An Update in Clinical Informatics

according to Moe & Larry

AMDIS PCC
Ojai, California
June 2024

Colin Banas, CMO, *Dr. First*
Bill Galanter, xMDIS, xAssocCHIO, Assoc Prof,
University of Illinois Chicago

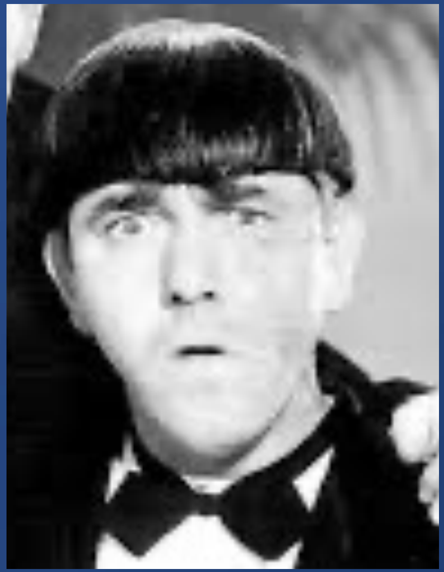
DrFirst
Association of Medical Directors of Information Systems AMDIS
University of Illinois Hospital & Health Sciences System
Changing medicine. For good.

Conflicted Stooges

Dr. Howard Dr. Fine

Dr. First Dr. Nothing

Review Methodology



What struck him as cool last year
(during the times of sobriety)
Random Thoughts

Clinical studies in last year
Longitudinal Analysis of Publications
Analysis of Abstract Content
Self Promotion

Bills methodology to find articles

Select CI MESH headings from major MESH headers of "clinical", "medical", "nursing", "dental", "health“. Did not include straight technology(i.e. "Biomedical Engineering", "Biomedical technology", "Electronics, Medical" etc..)

This gives the following concepts;

Adverse Drug Reaction Reporting Systems, Ambulatory Care Information Systems, Artificial Intelligence, Biological Ontologies, clinical informatics , Clinical Laboratory Information Systems, Clinical Pharmacy Information Systems, Community Networks, Consumer Health Informatics, Decision Making, Computer-Assisted, Decision Support Systems, Clinical, Decision Support Techniques, Dental Informatics, Diagnosis, Computer-Assisted, Drug Therapy, Computer-Assisted, Electronic Prescribing, Geographic Information Systems, Health Information Exchange, Health Information Systems, Health Smart Cards, Hospital Information Systems, Image Interpretation, Computer-Assisted, Information Systems, Integrated Advanced Information Management Systems, Knowledge Bases, Medical Informatics Applications, Medical Informatics Computing, Medical Order Entry Systems, Medical Record Linkage, Medical Records Systems, Computerized, Nursing Informatics, Operating Room Information Systems, Patient Generated Health Data, Patient Portals, Point-of-Care Systems, Prescription Drug Monitoring Programs, Public Health Informatics, Radiology Information Systems, Radiotherapy, Computer-Assisted, Reminder Systems, Telemedicine

This is simplified by using proximal concepts that contain many of the sub-concepts of interest;

"Artificial Intelligence"[MESH] OR "Clinical Laboratory Information Systems"[MESH] OR "Consumer Health Informatics"[MESH] OR "Therapy, Computer-Assisted "[MESH] OR "Decision Support Techniques"[MESH] OR "Dental Informatics"[MESH] OR "Drug Information Services"[MESH] OR "Electronic Prescribing"[MESH] OR "Health Records, Personal"[MESH] OR "Hospital Information Systems"[MESH] OR "Information Systems"[MESH] OR "Medical Informatics"[MESH] OR "Medical Record Linkage"[MESH] OR "Medical Records Systems, Computerized"[MESH] OR "Nursing Informatics"[MESH] OR "Public Health Informatics"[MESH] OR "Radiology Information Systems"[MESH] OR "Reminder Systems"[MESH] OR "Telemedicine"[MESH]

Results of Analysis

From 6/1/23→5/31/24 this query gave 39,154 entries

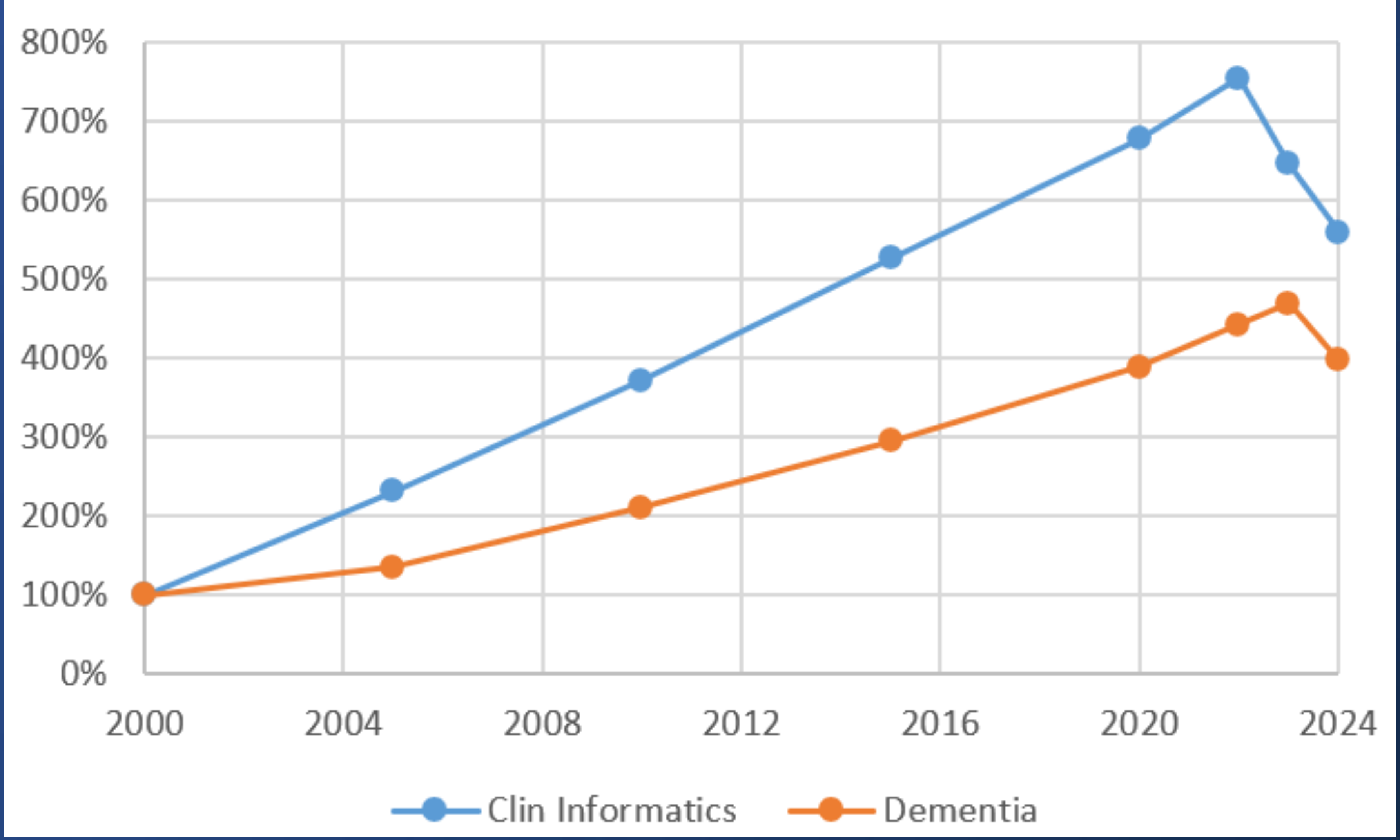
For the rest of the analysis, only titles with English (99%) abstracts were included: 35,714

From this 35,714 entries, the trials, 716, were reviewed and a sample selected for presentation

The selection was NOT random

Growth in Publications

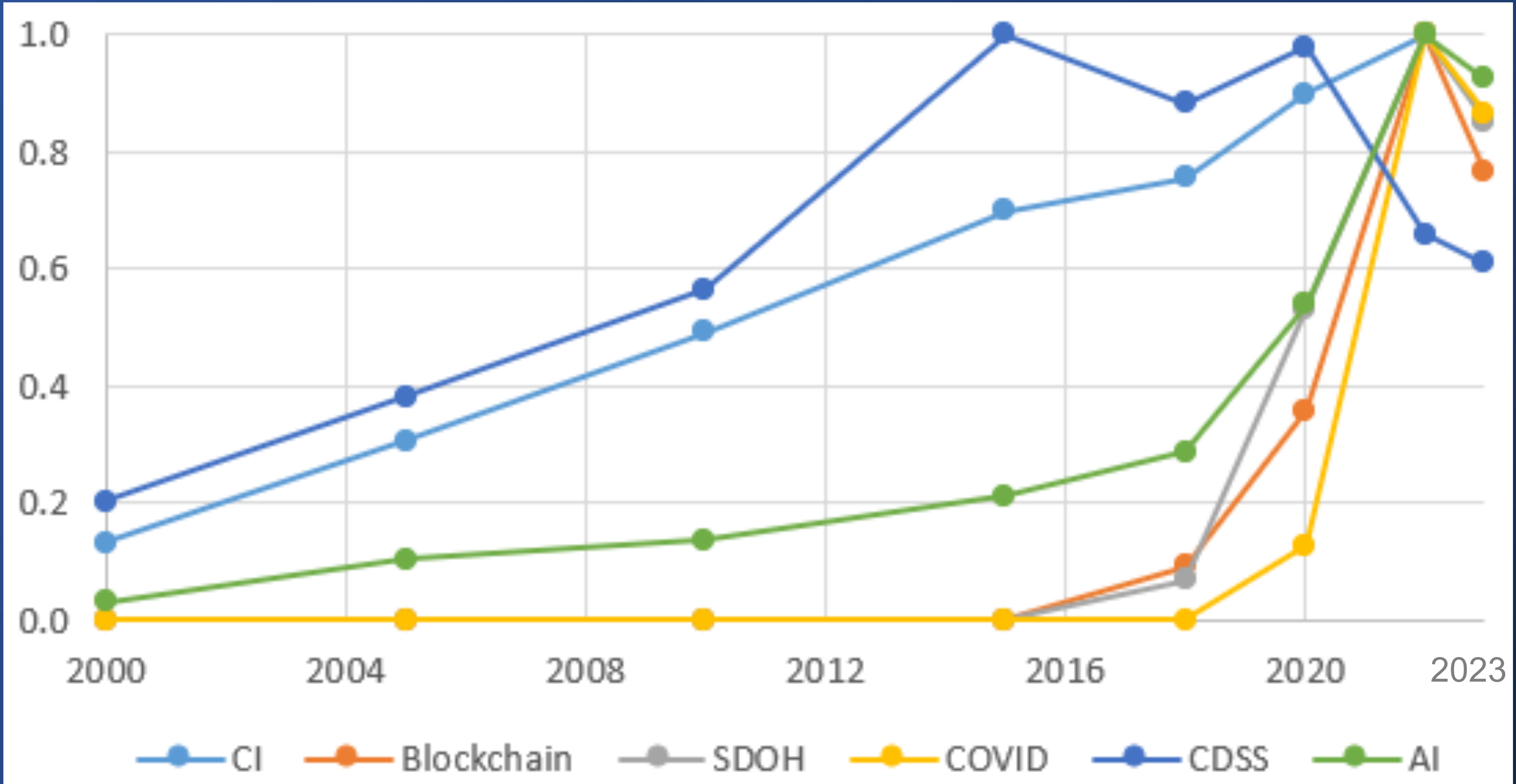
(compared to 2000, using Dementia as comparator)



35,714

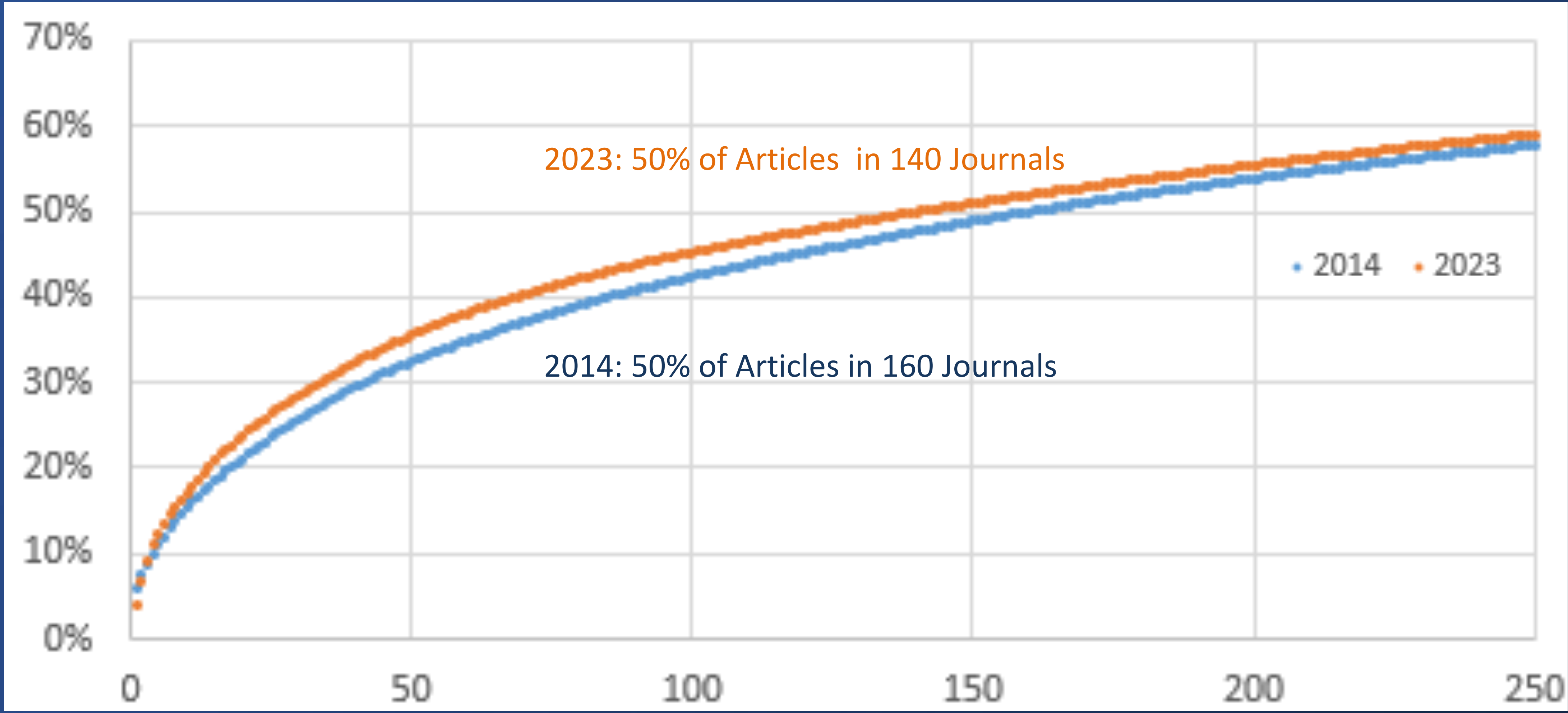
10,350

Growth in Publications



What Journals?

Proportion of Articles vs. # of Journals



# of Journals

What Journals?

PLoS One	**Sci Rep**	**Comput Biol Med**	2.1-3.5%
Sensors (Basel)	**Neural Netw**	**Stud Health Technol Inform**	1.0-1.9%
Med Phys	J Med Internet Res	Telemed J E Health	.8-1%
J Chem Inf Model	Comput Methods Programs Biomed	J Appl Clin Med Phys	0.75%
Annu Int Conf IEEE Eng Med Biol Soc	BMJ Open	Int J Med Inform	0.65%
J Robot Surg	Environ Sci Pollut Res Int	Eur Radiol	0.60%
Front Public Health	Nat Commun	Brief Bioinform	0.56%
IEEE J Biomed Health Inform	Phys Med Biol	Artif Intell Med	0.53%
Nucleic Acids Res	BMC Med Inform Decis Mak	Bioinformatics	0.50%
Med Image Anal	Medicine (Baltimore)	J Environ Manage	0.45%
Med Biol Eng Comput	Front Immunol	Int J Mol Sci	0.40%
Eur J Radiol	J Am Med Inform Assoc	J Biomed Inform	0.35%
Radiother Oncol	BMC Bioinformatics	Surg Endosc	0.32%
IEEE Trans Med Imaging	Sci Data	Sci Total Environ	0.30%
PLoS Comput Biol	BMC Health Serv Res	Acad Radiol	0.30%
J Cancer Res Clin Oncol	Chemosphere	J Digit Imaging	0.29%
JAMA Netw Open	Front Endocrinol (Lausanne)	Int J Comput Assist Radiol Surg	0.28%
Environ Monit Assess	Magn Reson Med	World Neurosurg	0.25%
J Affect Disord	Food Chem	IEEE Trans Neural Netw Learn Syst	0.23%
Phys Med	Ann Biomed Eng	J Telemed Telecare	0.22%

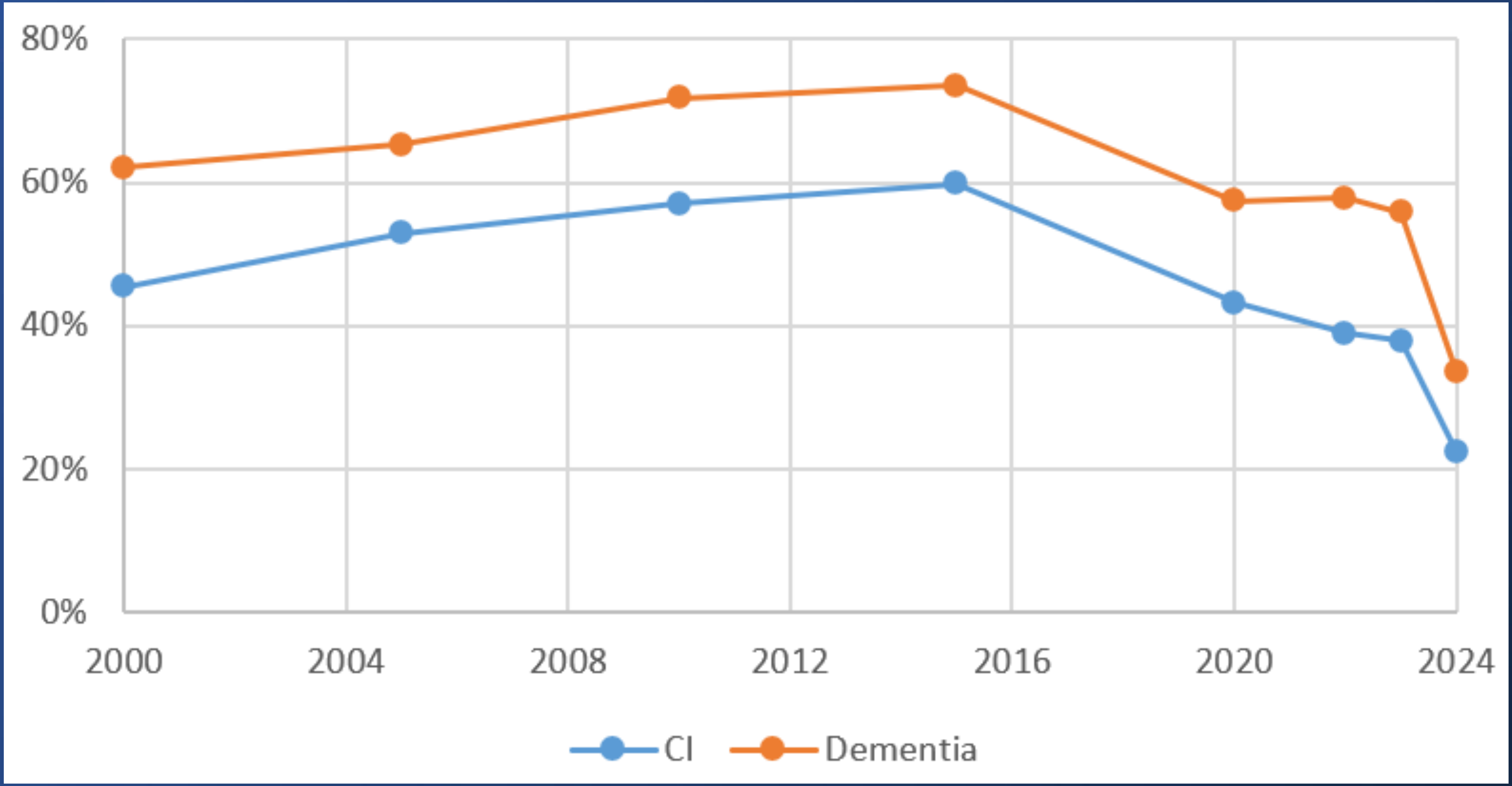
What Journals are used more or less?

2014→2024

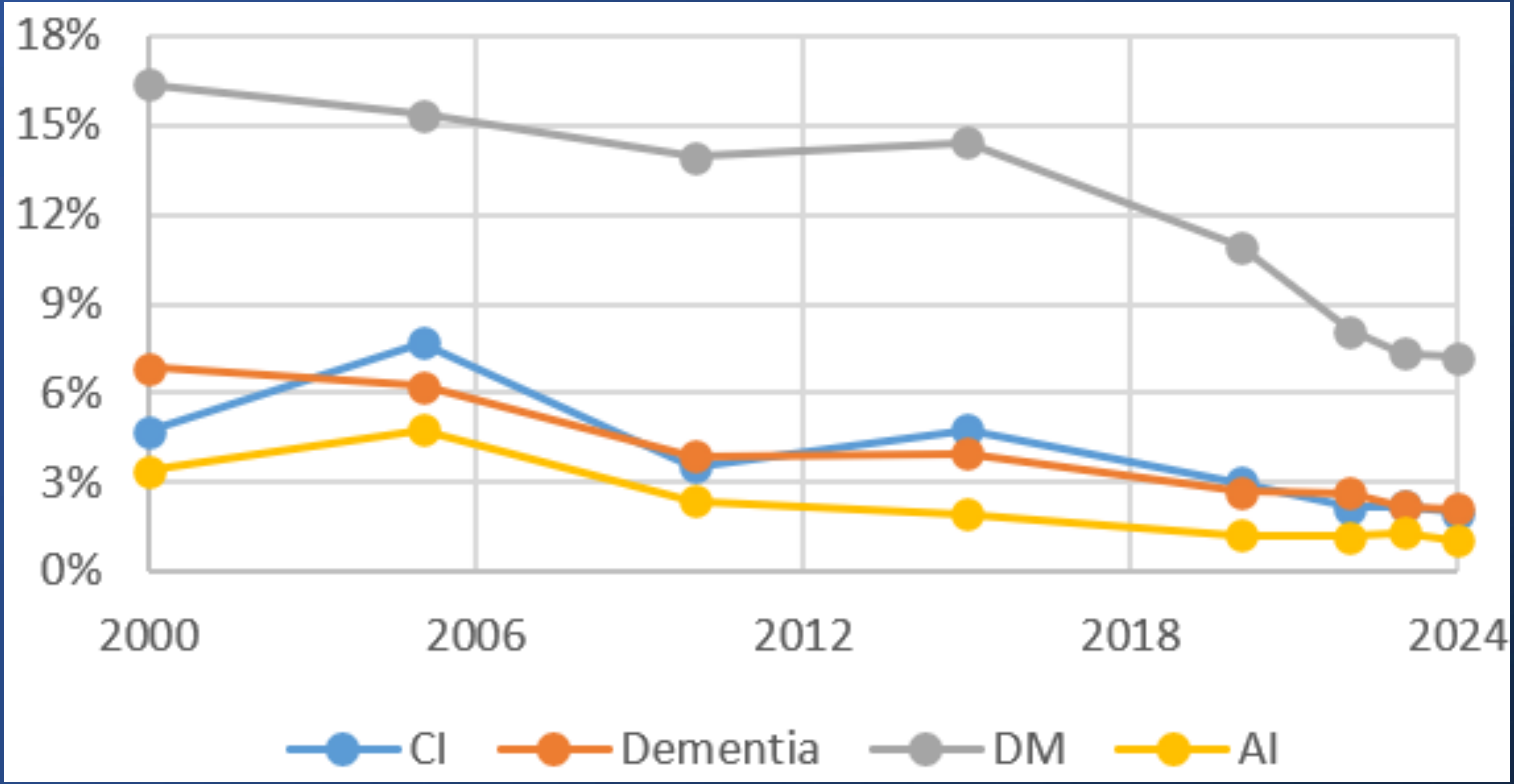
	↑Articles	AbsΔ	↓Articles	AbsΔ
*	PLoS One	2.58%	Sci Rep	-3.08%
*	Stud Health Technol Inform	1.04%	Comput Biol Med	-1.64%
*	Bioinformatics	0.73%	Sensors (Basel)	-1.54%
*	Magn Reson Med	0.59%	Neural Netw	-0.98%
	J Magn Reson Imaging	0.59%	J Robot Surg	-0.59%
*	Front Public Health (new)	0.58%	J Med Internet Res	-0.58%
	Nucleic Acids Res	0.54%	Nat Commun	-0.51%
	Database (Oxford)	0.46%	J Chem Inf Model	-0.48%
	AMIA Annu Symp Proc	0.43%	Environ Sci Pollut Res Int	-0.47%
*	Biomed Res Int	0.43%	Med Phys	-0.46%
*	Front Immunol (new)	0.39%	BMJ Open	-0.43%
	Stat Med	0.37%	Brief Bioinform	-0.42%
	Int J Radiat Oncol Biol Phys	0.37%	Eur Radiol	-0.41%
	Neuroimage	0.36%	J Appl Clin Med Phys	-0.41%
*	BMC Genomics	0.36%	Comput Methods Programs Biomed	-0.40%
	J Biomed Inform	0.32%	Medicine (Baltimore)	-0.38%
	J Am Med Inform Assoc	0.32%	Telemed J E Health	-0.38%
*	Comput Math Methods Med	0.31%	Artif Intell Med	-0.37%
*	IEEE Trans Biomed Eng	0.28%	Med Biol Eng Comput	-0.31%
*	JAMA Netw Open (New)	0.28%	J Cancer Res Clin Oncol	-0.28%

* Open Access

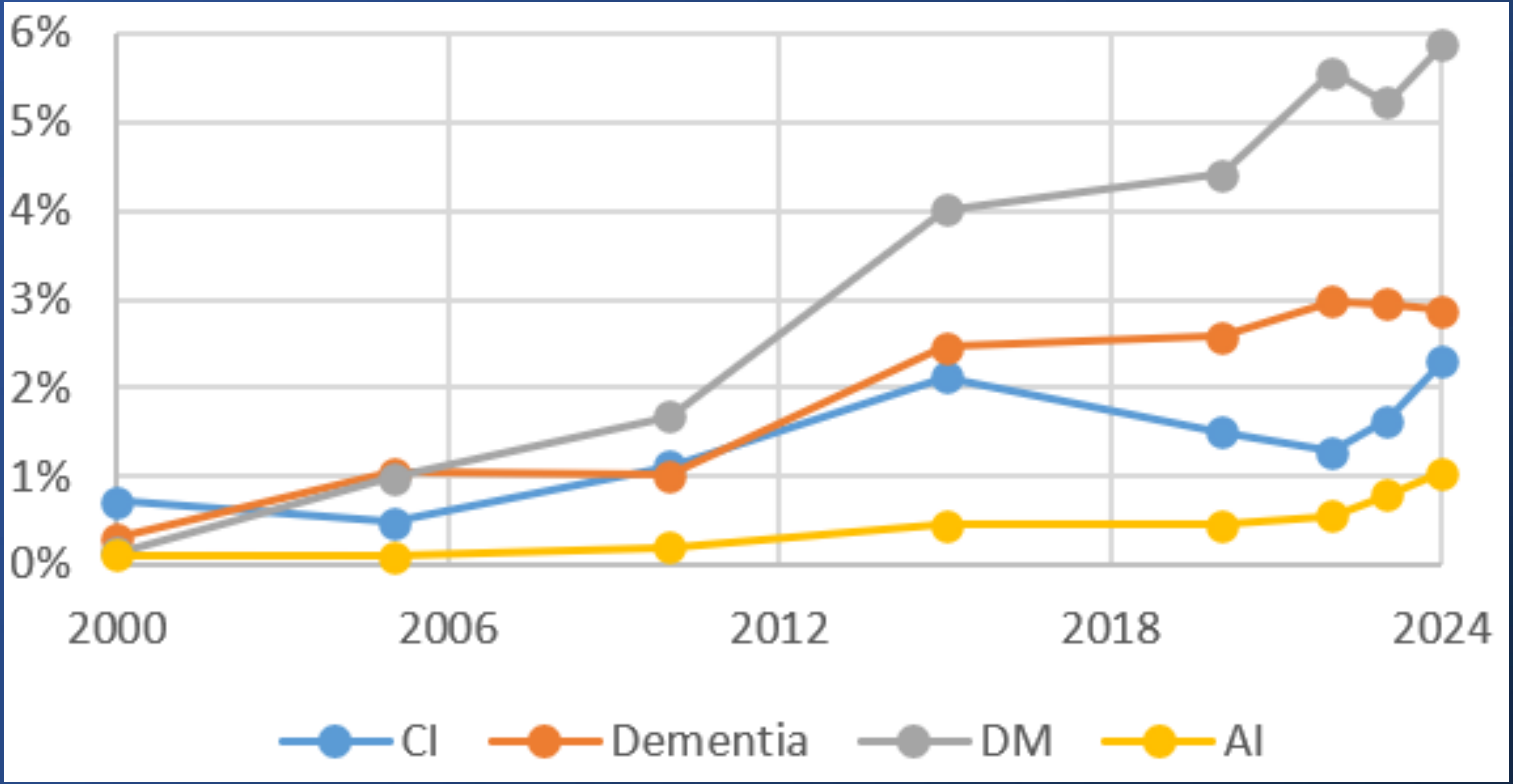
% of publications with any funding



Proportion of publications that are trials

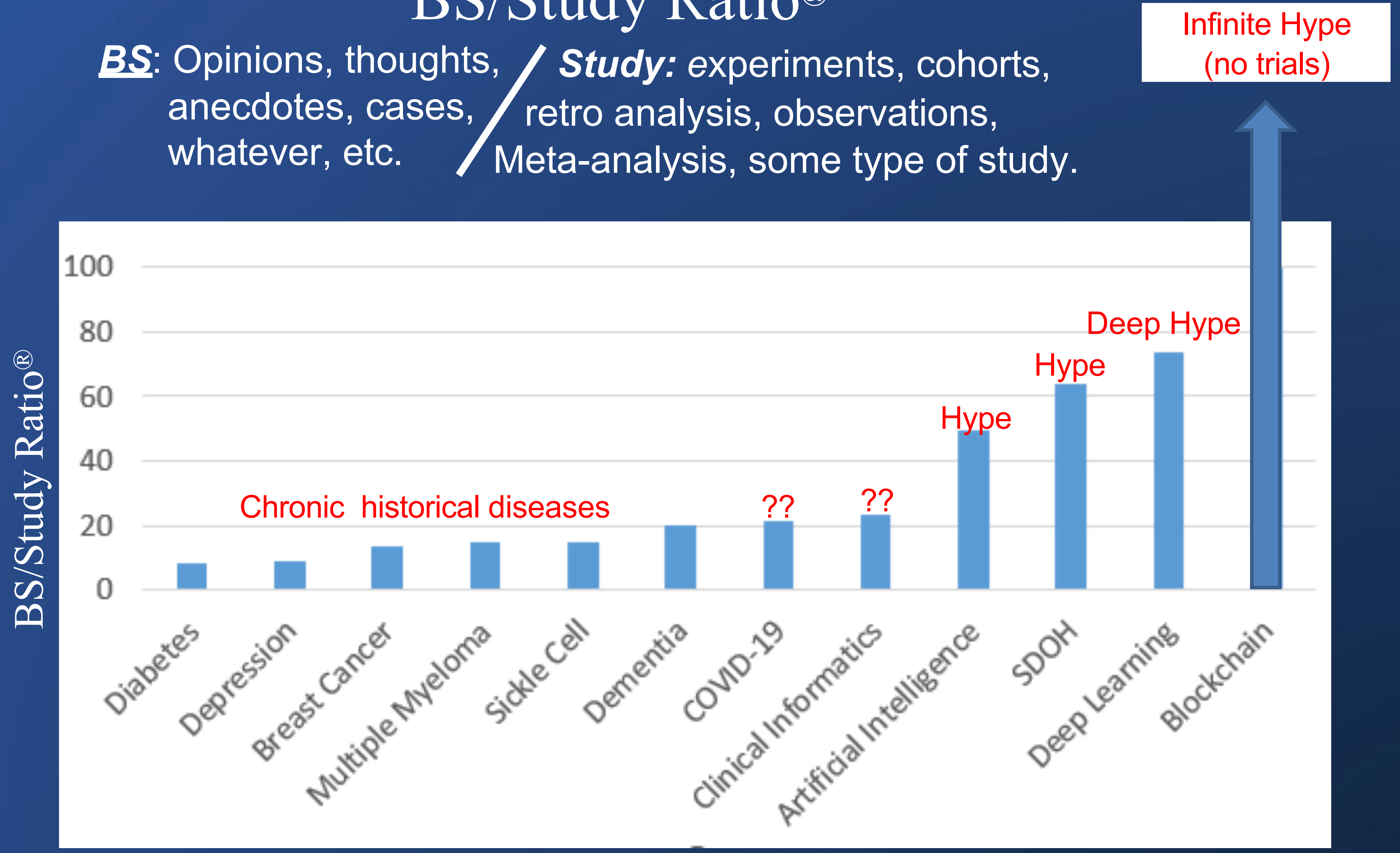


Proportion of Publications that are Meta-Analyses



BS/Study Ratio®

BS: Opinions, thoughts, anecdotes, cases, whatever, etc. / ***Study:*** *experiments*, cohorts, retro analysis, observations, Meta-analysis, some type of study.



