Explosión de artículos formuleros basados en NHANES

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



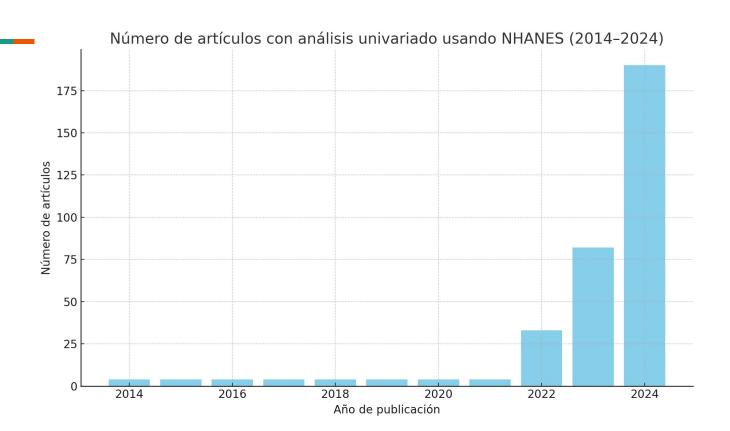
Contexto

- 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.

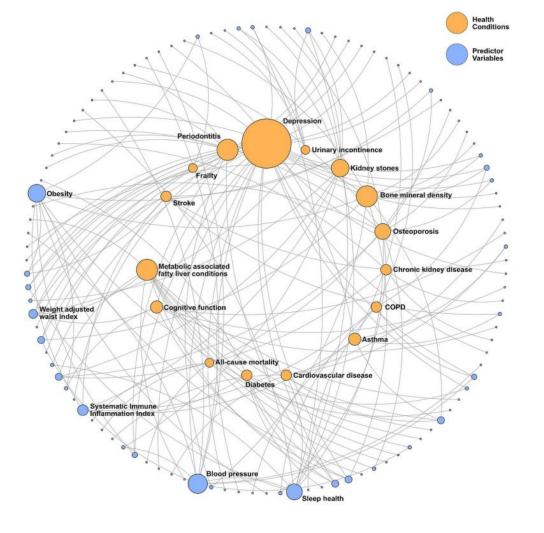
Entre 2014 y 2020, 2 de 25 manuscritos tenían un autor principal en China, en comparación con 292 de 316 manuscritos entre 2021 y 2024.



Frontiers Media SA			Springer		Elsevier	Elsevier		
Frontiers in Nutrition		Frontiers in Public Health	Scientific Reports	Descriptions until set between Acts Chaldeslegists Access	Journal of Affective Disorders	Preventive Medicine Reports Served water harvery		
				World Journal of Urology		Chemosphere Common Services		
		Frontiers in Medicine				Epilopia Balantania - Non-street Sente Sancol - Non-street Principal - Non-street Sentence - Non-street Senten		
				ention d'adequates Endocrine		hare Sales Sales Sales Sales Michigane		
Frontiers in Endocrinology		Province in Psychiatry - Provinces in Pedadrics		Once the manage	Heliyon			
					Ny of the formats	Heart and Long Remote Public Health Urology		
		Frontiers in Oncology	Lippincott Williams and Wilkins	John Wiley and Sons Inc	Academic Press Inc. Oxford U	niversity Press SAGE Publications Inc.		
		Frontiers in Neurology	Medicine (United States)	Clinical Cardiology Journal of Diabetes	Environmental Research	Dead fortige and the state of t		
BioMed Central Ltd				Secretary of Landson Control of C		Section of Decomposition Medicine		
BMC Public Health	BMC Oral He	alth		Jauma' of Remainships Preglatoric Obesity	Hindawi Limited Churchill Livingston	Dove Medical Press Ltd Discretized Survey of COTO Sinusconnectal Health Perspectives		
				MDPI	Central and Centra	and the contract of the contra		
	BHC Sessorial Descripts	Survey of trades, Nandative and Substitute		Nutrients	Tree Serve of Service Associates			
		BMC Urology BmC reduction BMC Natrition BMC Certains	BMJ Publishing Group		Taylor and Francis Ltd. And Continuous Arrest of Automit	Del Presidentes John Libbary		
	BMC Psychiatry		BMJ Open		S. Karger AG	SOUTCE Last and Uses		
Lipids in Health and Disease		EM: Remark Hubb. Mr. Stransman, Inc.		Public Library of Science	Oglehalmic Randarch	Karger European haumanage 100 methodorus Australianus		
		BRC Administry Hobidity		PLoS ONE	Academic Press Wolters Kluwer	Executive Characteristic Characteris		
		Management States and States States and States States and States States and S	8MJ Open Respiratory Research BMJ open		Int J Surg	Super Print, Id.		
						Sealth in Pagastra Cross Access Access to Apparture		

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.



