



Explosión de artículos **formularos** basados en NHANES

¿Estamos presenciando una crisis de calidad en la ciencia automatizada?



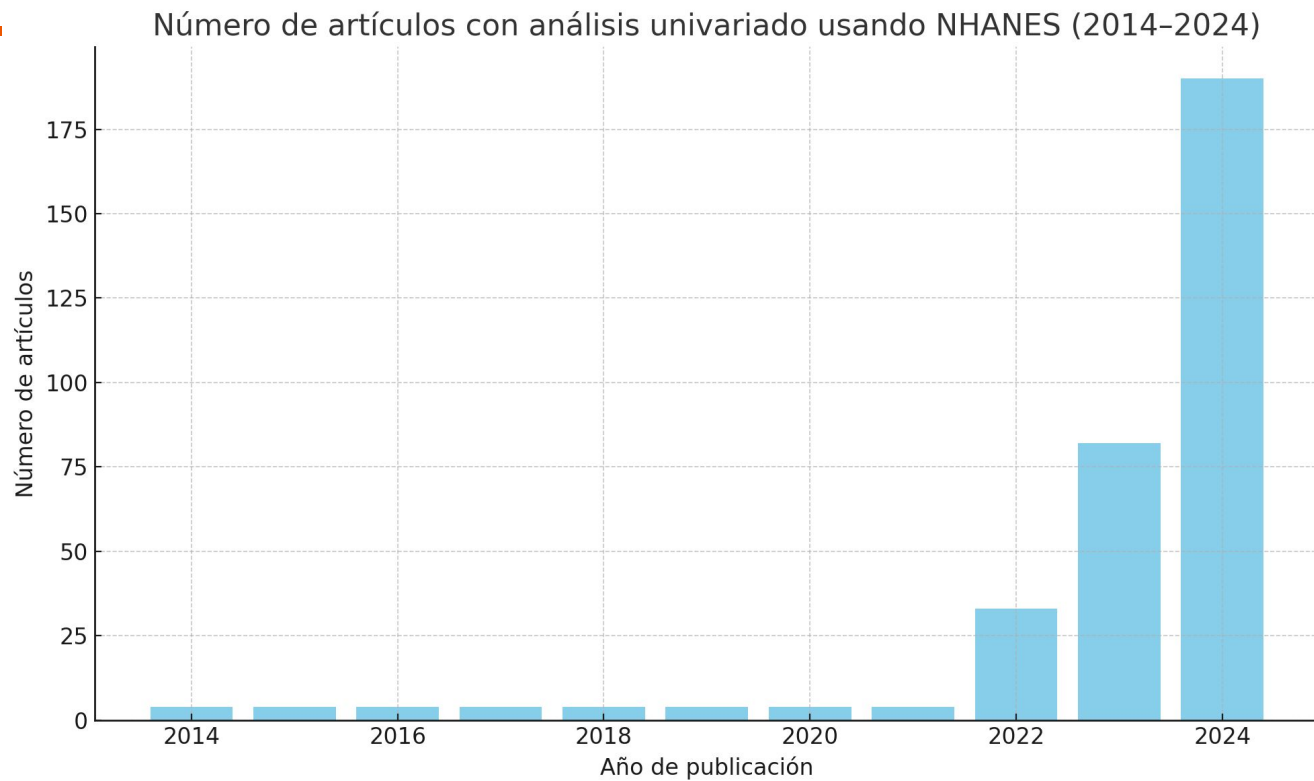
Contexto

- NHANES: base de datos pública con más de 700 variables de salud y nutrición.
- Alta accesibilidad y compatibilidad con herramientas de IA.
- Ideal para minería de datos masiva y automatizada.

El estudio (PLOS Biology, 2025)

- 341 artículos entre 2014–2024 analizados, 147 revistas.
- 190 publicados en 2024.
- Patrones repetitivos, ausencia de análisis multifactoriales y uso sospechoso de IA.