Text Review Methodology

Cohort

35,714 Medline results with abstracts in English

-include abstract and title
-7.5×10^6 "words“

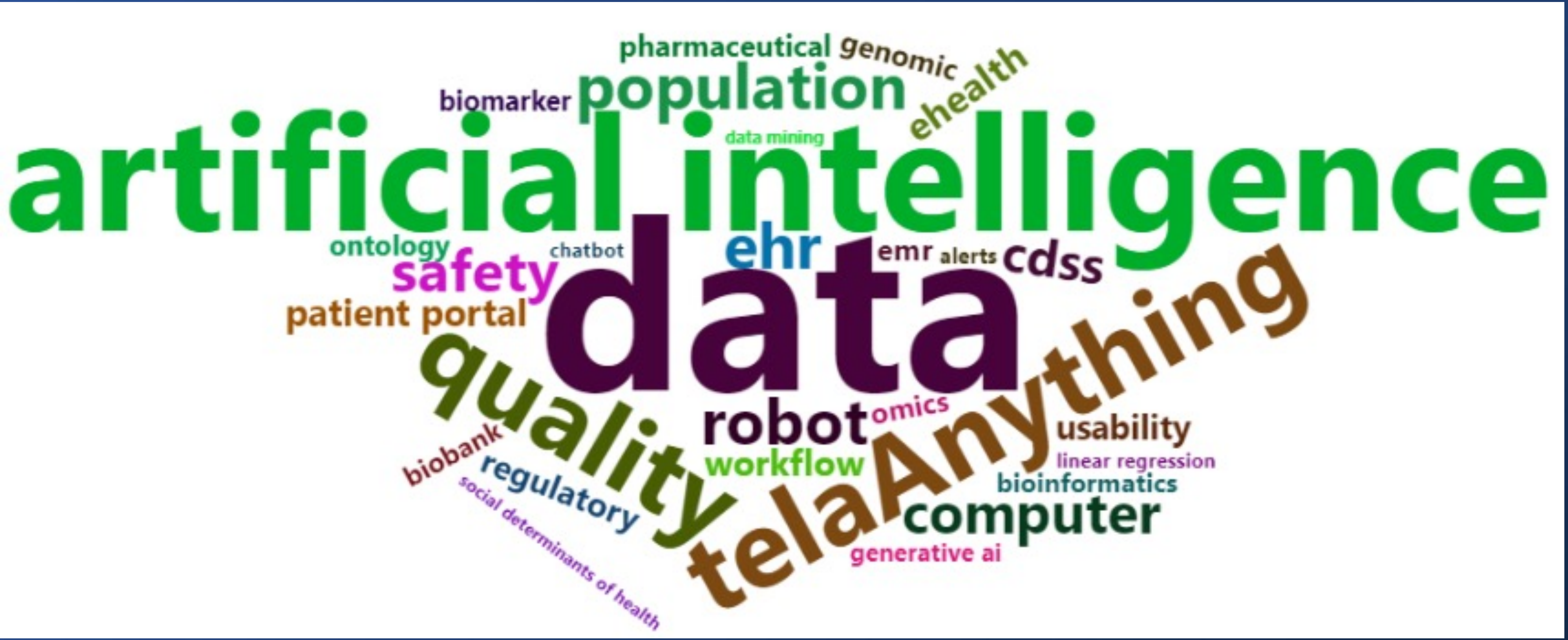
Phrases and words were counted in R with help from ChatGPT4, who again, was unable to come to the meeting this year.

All word maps were compressed to produce the freq-high/freq-low $\approx$ 10-20

Who is mentioned in the “Corpus”

What ***diseases*** are in the "Corpus"

What are *CI concepts* in the "Corpus"

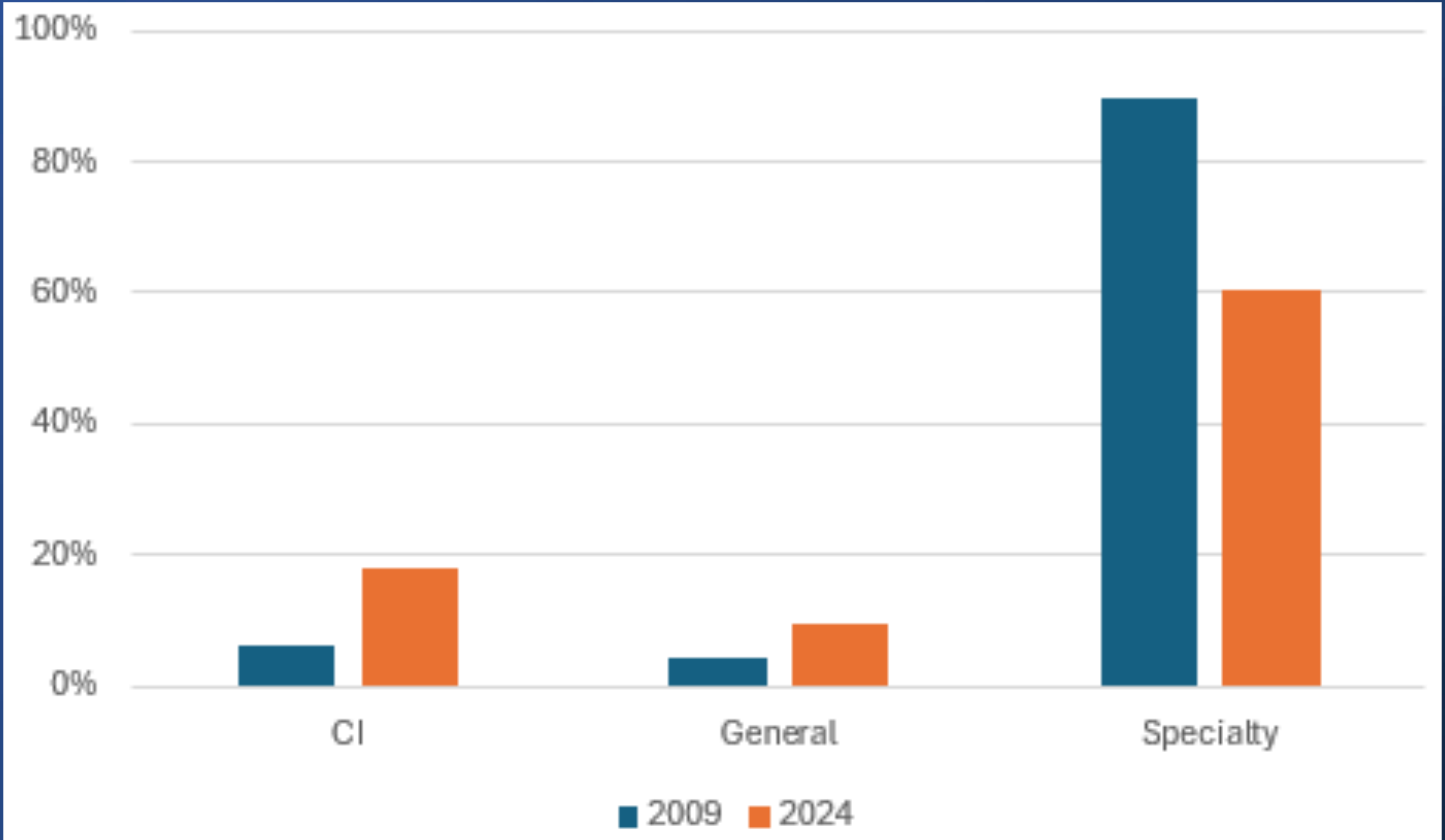


No: Blockchain, burnout, virtual reality, ROI, hie, personalized medicine

What's your favorite flavor of "tele"?

What a*nalytic concepts/tools* are in the Corpus?

Distribution of articles by type of journal in the *trials* cohort



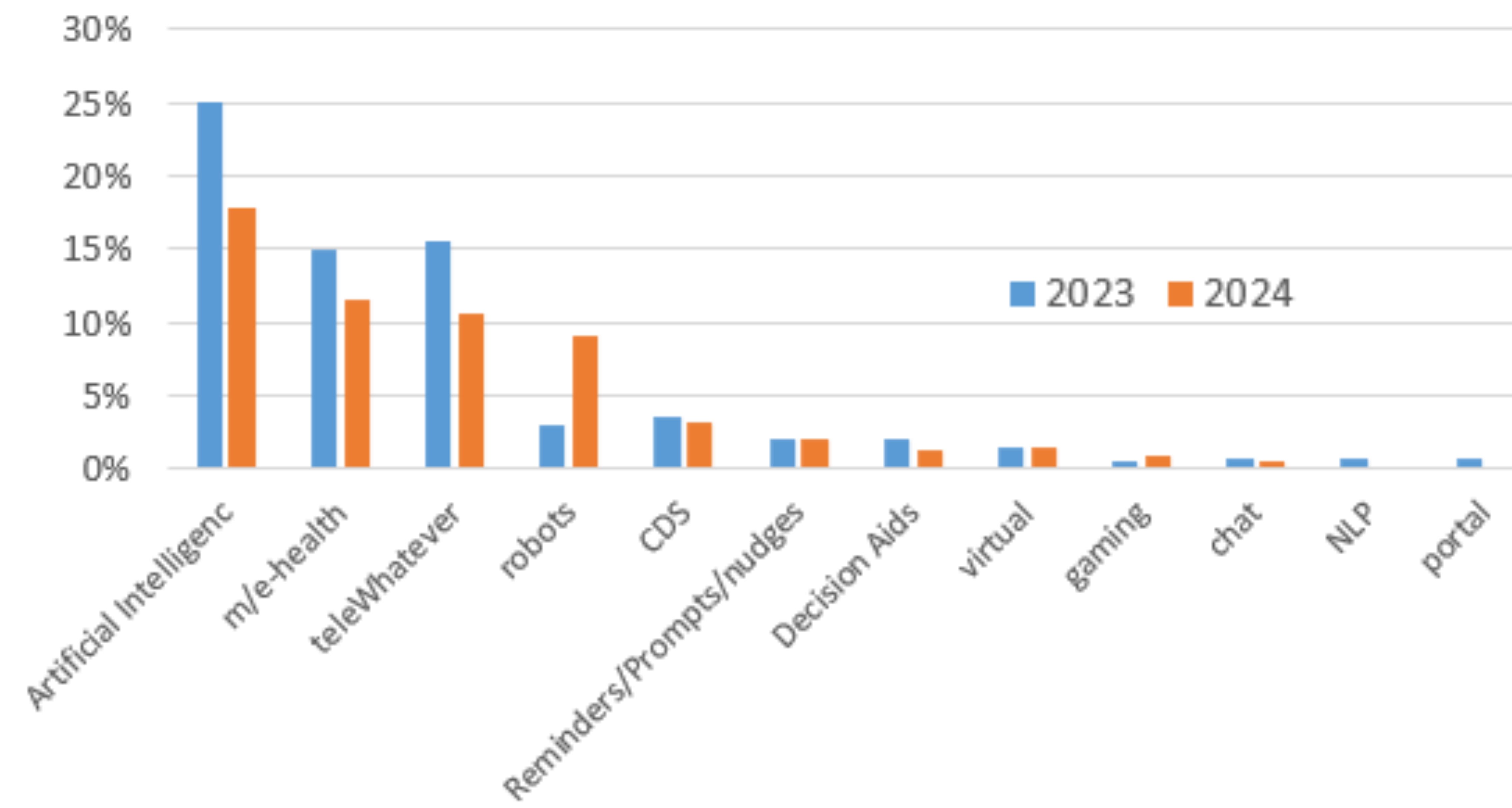
Top Informatics Journals publishing Trials

2009	2024
J Med Internet Res	J Med Internet Res
J Telemed Telecare	Telemed J E Health
J Am Med Inform Assoc	J Telemed Telecare
Int J Med Inform	JMIR Mhealth Uhealth
AMIA Annu Symp Proc	Lancet Digit Health
Artif Intell Med	Int J Med Inform
Stud Health Technol Inform	JMIR Res Protoc
Telemed J E Health	J Am Med Inform Assoc
Comput Methods Programs Biomed	JMIR Hum Factors
Hum Factors	Int J Med Robot
Inform Prim Care	Annu Int Conf IEEE Eng Med Biol Soc
Med Biol Eng Comput	Appl Clin Inform
Comput Inform Nurs	BMJ Health Care Inform
IEEE Trans Inf Technol Biomed	Comput Biol Med
Inform Health Soc Care	Comput Inform Nurs

Distribution of "general" journals in the trials cohort

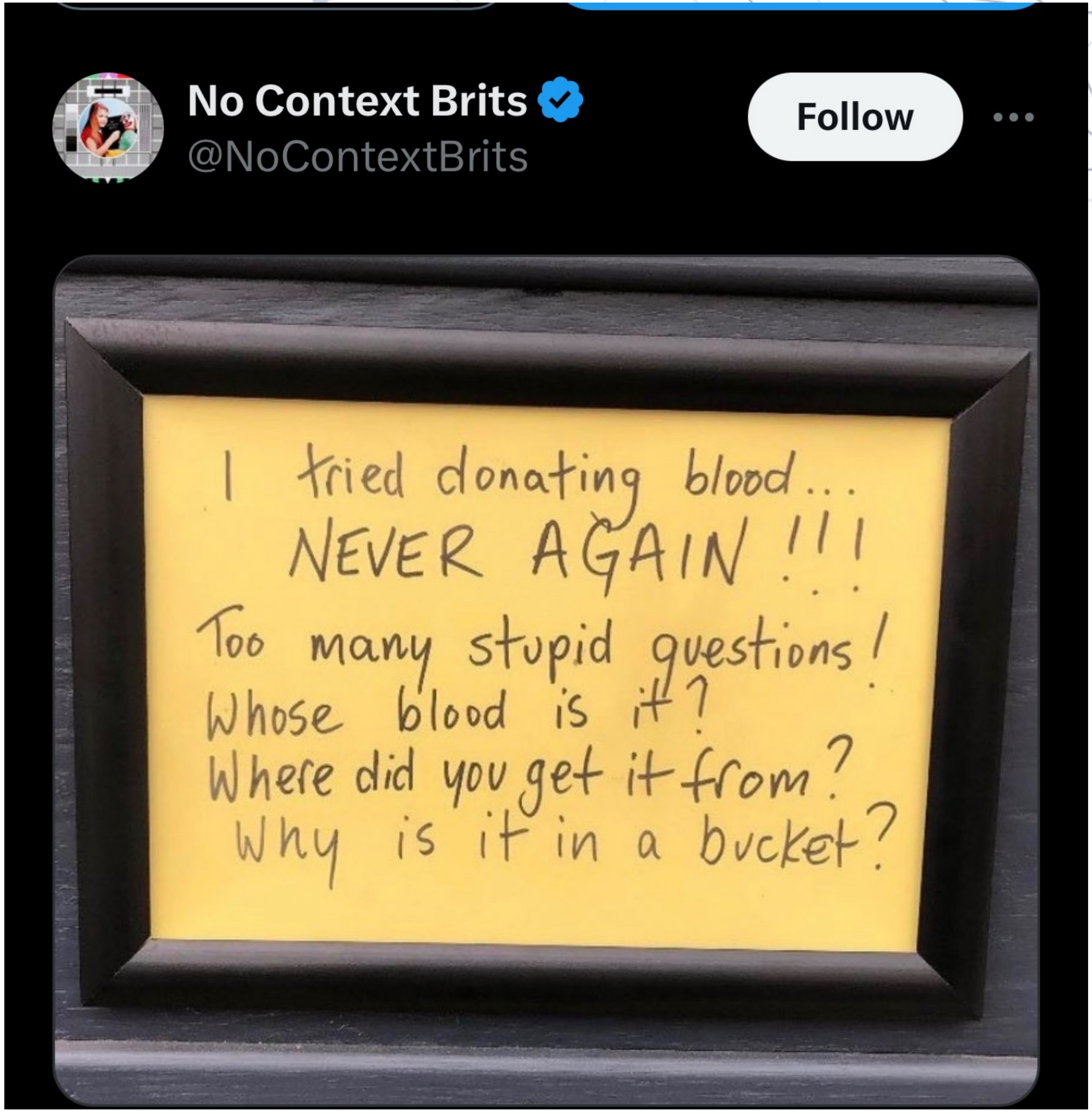
2009	2024
BMJ	PLoS One
Med Care	JAMA Netw Open
Medicina (Kaunas)	Sci Rep
BMC Med Inform Decis Mak	J Gen Intern Med
Health Expect	Front Public Health
J Med Syst	BMC Health Serv Res
Trials	BMC Med Educ
Acta Biomed	BMC Med Inform Decis Mak
Contemp Clin Trials	Eur J Med Res
Exp Biol Med (Maywood)	J Appl Clin Med Phys
Health Policy	JAMA
Isr Med Assoc J	Am J Prev Med
J Biomed Biotechnol	BMC Med
J Biopharm Stat	BMJ Open
J Med Assoc Thai	BMJ Open Qual
J Natl Med Assoc	Health Expect
JAMA	J Am Med Dir Assoc
Lancet	J Med Syst
Mayo Clin Proc	Lancet
N Engl J Med	Medicine (Baltimore)
N Z Med J	Nat Commun

Methods used in Trials



Colin's Turn !!!

The Singularity is Near Section



Chatbots Passing Exams

ChatGPT and the clinical informatics board examination: the end of unproctored maintenance of certification?

Yaa Kumah-Crystal[1,*], Scott Mankowitz[2], Peter Embi[3], and Christoph U. Lehmann[4]

Vanderbilt

Excluded questions with images

GPT 3.5 correct rate overall = 74%; no difference between categories statistically

Our (human) correct rate = ~60%

Takeaway – fire up AI when you're doing unproctored MOC !!!

Table 1. ChatGPT's performance on categories of CIBE questions

Clinical informatics category	Correct/total
Fundamental knowledge and skills	28/33 (85%)
Leadership and professionalism	52/68 (76%)
Data governance and data analytics	17/23 (74%)
Enterprise information systems	28/39 (72%)
Improving care delivery and outcomes	65/91 (71%)
Total	190/254 (75%)

ChatGPT-4 performance on USMLE Step 1 questions and its implications for medical education: A comparative study across systems and disciplines

Drexel

GPT 4

8 hours / 7 blocks / 40 questions per block

Takeaway – better than humans

System	AI Accuracy	Student Accuracy	Discipline	AI Accuracy	Student Accuracy
Behavioral Health	88%	70%	Anatomy	82%	59%
Biostats, Epidemiology/Population Health & Interpretation of Medical Literature	80%	62%	Biochemistry, Nutrition	80%	58%
Blood & Lymphoreticular System	84%	60%	Epidemiology, Biostatistics, and Medical Informatics	80%	61%
Cardiovascular System	90%	64%	Histology	90%	57%
Endocrine System	82%	61%	Microbiology & Virology	90%	55%
Gastrointestinal System	82%	60%	Molecular & Cell Biology	84%	52%
General Principles of Foundational Science	80%	50%	Pathology	84%	59%
Immune System	92%	56%	Pharmacology	92%	57%
Multisystem Processes & Disorders	84%	54%	Physiology	86%	56%
Musculoskeletal System	94%	61%	Prevention, Health Promotion	96%	73%
Nervous System & Special Senses	78%	59%			
Renal & Urinary Systems	82%	60%			
Reproductive System	96%	58%			
Respiratory System	90%	59%			
Skin & Subcutaneous Tissue	90%	59%			
Social Sciences	92%	83%			

https://doi.org/10.21203/rs.3.rs-3240108/v1

DATASETS, BENCHMARKS, AND PROTOCOLS

GPT versus Resident Physicians — A Benchmark Based on Official Board Scores

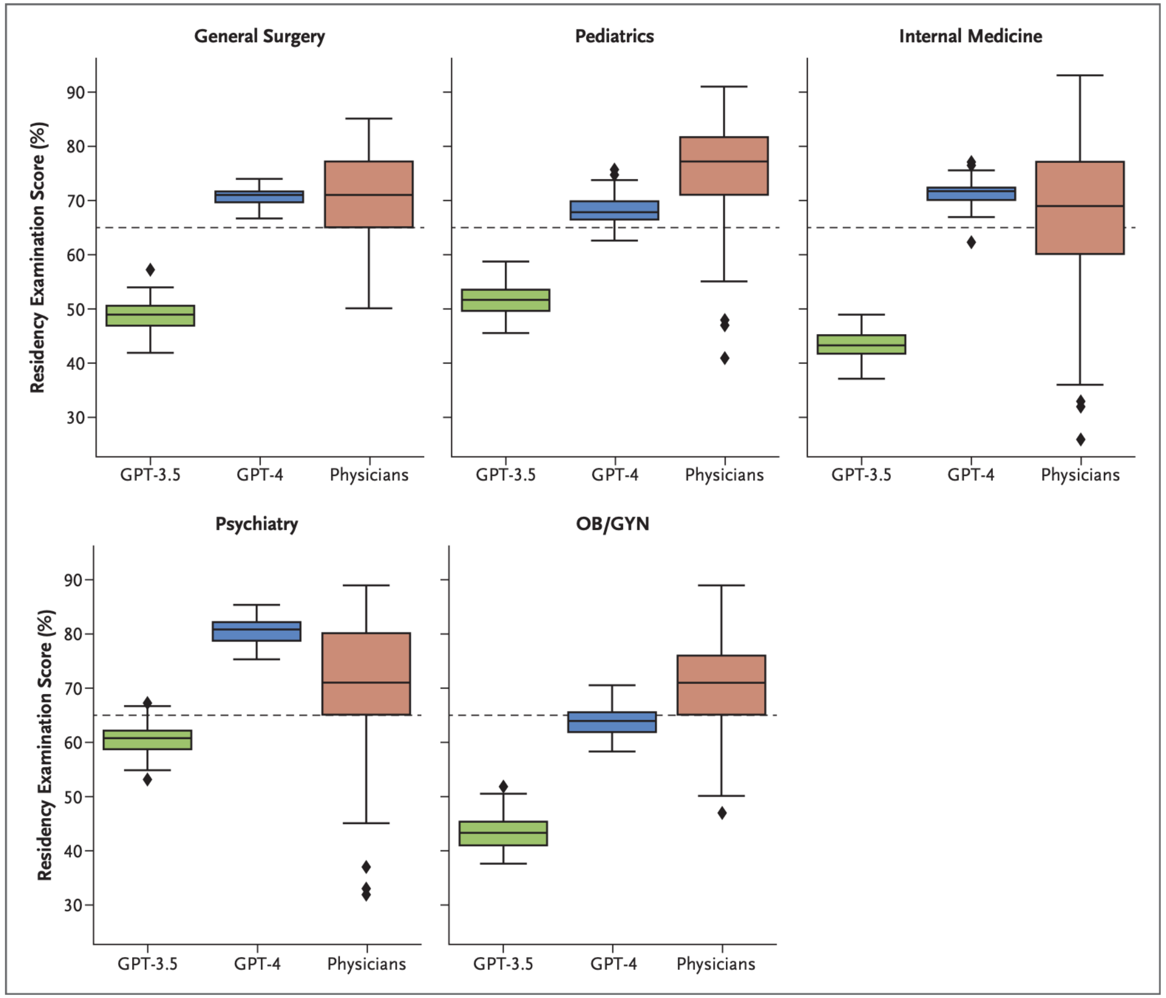
Uriel Katz, M.D.,[1] Eran Cohen, M.D.,[2,3] Eliya Shachar, M.D.,[2,4] Jonathan Somer, B.Sc.,[5] Adam Fink, M.D.,[6] Eli Morse, M.D.,[7] Beki Shreiber, B.Sc.,[8] and Ido Wolf, M.D.[2,3,4]

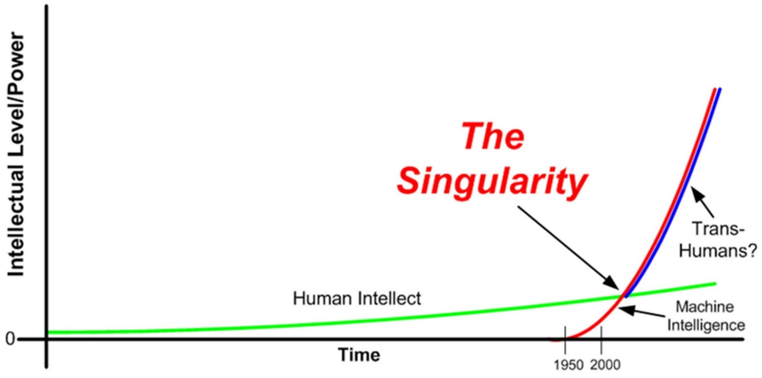
Received: October 18, 2023; Revised: January 31, 2024; Accepted: February 5, 2024; Published: April 12, 2024

Israel 2022

All residents taking specialty exams vs ChatGPT3.5 and then ChatGPT4

On par with humans and look at how the improvement is progressing





NEJM AI

DOI: 10.1056/AIdbp2300192

Chatbots and Pharmacists

BJCP British Journal of Clinical Pharmacology

ORIGINAL ARTICLE

Evaluating the performance of ChatGPT in clinical pharmacy: A comparative study of ChatGPT and clinical pharmacists

Xiaoru Huang, Dannya Estau, Xuening Liu, Yang Yu, Jiguang Qin, Zijian Li

First published: 25 August 2023 | https://doi.org/10.1111/bcp.15896 | Citations: 4

Zijian Li is the principal investigator of this study.

Contents lists available at ScienceDirect

Journal of the American Pharmacists Association

journal homepage: www.japha.org

BRIEF REPORT

Accuracy of a chatbot in answering questions that patients should ask before taking a new medication

Bernadette R. Cornelison*, Brian L. Erstad, Christopher Edwards

A Cautionary Study Reveals ChatGPT's Limitations in Providing Accurate Drug Information

Story by Glory Kaburu • 5mo • 3 min read

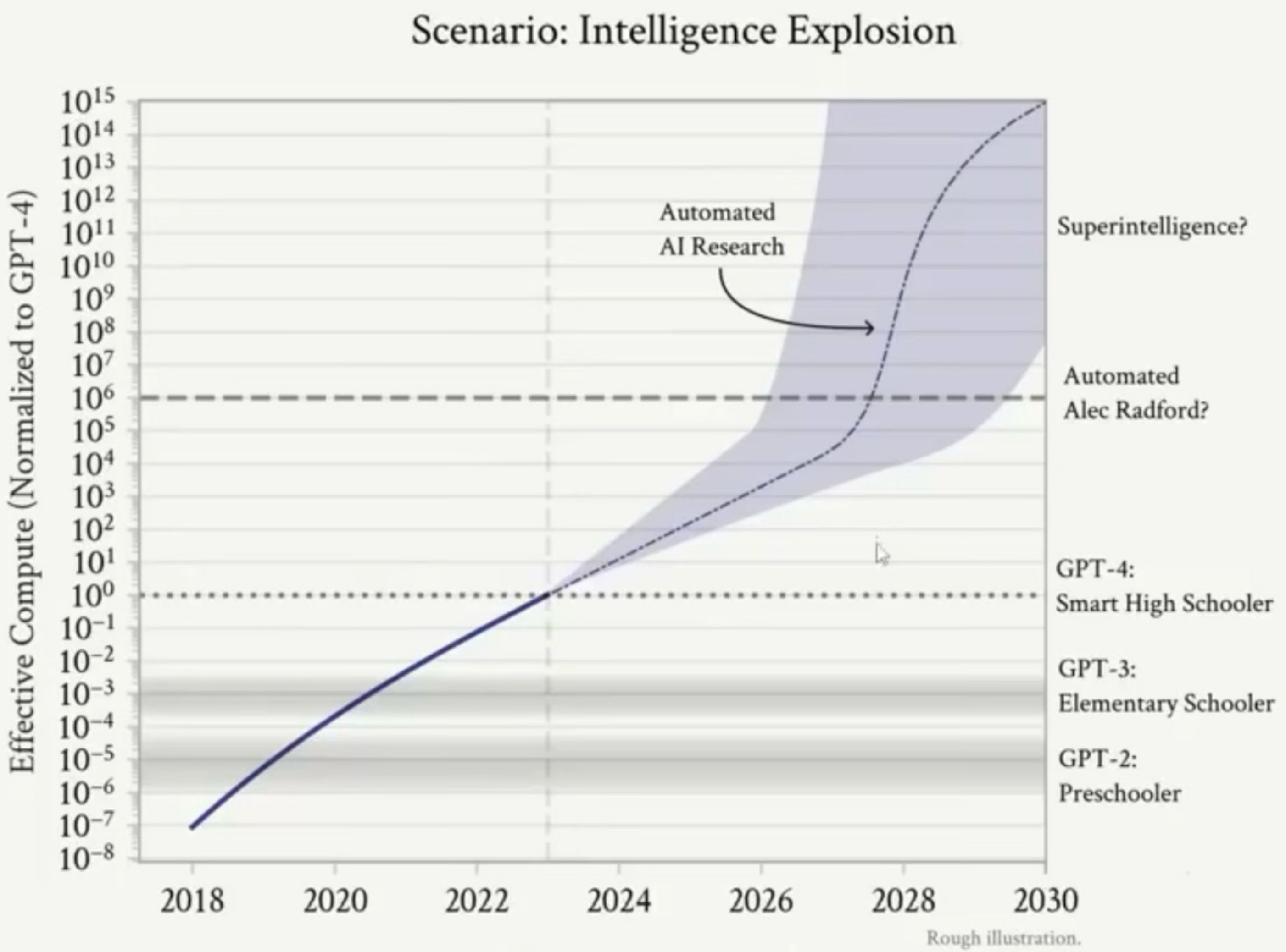
https://doi.org/10.1111/bcp.15896

Not just clinical exams...

Performance on common exams
(percentile compared to human test-takers)

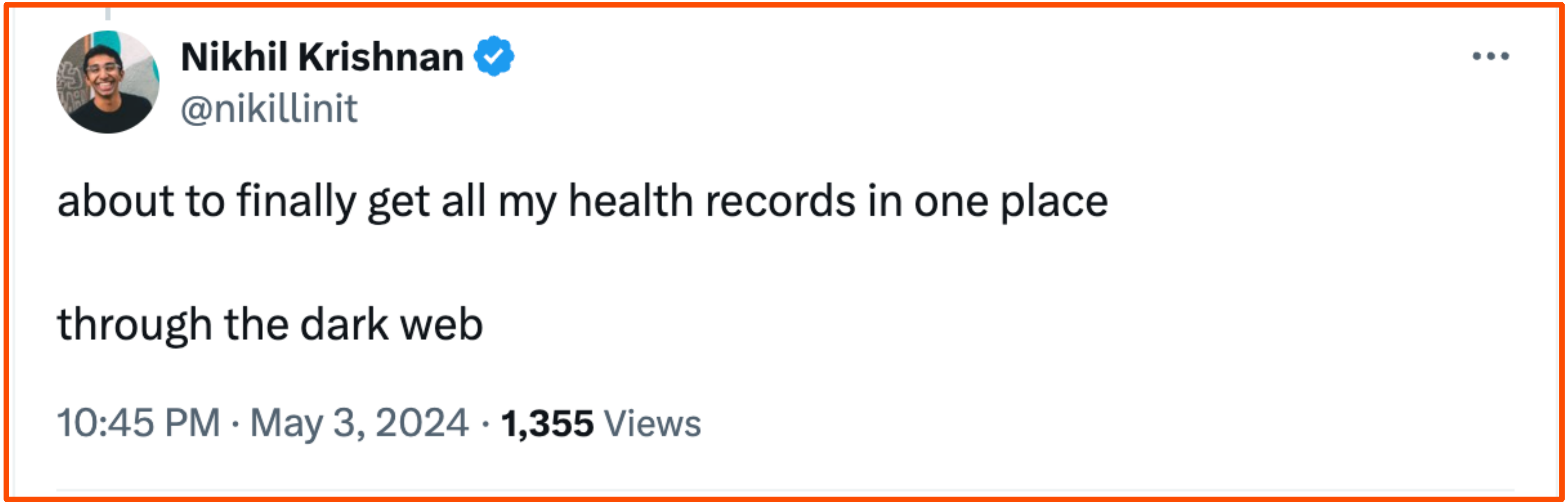
	GPT-4 (2023)	GPT-3.5 (2022)
Uniform Bar Exam	90th	10th
LSAT	88th	40th
SAT	97th	87th
GRE (Verbal)	99th	63rd
GRE (Quantitative)	80th	25th
US Biology Olympiad	99th	32nd
AP Calculus BC	51st	3rd
AP Chemistry	80th	34th
AP Macroeconomics	92nd	40th
AP Statistics	92nd	51st

SITUATIONAL AWARENESS | Leopold Aschenbrenner



SITUATIONAL AWARENESS | Leopold Aschenbrenner

Transition Comedy Slide:
Too soon? Back to you Bill !!



