narrativa, no sistemática, con un enfoque descriptivo y conceptual de la literatura científica y de las herramientas disponibles.

Se consultaron las bases de datos PubMed, arXiv y Google Scholar entre enero de 2024 y enero de 2025. Este período se seleccionó para capturar la evidencia más reciente sobre modelos de lenguaje grande (LLM) en bioestadística, un campo en rápida evolución; la literatura previa seminal se incluye mediante citas directas en el texto. Se excluyeron Embase y Cochrane Library porque la cadena de búsqueda empleada—orientada a preprints y publicaciones tecnológicas—ofrece cobertura limitada en esas plataformas para el dominio de IA generativa. Se utilizó la cadena de búsqueda (“*large language models*” OR “*generative AI*”) AND (“*biostatistics*” OR “*medical research*”).

La búsqueda identificó 300 registros; se priorizaron estudios sobre sesgos algorítmicos y validación clínica. Tras eliminar 50 duplicados o textos sin acceso, se cribaron 250 referencias y se excluyeron 205 por no cumplir al menos uno de los siguientes criterios operativos de inclusión: (a) relevancia directa para

el uso de LLM o herramientas de IA en análisis estadístico biomédico; (b) disponibilidad del texto completo; (c) aporte de evidencia empírica, evaluación comparativa de herramientas o revisión metodológica con aplicabilidad clínica documentada. De los 45 artículos evaluados a texto completo, se descartaron 30 por limitaciones metodológicas mayores—ausencia de validación externa, resultados no replicables o muestras no representativas de población clínica—o por falta de aplicabilidad directa al contexto de la medicina de familia, y se seleccionaron

Las herramientas analizadas se seleccionaron bajo criterios de: accesibilidad para no-programadores (interfaces *no-code* o lenguaje natural), relevancia en publicaciones recientes y disponibilidad de alternativas de código abierto (*Open Source*) para garantizar la equidad de acceso frente a soluciones comerciales (anexo 1).

Resultados

El ecosistema de análisis de datos ha evolucionado desde herramientas estándar como SPSS, que ahora integra módulos de IA (watsonx.ai), hacia el paradigma de los modelos de lenguaje grande (LLM), como GPT-5, Gemini y Claude⁵⁻⁷. Estas herramientas conversacionales permiten ejecutar análisis mediante lenguaje natural. No obstante, para evitar sesgos comerciales y barreras de acceso, es crucial destacar alternativas de código abierto, como JASP o Jamovi, que integran interfaces intuitivas con la potencia de R, y bibliotecas de Python como *PandasAI* que ofrecen capacidades similares a las herramientas de pago sin coste de licencia.

Grambow y cols revelan una rápida adopción de LLM: el 63,8 % de bioestadísticos los utiliza, principalmente para depuración de código. Sin embargo, el 70,7 % encontró errores significativos, lo que refuerza la necesidad de verificación humana⁸. Además, el soporte institucional es limitado, y existe un vacío en directrices formales.

Las aplicaciones de la IA incluyen predicción de riesgo y algoritmos diagnósticos^{9,10}. Sin embargo, la simplificación del análisis conlleva riesgos que deben gestionarse activamente.

El uso de herramientas en la nube (SaaS), como Julius AI o ChatGPT, plantea desafíos clave bajo el Reglamento General de Protección de Datos (RGPD, Reglamento UE 2016/679). Es recomendable que los investigadores anonimicen estrictamente los datasets antes de subirlos a estas plataformas, o

Anexo 1 – Ecosistema de herramientas para análisis biomédico (selección representativa).			
Categoría	Herramienta	Tipo	Uso principal y ventaja
IA conversacional	GPT-5.1/Gemini	Pago/Free	Generación de código R/Python. Alta versatilidad
	Llama 3/DeepSeek	Open Source	Ejecución local para privacidad total de datos sensibles
Asistentes de datos	Julius AI	Pago	Análisis “no-code” automatizado y gráficos rápidos
	PandasAI (Python)	Open Source	Alternativa gratuita programable para análisis conversacional
Software estadístico	SPSS + watsonx	Pago (Alto)	Estándar clásico con integración de IA reciente
	JASP/Jamovi	Open Source	Interfaz “point-and-click” gratuita, ideal para Medicina de Familia
Machine learning visual	Orange Data Mining	Open Source	Minería de datos visual sin código. Ideal docencia/exploración
	KNIME	Freemium	Flujos de trabajo complejos (ETL) y automatización
Captura y visualización	REDCap	Académico	Estándar “gold” para recolección segura de datos clínicos
	Flourish Studio	Freemium	Visualización interactiva web

La inclusión de herramientas no implica recomendación ni aval por parte de SEMG.
Se incluyen alternativas Open Source (OS) para fomentar la accesibilidad.

prioricen el uso de LLM locales (como Llama 3 o Gema 4) para datos sensibles, con el fin de evitar la exposición de información de salud protegida. En España, la Estrategia de Salud Digital del Sistema Nacional de Salud establece marcos para la transformación digital que incluyen el uso responsable de IA en investigación¹¹.

Existe el riesgo de amplificar sesgos raciales o socioeconómicos presentes en los datos de entrenamiento, lo que perpetúa disparidades en salud^{12,13}. Las “alucinaciones” de los LLM exigen un escepticismo metodológico y validación constante⁸.

Se identifican herramientas facilitadoras (ver anexo 1 para comparativa detallada):

- Análisis conversacional. *Julius AI* permite el análisis en lenguaje natural, útil para comparar variables clínicas (HbA1c, por ejemplo). Como alternativa gratuita y segura, se recomienda el uso de *scripts* de R generados por IA, pero ejecutados en entornos locales seguros.
- Visualización. *Flourish Studio* destaca en *storytelling* de datos.
- *Machine learning visual*: *Orange Data Mining (Open Source)* ofrece minería de datos mediante widgets, ideal para detectar patrones en bases de datos clínicas sin coste alguno.