Entre 2014 y 2020, 2 de 25 manuscritos tenían un autor principal en China, en comparación con 292 de 316 manuscritos entre 2021 y 2024.



Problemas metodológicos detectados

- Modelos univariados en condiciones multifactoriales.
- No se aplican correcciones por falsos positivos.
- Selección arbitraria de subconjuntos de datos.
- Causalidad invertida entre variables.

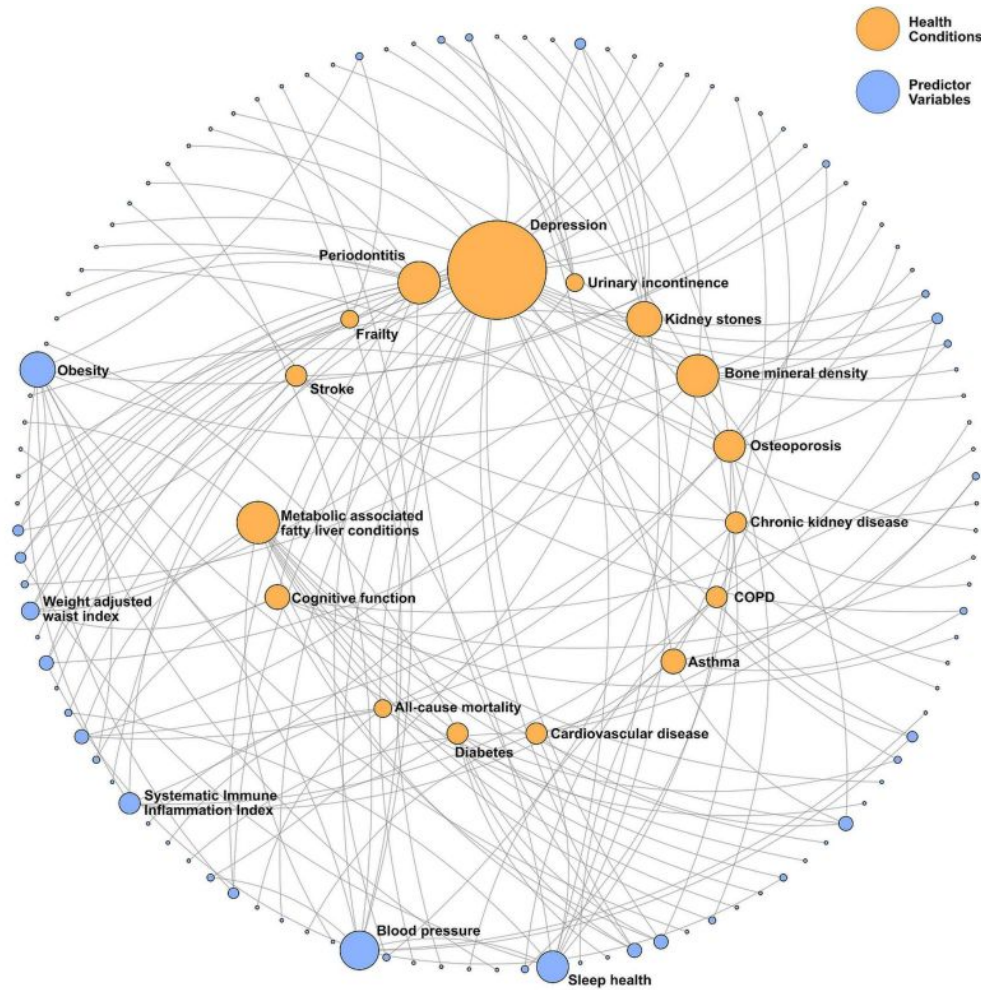


Table 2. Summary of papers identifying associations with depression. All manuscripts use NHANES data, 2014-2024. Articles listed by date of publication.