Nikhil Krishnan
@nikillinit

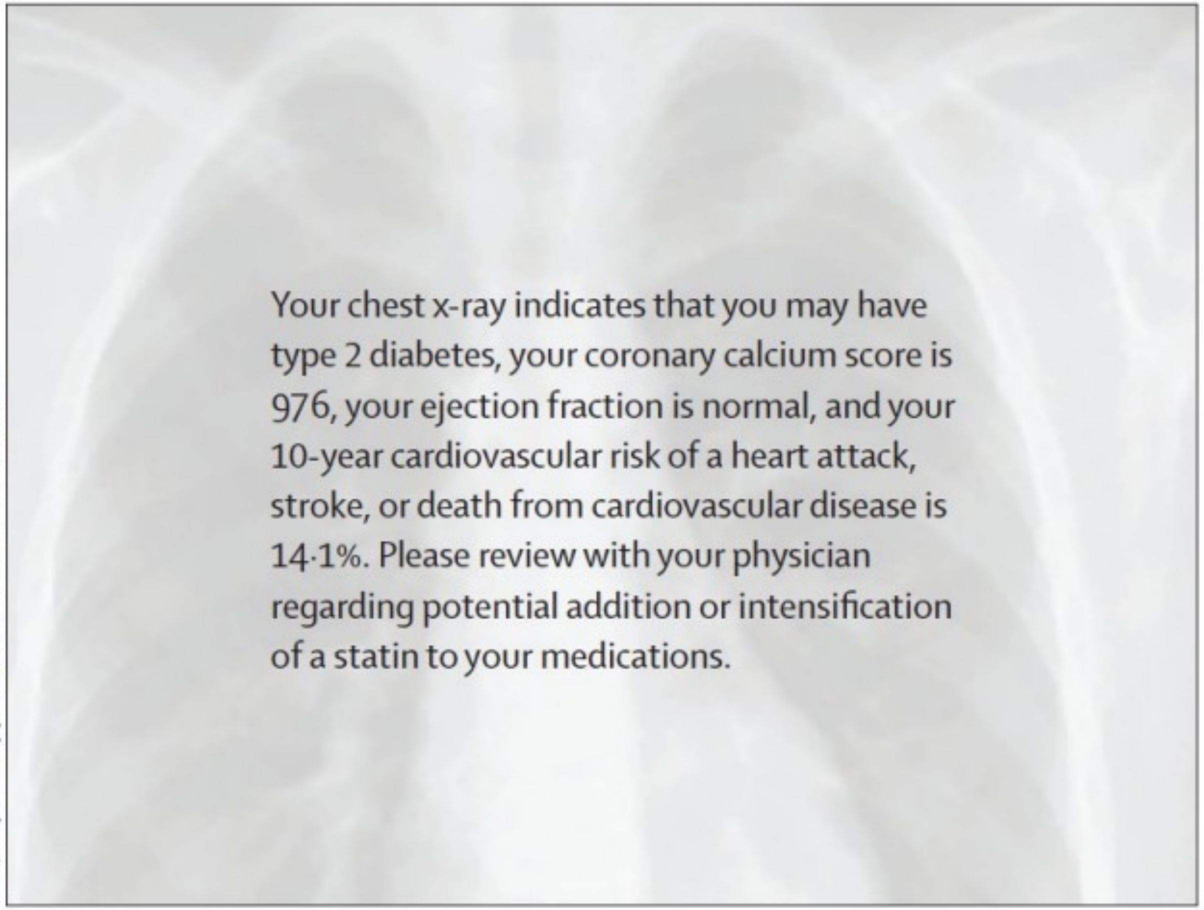
about to finally get all my health records in one place

through the dark web

10:45 PM · May 3, 2024 · **1,355** Views

Digital medicine

AI-enabled opportunistic medical scan interpretation



Eric J Topol/Scripps Research Translational Institute

Figure: **Hypothetical AI-enabled chest x-ray report**

Topol EJ. AI-enabled opportunistic medical scan interpretation. Lancet. 2024 May 11;403(10439):1842. doi: 10.1016/S0140-6736(24)00924-3. PMID: 38735291.

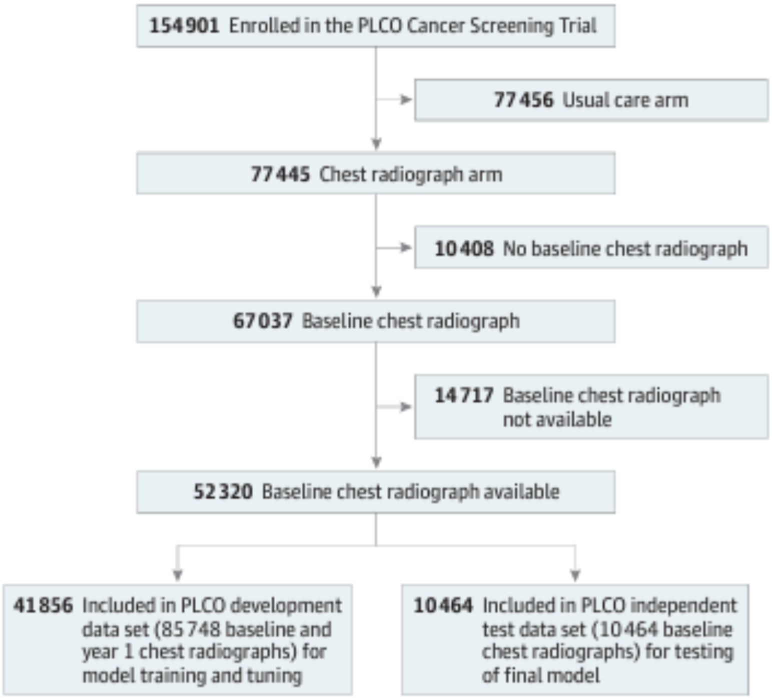
JAMA Network Open

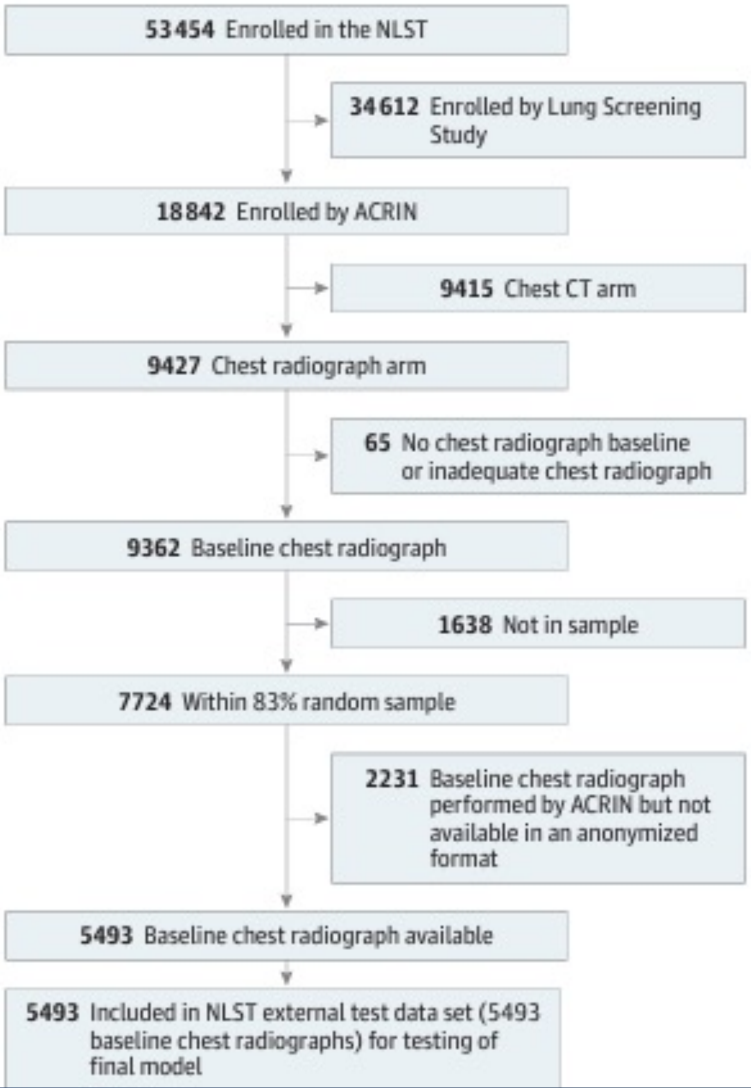
Original Investigation | Health Informatics

Deep Learning to Assess Long-term Mortality From Chest Radiographs

Michael T. Lu, MD, MPH; Alexander Ivanov, BS; Thomas Mayrhofer, PhD; Ahmed Hosny, MS; Hugo J. W. L. Aerts, PhD; Udo Hoffmann, MD, MPH

Figure 1. Data Sets for Deep Learning Model Development and Testing





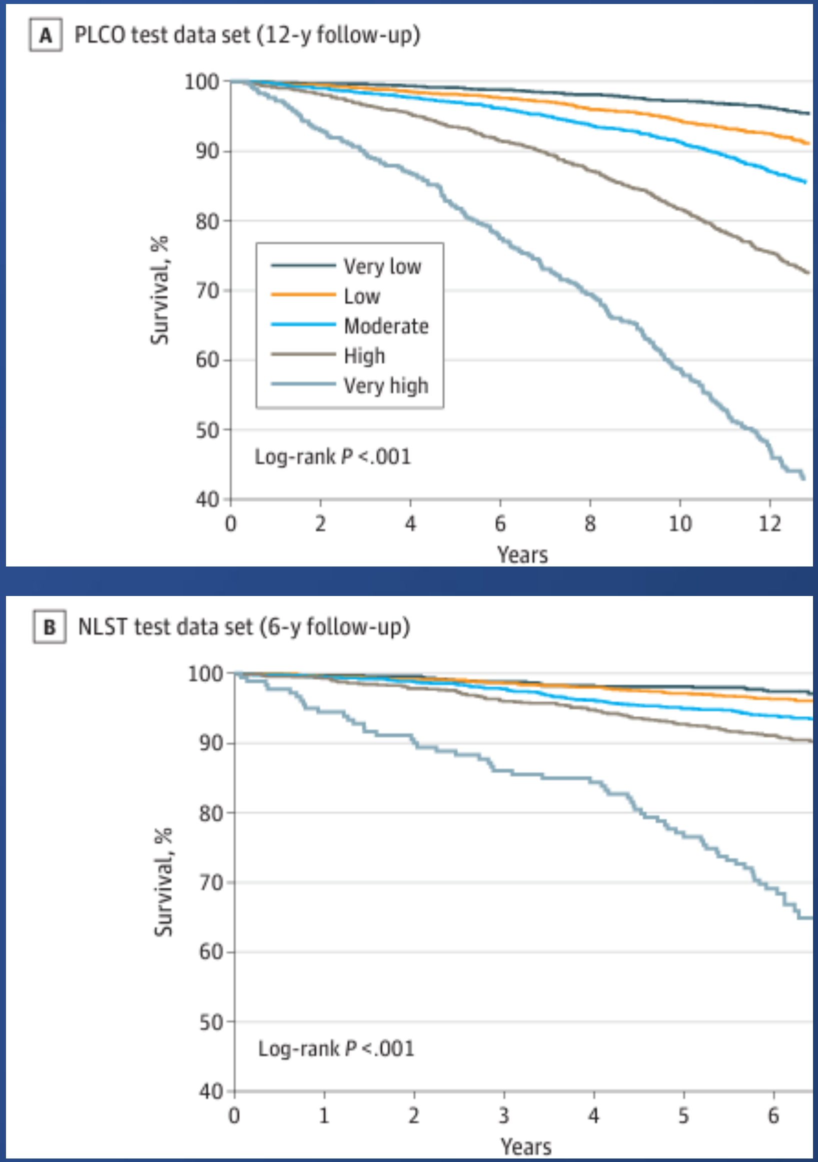
Development   Internal Test

Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO)

External Test

National Lung Screening Trial (NLST)

Lu MT, Ivanov A, Mayrhofer T, Hosny A, Aerts HJWL, Hoffmann U. Deep Learning to Assess Long-term Mortality From Chest Radiographs. JAMA Netw Open. 2019 Jul 3;2(7):e197416. doi: 10.1001/jamanetworkopen.2019.7416. PMID: 31322692; PMCID: PMC6646994.



eTable 5. Area Under the Receiver Operating Characteristic Curve (AUC) and Continuous Net Reclassification Index (NRI) for All-Cause Mortality

	PLCO Test (n=10,464, 12-year follow-up)					NLST Test (n=5,493, 6-year follow-up)				
	AUC alone (95% CI)	AUC with CXR-risk (95% CI)	P for ΔAUC	NRI (95% CI)	P for NRI	AUC alone (95% CI)	AUC with CXR-risk (95% CI)	P for ΔAUC	NRI (95% CI)	P for NRI
CXR-risk	0.75 (0.73-0.76)	NA	NA	NA	NA	0.68 (0.65-0.71)	NA	NA	NA	NA
Radiograph findings	0.58 (0.57-0.59)	0.74 (0.73-0.76)	<0.001	0.59 (0.53-0.65)	<0.001	0.59 (0.56-0.62)	0.70 (0.67-0.73)	<0.001	0.44 (0.33-0.55)	<0.001
Clinical risk factors	0.76 (0.75-0.78)	0.78 (0.77-0.79)	<0.001	0.21 (0.15-0.28)	<0.001	0.68 (0.65-0.71)	0.72 (0.69-0.75)	<0.001	0.32 (0.20-0.43)	<0.001
Risk factors + findings	0.76 (0.75-0.78)	0.78 (0.77-0.79)	<0.001	0.20 (0.13-0.27)	<0.001	0.70 (0.67-0.73)	0.73 (0.70-0.76)	<0.001	0.28 (0.17-0.41)	<0.001

Chest radiograph (CXR) findings include lung nodule, major atelectasis, pleural plaque or effusion, lymphadenopathy, chest wall or bony lesion, COPD/emphysema, cardiomegaly or other cardiovascular abnormality, and lung fibrosis. Risk factors include age, sex, smoking category, diabetes, hypertension, obesity, underweight, past myocardial infarction, past stroke, and past cancer.

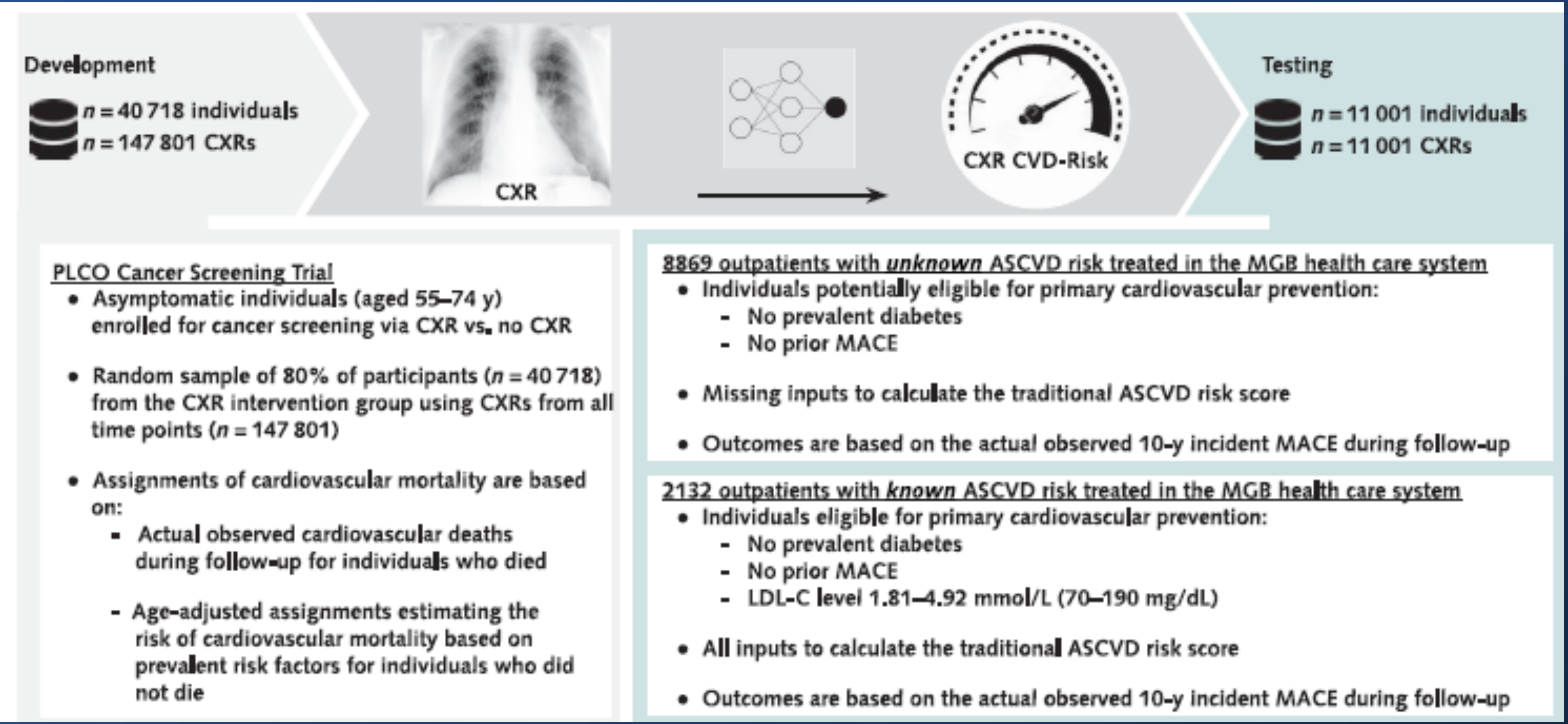
Lu MT, Ivanov A, Mayrhofer T, Hosny A, Aerts HJWL, Hoffmann U. Deep Learning to Assess Long-term Mortality From Chest Radiographs. JAMA Netw Open. 2019 Jul 3;2(7):e197416. doi: 10.1001/jamanetworkopen.2019.7416. PMID: 31322692; PMCID: PMC6646994.

Annals of Internal Medicine ORIGINAL RESEARCH

Deep Learning to Estimate Cardiovascular Risk From Chest Radiographs

A Risk Prediction Study

Jakob Weiss, MD*; Vineet K. Raghu, PhD*; Kaavya Paruchuri, MD; Aniket Zinzuwadia, AB; Pradeep Natarajan, MD, MMSc; Hugo J.W.L. Aerts, PhD; and Michael T. Lu, MD, MPH



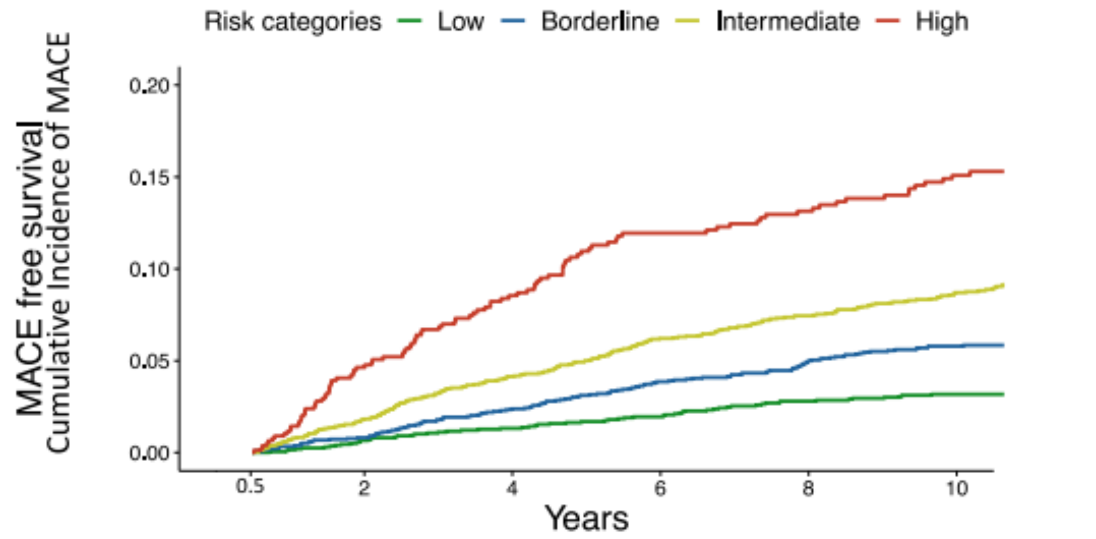
Weiss J, Raghu VK, Paruchuri K, Zinzuwadia A, Natarajan P, Aerts HJWL, Lu MT. Deep Learning to Estimate Cardiovascular Risk From Chest Radiographs : A Risk Prediction Study. Ann Intern Med. 2024 Apr;177(4):409-417. doi: 10.7326/M23-1898. Epub 2024 Mar 26. PMID: 38527287.

Annals of Internal Medicine ORIGINAL RESEARCH

Deep Learning to Estimate Cardiovascular Risk From Chest Radiographs

A Risk Prediction Study

Cumulative incidence curves for CXR CVD-Risk in outpatients with unknown ASCVD risk, excluding individuals experiencing MACE <6 months after the CXR



Supplement Table 3: Frequencies of risk categories calculated using CXR CVD-Risk and the traditional ASCVD Risk score respectively as well as the median risk within the different risk categories and observed MACE rate

	CXR CVD-Risk			ASCVD Risk Score		
Risk categories	Number of people	Median predicted risk (IQR)	Observed MACE rate	Number of people	Median predicted risk (IQR)	Observed MACE rate
Low	29.5% (628/2,132)	4.3% (4.0-4.6)	1.9% (12)	46.8% (998/2,132)	2.6% (1.5-3.6)	2.0% (20)
Borderline	33.6% (717/2,132)	6.0% (5.5-6.8)	4.2% (30)	16.5% (352/2,132)	6.1% (5.6-6.7)	4.0% (14)
Intermediate	31.6% (673/2,132)	10.3% (8.7-13.5)	7.3% (49)	32.3% (689/2,132)	11.4% (9.3-14.5)	7.8% (54)
High	5.3% (114/2,132)	27.3% (22.7-34.7)	9.6% (11)	4.4% (93/2,132)	23.1% (21.5-27.1)	15.1% (14)
Total	100% (2,132/2,132)	6.3% (4.8-9.3)	4.8% (102/2132)	100% (2,132/2,132)	5.5% (2.7-10.1)	4.8% (102/2,132)

Low: <5%, Borderline: ≥5% and <7.5%, Intermediate ≥7.5% and < 20%, High: ≥20% risk

Unknown ASCVD Test

Known ASCVD test

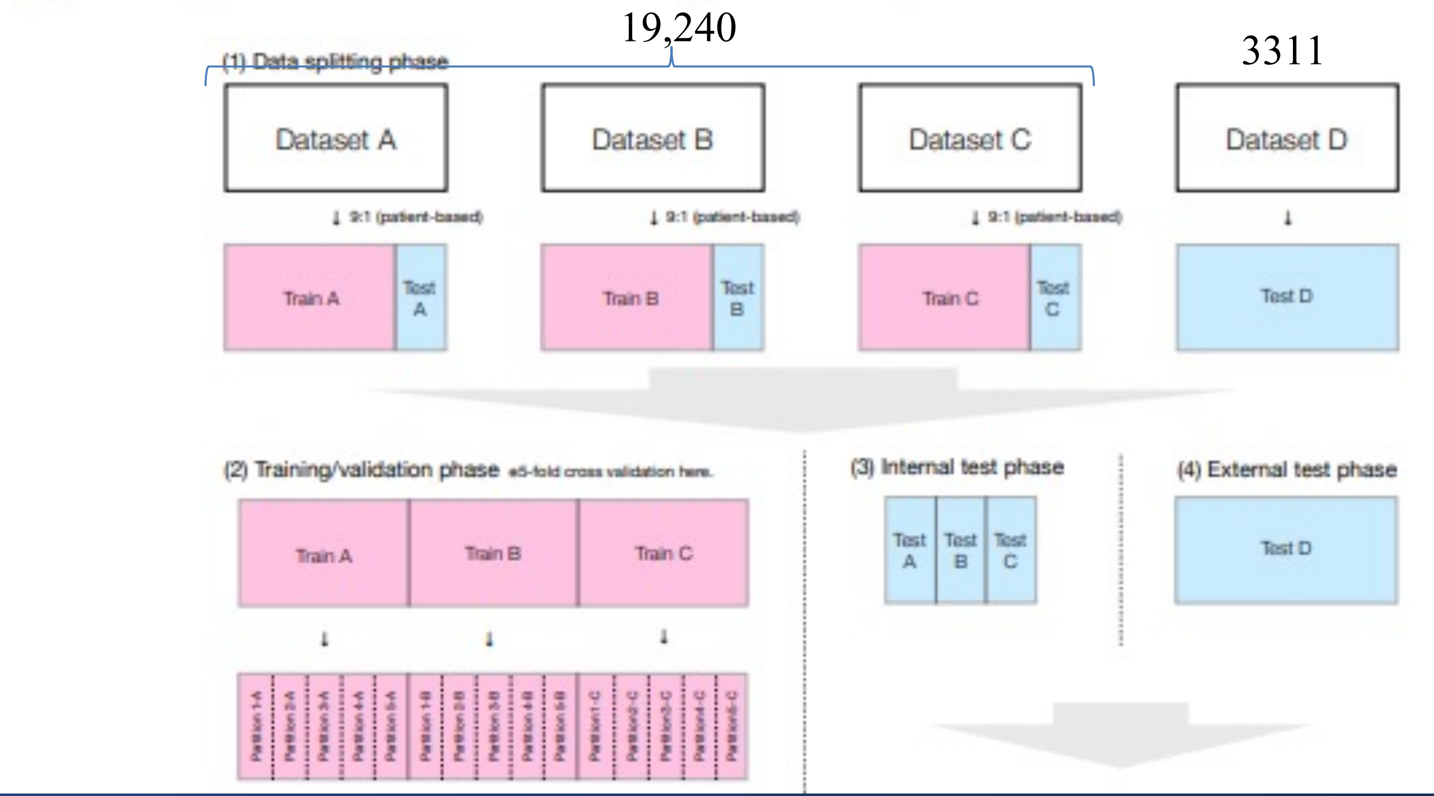
AUC’s CXR 0.67 [0.61-0.73] ASCVD 0.72 [0.66-0.78]

CXR & ASCVD 0.73 [0.67-0.79]

Weiss J, Raghu VK, Paruchuri K, Zinzuwadia A, Natarajan P, Aerts HJWL, Lu MT. Deep Learning to Estimate Cardiovascular Risk From Chest Radiographs : A Risk Prediction Study. Ann Intern Med. 2024 Apr;177(4):409-417. doi: 10.7326/M23-1898. Epub 2024 Mar 26. PMID: 38527287.

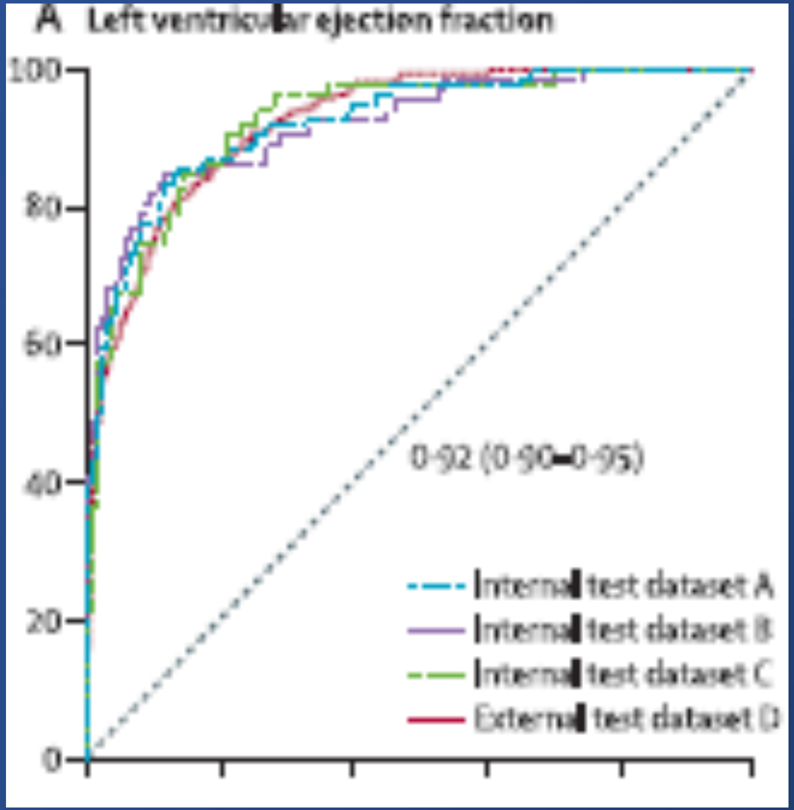
Artificial intelligence-based model to classify cardiac functions from chest radiographs: a multi-institutional, retrospective model development and validation study

Appendix Figure 1: Definitions of the training, internal test, and external test

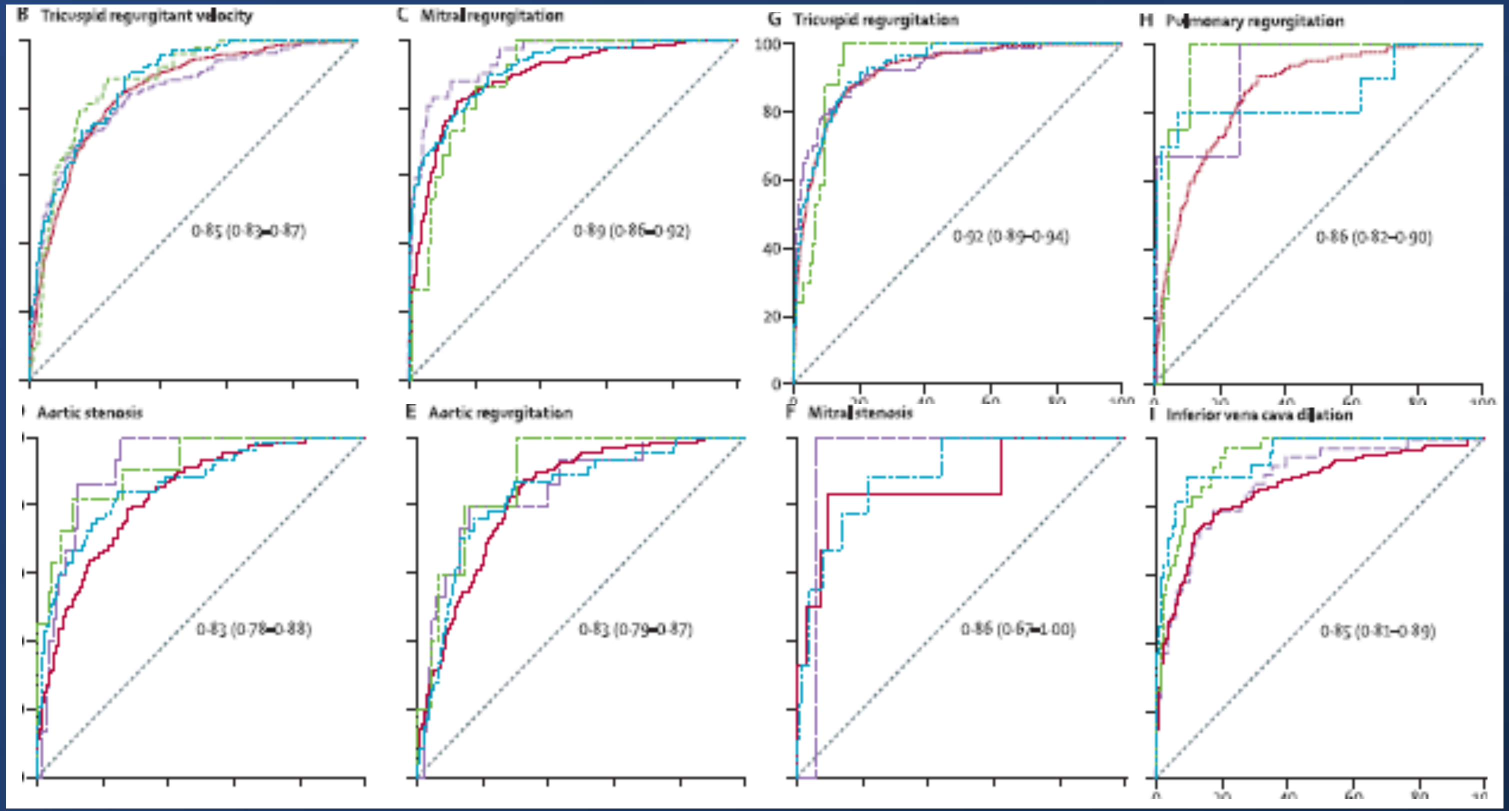


Ueda D, Matsumoto T, Ehara S, Yamamoto A, Walston SL, Ito A, Shimono T, Shiba M, Takeshita T, Fukuda D, Miki Y. Artificial intelligence-based model to classify cardiac functions from chest radiographs: a multi-institutional, retrospective model development and validation study. Lancet Digit Health. 2023 Aug;5(8):e525-e533. doi: 10.1016/S2589-7500(23)00107-3. Epub 2023 Jul 6. PMID: 37422342.