Year Published Predictor variable associated with depression Reported or derived p-value FDR p-value **PMID**

39211033	2024	Serum neuroniament light chains	0.039	0.101
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

Non-HDL cholesterol to HDL cholesterol ratio (NHHR)

Per- and polyfluoroalkyl substances (PFAS)

Non-high-density lipoprotein cholesterol

Organophosphorus pesticide exposure

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

38579547

39044344

38154580

38220117

38910137

37838268

37775007

39227001

37474898

37236270

37447273

36462606

37340352

36265728

36008859

28395506

24636212

33292309 [17]

2024

2024

2024

2024

2024

2024

2023

2023

2023

2023

2023

2023

2023

2023

2022

2020

2017

2014

Physical activity

Blood cadmium

Serum a-Klotho

Serum Vitamin D

Serum albumin

Physical activity

Poor dental health

Lipid accumulation products

Weight-Adjusted Waist Index

Dietary anthocyanidins intake

Dietary Inflammatory Index (DII)

Ovariectomy-reduced hormones

Inflammatory bowel disease

Ethylene oxide levels

20277022

2024 Corum nourofilament light chains 0.020 0 161

publication.

0.032

0.031

0.021

0.008

< 0.001

< 0.001

0.023

0.020

0.022

0.032

< 0.001

0.037

0.040

< 0.001

< 0.001

0.002

< 0.001

0.031

0.153

0.153

0.133

0.059

< 0.001

0.153

< 0.001

0.133

0.133

0.133

0.153

< 0.001

0.161

0.161

< 0.001

< 0.001

0.017

< 0.001

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

Table 3. Summary of papers analyzing SII. All manuscripts use NHANES data, 2014–2024. Articles listed by date of publication. **PMID** Year Health conditions associated with Years analyzed in 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)

paper

36904176 2023 37883981 2023 38027115 2023 Diabetes

2023

2023

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

2024

2024

2024

2024

2024

39069464

39076553

38988356

39239395

39267962

38076255

36895572

published

SII

Obesity

Stroke

Asthma

(COPD)

Kidney stones

Atopic dermatitis Hyperlipidemia Serum ferritin levels

Chronic obstructive pulmonary disease

Serum neurofilament light chain

1999-2020 2009-2018 2013-2014

2011-2018

2001-2006

2015-2020

2015-2018

2017-2020

2013-2020

2007-2018

1999-2020 (MCQ160f/F) 2013-2014 (SSSNFL)

2011-2020 (MCQ160g/o)

2007-2020

2005-2020 (KIQ480/KIQ044) 1999-2020 (BMXBMI/ BMXWAIST)

2001-2020 (MCQ010/25/35/40/50)

2001-2006 (DED061/AGQ180)

1999-2020 (LAB13/TCHOL)

1999-2002, 2005-2010, 2015-2020 (LBXFER) 1999-2020

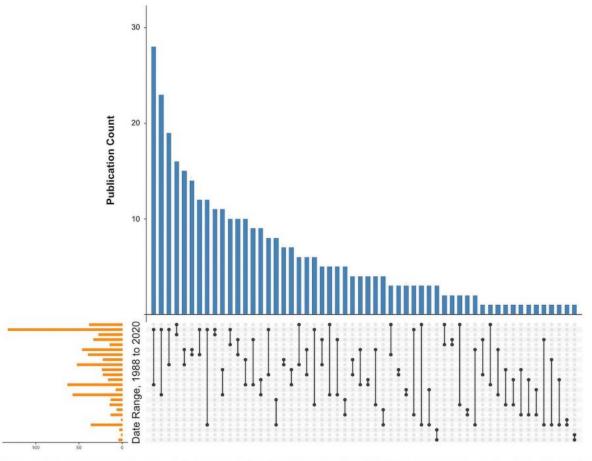
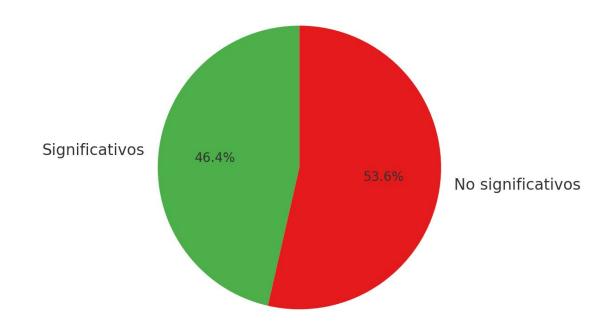


Fig 5. UpSet plot illustrating selections of annual data taken from NHANES that were employed for the 341 publications analyzed in this work. The data are taken from S1 Data Table A, using the start dates to the end dates of subsets of NHANES data analyzed by each manuscript.

¿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.

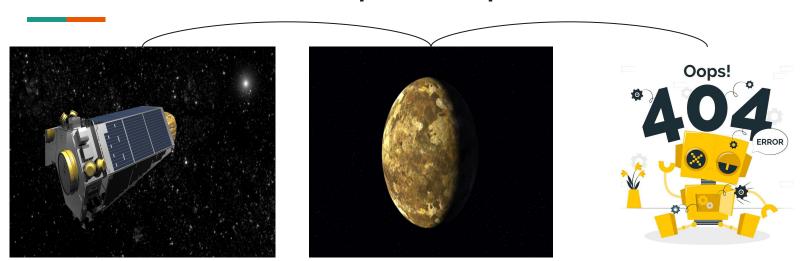
Su modelo clasificó como planetas varios falsos positivos (p. ej., eclipses binarios, artefactos instrumentales).

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 Spick₆₀^{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