PMID	Year Published	Predictor variable associated with depression	Reported or derived p-value	FDR p-value
39277033	2024	Serum neurofilament light chains	0.039	0.161
39355374	2024	Blood pressure ^a	0.044	0.173
38869164	2024	Life's Essential 8	<0.001	<0.001
39121312	2024	Triglyceride-glucose index	<0.001	0.004
38970091	2024	Adherence to Mediterranean diet	0.039	0.161
38355455	2024	Geriatric nutritional risk index	<0.001	<0.001
38642902	2024	Overactive bladder	<0.001	<0.001
39029685	2024	Muscle mass	0.008	0.059
38710330	2024	Dietary vitamin C intake	<0.001	<0.001
38944294	2024	Immune-inflammation-based prognostic index	<0.001	<0.001
38579547	2024	Physical activity	0.032	0.153
39044344	2024	Lipid accumulation products	0.031	0.153
38154580	2024	Ethylene oxide levels	0.021	0.133
38220117	2024	Weight-Adjusted Waist Index	0.008	0.059
38910137	2024	Blood cadmium	<0.001	<0.001
37838268	2024	Non-HDL cholesterol to HDL cholesterol ratio (NHHR)	0.031	0.153
37775007	2023	Per- and polyfluoroalkyl substances (PFAS)	<0.001	<0.001
39227001	2023	Non-high-density lipoprotein cholesterol	0.023	0.133
37474898	2023	Dietary anthocyanidins intake	0.020	0.133
37236270	2023	Serum α -Klotho	0.022	0.133
37447273	2023	Serum Vitamin D	0.032	0.153
36462606	2023	Dietary Inflammatory Index (DII)	<0.001	<0.001
37340352	2023	Serum albumin	0.037	0.161
36265728	2023	Organophosphorus pesticide exposure	0.040	0.161
36008859	2022	Physical activity	<0.001	<0.001
33292309 [17]	2020	Ovariectomy-reduced hormones	<0.001	<0.001
28395506	2017	Inflammatory bowel disease	0.002	0.017
24636212	2014	Poor dental health	<0.001	<0.001

^aStudy separated systolic and diastolic blood pressure and reported a p-value 0.04 for diastolic and 0.09 for systolic blood pressure.

Table 3. Summary of papers analyzing SII. All manuscripts use NHANES data, 2014–2024. Articles listed by date of publication.

PMID	Year published	Health conditions associated with SII	Years analyzed in paper	Years available (NHANES table/var code)
39052624	2024	Liver injury	2017–2020	(ALT): 1999–2020 (LBXSATSI)
				(AST): 1999–2020 (LBXSASSI)
				(GGT): 1999–2020 (LBXSGTSI)
				(ALP): 2003–2020 (LBXSAPSI)
39312354	2024	Hearing loss	2005–2018	1999–2020 (AUX)
38974989	2024	Diabetes	2003–2018	1999–2020 (DIQ)
38822015	2024	Overactive bladder	2005–2018	2005–2020 (KIQ480/KIQ044)
39069464	2024	Obesity	2011–2018	1999–2020 (BMXBMI/ BMXWAIST)
39076553	2024	Stroke	1999–2020	1999–2020 (MCQ160f/F)
38988356	2024	Asthma	2009–2018	2001–2020 (MCQ010/25/35/40/50)
39239395	2024	Serum neurofilament light chain	2013–2014	2013–2014 (SSSNFL)
39267962	2024	Atopic dermatitis	2001–2006	2001–2006 (DED061/AGQ180)
36904176	2023	Hyperlipidemia	2015–2020	1999–2020 (LAB13/TCHOL)
37883981	2023	Serum ferritin levels	2015–2018	1999–2002, 2005–2010, 2015–2020 (LBXFER)
38027115	2023	Diabetes	2017–2020	1999–2020
38076255	2023	Chronic obstructive pulmonary disease (COPD)	2013–2020	2011–2020 (MCQ160g/o)
36895572	2023	Kidney stones	2007–2018	2007–2020