Artificial intelligence-based model to classify cardiac functions from chest radiographs: a multi-institutional, retrospective model development and validation study



LV Function < 40%



Valvular Function, IVC Dilatation

Ueda D, Matsumoto T, Ehara S, Yamamoto A, Walston SL, Ito A, Shimono T, Shiba M, Takeshita T, Fukuda D, Miki Y. Artificial intelligence-based model to classify cardiac functions from chest radiographs: a multi-institutional, retrospective model development and validation study. Lancet Digit Health. 2023 Aug;5(8):e525-e533. doi: 10.1016/S2589-7500(23)00107-3. Epub 2023 Jul 6. PMID: 37422342.

nature communications

Article

https://doi.org/10.1038/s41467-023-39631-x

Opportunistic detection of type 2 diabetes using deep learning from frontal chest radiographs

Received: 14 December 2022

Accepted: 19 June 2023

Published online: 07 July 2023

Check for updates

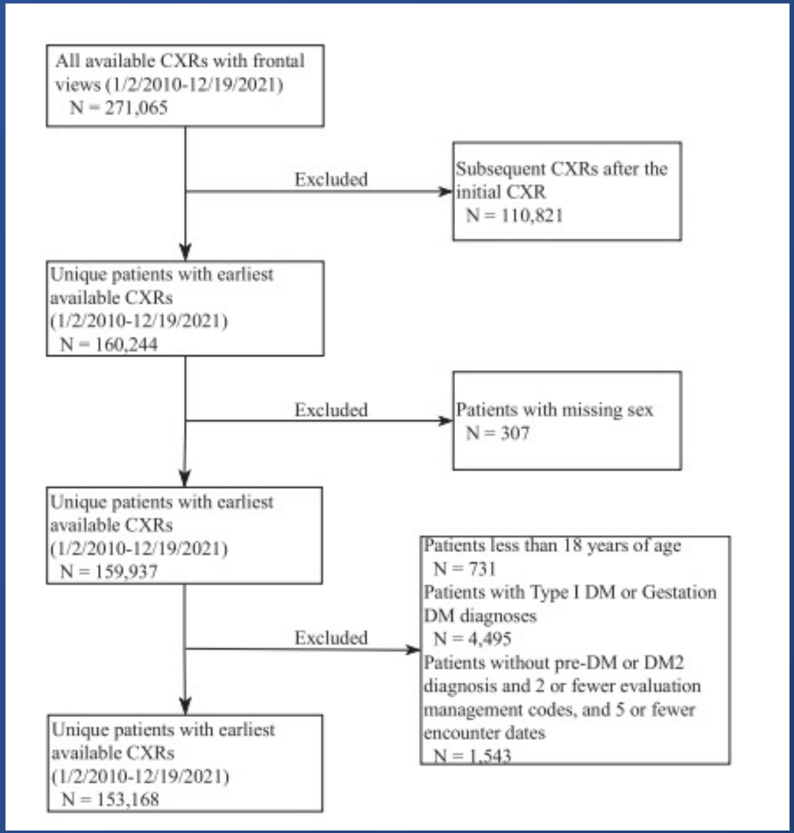
Ayis Pyrros[1,2,20], Stephen M. Borstelmann[3,20], Ramana Mantravadi[4], Zachary Zaiman[5], Kaesha Thomas[5], Brandon Price[6], Eugene Greenstein[7], Nasir Siddiqui[1], Melinda Willis[1], Ihar Shulhan[8], John Hines-Shah[1], Jeanne M. Horowitz[9], Paul Nikolaidis[9], Matthew P. Lungren[10,11,12], Jorge Mario Rodríguez-Fernández[13], Judy Wawira Gichoya[5], Sanmi Koyejo[14], Adam E Flanders[15], Nishith Khandwala[16], Amit Gupta[17], John W. Garrett[18], Joseph Paul Cohen[11], Brian T. Layden[19], Perry J. Pickhardt[18] & William Galanter[19]

Upright, PA, frontal CXR

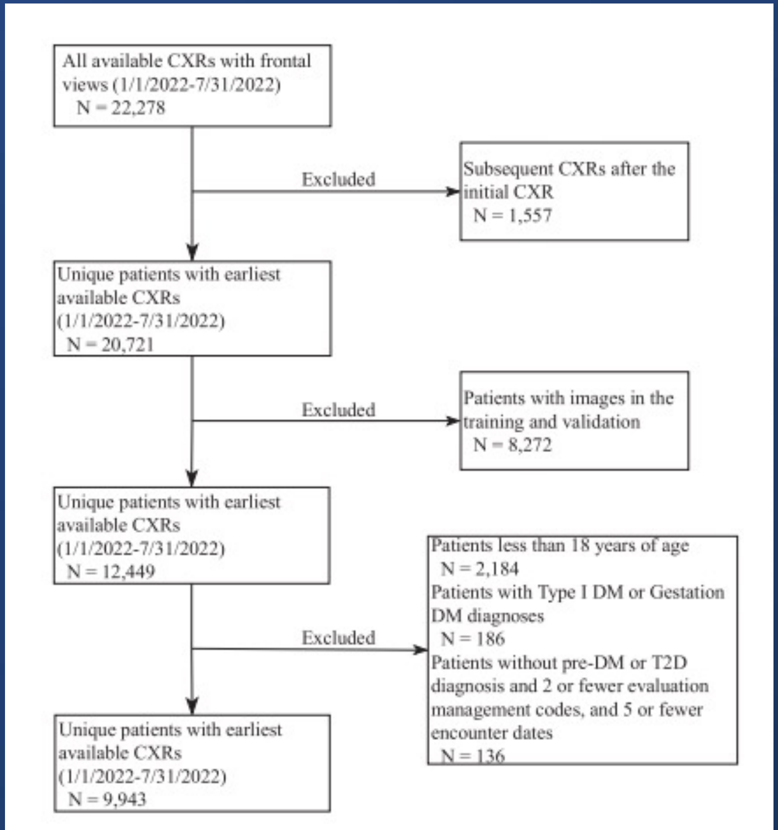
Neural Network trained on presence or absence of DM from 1/1/2010 → 12/31/2020

DM defined by ICD Codes present at any time

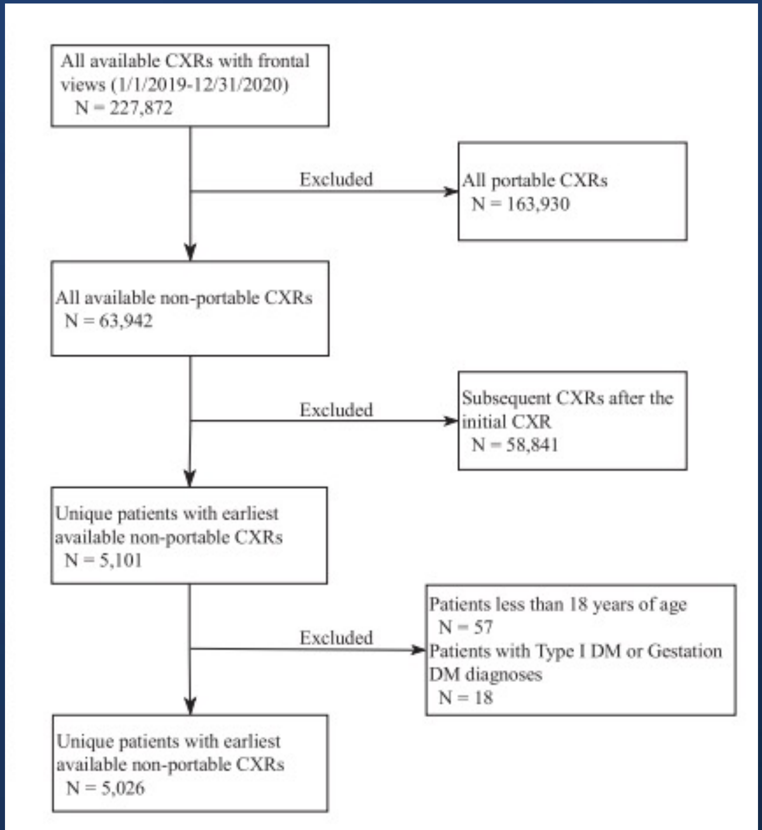
Pyrros A, Borstelmann SM, Mantravadi R, Zaiman Z, Thomas K, Price B, Greenstein E, Siddiqui N, Willis M, Shulhan I, Hines-Shah J, Horowitz JM, Nikolaidis P, Lungren MP, Rodríguez-Fernández JM, Gichoya JW, Koyejo S, Flanders AE, Khandwala N, Gupta A, Garrett JW, Cohen JP, Layden BT, Pickhardt PJ, Galanter W. Opportunistic detection of type 2 diabetes using deep learning from frontal chest radiographs. Nat Commun. 2023 Jul 7;14(1):4039. doi: 10.1038/s41467-023-39631-x. PMID: 37419921; PMCID: PMC10328953.



Development Chicago Suburban
N=153,168



Test Chicago Suburban
N = sequentially next 9,943



External Test Atlanta
N = 5,026

Pyrros A, Borstelmann SM, Mantravadi R, Zaiman Z, Thomas K, Price B, Greenstein E, Siddiqui N, Willis M, Shulhan I, Hines-Shah J, Horowitz JM, Nikolaidis P, Lungren MP, Rodríguez-Fernández JM, Gichoya JW, Koyejo S, Flanders AE, Khandwala N, Gupta A, Garrett JW, Cohen JP, Layden BT, Pickhardt PJ, Galanter W. Opportunistic detection of type 2 diabetes using deep learning from frontal chest radiographs. Nat Commun. 2023 Jul 7;14(1):4039. doi: 10.1038/s41467-023-39631-x. PMID: 37419921; PMCID: PMC10328953.

Model Score with/without DM

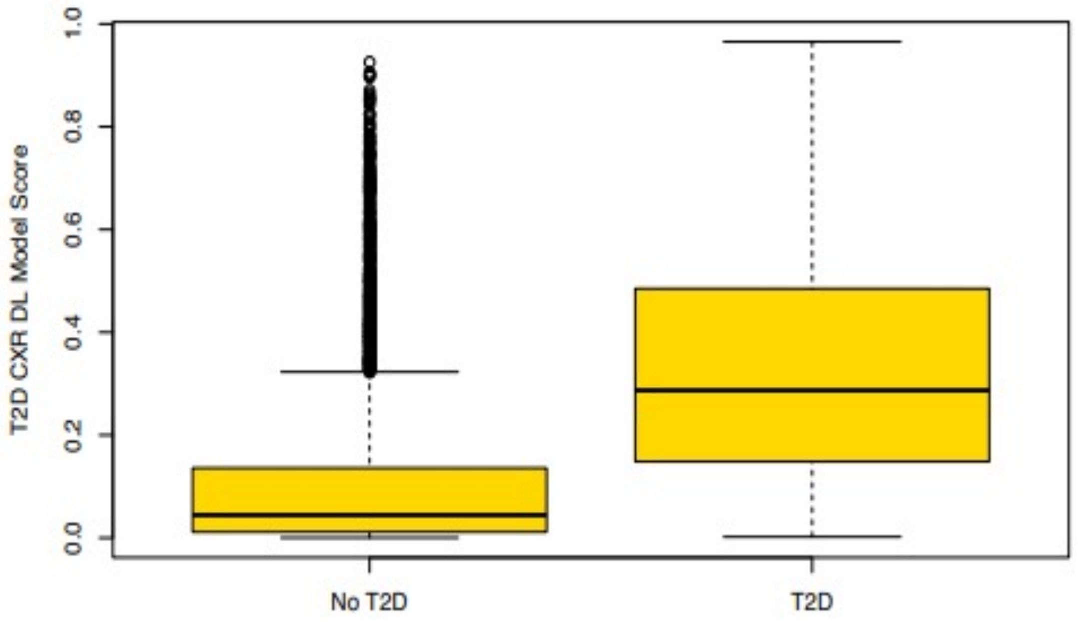


Fig. 3 | CXR DL model scores for prospective cohort positive and negative for T2D.

Model Performance
CXR-AI vs. Clinical vs. Both

Clinical logistic regression model[a]

N	N_{total}	Sens.[b]	Spec.[b]	NPV	PPV[b]	Prev. %	F1	AUC	AUC CI
1554	8126	0.79	0.65	0.94	0.3	16.1	0.43	0.79	(0.77, 0.80)

CXR deep learning model

N	N_{total}	Sens.[b]	Spec.[b]	NPV	PPV[b]	Prev. %	F1	AUC	AUC CI
1561	8382	0.85	0.68	0.96	0.33	15.7	0.48	0.84	(0.83, 0.85)

Deep learning with logistic regression model[a]

N	N_{total}	Sens.[b]	Spec.[b]	NPV	PPV[b]	Prev. %	F1	AUC	AUC CI
1,554	8,126	0.78	0.76	0.95	0.38	16.1	0.51	0.85	(0.84, 0.85)

Pyrros A, Borstelmann SM, Mantravadi R, Zaiman Z, Thomas K, Price B, Greenstein E, Siddiqui N, Willis M, Shulhan I, Hines-Shah J, Horowitz JM, Nikolaidis P, Lungren MP, Rodríguez-Fernández JM, Gichoya JW, Koyejo S, Flanders AE, Khandwala N, Gupta A, Garrett JW, Cohen JP, Layden BT, Pickhardt PJ, Galanter W. Opportunistic detection of type 2 diabetes using deep learning from frontal chest radiographs. Nat Commun. 2023 Jul 7;14(1):4039. doi: 10.1038/s41467-023-39631-x. PMID: 37419921; PMCID: PMC10328953.

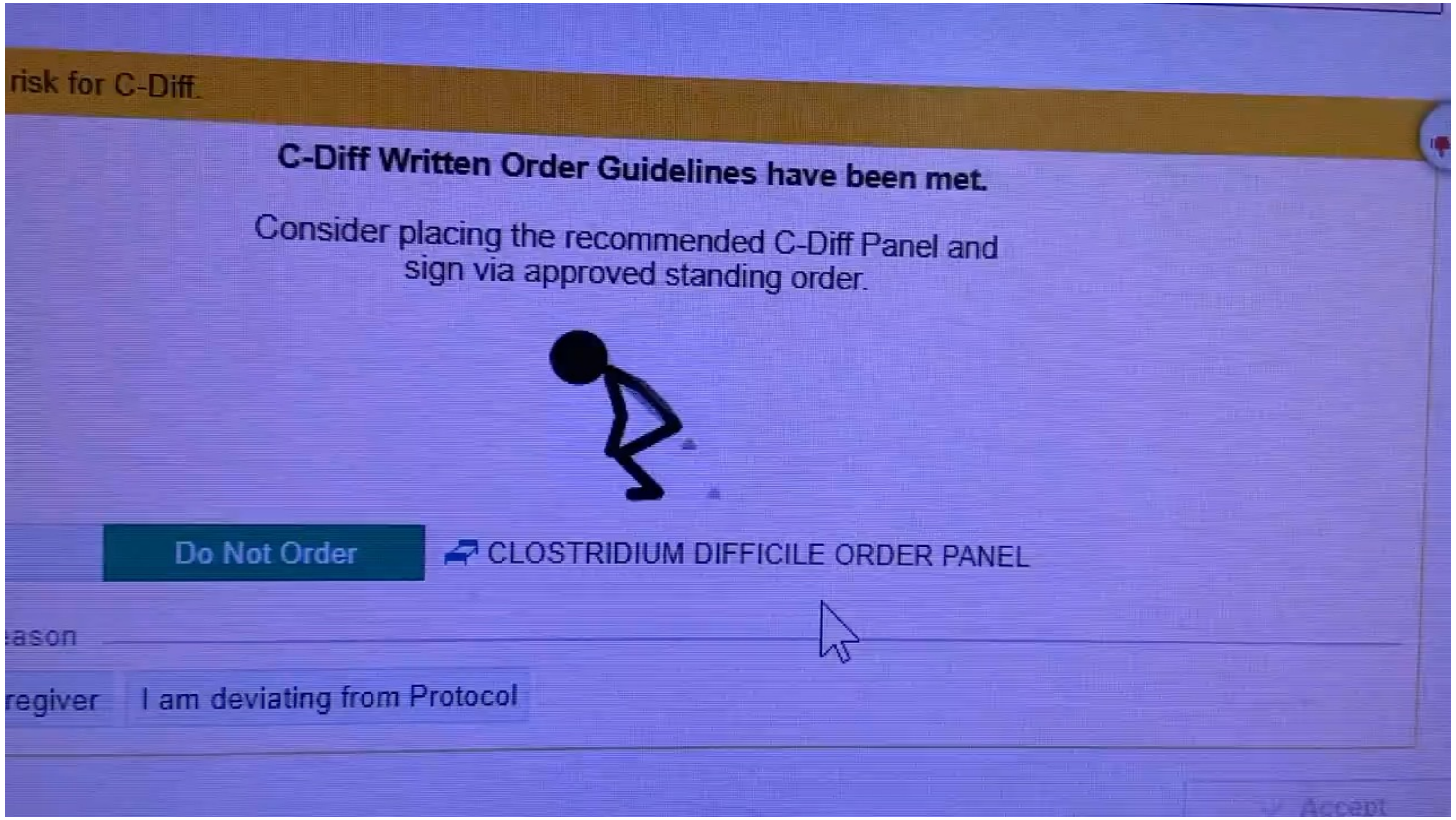
Table 3 | Area under the receiver operating characteristic curve for evaluation of model equity

Characteristic	Cases	Controls	AUC (95% CI) (Delong)	Prevalence	NPV	PPV	Sensitivity	Specificity
Sex*								
Male	817	3,485	0.83* (0.82, 0.84)	19	0.93	0.41	0.76	0.74
Female	744	4,897	0.85 (0.84, 0.86)	13	0.97	0.29	0.87	0.68
Race/Ethnicity**								
Asian	153	522	0.86 (0.83, 0.89)	23	0.95	0.49	0.86	0.74
Black	164	513	0.80 (0.77, 0.84)	24	0.89	0.47	0.72	0.74
Hispanic	153	723	0.84 (0.81, 0.87)	18	0.96	0.37	0.88	0.68
White	985	5,905	0.84 (0.83, 0.86)	14	0.97	0.3	0.87	0.66
Unknown/Other	106	719	0.84 (0.81, 0.88)	13	0.97	0.32	0.82	0.74

Pyrros A, Borstelmann SM, Mantravadi R, Zaiman Z, Thomas K, Price B, Greenstein E, Siddiqui N, Willis M, Shulhan I, Hines-Shah J, Horowitz JM, Nikolaidis P, Lungren MP, Rodríguez-Fernández JM, Gichoya JW, Koyejo S, Flanders AE, Khandwala N, Gupta A, Garrett JW, Cohen JP, Layden BT, Pickhardt PJ, Galanter W. Opportunistic detection of type 2 diabetes using deep learning from frontal chest radiographs. Nat Commun. 2023 Jul 7;14(1):4039. doi: 10.1038/s41467-023-39631-x. PMID: 37419921; PMCID: PMC10328953.

Transition Comedy Slide

Colin's CDS Section



As seen on Twitter ! https://x.com/Dr_Oubre/status/1787982558526853462

Cefepime vs Piperacillin-Tazobactam in Adults Hospitalized With Acute Infection

Vanderbilt

The Setup:

Empiric antibiotics in sepsis, Pip-Tazo causes AKI but Cefipime causes neurological dysfunction. What's an intensivitist to do?

Real-time enrollment into the study via EHR randomization

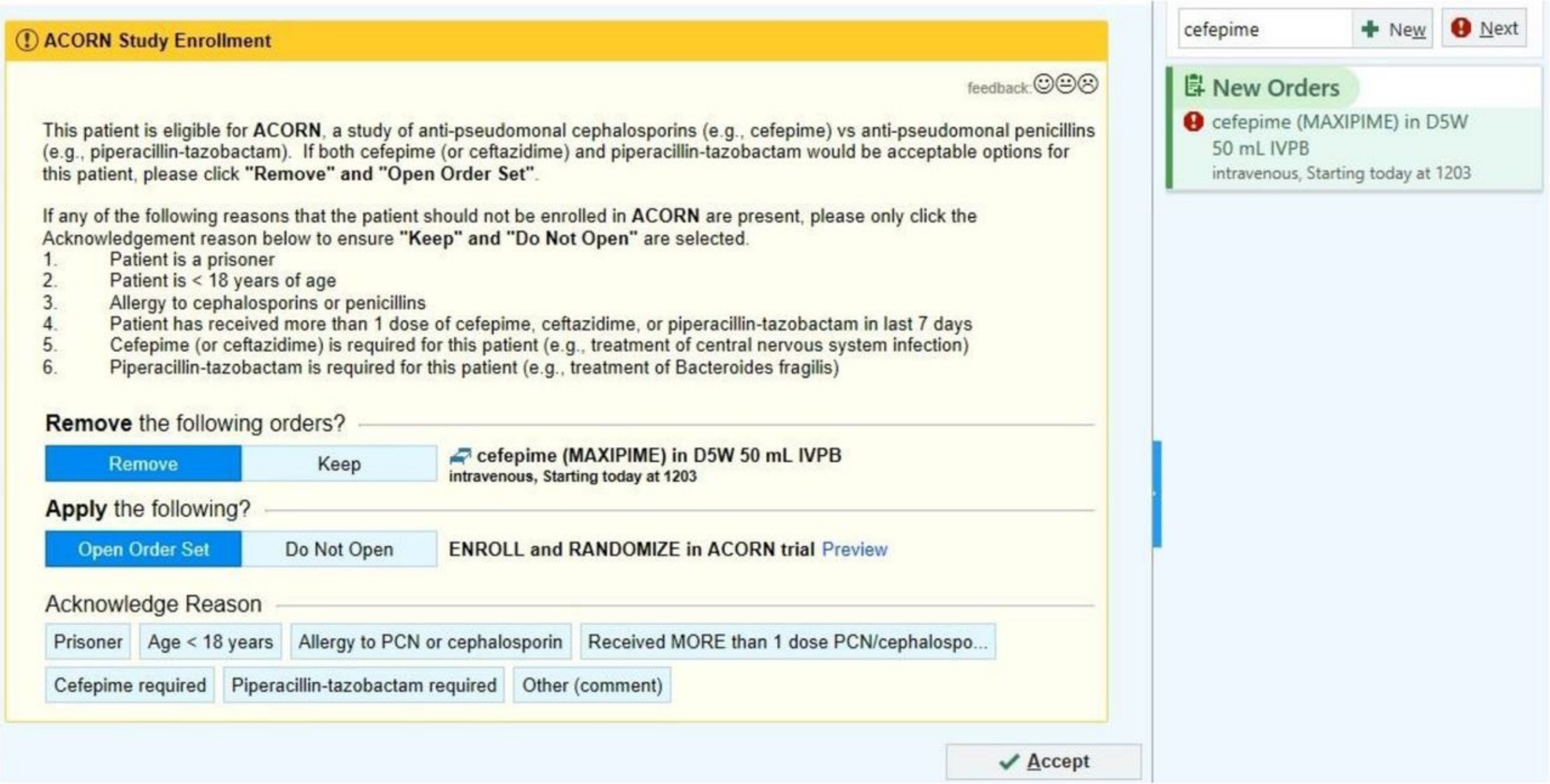
YouTube Link – Vanderbilt Informatics

https://www.youtube.com/watch?v=edS4XmCBQ9A&ab_channel=DepartmentofBiomedicalInformatics

"Hey you're about to order an anti-pseudomonal in a septic patient, wanna enroll in the ACORN trial? "

Orderset would then appropriately randomize and help guide appropriate dosing of the antibiotic

doi: 10.1001/jama.2023.20583


ACORN Study Enrollment
feedback:
This patient is eligible for ACORN, a study of anti-pseudomonal cephalosporins (e.g., cefepime) vs anti-pseudomonal penicillins (e.g., piperacillin-tazobactam). If both cefepime (or ceftazidime) and piperacillin-tazobactam would be acceptable options for this patient, please click "Remove" and "Open Order Set".
If any of the following reasons that the patient should not be enrolled in ACORN are present, please only click the Acknowledgement reason below to ensure "Keep" and "Do Not Open" are selected.
1. Patient is a prisoner
2. Patient is < 18 years of age
3. Allergy to cephalosporins or penicillins
4. Patient has received more than 1 dose of cefepime, ceftazidime, or piperacillin-tazobactam in last 7 days
5. Cefepime (or ceftazidime) is required for this patient (e.g., treatment of central nervous system infection)
6. Piperacillin-tazobactam is required for this patient (e.g., treatment of Bacteroides fragilis)
Remove the following orders?
Remove
Keep
cefepime (MAXIPIME) in D5W 50 mL IVPB
intravenous, Starting today at 1203
Apply the following?
Open Order Set
Do Not Open
ENROLL and RANDOMIZE in ACORN trial Preview
Acknowledge Reason
Prisoner
Age < 18 years
Allergy to PCN or cephalosporin
Received MORE than 1 dose PCN/cephalospo...
Cefepime required
Piperacillin-tazobactam required
Other (comment)
Accept
cefepime
New
Next
New Orders
cefepime (MAXIPIME) in D5W
50 mL IVPB
intravenous, Starting today at 1203

JAMA

QUESTION Does the choice between cefepime and piperacillin-tazobactam affect the risks of acute kidney injury or neurological dysfunction in adults hospitalized with acute infection?

CONCLUSION Among hospitalized adults, the risk of acute kidney injury did not differ between cefepime and piperacillin-tazobactam, but neurological dysfunction was more common with cefepime.

POPULATION

1439 Men **1071** Women

Adults hospitalized with acute infection

Median age: **58** years

LOCATION

1
Medical center in Nashville, Tennessee

INTERVENTION

2634 Patients randomized
2511 Patients analyzed

1214
Cefepime
Administered as an intravenous push over 5 minutes

1297
Piperacillin-tazobactam
Administered as a bolus for the initial administration and then extended infusion over 4 hours for subsequent doses

PRIMARY OUTCOME

Highest stage of acute kidney injury or death by day 14 (measured on a 5-level ordinal scale; range: no acute kidney injury to death)

FINDINGS

Highest stage of acute kidney injury or death by day 14

Cefepime

Survived with stage 3 acute kidney injury	**7.0%** (85 of 1214 patients)
Died	**7.6%** (92 of 1214 patients)

Piperacillin-tazobactam

Survived with stage 3 acute kidney injury	**7.5%** (97 of 1297 patients)
Died	**6.0%** (78 of 1297 patients)

There was no significant between-group difference:
Odds ratio, 0.95 (95% CI, 0.80 to 1.13); *P* = .56

Qian ET, Casey JD, Wright A, et al; Vanderbilt Center for Learning Healthcare and the Pragmatic Critical Care Research Group. Cefepime vs piperacillin-tazobactam in adults hospitalized with acute infection: the ACORN randomized clinical trial. *JAMA*. Published online October 14, 2023. doi:10.1001/jama.2023.20583

Case Report

Integrating clinical research into electronic health record workflows to support a learning health system

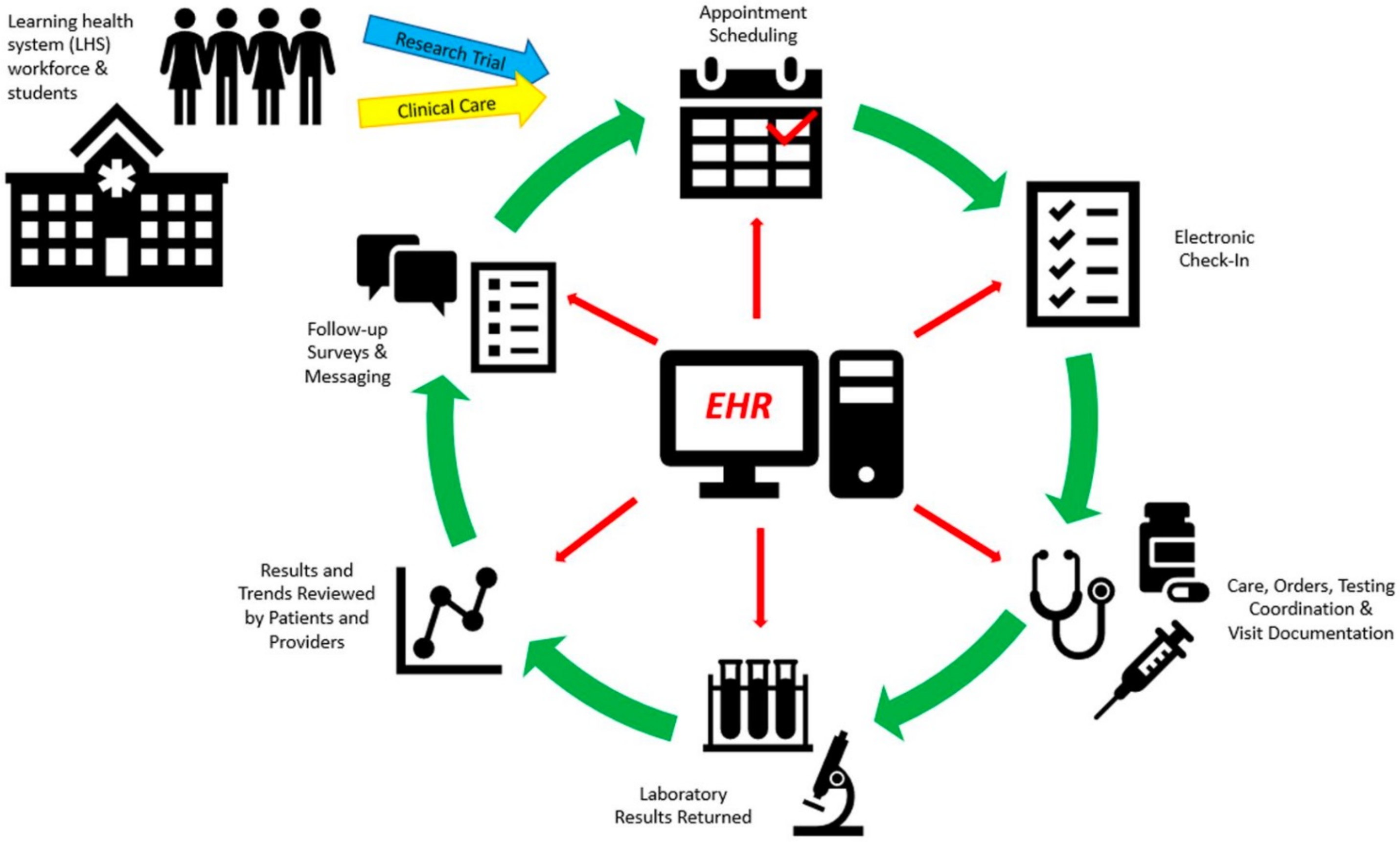
Nicole H. Goldhaber, MD, MA[1,*], Marni B. Jacobs, PhD, MPH[2], Louise C. Laurent, MD, PhD[2], Rob Knight, PhD[3,4,5], Wenhong Zhu, PhD[6], Dean Pham, PharmD, MBA[6], Allen Tran, PharmD[6], Sandip P. Patel, MD[7], Michael Hogarth, MD[8], Christopher A. Longhurst, MD, MS[3,8]

UCSD

"AHRQ defines LHSs as those in which internal data and experience are systematically integrated with external evidence to facilitate data-driven intervention."