Comentarios y conclusiones

La integración de la IA en la investigación biomédica es transformadora, pero requiere precaución. Lo ideal es tener un modelo de colaboración sinérgica: la IA debe emplearse en la fase exploratoria y en la generación de código, mientras que la validación final debe anclarse en la estadística tradicional.

La facilidad para generar múltiples análisis con IA aumenta el riesgo de *data dredging* (tortura de datos) o *p-hacking*. Los investigadores deben pre-registrar sus hipótesis y no utilizar la IA para buscar correlaciones espurias indiscriminadamente hasta obtener una “p” significativa.

Para los médicos de familia es indispensable una formación en IA que incluya comprensión conceptual y ética. Fomentar una cultura de validación continua y gobernanza de datos es esencial para garantizar avances científicos reproducibles y seguros¹⁴.

Limitaciones

El diseño narrativo no sistemático limita la reproducibilidad del proceso de selección bibliográfica e impide establecer inferencias causales o cuantificar la magnitud de los efectos descritos. La restricción temporal de la búsqueda (enero 2024-enero 2025), justificada por la naturaleza emergente del campo, excluye trabajos seminales anteriores no referenciados explícitamente en el texto. La selección de herramientas descrita en el Anexo 1 es representativa, no exhaustiva, y refleja el estado del ecosistema en el momento de elaboración del manuscrito; la velocidad de evolución tecnológica en este dominio puede limitar la vigencia de algunas recomendaciones. Finalmente, las afirmaciones sobre rendimiento de herramientas de IA se basan en la evidencia disponible en las fuentes primarias citadas y no en una evaluación experimental directa por parte del autor.

Conflicto de intereses

El autor declara no tener ningún conflicto de intereses. Se utilizaron Claude Sonnet 4.6 (Anthropic), Gemini 3.1 (Google) y GPT-5.1 (OpenAI) exclusivamente para asistencia en estructuración y corrección de estilo, en cumplimiento con la política de SEMG sobre uso de IA en publicaciones¹⁵.

El autor asume total responsabilidad por el contenido.

BIBLIOGRAFÍA

1. Stephens ZD, Lee SY, Faghri F, Campbell RH, Zhai C, Efron MJ, et al. Big Data: Astronomical or Genomical? *PLoS Biol.* 2015;13(7):e1002195. <https://doi.org/10.1371/journal.pbio.1002195>
2. Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. *Nat Methods.* 2018;5(4):233-4. <https://doi.org/10.1038/nmeth.4642>
3. Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. *N Engl J Med.* 2019;380(14):347-58. <https://doi.org/10.1056/NEJMr1814259>

4. Haug CJ, Drazen JM. Artificial Intelligence and Machine Learning in Clinical Medicine, 2023. *N Engl J Med.* 2023;388(13):1201-8. <https://doi.org/10.1056/NEJMra2302038>
5. Lee P, Bubeck S, Petro J. Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine. *N Engl J Med.* 2023;388(13):1233-9. <https://doi.org/10.1056/NEJMr2214184>
6. Gemini Team Google. Gemini: A family of highly capable multimodal models [Technical Report]. Google; 2025. Disponible en: https://storage.googleapis.com/deepmind-media/gemini/gemini_1_report.pdf
7. Anthropic. Claude 3 Model Card [Technical Report]. Anthropic; 2024. Disponible en: <https://www-cdn.anthropic.com/9ff93d-fa8f445c932415d335c88852ef47f1201e.pdf>
8. Grambow SC, Desai M, Weinfurt KP, Lindsell CJ, Pencina MJ, Rende L, et al. Integrating large language models in biostatistical workflows for clinical and translational research. *J Clin Transl Sci.* 2025;9(1):e131. <https://doi.org/10.1017/cts.2025.10064>
9. Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS One.* 2017;12(4):e0174944. <https://doi.org/10.1371/journal.pone.0174944>
10. Shimabukuro DW, Barton CW, Feldman MD, Mataraso SJ, Das R. Effect of a ML-based severe sepsis prediction algorithm. *BMJ Open Respir Res.* 2017;4(1):e000234. <https://doi.org/10.1136/bmjresp-2017-000234>
11. Ministerio de Sanidad. Estrategia de Salud Digital del Sistema Nacional de Salud 2021-2026. Madrid: Gobierno de España; 2021. Disponible en: https://www.sanidad.gob.es/areas/saludDigital/doc/Presentacion_avances_de_la_Estrategia_de_Salud_Digital_web.pdf
12. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science.* 2019;366(6464):447-53. <https://doi.org/10.1126/science.aax2342>
13. Haider SA, Borna S, Gómez-Cabello CA, Pressman SM, Haider CR, Forte AJ The algorithmic divide: AI-driven racial disparities in healthcare. *J Racial Ethn Health Disparities.* 2026;13(1):188-217. <https://doi.org/10.1007/s40615-024-02237-0>
14. European Medicines Agency (EMA). Reflection paper on the use of AI in the medicinal product lifecycle [Internet]. EMA; 2023. Disponible en: <https://www.ema.europa.eu/en/news/reflection-paper-use-artificial-intelligence-lifecycle-medicines>
15. Rivas Fernández AM, Ramos Ortiz C. La bioética en la era de la inteligencia artificial: desafíos y perspectivas. *Med Gen Fam.* 2024;13(1):1-3. <https://doi.org/10.24038/mgyf.2024.001>