<https://doi.org/10.1371/journal.pbio.3003152.t003>

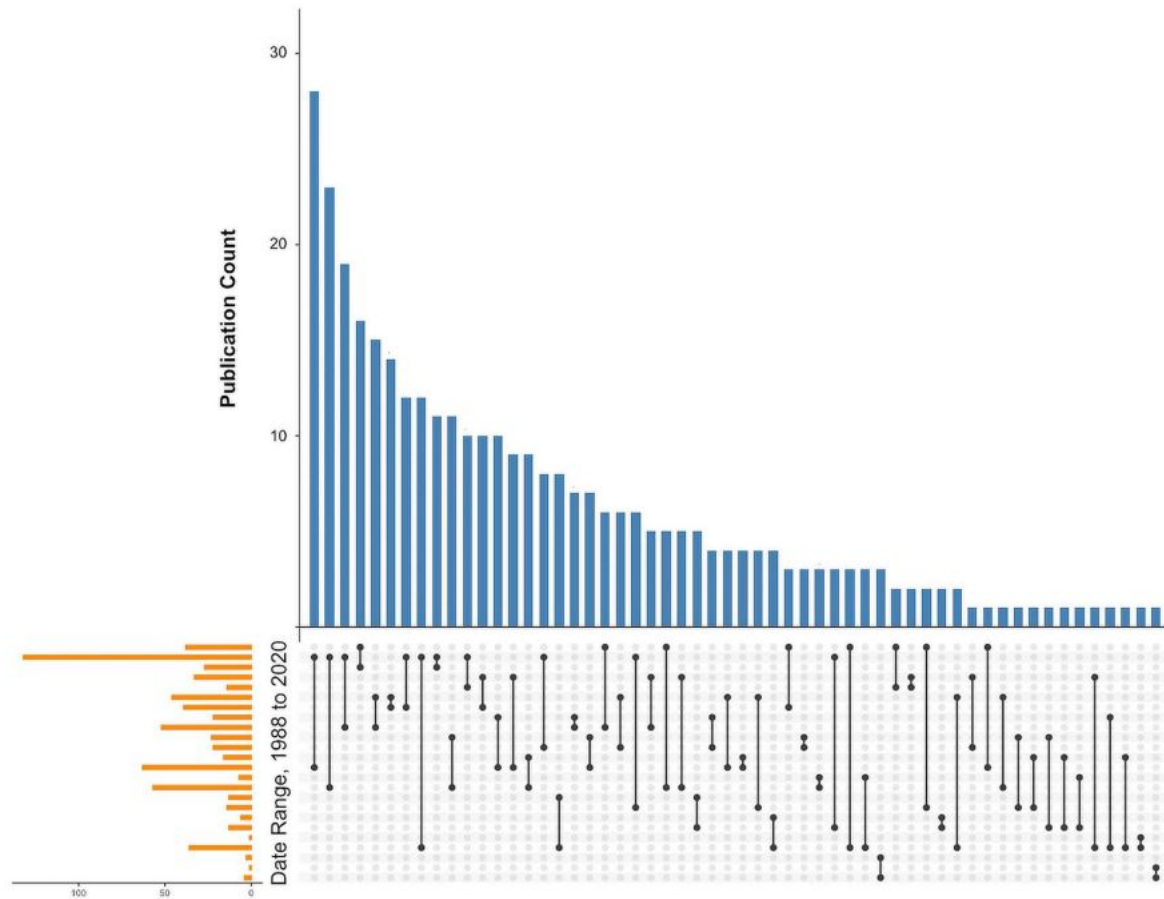


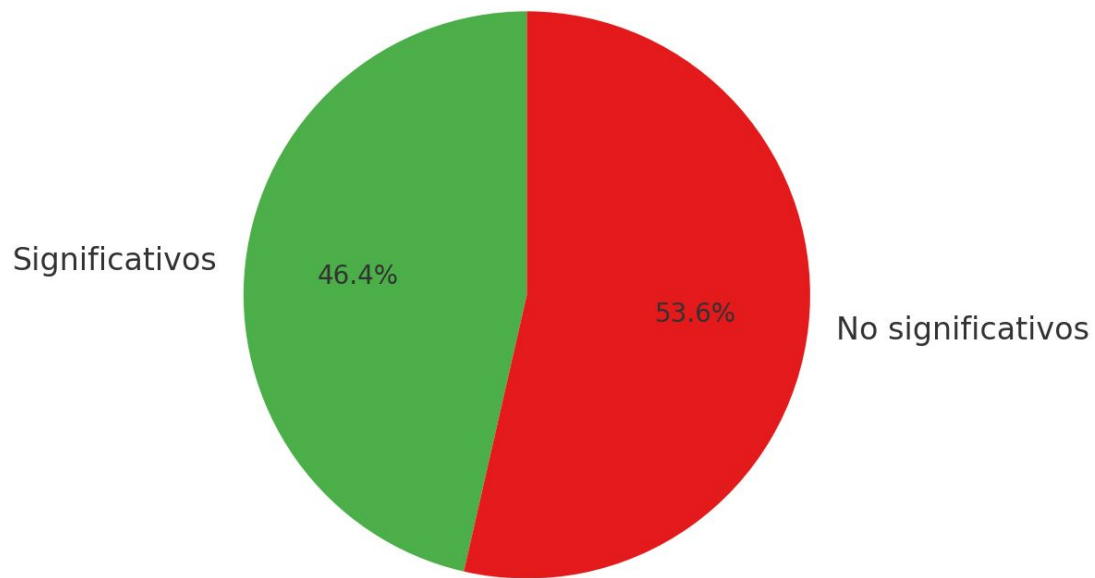
Fig 5. UpSet plot illustrating selections of annual data taken from NHANES that were employed for the 341 publications analyzed in this work. The data are taken from [S1 Data](#) Table A, using the start dates to the end dates of subsets of NHANES data analyzed by each manuscript.