Let's do a COVID study utilizing multiple EHR tools in the spirits of LHS

- Emailed QR codes to link to MyChart
- Questionnaires / Consent / Scheduling / eCheck-In
- Lab orders and results



https://doi.org/10.1093/jamiaopen/ooae023

ARTICLE OPEN

Check for updates

Impact of a deep learning sepsis prediction model on quality of care and survival

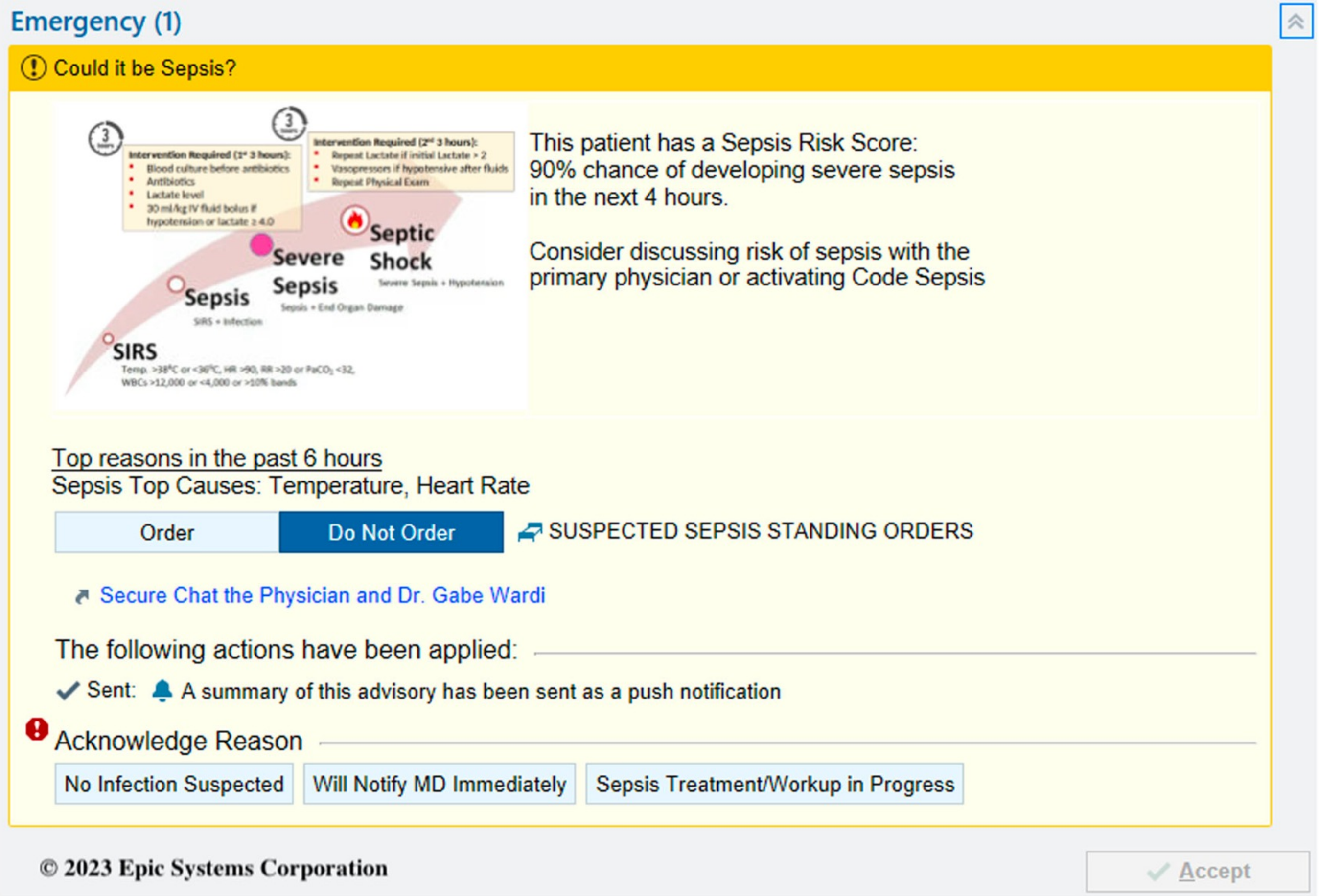
Aaron Boussina[1,4], Supreeth P. Shashikumar[1,4], Atul Malhotra[1], Robert L. Owens[1], Robert El-Kareh[1,2], Christopher A. Longhurst[1,2], Kimberly Quintero[2], Allison Donahue[3], Theodore C. Chan[3], Shamim Nemati[1,3,5] and Gabriel Wardi[1,3,5]✉

UCSD: Pre/Post Sepsis Interventions

COMPOSER

"It is a feed-forward neural network model that incorporates routinely collected laboratory and vital signs as well as patient demographics (age and sex), comorbidities, and concomitant medications to output a risk score for the onset of sepsis within the next 4 h."

BPA in workflow alert presented to nursing



npj | digital medicine

COMPOSER continued…

Editorial | Open access | Published: 01 March 2024

Integrating artificial intelligence into healthcare systems: more than just the algorithm

Jethro C. C. Kwong, Grace C. Nickel, Serena C. Y. Wang & Joseph C. Kvedar

npj Digital Medicine 7, Article number: 52 (2024) | Cite this article

Improvement in outcomes as well as performance metrics

- Reduction mortality
  - 1.9% absolute
  - 17% relative
- Improvement in Bundle Compliance rates including abx/ IVF / etc.
  - 5% absolute
  - 10% relative increase

Table 2. Observed outcomes in the pre-intervention period, the expected counterfactual values from causal impact analysis, and the actual post-intervention values.

Outcome	Pre-intervention value	Expected post-intervention value (95% CI)	Actual post-intervention value
In-hospital mortality %	10.3%	11.4% (9.8%–13.0%)	**9.5%**
Average 72-h Change in SOFA	3.71	3.71 (3.6–3.8)	**3.56**
Sepsis bundle compliance rate	48.3%	48.4% (45.5%–51.0%)	**53.4%**
Blood cultures prior to antibiotics compliance rate	71.1%	72.0% (69.9%–73.9%)	73.9%
Rate of antibiotics administered within 24 h prior and 3 h after severe sepsis onset.	82.8%	82.8% (81.3%–84.4%)	**84.6%**
Rate of lactate measured within 6 h prior and 3 h after severe sepsis onset	83.5%	83.4% (81.3%–85.8%)	85.6%
Rate of repeat lactate measured within 6 h after severe sepsis onset if initial lactate is elevated	97.8%	97.3% (96.2%–98.4%)	**98.6%**
Rate of administration of vasoactive medications within 6 h of septic shock	58.0%	57.5% (46.7%–68.2%)	55.5%
Rate of administration of 30cc/kg of fluids within 3 h of presentation of septic shock or hypotension	54.2%	53.9% (48.9%–58.8%)	**59.3%**
ICU transfer rate	32.6%	32.5% (30.7%–34.2%)	31.8%
Average ICU-free days	25.4	25.1 (24.6–25.6)	25.6

Significant post-intervention values against the 95% confidence interval are bolded.

Why do users override alerts? Utilizing large language model to summarize comments and optimize clinical decision support

Siru Liu, PhD*[,1,2], Allison B. McCoy, PhD[1], Aileen P. Wright, MD, MS[1,3], Scott D. Nelson, PharmD, MS[1], Sean S. Huang, MD[1,3], Hasan B. Ahmad, DO, MBA[4], Sabrina E. Carro, MD[5], Jacob Franklin, MD[3], James Brogan, MD, MS[3], Adam Wright, PhD[1,3]

Vanderbilt / Clickbusters

Let's take the 4000 user comments from 8 alerts and have human summaries compared to GPT-4 summaries and subsequent suggested changes

Let's rank the quality of these summaries' human vs LLM

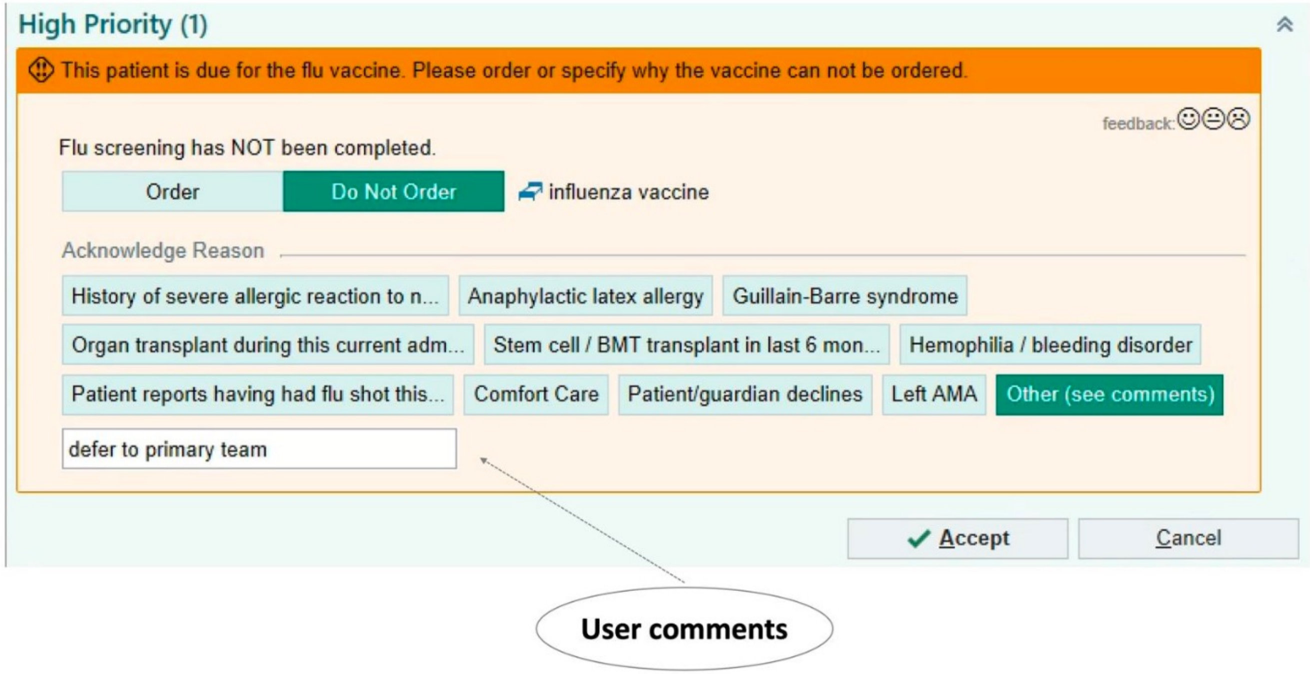
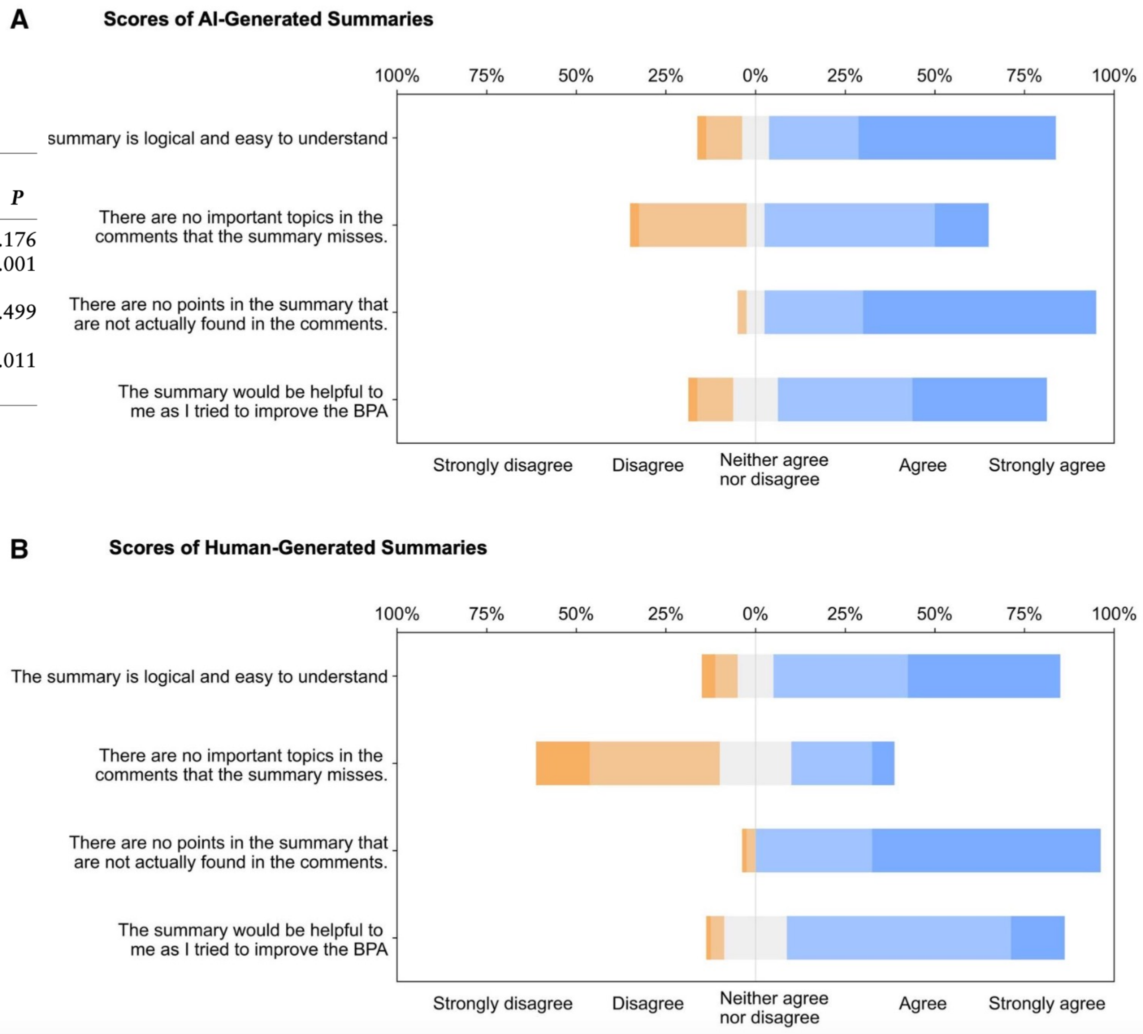


Table 3. Means and SD for survey questions rated on a 5-point Likert scale, with 1 indicating "strongly disagree" and 5 indicating "strongly agree."

	AI-generated summaries mean (SD)	Human-generated summaries mean (SD)	*P*
Clarity: The summary is logical and easy to understand.	4.2 (1.1)	4.1 (1.1)	.176
Completeness: There are no important topics in the comments that the summary misses.	3.4 (1.2)	2.7 (1.2)	.001
Accuracy: There are no points in the summary that are not actually found in the comments.	4.5 (0.7)	4.5 (0.7)	.499
Usefulness: The summary would be helpful to me as I tried to improve the alert.	4.0 (1.1)	3.9 (0.8)	.011



ChatGPT won.....

Let's save time scouring over thousands of comments and let the LLM take first pass prior to sending it committee for possible alterations in the CDSS

JOURNAL ARTICLE

Leveraging explainable artificial intelligence to optimize clinical decision support

Siru Liu, PhD, Allison B McCoy, PhD, Josh F Peterson, MD, MPH, Thomas A Lasko, MD, PhD, Dean F Sittig, PhD, Scott D Nelson, PharmD, MS, Jennifer Andrews, MD, Lorraine Patterson, MSN, Cheryl M Cobb, MD, David Mulherin, PharmD ... Show more

JAMIA
A SCHOLARLY JOURNAL OF INFORMATICS IN HEALTH AND BIOMEDICINE

Vanderbilt / Clickbusters

Natural progression to prior papers regarding the utility of ML in assessing and improving CDS best practice alerts (BPAs)

Let's improve CDS using Explainable AI (XAI) local models, take into account multiple variables at our disposal for triggering and exclusion

"Do not fire breast cancer screening alerts for patients in the hospice unit."

- Epic Clarity data for BPA firing and user responses / feedback for 2 years (2019-2021)
- **Explainable AI (XAI)** refers to the ability of an artificial intelligence (AI) system or model to provide clear and understandable explanations for its actions or decisions. In other words, XAI is about making AI transparent and interpretable to humans[1]
  - They tried 4 different XAI models
- Compare the ML generated suggestions against the change-log for the BPAs over the years, if a suggested change was already implemented.
- Give the ML suggestion "credit" if it suggested something that had been suggested and implemented by the CDS team

Insert fancy XAI math here....

Essentially taking into account the suggestions from XAI would account for 10% reduction in unnecessary firings and thus boost up the acceptance rate

Once again, using AI to *augment* the CDS review process, especially in resource constrained organizations

Table 3. Examples of generated suggestions and feedback from clinicians.

BPA	Generated suggestion	Comment
This patient has one or more Shared Plan of Care FYI flags which may require your attention. [High Priority]	Do not fire when: Encounter Type = Documentation	Already changed, the same exclusion was added on March 16, 2023.
This patient is due for the flu vaccine. Please order or specify why the vaccine cannot be ordered. [High Priority]	Do not fire when: Patient Department = VPH ADULT PARTIAL HOSPITALIZATION	Already changed, the same exclusion was added on December 17, 2020.
Contraindicated—NSAIDs and Pregnancy [Important]	Do not fire when: Patient Department = VUH 4E POST PARTUM	Already changed, add exclusion criteria: exclude Department = VUH 4E POST PARTUM on August 6, 2020.

Randomized Comparison of Electronic Health Record Alert Types in Eliciting Responses about Prognosis in Gynecologic Oncology Patients

Robert Clayton Musser[1,2] Rashaud Senior[2,3] Laura J. Havrilesky[4] Jordan Buuck[2] David J. Casarett[5]
Salam Ibrahim[6] Brittany A. Davidson[4]

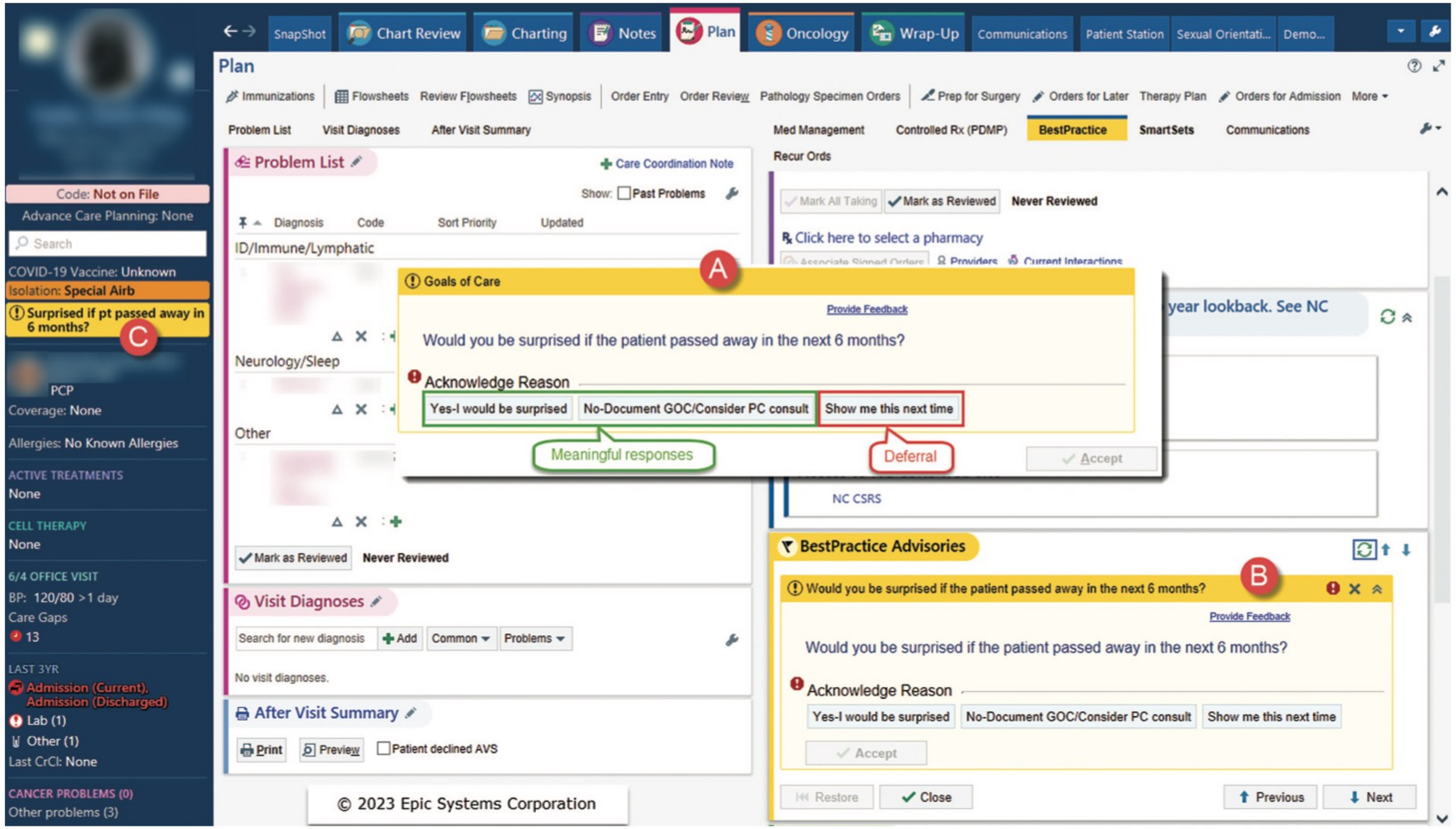
Duke

"Would you be surprised if this patient passed away in the next 6 months?"

Measure response rates:

- Passive Storyboard
- Chart Open – interruptive
- Chart Close – interruptive

Why? To increase GOC conversations in cancer patients



Interruptive still wins, 5x more.

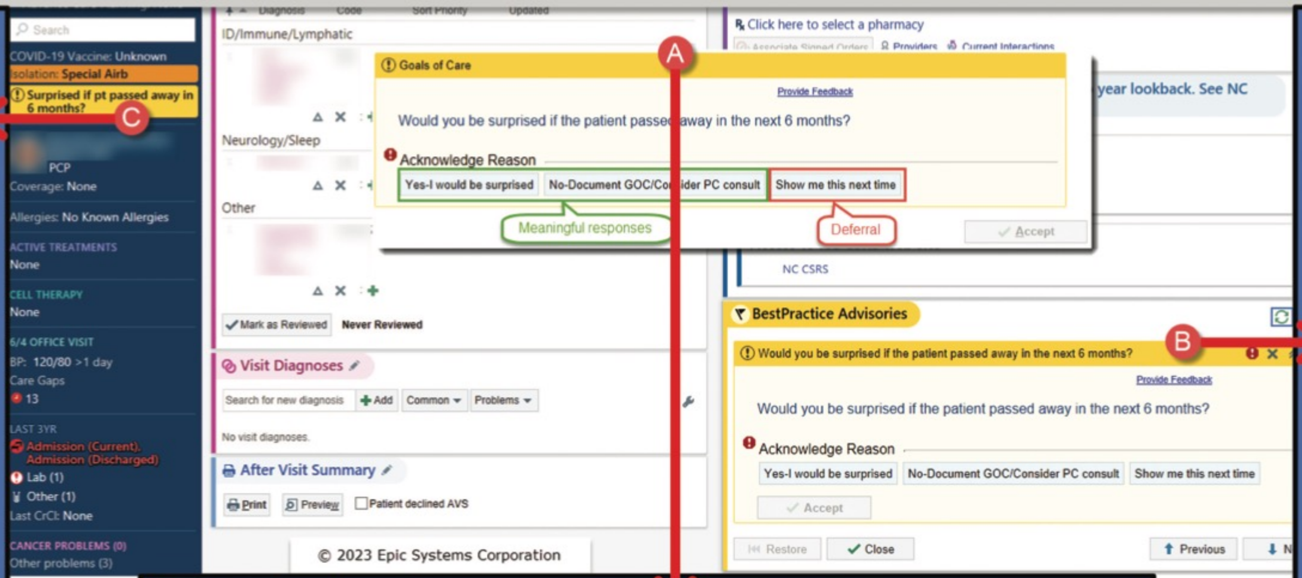
Passive led to the most "No – not surprised answers" though

Surprised?

C. Optional Persistent

(Storyboard BPA; 203 patients)

- Elicited **fewest responses**, both meaningful* (19.7% of encounters) and total (19.9%)
- **Deferred the least** (98.3% responses meaningful*)
- **Most likely to say "No"** (13.6% of Y/N responses)



B. Required on Close

(Navigator BPA; 279 patients**)

- Alerted 2nd most (1.3 alerts/encounter)
- Elicited 2nd most responses, both meaningful* (90.1% of encounters) & total (96.1%)
- Deferred 2nd most (10.9% of encounters)
- **Least likely to say "No"** (7.0% of Y/N responses)

A. Required on Open

(pop-up BPA on opening chart; 203 patients)

- Alerted most (1.5 alerts/encounter)
- Elicited **most responses**, both meaningful* (94.8% of encounters) & total (100%)
- **Deferred the most** (31.7% of encounters)
- 2nd most likely to say "No" (10.3% of Y/N responses)

*Meaningful responses = "Yes" or "No", but not deferral (see green annotations, image A).

**Patient counts differ due to 4:3:3 randomization.

Musser et al. Randomized Comparison of Electronic Health Record Alert Types in Eliciting Responses about Prognosis in Gynecologic Oncology Patients. 2023.

Original Investigation | Medical Education

JAMA Network Open

An Automated System for Physician Trainee Procedure Logging via Electronic Health Records

Brian Kwan, MD, MS; Jeffery Engel, BA; Brian Steele, MCM; Leslie Oyama, MD; Christopher A. Longhurst, MD, MS; Robert El-Kareh, MD, MPH; Michelle Daniel, MD, MHPE; Charles Goldberg, MD, MS; Brian Clay, MD

UCSD

Epic and homegrown magic

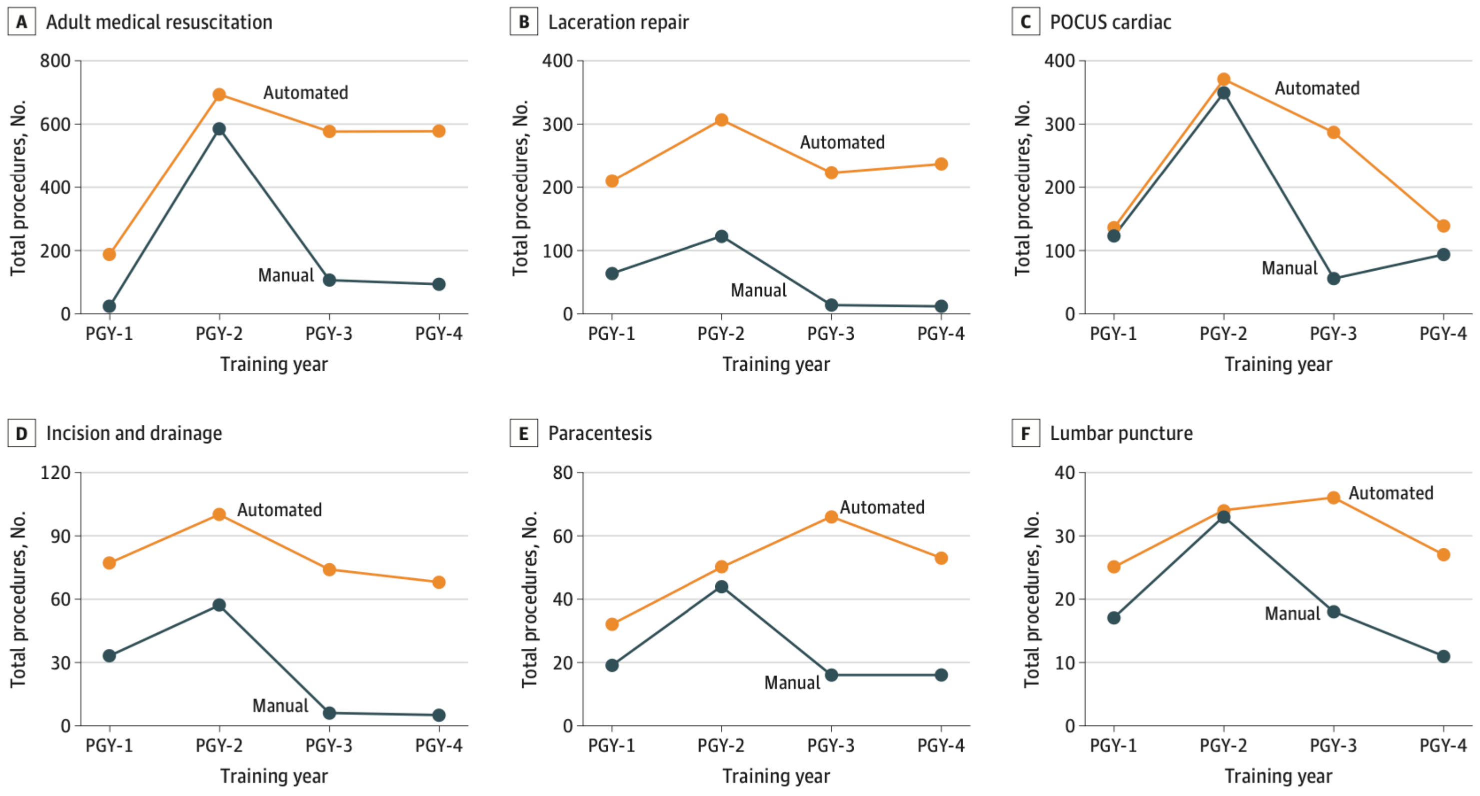
Surgical Residents PGY1 -4

1 year's worth of procedures

Let's design a method to extract procedures performed from the EHR, let's compare it to how awful we are at logging it manually

Where was this 20 years ago???