<https://doi.org/10.1371/journal.pbio.3003152.g005>

¿Está la IA detrás de esto?

- Redacción automatizada y papers con estructura repetitiva.
- Combinaciones masivas de variables para buscar significancia.
- Sospecha de paper mills usando NHANES como blanco.

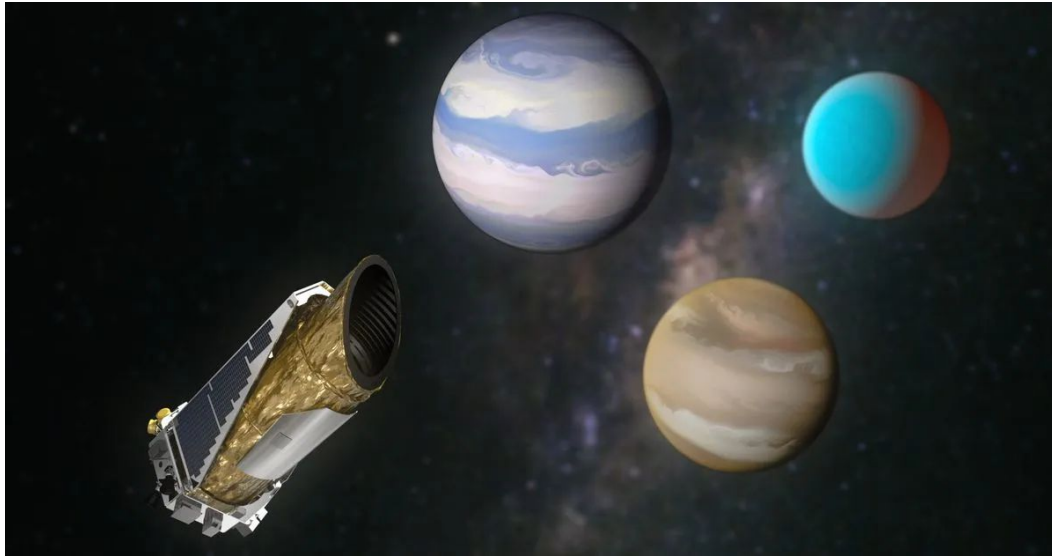
Resultados tras corrección FDR en estudios sobre depresión (n=28)