Figure 1. Comparison of Automated and Manual Capture of Procedure Counts for Selected Procedures by Training Year

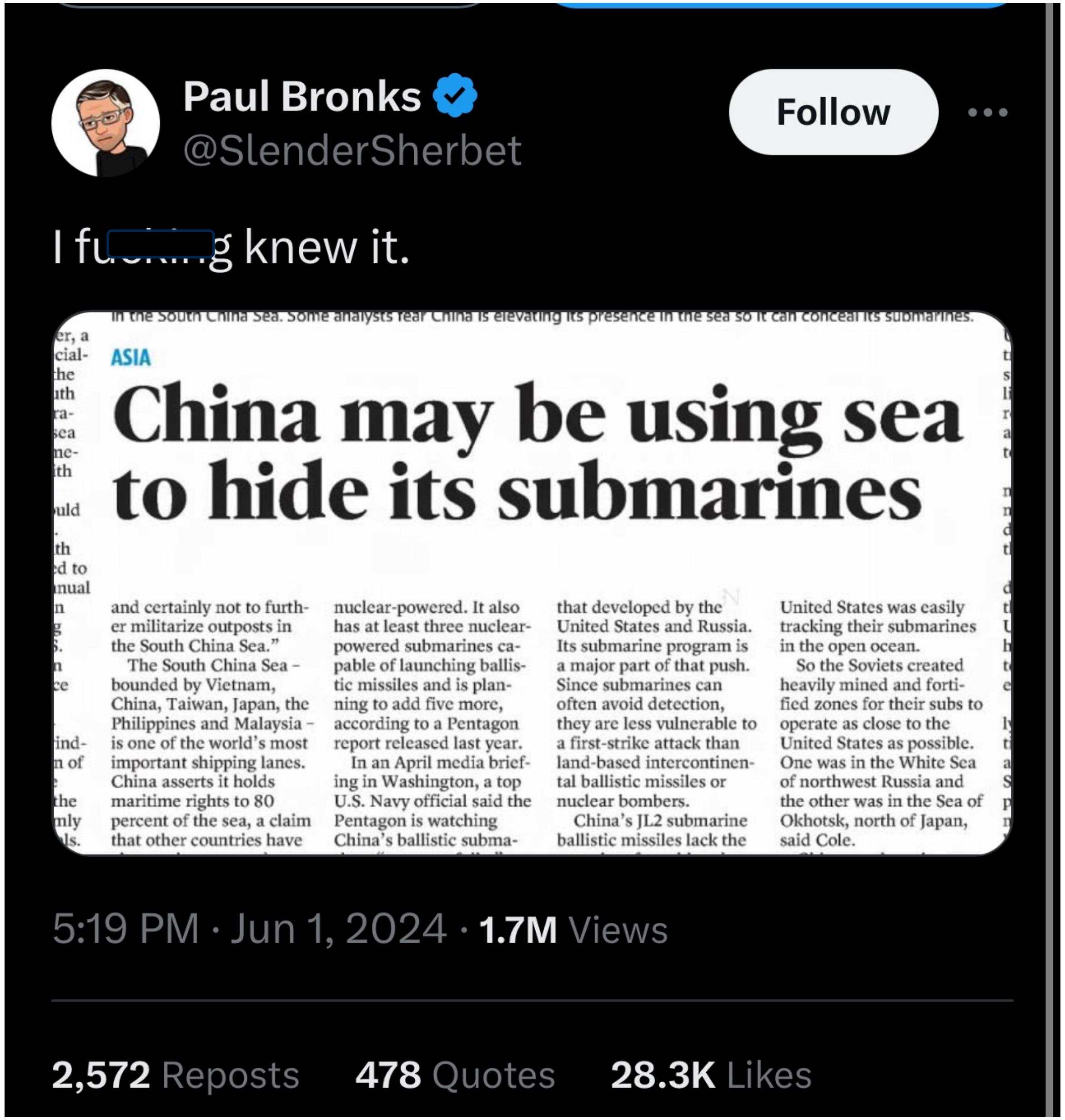


doi:10.1001/jamanetworkopen.2023.52370

Transition Comedy Slide

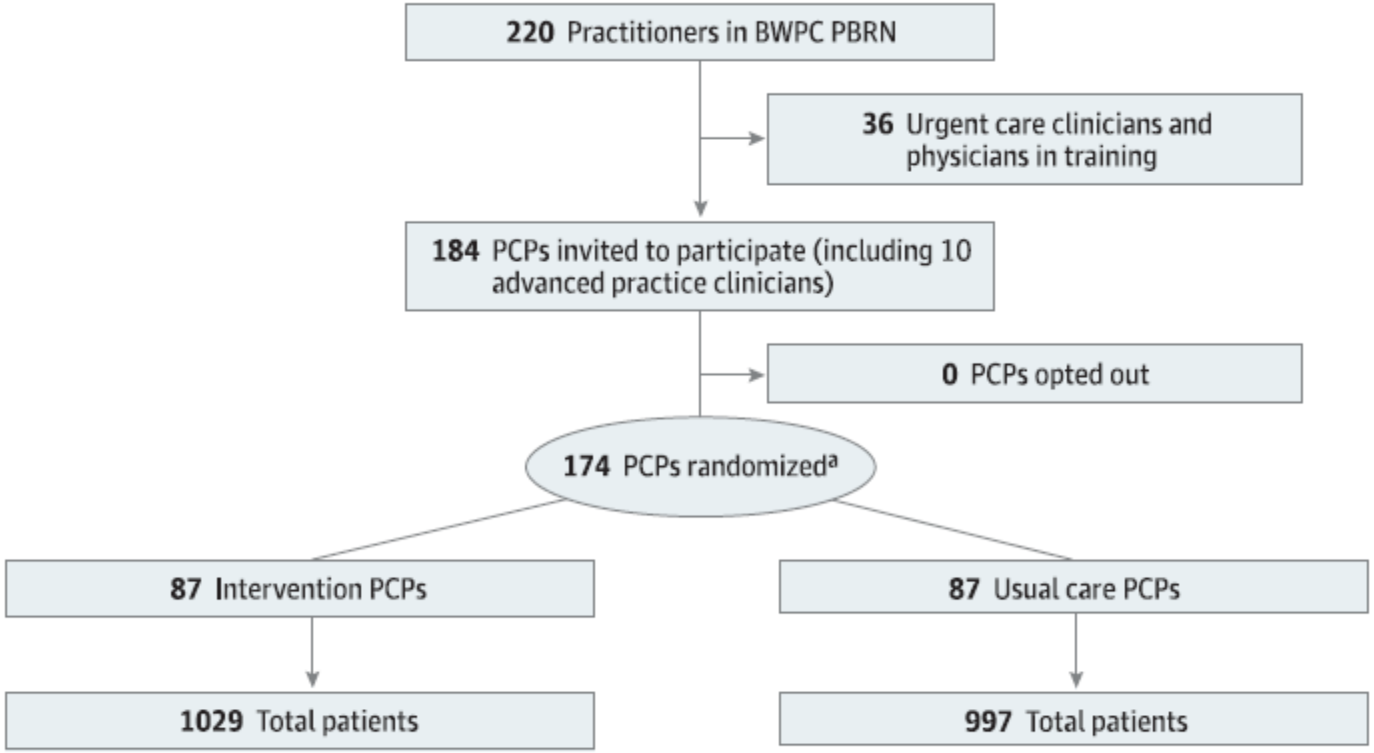
Back to you Bill !

More CDS and more AI



JAMA Internal Medicine | Original Investigation

Clinical Decision Support for Hypertension Management in Chronic Kidney Disease
A Randomized Clinical Trial



Samal L, Kilgallon JL, Lipsitz S, Baer HJ, McCoy A, Gannon M, Noonan S, Dunk R, Chen SW, Chay WI, Fay R, Garabedian PM, Wu E, Wien M, Blecker S, Salmasian H, Bonventre JV, McMahon GM, Bates DW, Waikar SS, Linder JA, Wright A, Dykes P. Clinical Decision Support for Hypertension Management in Chronic Kidney Disease: A Randomized Clinical Trial. JAMA Intern Med. 2024 May 1;184(5):484-492. doi: 10.1001/jamainternmed.2023.8315. Erratum in: JAMA Intern Med. 2024 Jun 10. doi: 10.1001/jamainternmed.2024.2589. PMID: 38466302;

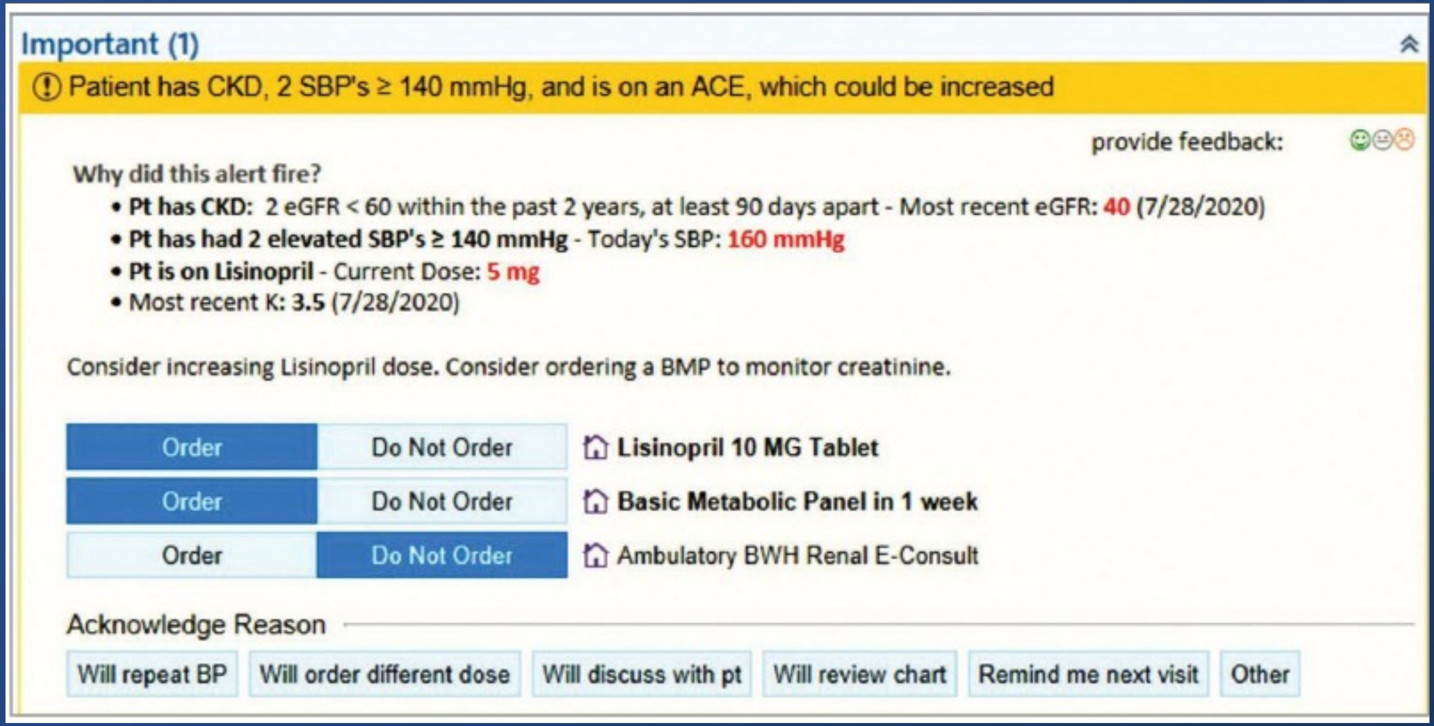


Table 2. SBP at Baseline and 180 Days, Change in Mean SBP from Baseline, and BP Control

Measurement variable	Intervention	Usual care	*P* value
Baseline SBP, mean (SD), mm Hg[a]	154.3 (14.2)	153.7 (14.4)	.54
SBP at 180 d, mean (SD), mm Hg[a]	139.5 (19.7)	142.1 (19.9)	.009
Change in SBP, % (95% CI), mm Hg[b]	−14.6 (−13.1 to −16.0)	−11.7 (−10.2 to −13.1)	.005
BP control, % (95% CI)[c]	50.4 (46.5 to 54.3)	47.1 (43.3 to 51.0)	.23

Samal L, Kilgallon JL, Lipsitz S, Baer HJ, McCoy A, Gannon M, Noonan S, Dunk R, Chen SW, Chay WI, Fay R, Garabedian PM, Wu E, Wien M, Blecker S, Salmasian H, Bonventre JV, McMahon GM, Bates DW, Waikar SS, Linder JA, Wright A, Dykes P. Clinical Decision Support for Hypertension Management in Chronic Kidney Disease: A Randomized Clinical Trial. JAMA Intern Med. 2024 May 1;184(5):484-492. doi: 10.1001/jamainternmed.2023.8315. Erratum in: JAMA Intern Med. 2024 Jun 10. doi: 10.1001/jamainternmed.2024.2589. PMID: 38466302;

Article
https://doi.org/10.1038/s41591-024-02961-4

AI-enabled electrocardiography alert intervention and all-cause mortality: a pragmatic randomized clinical trial

Received: 16 April 2023

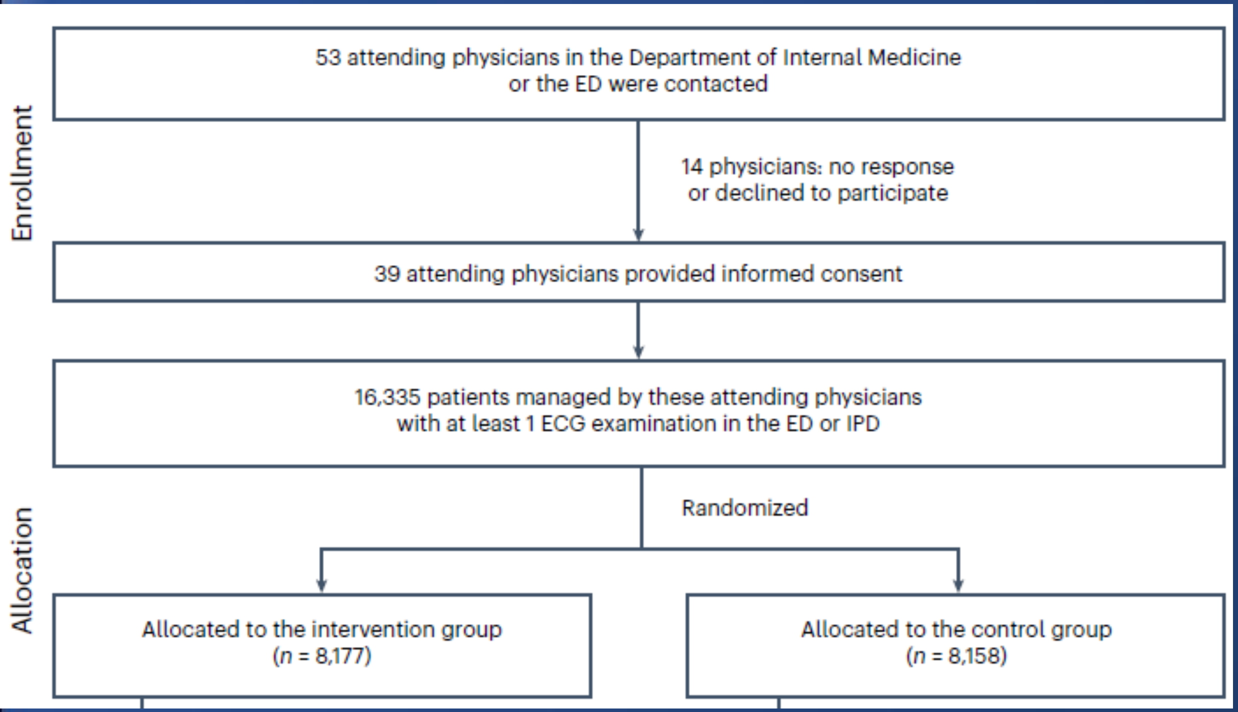
Accepted: 29 March 2024

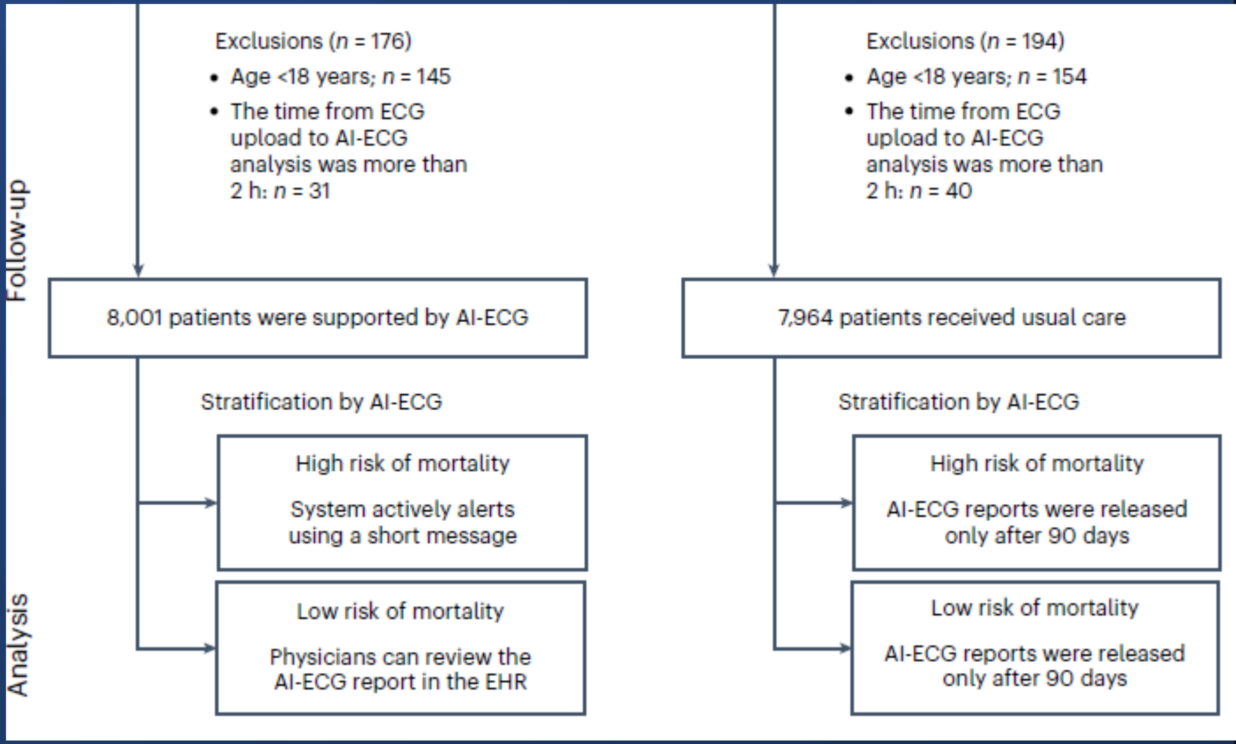
Published online: 29 April 2024

Chin-Sheng Lin[1,2], Wei-Ting Liu[1], Dung-Jang Tsai[2,3,4], Yu-Sheng Lou[3], Chiao-Hsiang Chang[1], Chiao-Chin Lee[1], Wen-Hui Fang[3,5], Chih-Chia Wang[5], Yen-Yuan Chen[6], Wei-Shiang Lin[1], Cheng-Chung Cheng[1], Chia-Cheng Lee[7,8], Chih-Hung Wang[9,10], Chien-Sung Tsai[11], Shih-Hua Lin[12] & Chin Lin[2,3,13,14]

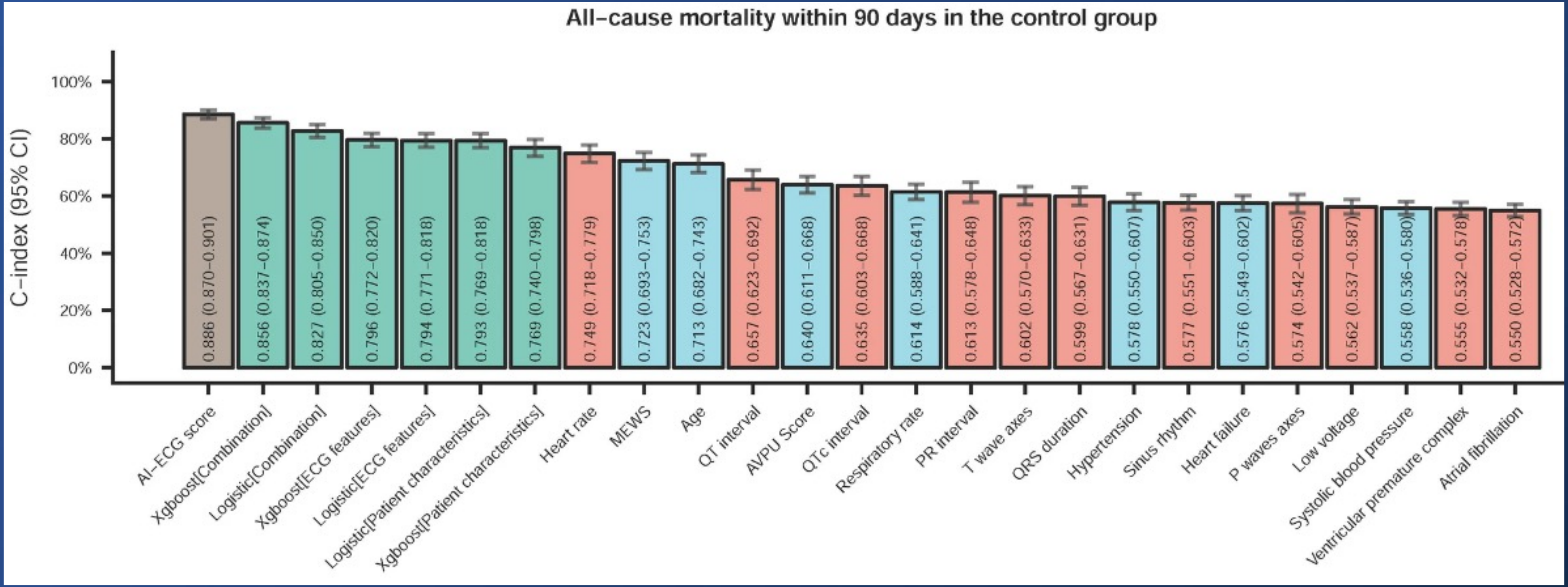
A gift from Colin

Lin CS, Liu WT, Tsai DJ, Lou YS, Chang CH, Lee CC, Fang WH, Wang CC, Chen YY, Lin WS, Cheng CC, Lee CC, Wang CH, Tsai CS, Lin SH, Lin C. AI-enabled electrocardiography alert intervention and all-cause mortality: a pragmatic randomized clinical trial. Nat Med. 2024 May;30(5):1461-1470. doi: 10.1038/s41591-024-02961-4. Epub 2024 Apr 29. PMID: 38684860.





Lin CS, Liu WT, Tsai DJ, Lou YS, Chang CH, Lee CC, Fang WH, Wang CC, Chen YY, Lin WS, Cheng CC, Lee CC, Wang CH, Tsai CS, Lin SH, Lin C. AI-enabled electrocardiography alert intervention and all-cause mortality: a pragmatic randomized clinical trial. Nat Med. 2024 May;30(5):1461-1470. doi: 10.1038/s41591-024-02961-4. Epub 2024 Apr 29. PMID: 38684860.

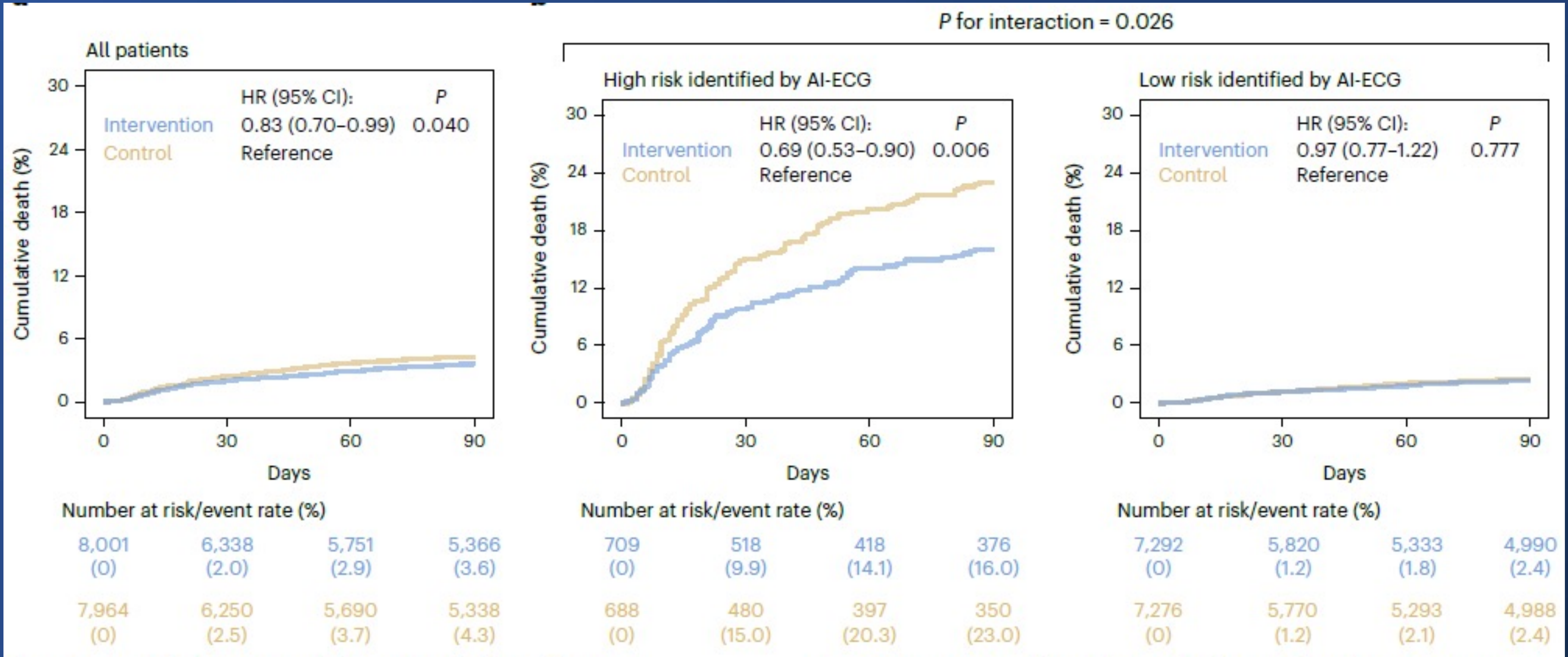


Lin CS, Liu WT, Tsai DJ, Lou YS, Chang CH, Lee CC, Fang WH, Wang CC, Chen YY, Lin WS, Cheng CC, Lee CC, Wang CH, Tsai CS, Lin SH, Lin C. AI-enabled electrocardiography alert intervention and all-cause mortality: a pragmatic randomized clinical trial. Nat Med. 2024 May;30(5):1461-1470. doi: 10.1038/s41591-024-02961-4. Epub 2024 Apr 29. PMID: 38684860.

Screenshot for the intervention group

Screenshot for the control group

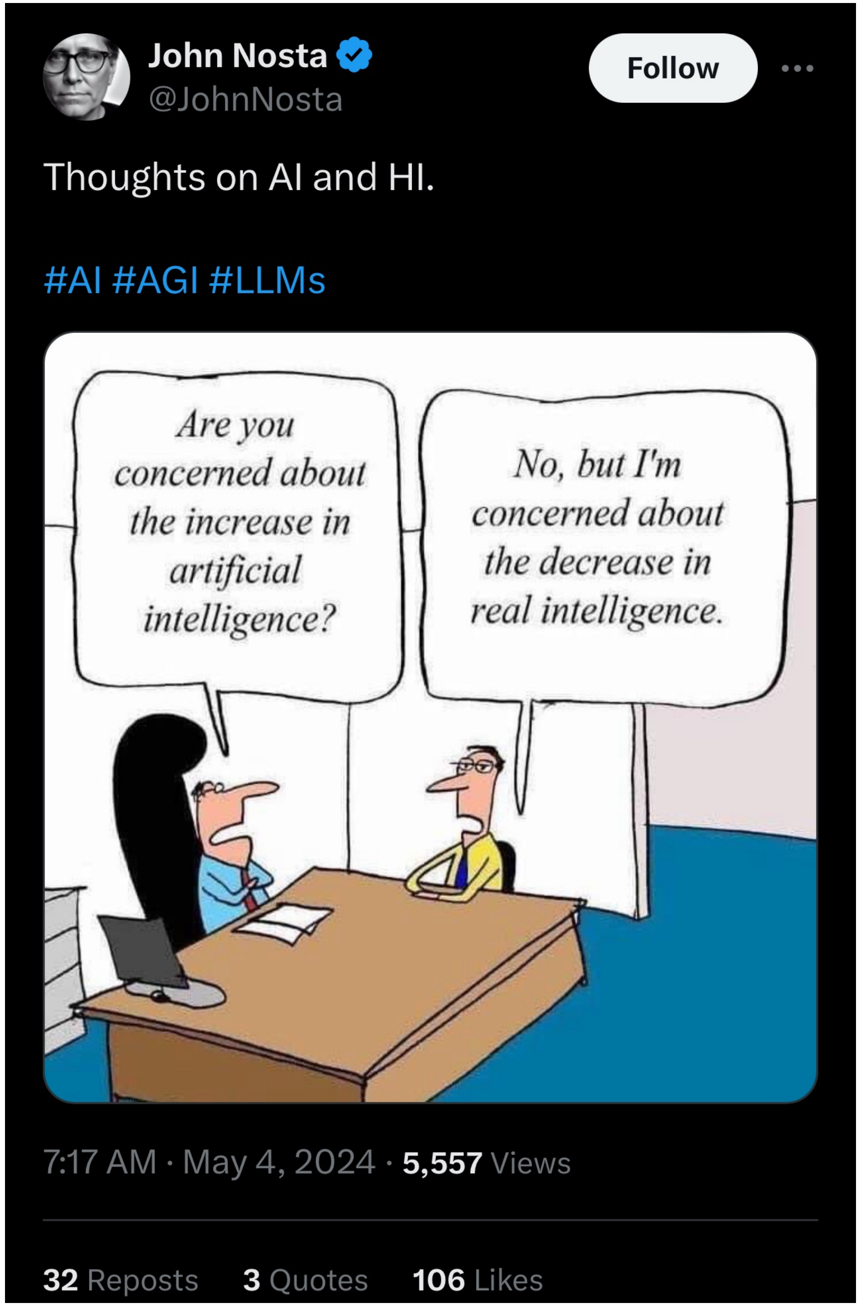
Lin CS, Liu WT, Tsai DJ, Lou YS, Chang CH, Lee CC, Fang WH, Wang CC, Chen YY, Lin WS, Cheng CC, Lee CC, Wang CH, Tsai CS, Lin SH, Lin C. AI-enabled electrocardiography alert intervention and all-cause mortality: a pragmatic randomized clinical trial. Nat Med. 2024 May;30(5):1461-1470. doi: 10.1038/s41591-024-02961-4. Epub 2024 Apr 29. PMID: 38684860.



Lin CS, Liu WT, Tsai DJ, Lou YS, Chang CH, Lee CC, Fang WH, Wang CC, Chen YY, Lin WS, Cheng CC, Lee CC, Wang CH, Tsai CS, Lin SH, Lin C. AI-enabled electrocardiography alert intervention and all-cause mortality: a pragmatic randomized clinical trial. Nat Med. 2024 May;30(5):1461-1470. doi: 10.1038/s41591-024-02961-4. Epub 2024 Apr 29. PMID: 38684860.