Predicción automatizada de exoplanetas con IA y falsos positivos



Los autores desarrollan una red neuronal convolucional para **detectar exoplanetas** en los datos del telescopio Kepler. El modelo logró descubrir nuevos candidatos planetarios como Kepler-90i. Sin embargo, el mismo trabajo reconoce limitaciones importantes que reflejan los problemas del primer artículo:





Reflejo en Shallue & Vanderburg (astronomía + IA)

Se usaron datos del pipeline Kepler ya preprocesados, pero admiten que hay sesgos ópticos y de clasificación.

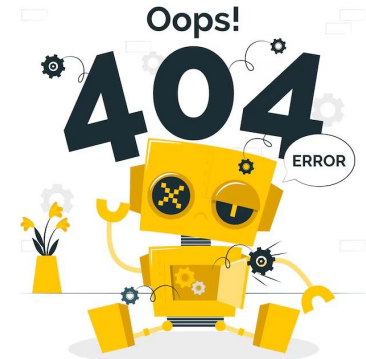
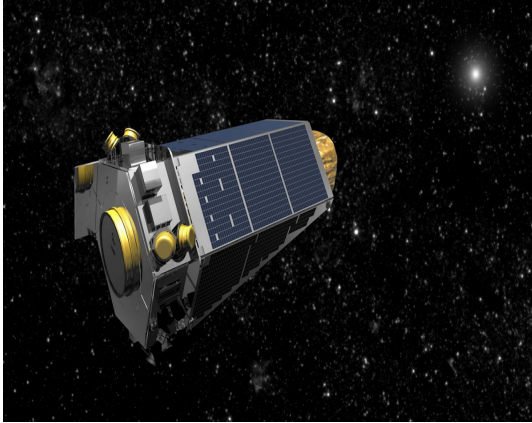
Su modelo clasificó como planetas varios falsos positivos (p. ej., eclipses binarios, artefactos instrumentales).

Reconocen que no incluyeron suficientes datos simulados en el entrenamiento ni validación robusta de falsos positivos.

El modelo identifica tránsitos sin contexto astrofísico (por ejemplo, sin datos de posición en el CCD), y clasificó señales espurias como exoplanetas.

Aunque no es un “paper mill”, advierten que si no se controla, el modelo puede ser usado ciegamente para poblar catálogos.

Cuando tu IA encuentra un planeta... que en realidad es un error...



Shallue, C. J., & Vanderburg, A. (2018). Identifying Exoplanets with Deep Learning: A Five-planet Resonant Chain Around Kepler-80 and an Eighth Planet Around Kepler-90. *The Astronomical Journal*, 155(2), 94.

Yu, L., Petigura, E. A., & Howard, A. W. (2019). *"Reevaluating exoplanet candidate Kepler-90i: a false positive from a convolutional neural network."* *The Astronomical Journal*, 158(1), 25.
<https://doi.org/10.3847/1538-3881/ab21d6>

Repercusiones y recomendaciones

- Riesgo de saturación de la literatura con asociaciones espurias.
- Necesidad de pre registro, revisión estadística y rechazo editorial preventivo.
- Discusión post-publicación para depurar la evidencia científica.
- Protección ética del uso de IA en ciencia.

META-RESEARCH ARTICLE

Explosion of formulaic research articles, including inappropriate study designs and false discoveries, based on the NHANES US national health database

Tulsi Suchak¹, Anietie E. Aliu¹, Charlie Harrison², Reyer Zwiggelaar², Nophar Geifman¹, Matt Spick^{1*}

1 School of Health Sciences, Faculty of Health and Medical Sciences, University of Surrey, Guildford, United Kingdom, **2** Department of Computer Science, Aberystwyth University, Ceredigion, United Kingdom