Transition Comedy Slide:

The LLM and Inbox Section



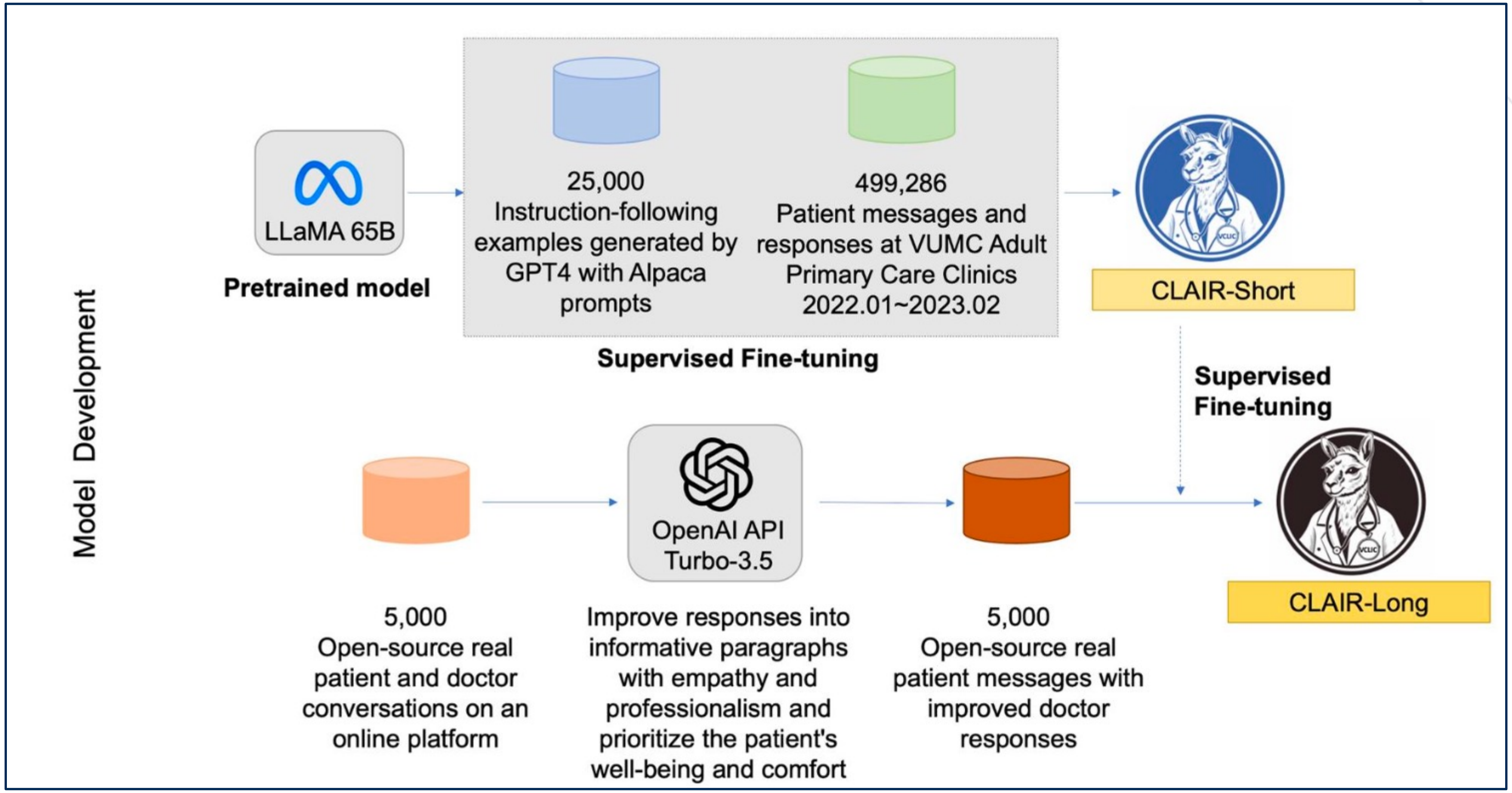
Leveraging large language models for generating responses to patient messages—a subjective analysis

Siru Liu, PhD*[,1], Allison B. McCoy, PhD[1], Aileen P. Wright, MD, MS[1,2], Babatunde Carew, MD[3], Julian Z. Genkins, MD[4], Sean S. Huang, MD[1,2], Josh F. Peterson, MD, MPH[1,2], Bryan Steitz, PhD[1], Adam Wright, PhD[1]

Vanderbilt

Pre-trained models (LLaMA) and GPT 3.5, subsequently fine tuned with real world VUMC patient messages and real world open source patient to doctor conversations

Let's rate and compare 4 different LLM models and human responses against 10 real world scenarios

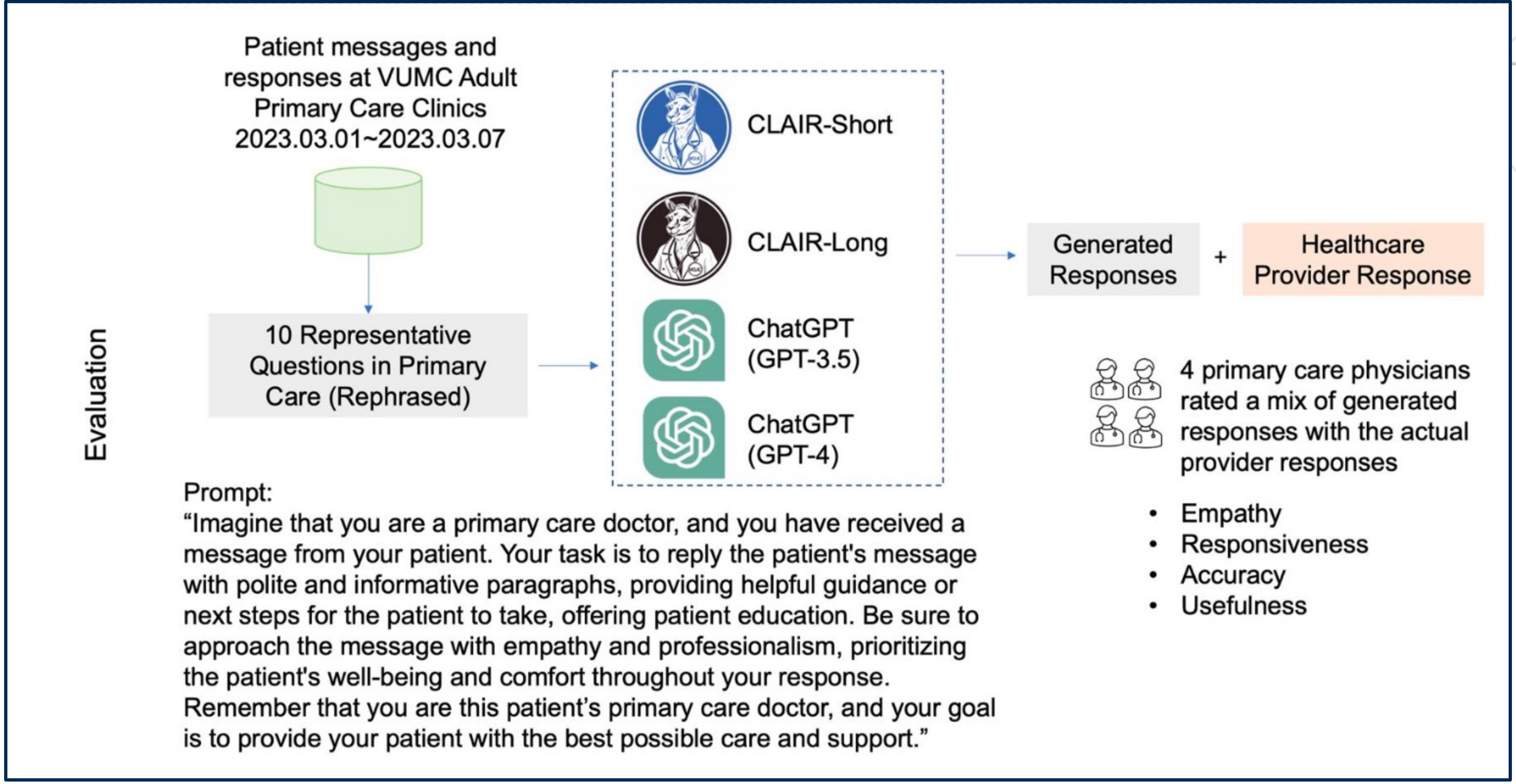


https://doi.org/10.1093/jamia/ocae052

Simulated patient interactions run through each of the engines.

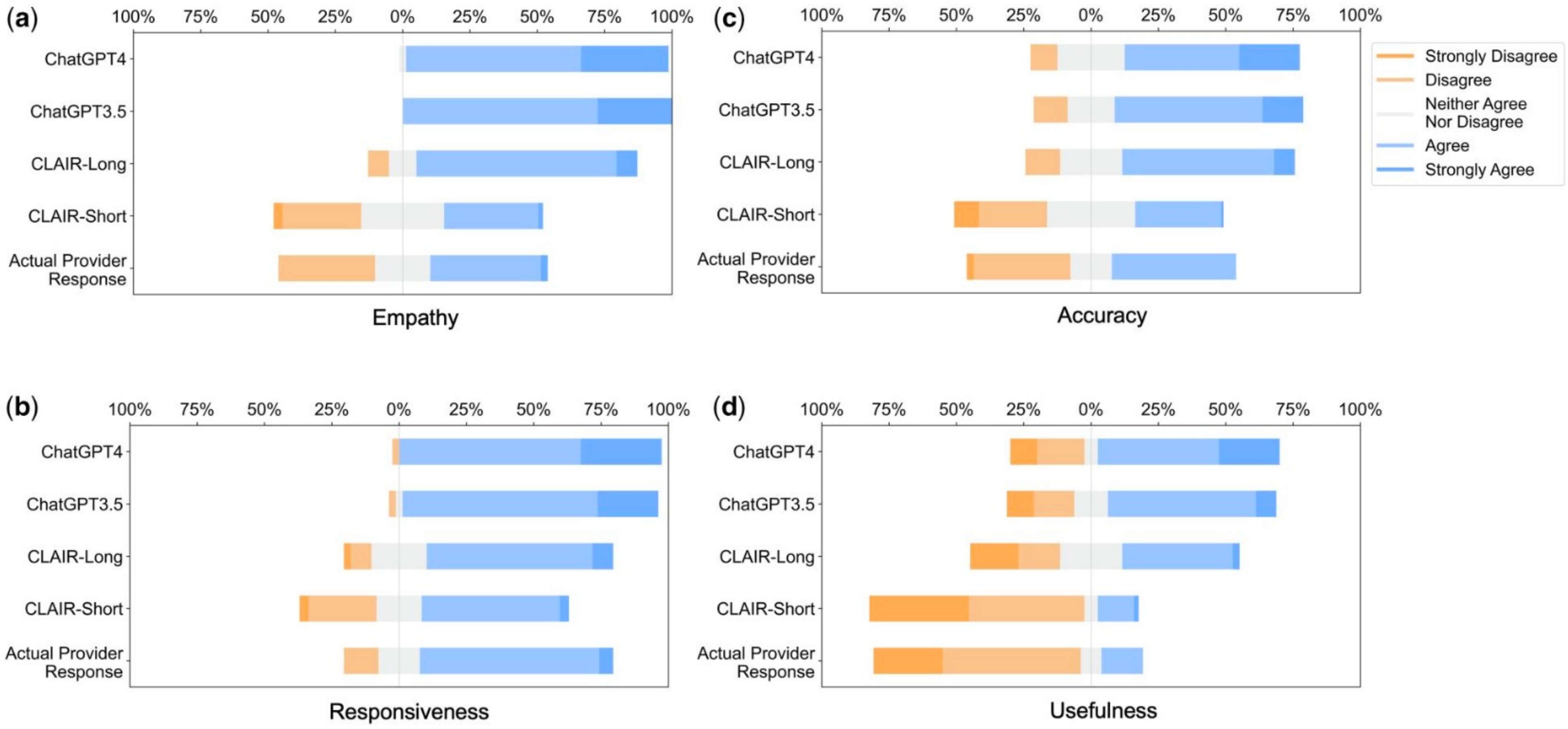
Responses were then rated by PCPs:

- Empathy
- Responsiveness
- Accuracy
- Usefulness



Interestingly fine tuning didn't actually outperform regular ChatGPT

Fine tuning with local data and then again with open source data (patient – provider interactions) was better than local data tuning alone



The effect of using a large language model to respond to patient messages

Lancet Digit Health 2024

Brigham 2023

Radiation Oncologist scenarios based on real world patient questions and answers

200+ cases, simulated

Manual responses (6 oncologists) vs LLM only vs LLM + oncologist edit

Third party survey responses for:

- Helpfulness, Content, Quality, Safety

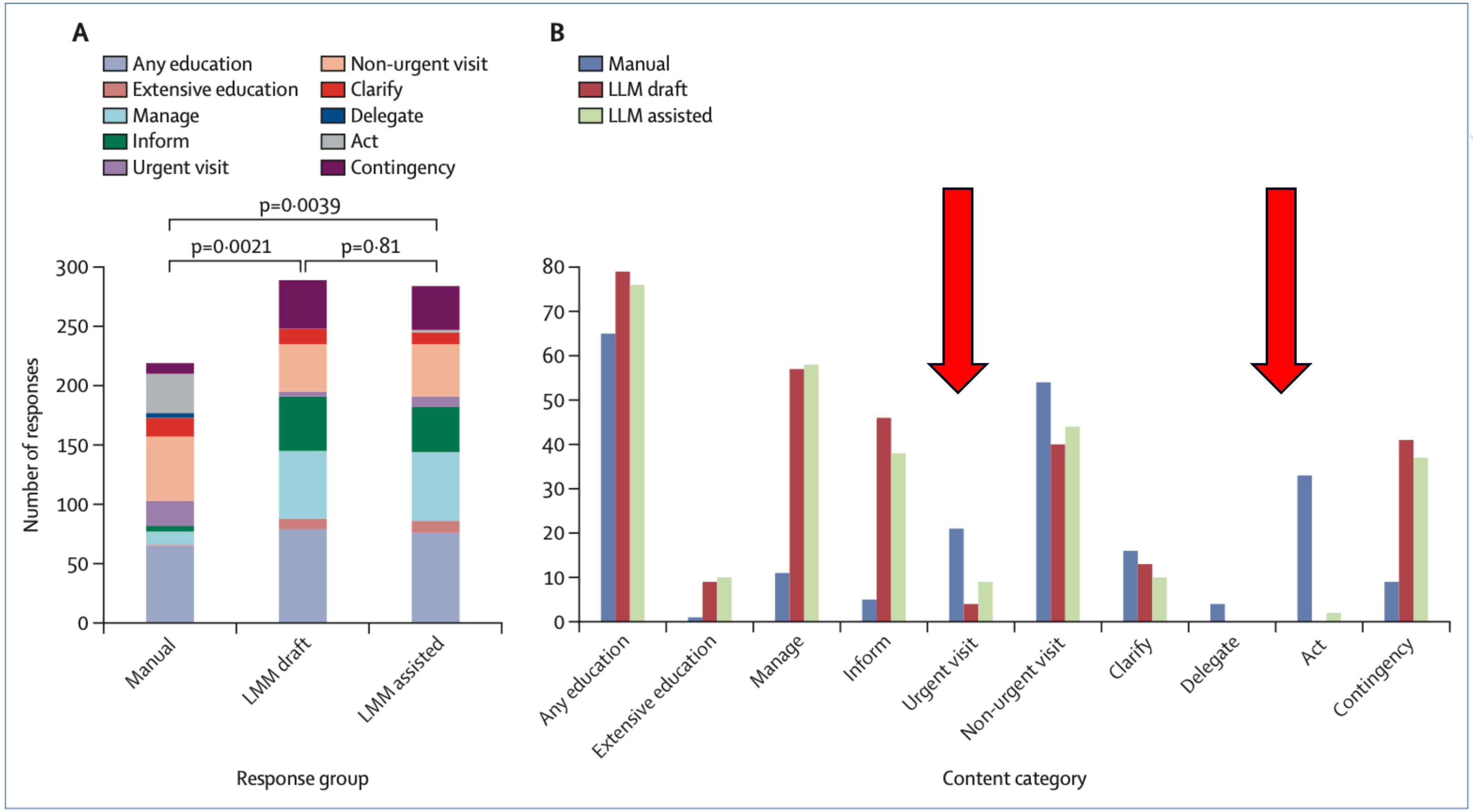


Figure: **Response content comparisons**

Total number of responses that included each content category for manual, LLM draft, and LLM-assisted responses. (A) The overall distribution of content categories present in each response type. Pairwise comparisons of the overall distributions according to response type were done using Mann–Whitney U tests. (B) Visualisation of the total count of each category for the three response types. LLM=large language model.

https://doi.org/10.1016/S2589-7500(24)00060-8

Results

- Manual responses were shortest, 34 words
- LLM longest, 165 words
- LLM + edit, 160 words

But….

- LLM misses acuity and appropriate recommendations – a lot…..

Takeaway – you'd better check that bad boy

> It was felt by the assessing physicians that the LLM drafts posed a risk of severe harm in 11 (7·1%) of 156 survey responses, and death in one (0·6%) survey response. The majority of harmful responses were due to incorrectly determining or conveying the acuity of the scenario and recommended action.

Original Investigation | Health Informatics

Artificial Intelligence-Generated Draft Replies to Patient Inbox Messages

Patricia Garcia, MD; Stephen P. Ma, MD, PhD; Shreya Shah, MD; Margaret Smith, MBA; Yejin Jeong, BA; Anna Devon-Sand, MPH; Ming Tai-Seale, PhD, MPH; Kevin Takazawa, BBA; Danyelle Clutter, MBA; Kyle Vogt, BA; Carlene Lugtu, MCiM; Matthew Rojo, MS; Steven Lin, MD; Tait Shanafelt, MD; Michael A. Pfeffer, MD; Christopher Sharp, MD

Stanford

5 week prospective study July-Aug 2023

162 providers

ChatGPT 4 (not fine tuned for medical)

PCPs and GI/Hep

75% of the time a Chat response was available

Only ~20% acceptance rate of the draft (!!)

No change in response time, write time, read time !!

<u>BUT…. Statistically significant reduction in perceived burden and work exhaustion</u>

Table 2. Draft Utilization per Clinician Stratified by Specialty and Role

Specialty and role	Mean (SD)			
	Reply action count	Reply action count with draft available	Draft used count	Draft utilization rate
Overall	79.3 (95.5)	59.4 (72.6)	8.6 (16.9)	0.203 (0.268)
Primary care	98.5 (84.4)	74.1 (62.9)	9.3 (11.3)	0.176 (0.212)
Physician and APP	102.0 (75.5)	78.5 (61.0)	9.9 (11.9)	0.153 (0.185)
Nurse	164.8 (215.0)	97.0 (109.0)	5.0 (6.8)	0.111 (0.136)
Clinical pharmacist	29.5 (26.0)	17.4 (15.9)	5.1 (3.8)	0.444 (0.317)
Gastroenterology and hepatology	52.8 (103.9)	39.1 (80.3)	7.6 (22.6)	0.250 (0.342)
Physician and APP	19.3 (33.2)	12.9 (20.6)	1.1 (1.8)	0.240 (0.365)
Nurse	246.5 (156.3)	191.1 (123.5)	45.0 (44.2)	0.293 (0.219)

JAMA Network Open

JAMA Network Open. 2024;7(3):e243201. doi:10.1001/jamanetworkopen.2024.3201

Theme	Representative quotations
Draft message voice and/or tone	Positive: “I was impressed by the tone that varied based on patient’s concerns and questions, and felt messaging was overall very professional and clear.” Negative: “I think the drafts are great but can further be improved if it did not sound robotic and had a more personable touch.”
Future use	Positive: “Please continue to allow us to utilize this tool and spread to other SHC clinics!” Negative: “I still think it’s a good idea but not ready for real life situations.”
Impact on time	Positive: “It helped save me a lot of time starting from scratch.” Negative: “Right now, it is just piling on top of the work that we are already doing, and it is faster for me to type a prose response that I have generated myself.”
Draft message length and/or brevity	Positive: “However, the responses are very thorough. I had a patient that needed a refill and the draft wrote out almost a whole letter when I typically would maybe just write a short sentence saying ‘Yes, I will send!’” Negative: “Overall the responses seemed unnecessarily wordy in noncontributory ways.”

This seems a bit at odds with the various other studies that mentioned improved empathy and appropriateness….. Hmmm….

Original Investigation | Health Informatics

AI-Generated Draft Replies Integrated Into Health Records and Physicians' Electronic Communication

Ming Tai-Seale, PhD, MPH; Sally L. Baxter, MD, MSc; Florin Vaida, PhD; Amanda Walker, MS; Amy M. Sitapati, MD; Chad Osborne, MD; Joseph Diaz, MD; Nimit Desai, BS; Sophie Webb, MS; Gregory Polston, MD; Teresa Helsten, MD; Erin Gross, MD; Jessica Thackaberry, MD; Ammar Mandvi, MD; Dustin Lillie, MD; Steve Li, MD; Geneen Gin, DO; Suraj Achar, MD; Heather Hofflich, DO; Christopher Sharp, MD; Marlene Millen, MD; Christopher A. Longhurst, MD, MS

JAMA Network Open

UCSD

Cross-Pollinated study with UCSF LLM paper

122 PCPs, 50 with access to GenAI inbox replies, 72 were control

Mesure:

- Reading Time 22% ↑
- Response Length 18% ↑
- Reply Time →
- Perceived benefits - Yup

Likelihood to recommend	Quote
Likelihood: 9 or 10	
Tone and value: acknowledgment of the robotic tone of AI replies, recognizing their role in initiating patient interactions, and serving as a valuable baseline.	"Though the replies sound very robotic still, they're extremely helpful for generating the baseline response to what you'd want to say to a patient."
Potential for improvement and mimicry of physician language: anticipation for AI-generated replies to improve and emulate the communication style of individual physicians, enhancing personalization and human-like interactions.	"I can't wait for them to get even better, to the point where they can mimic each physician's language/tone."
AI replies' place and role: recognition of AI-generated replies' valuable role in health care workflows, aiding in workload management and effective patient communication, and contributing to workflow efficiency.	"I think AI responses have its place. [I] worry about inaccuracies that I may miss due to busy workload. I have been very impressed with [a] few of the responses."
Hope for reduced supervision: expressing hope for AI advancements leading to reduced supervision, envisioning a future where AI can function autonomously while maintaining high-quality patient communication.	"Great initiative which requires supervision. Hopefully there would be time when minimal supervision would be needed."
Likelihood: 7 or 8	
Tone and empathy: recognition of AI-generated replies for their kind and empathetic tone, aiding in maintaining respectful and caring interactions with patients.	"Helpful in drafting responses, provides more empathy into a response without me taking time to type it all out."
Time savings and future expectations: appreciation for saving time and enhancing efficiency by initiating tailored drafts swiftly, compared with starting from scratch; optimism about future enhancements.	"Not perfect but decreases time I spend on it and has a kind tone." "While not perfect, I think there have been a good number of cases where I use the draft as a starting point. I expect the AI responses to get better over time."
Alleviating pressure to address patient concerns online only: perceived relief from the pressure of responding solely through MyChart, with AI aiding appropriate recommendations for in-person evaluations when necessary, thus lessening workload.	"AI generated messages often appropriately recommend that the patient be evaluated in person for specific concerns. Sometimes clinicians feel pressured to deal with patient concerns by MyChart alone. The use of AI generated messages can take away this pressure."
Recommendation to colleagues: general endorsement of AI-generated draft replies to colleagues, emphasizing its potential benefits in starting drafts and infusing empathy into responses, despite the need for some editing.	"I generally would recommend auto-generated draft replies to colleagues because it seems to be net even—it may be helpful to start a draft, but most of the time, I am editing the replies, so it is not completely/automatically helpful."
Likelihood: 0-6	
Tone and language: critique of AI-generated replies for being excessively polite, formal, impersonal, and not aligning with the desired direct and concise tone in patient interactions.	"Messages were too nice and wordy, but sometimes offered good advice." "Not personalized to the specific patient. I tend to personalize my response in different ways depending on the patient."
Efficiency and time savings: recognition of AI replies as helpful starting points, but often requiring extensive editing, reducing potential time savings compared with drafting from scratch.	"I found I used them most when I was covering for other providers and did not know the patients as well so did not need to provide as customized of a message."
Challenges in clinic population and workload: noting challenges in AI replies' applicability, particularly regarding urgent and specific patient needs, where recommending appointments may not be appropriate or feasible.	"It ended up creating more work for me and ultimately always recommended scheduling an appointment. Our clinic is largely populated with persons who are struggling with multimorbidity and SDOH [social determinants of health] and need urgent assistance rather than a future appointment."
Improvement and future potential: acknowledgment of AI potential, urging significant improvement in understanding patient queries, offering accurate information, and considering context for appropriate responses.	"I think it has potential, but is not anywhere near where it needs to be to be useful."

Association of Primary Care Physicians' Electronic Inbox Activity Patterns with Patients' Likelihood to Recommend the Physician

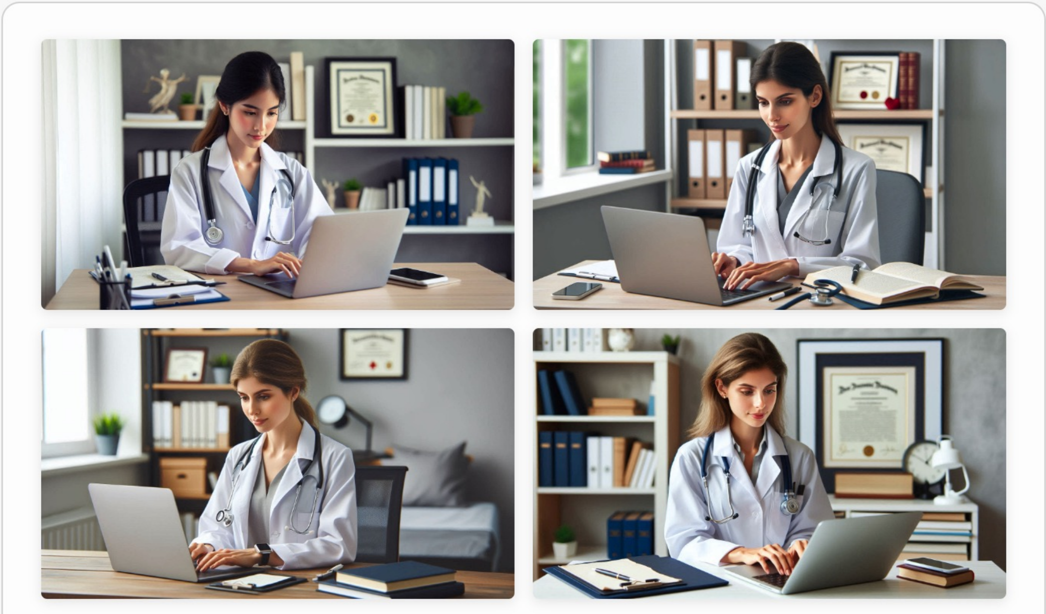
For Bill !!!!

The Brigham / Epic and Clarity

Question: Can we correlate inbox activity (responsiveness, message volume, amount of time spent in the inbox) to a patient's Likeliness to Recommend (LTR)

Answer – Yup, and women providers are better at it.

Also – messages beget more messages, and if you take too long to respond patients don't like it



DOI: 10.1007/s11606-023-08417-8

Large Language Models Are Poor Medical Coders — Benchmarking of Medical Code Querying

Authors: Ali Soroush, M.D., M.S., Benjamin S. Glicksberg, Ph.D., Eyal Zimlichman, M.D., M.Sc., Yiftach Barash, M.D., M.Sc., Robert Freeman, R.N., M.S.N., N.E.-B.C., Alexander W. Charney, M.D., Ph.D., Girish N Nadkarni, M.D., M.P.H., and Eyal Klang, M.D. Author Info & Affiliations

Published April 19, 2024 | NEJM AI 2024;1(5) | DOI: 10.1056/AIdbp2300040 | VOL. 1 NO. 5

NEJM AI

Mount Sinai

GPT 3.5, GPT 4, Gemini Pro, LLAMA 2-70b

GPT4 wins… but it's not great. ICD10 still 25% incorrect and CPT 95% incorrect

Same takeaway – don't use it yet, needs fine tuning

Coding System	Metric	GPT-3.5 Turbo (Nov)†	GPT-4 (Nov)†	Gemini Pro†	Llama2-70b Chat†
ICD-9-CM (n=200)	Incorrect codes, n (% of total)	67 (33.5%)	43 (21.5%)	131 (65.5%)	191 (95.5%)
	Valid code, % (95% CI)	95.5% (89.6%–100.0%)	93.0% (83.7%–100.0%)	82.4% (75.6%–88.5%)	55.0% (48.2%–61.8%)
	Billable code, % (95% CI)	91.0% (83.6%–97.0%)	83.7% (72.1%–93.0%)	62.6% (54.2%–71.0%)	44.5% (37.7%–51.3%)
	Equivalent match, % (95% CI)	3.0% (0.0%–7.5%)	7.0% (0.0%–16.3%)	4.6% (1.5%–8.4%)	0.5% (0.0%–1.6%)
	Generalized match, % (95% CI)	29.9% (19.4%–40.3%)	18.6% (7.0%–30.2%)	9.2% (4.6%–14.5%)	1.6% (0.0%–3.7%)
	Nonbillable code, % (95% CI)	4.5% (0.0%–10.4%)	9.3% (2.3%–18.6%)	19.8% (13.0%–26.7%)	10.5% (6.3%–15.2%)
	Fabricated code, % (95% CI)	4.5% (0.0%–10.4%)	7.0% (0.0%–16.3%)	17.6% (11.5%–24.4%)	45.0% (38.2%–52.4%)
	CodeSTS score, mean (95% CI)	1.9 (1.6–2.1)	1.9 (1.5–2.3)	1.3 (1.1–1.5)	0.4 (0.3–0.5)
ICD-10-CM (n=200)	Incorrect codes, n (% of total)	81 (40.5%)	46 (23%)	144 (72%)	173 (86.5%)
	Valid code, % (95% CI)	87.7% (80.2%–93.8%)	84.8% (73.9%–93.5%)	63.9% (56.2%–71.5%)	79.2% (72.8%–85.0%)
	Billable code, % (95% CI)	76.5% (66.7%–85.2%)	65.2% (52.2%–78.3%)	47.9% (39.6%–56.2%)	49.1% (41.6%–56.6%)
	Equivalent match, % (95% CI)	4.9% (1.2%–9.9%)	10.9% (2.2%–19.6%)	0.7% (0.0%–2.1%)	2.3% (0.6%–4.6%)
	Generalized match, % (95% CI)	18.5% (9.9%–27.2%)	13.0% (4.3%–23.9%)	5.6% (2.1%–9.7%)	7.5% (4.0%–11.6%)
	Nonbillable code, % (95% CI)	11.1% (4.9%–18.5%)	19.6% (8.7%–30.4%)	16.0% (10.4%–22.2%)	30.1% (23.1%–37.0%)
	Fabricated code, % (95% CI)	12.3% (6.2%–19.8%)	15.2% (6.5%–26.1%)	36.1% (28.5%–44.4%)	20.8% (15.0%–26.6%)
	CodeSTS score, mean (95% CI)	1.7 (1.5–2.0)	1.8 (1.4–2.2)	0.9 (0.8–1.1)	1.1 (1.0–1.3)
CPT (n=200)	Incorrect codes, n (% of total)	94.6% (89.2%–98.6%)	84.8% (72.7%–97.0%)	86.1% (80.6%–91.7%)	74.2% (67.6%–80.2%)
	Valid code, % (95% CI)	0.0% (0.0%–0.0%)	0.0% (0.0%–0.0%)	0.0% (0.0%–0.0%)	0.0% (0.0%–0.0%)
	Equivalent match, % of (95% CI)	6.8% (1.4%–13.5%)	15.2% (3.0%–27.3%)	10.4% (5.6%–16.0%)	2.7% (0.5%–5.5%)
	Fabricated code, % of (95% CI)	5.4% (1.4%–10.8%)	15.2% (3.0%–27.3%)	13.9% (8.3%–19.4%)	25.8% (19.8%–32.4%)
	CodeSTS score, mean (95% CI)	1.2 (1.0–1.4)	1.8 (1.4–2.1)	1.0 (0.8–1.2)	(0.3–0.6)

DOI: 10.1056/AIdbp2300040

Back to Bill!

More CDS !!

JAMA Internal Medicine | Original Investigation

Text vs Patient Portal Messaging to Improve Influenza Vaccination Coverage
A Health System–Wide Randomized Clinical Trial

Peter G. Szilagyi, MD, MPH; O. Kenrik Duru, MD, MSHS; Alejandra Casillas, MD, MSHS; Michael K. Ong, MD, PhD; Sitaram Vangala, MS; Chi-Hong Tseng, PhD; Christina Albertin, BSN, MPH; Sharon G. Humiston, MD, MPH; Emma Clark, MS; Mindy K. Ross, MD, MBA; Sharon A. Evans; Michael Sloyan, MPH; Craig R. Fox, PhD; Carlos Lerner, MD, MPhil

Szilagyi PG, Duru OK, Casillas A, Ong MK, Vangala S, Tseng CH, Albertin C, Humiston SG, Clark E, Ross MK, Evans SA, Sloyan M, Fox CR, Lerner C. Text vs Patient Portal Messaging to Improve Influenza Vaccination Coverage: A Health System-Wide Randomized Clinical Trial. JAMA Intern Med. 2024 May 1;184(5):519-527. doi: 10.1001/jamainternmed.2024.0001. PMID: 38497955; PMCID: PMC10949147.

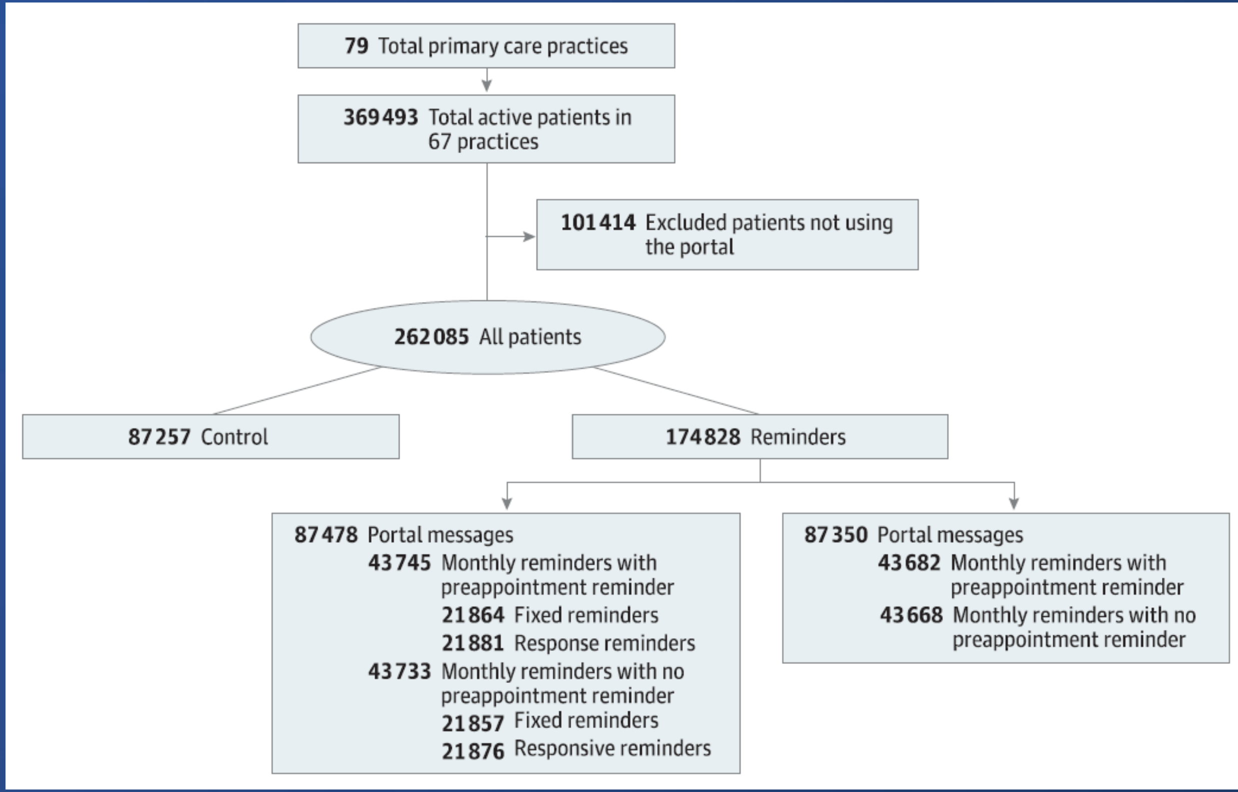


Table 3. Adjusted RRs for Influenza Vaccination by Study Group and Patient Characteristics, Using Mixed-Effects Poisson Regression Models of Vaccination Status

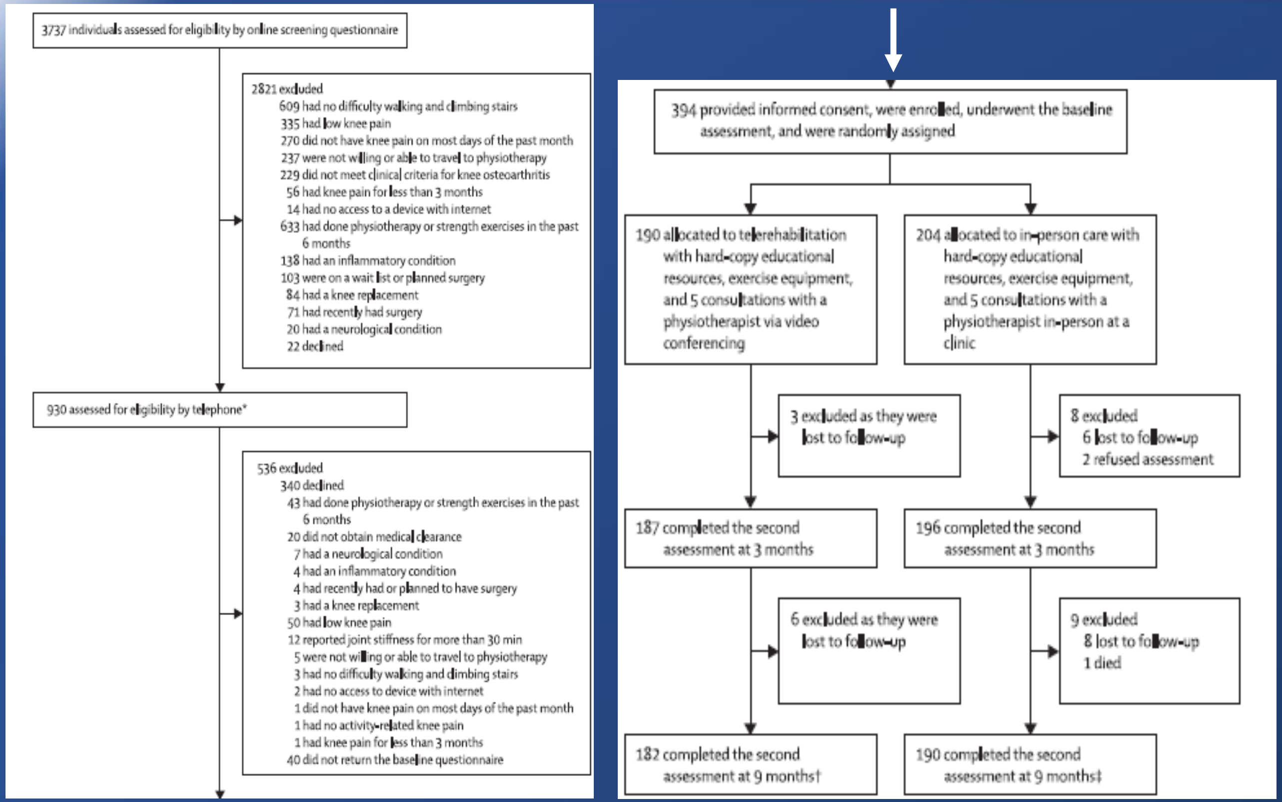
Comparison	Adjusted RR (95% CI)
Modality (reference group, control)	
Portal	0.99 (0.98-1.01)
Text	1.00 (0.98-1.01)
Preappointment reminder: yes compared with no, portal and text groups combined	1.01 (1.00-1.02)
Interactive: responsive compared with fixed (portal group only)	1.00 (0.99-1.01)

Szilagyi PG, Duru OK, Casillas A, Ong MK, Vangala S, Tseng CH, Albertin C, Humiston SG, Clark E, Ross MK, Evans SA, Sloyan M, Fox CR, Lerner C. Text vs Patient Portal Messaging to Improve Influenza Vaccination Coverage: A Health System-Wide Randomized Clinical Trial. JAMA Intern Med. 2024 May 1;184(5):519-527. doi: 10.1001/jamainternmed.2024.0001. PMID: 38497955; PMCID: PMC10949147.

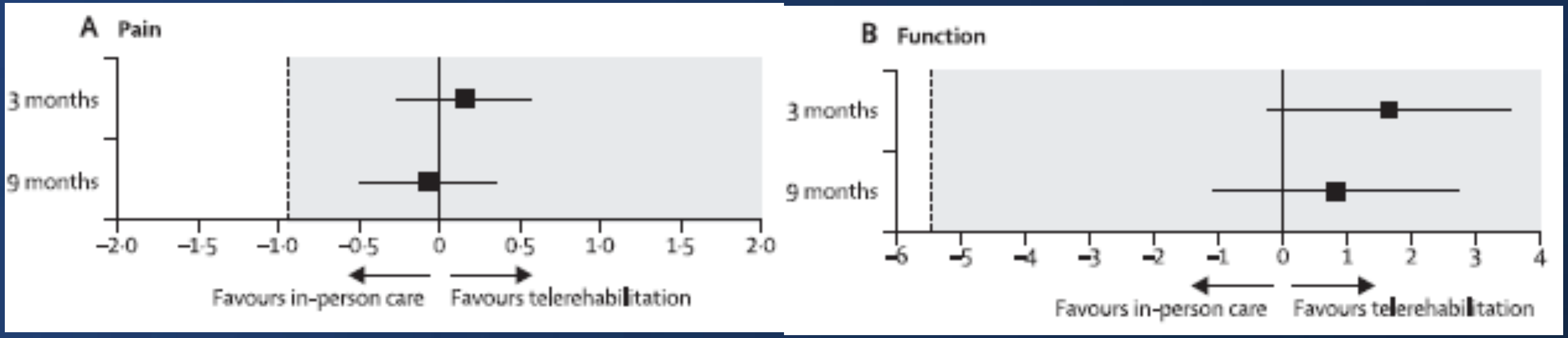
Telerehabilitation consultations with a physiotherapist for chronic knee pain versus in-person consultations in Australia: the PEAK non-inferiority randomised controlled trial

Rana S Hinman, Penny K Campbell, Alexander J Kimp, Trevor Russell, Nadine E Foster, Jessica Kasza, Anthony Harris, Kim L Bennell

Hinman RS, Campbell PK, Kimp AJ, Russell T, Foster NE, Kasza J, Harris A, Bennell KL. Telerehabilitation consultations with a physiotherapist for chronic knee pain versus in-person consultations in Australia: the PEAK non-inferiority randomised controlled trial. Lancet. 2024 Mar 30;403(10433):1267-1278. doi: 10.1016/S0140-6736(23)02630-2. Epub 2024 Mar 7. PMID: 38461844.



Results



Hinman RS, Campbell PK, Kimp AJ, Russell T, Foster NE, Kasza J, Harris A, Bennell KL. Telerehabilitation consultations with a physiotherapist for chronic knee pain versus in-person consultations in Australia: the PEAK non-inferiority randomised controlled trial. Lancet. 2024 Mar 30;403(10433):1267-1278. doi: 10.1016/S0140-6736(23)02630-2. Epub 2024 Mar 7. PMID: 38461844.

The effect of computerised decision support alerts tailored to intensive care on the administration of high-risk drug combinations, and their monitoring: a cluster randomised stepped-wedge trial

Tinka Bakker, Joanna E Klopotowska*, Dave A Dongelmans, Saeid Eslami, Wytze J Vermeijden, Stefaan Hendriks, Julia ten Cate, Attila Karakus, Ilse M Purmer, Sjoerd H W van Bree, Peter E Spronk, Martijn Hoeksema, Evert de Jonge, Nicolette F de Keizer, Ameen Abu-Hanna, on behalf of the SIMPLIFY study group†*

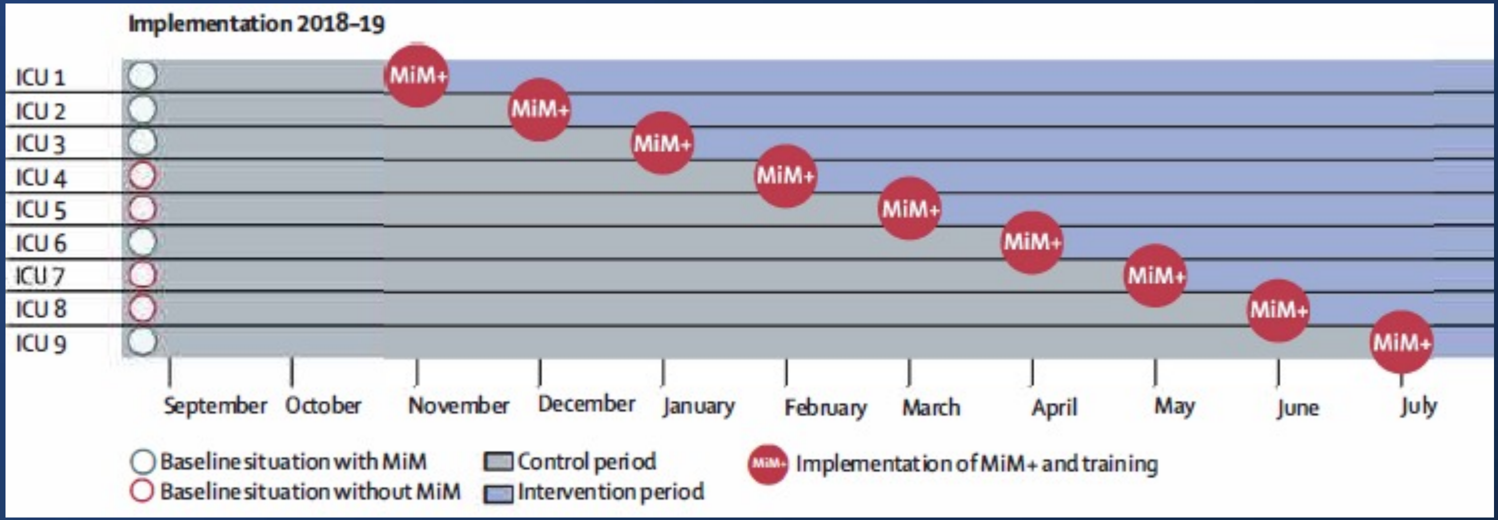
(Netherlands)

1 Analysis of potential DDIs in the ICU[15]

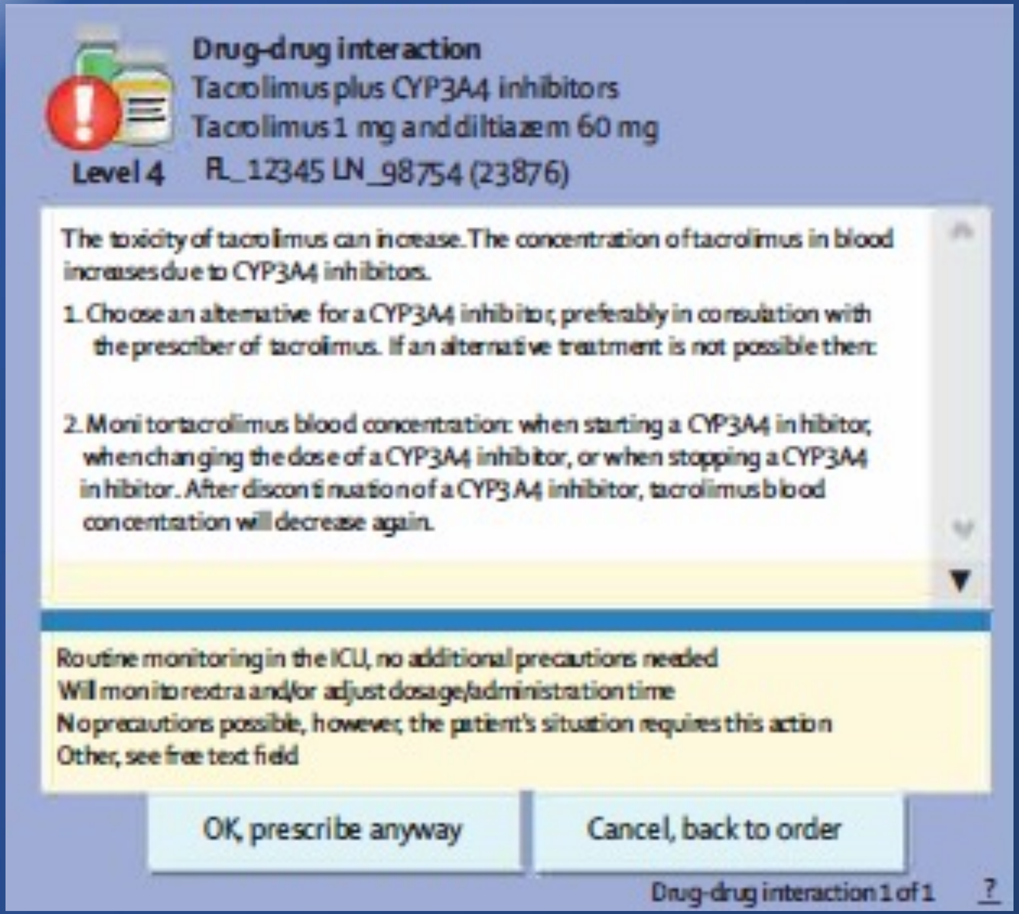
- Analysis of 228 2974 medication administrations of 103871 admissions in 13 ICUs
- Number of potential DDIs
  - Per 1000 medication administrations=mean 70·1 (SD 90·5)
  - Per admission=mean 2·2 (SD 4·1)
- Detected number of potential DDI types=270

2 Defining the clinical relevance of potential DDIs for the ICU setting[16]

- Modified Delphi procedure
  - Two rounds
  - Intensivists and hospital pharmacists
- Assessing the clinical relevance of 148 potential DDI types
  - √ With agreement=139 (94%) of 148 potential DDI types
  - √ Low-yield potential DDI types=53 (38%) of 139



Bakker T, Klopotowska JE, Dongelmans DA, Eslami S, Vermeijden WJ, Hendriks S, Ten Cate J, Karakus A, Purmer IM, van Bree SHW, Spronk PE, Hoeksema M, de Jonge E, de Keizer NF, Abu-Hanna A; SIMPLIFY study group. The effect of computerised decision support alerts tailored to intensive care on the administration of high-risk drug combinations, and their monitoring: a cluster randomised stepped-wedge trial. Lancet. 2024 Feb 3;403(10425):439-449. doi: 10.1016/S0140-6736(23)02465-0. Epub 2024 Jan 20. PMID: 38262430.



MetaVision ICU (iMDSoft) (the Netherlands)

	Variable	Estimated incidence rate ratio	95% CI lower bound	95% CI upper bound	p value
Unadjusted M0	MiM+	0·88	0·81	0·94	0·0004*
Adjusted M1	MiM+	0·86	0·80	0·92	<0·0001*
Adjusted M2	MiM+	0·88	0·82	0·95	0·0008*

Model M1 was adjusted for admission type (medical, emergency surgical, or elective surgical) and the presence of chronic obstructive pulmonary disease. Model M2 was adjusted for age, sex, admission type, Acute Physiology And Chronic Health Evaluation IV score, presence of cardiovascular disease, and presence of immunodeficiency. The result was considered significant when $p<0·05$. MiM=Medication Interaction Module. *Significant result.

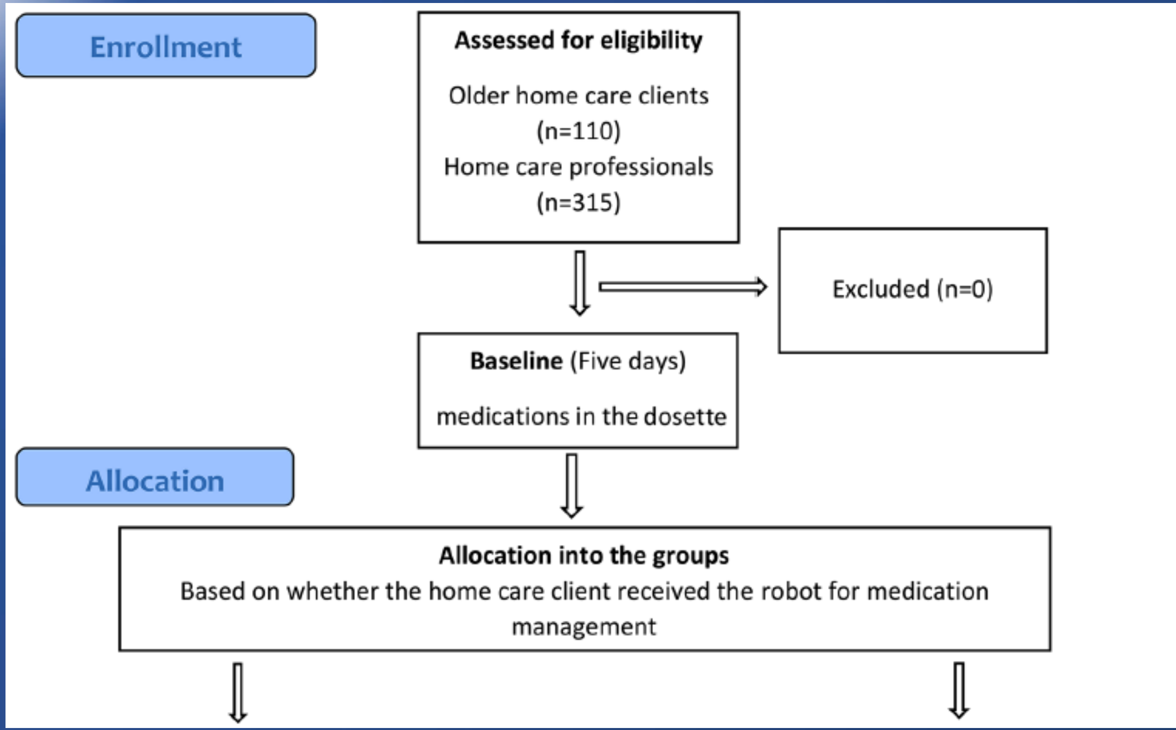
Table 2: **Output for the unadjusted and adjusted generalised linear mixed-effect models**

Bakker T, Klopotowska JE, Dongelmans DA, Eslami S, Vermeijden WJ, Hendriks S, Ten Cate J, Karakus A, Purmer IM, van Bree SHW, Spronk PE, Hoeksema M, de Jonge E, de Keizer NF, Abu-Hanna A; SIMPLIFY study group. The effect of computerised decision support alerts tailored to intensive care on the administration of high-risk drug combinations, and their monitoring: a cluster randomised stepped-wedge trial. Lancet. 2024 Feb 3;403(10425):439-449. doi: 10.1016/S0140-6736(23)02465-0. Epub 2024 Jan 20. PMID: 38262430.

RESEARCH **Open Access**

Effect of robot for medication management on home care professionals' use of working time in older people's home care: a non-randomized controlled clinical trial

Satu Kajander-Unkuri[1,2], Mojtaba Vaismoradi[3,7*], Jouko Katajisto[4], Mari Kangasniemi[1,5] and Riitta Turjamaa[6]



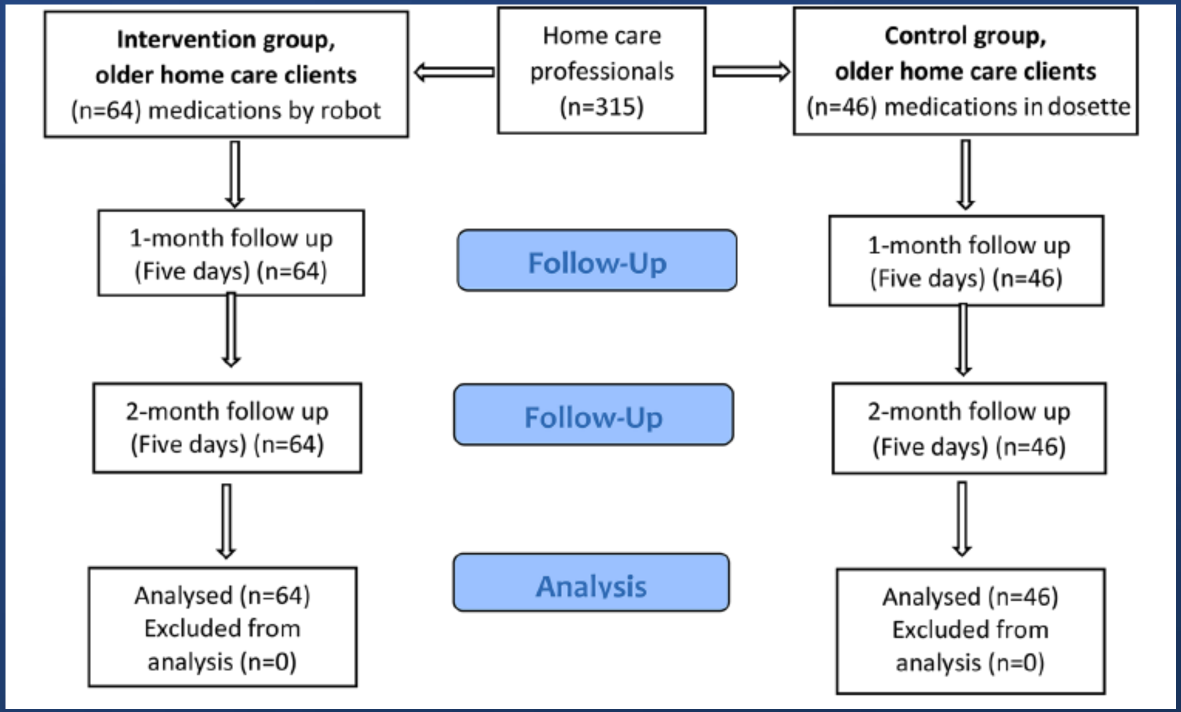


Table 3 The total working time (in minutes) used for medication management considering the number of visits per day analyzed with analysis of covariance (Sidak multiple comparisons)

Variable	IG (n = 64)					CG (n = 46)				
	Baseline (T1) mean[a] (95% CI)	1-month (T2) mean[a] (95% CI)	2-months (T3) mean[a] (95% CI)	Difference T2-T1 mean[a] (p-value)	Difference T3-T1 mean[a] (p-value)	Baseline (T1) mean[a] (95% CI)	1-month (T2) mean[a] (95% CI)	2-months (T3) mean[a] (95% CI)	Difference T2-T1 mean[a] (p-value)	Difference T3-T1 mean[a] (p-value)
The total working time used for medication management	54.2 (49.6–58.8)	40.8 (37.4–44.3)	34.9 (31.4–38.3)	-13.4 (< 0.001)*	-19.3 (< 0.001)*	75.2 (70.1–80.4)	65.0 (59.8–70.1)	74.3 (69.2–79.4)	-10.2 (0.02)*	-0.95 (0.992)

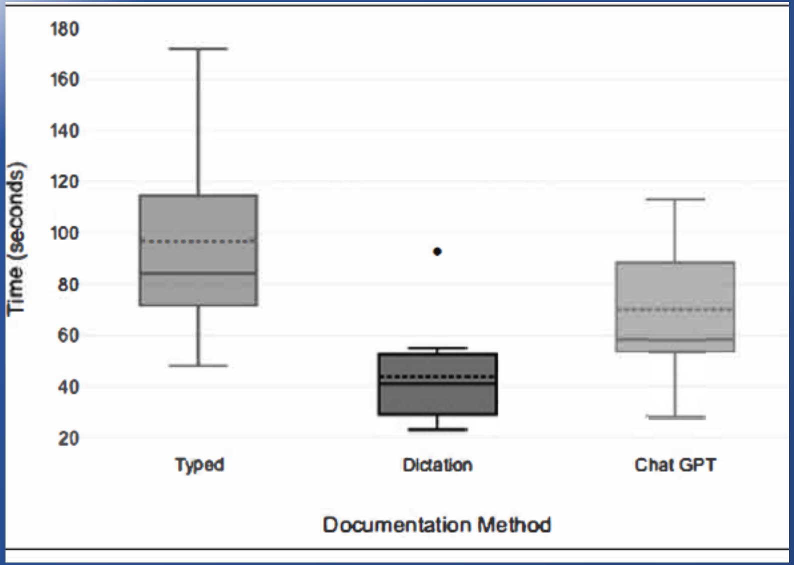
Kajander-Unkuri S, Vaismoradi M, Katajisto J, Kangasniemi M, Turjamaa R. Effect of robot for medication management on home care professionals' use of working time in older people's home care: a non-randomized controlled clinical trial. BMC Health Serv Res. 2023 Dec 2;23(1):1344. doi: 10.1186/s12913-023-10367-0. Erratum in: BMC Health Serv Res. 2024 Jan 15;24(1):75. doi: 10.1186/s12913-024-10584-1. PMID: 38042773; PMCID: PMC10693699.

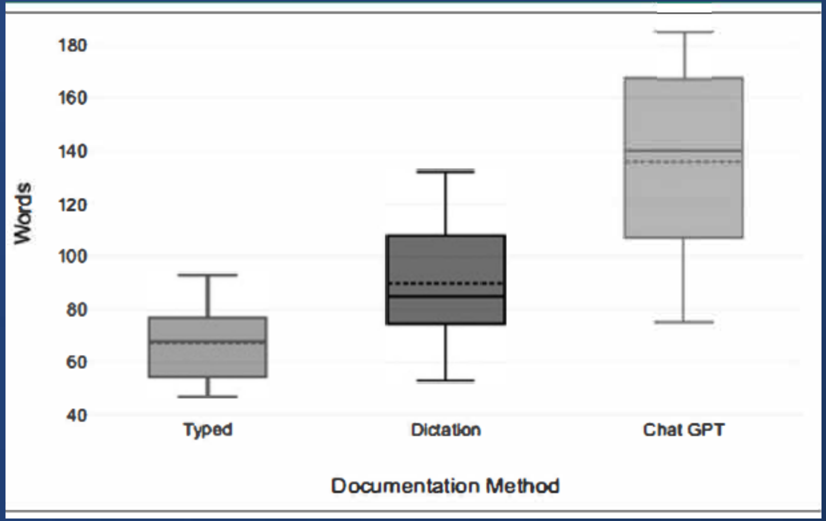
Research Article

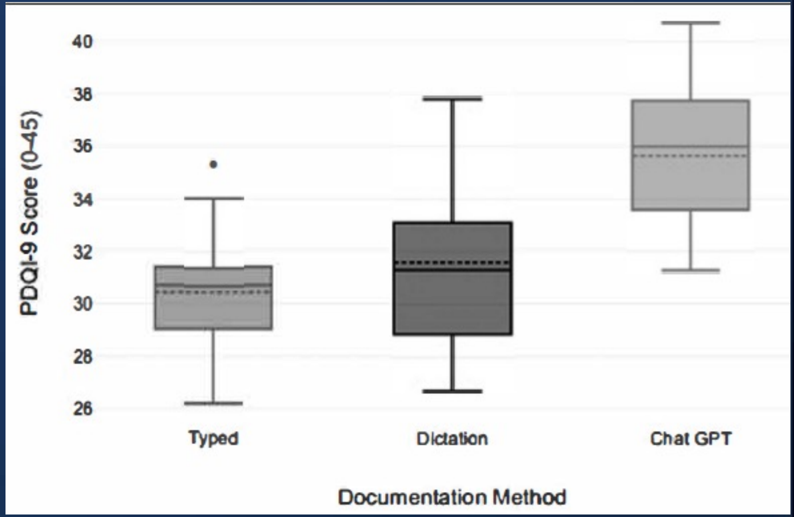
ChatGPT's Ability to Assist with Clinical Documentation: A Randomized Controlled Trial

- Orthopedics
- 4 med students, 2 PGY3, 2 PGY4, 3 Attendings
- HPI from standardized patients
- ChatGPT vs. Dictation vs. Typing

Baker HP, Dwyer E, Kalidoss S, Hynes K, Wolf J, Strelzow JA. ChatGPT's Ability to Assist with Clinical Documentation: A Randomized Controlled Trial. J Am Acad Orthop Surg. 2024 Feb 1;32(3):123-129. doi: 10.5435/JAAOS-D-23-00474. Epub 2023 Nov 17. PMID: 37976385.







Baker HP, Dwyer E, Kalidoss S, Hynes K, Wolf J, Strelzow JA. ChatGPT's Ability to Assist with Clinical Documentation: A Randomized Controlled Trial. J Am Acad Orthop Surg. 2024 Feb 1;32(3):123-129. doi: 10.5435/JAAOS-D-23-00474. Epub 2023 Nov 17. PMID: 37976385.

Telehealth-Guided Virtual Reality for Recovery of Upper Extremity Function Following Stroke

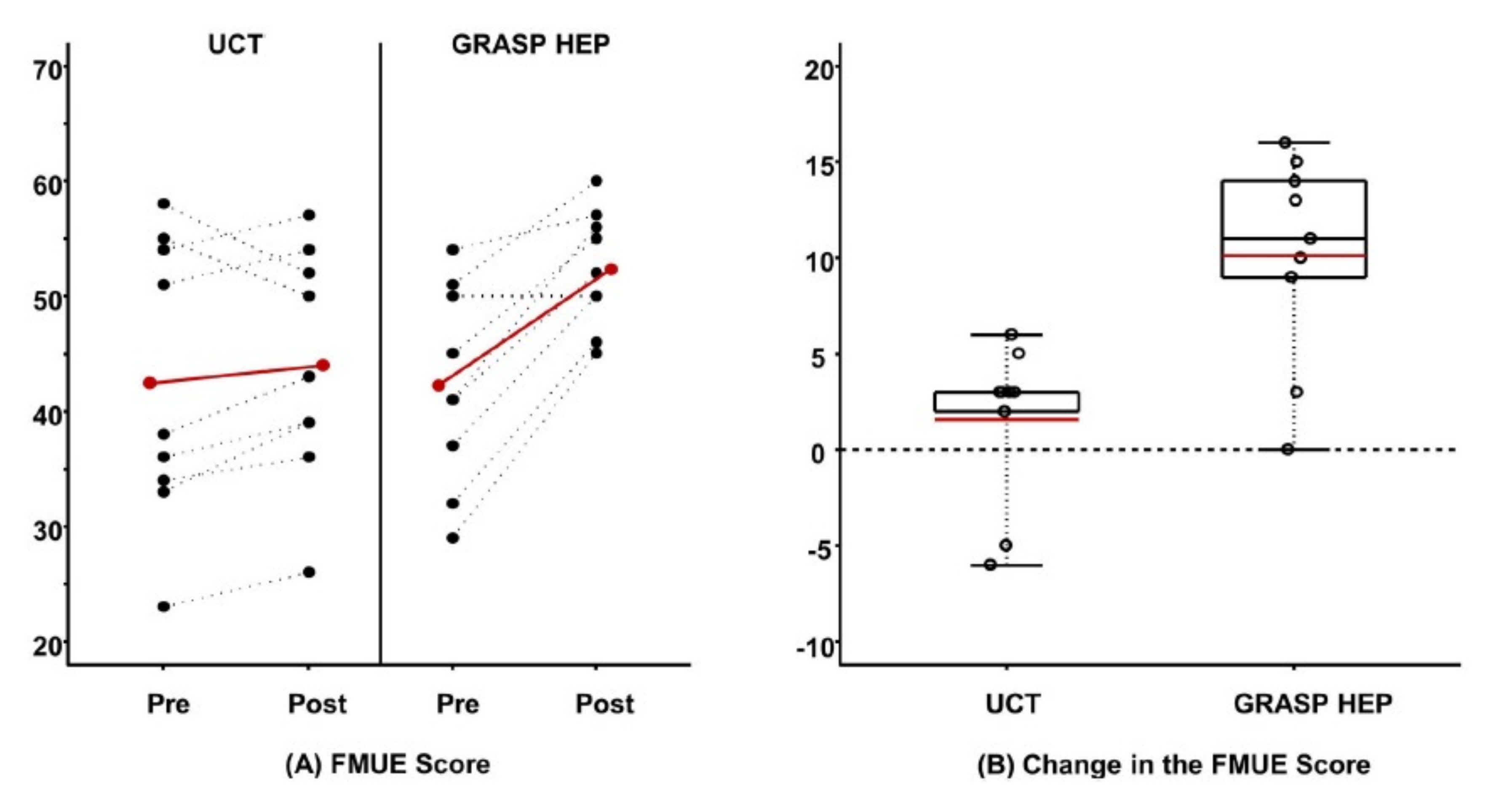
Richard J. Adams, PhD[1], Allison L. Ellington, OTD, OTR/L[2], Kate A. Kuccera, MSOT, OTR/L[3], Hannah Leaman, OTD, OTR/L[4], Catherine Smithson, OTL[5], James T. Patrie, MS[6]

Adams RJ, Ellington AL, Kuccera KA, Leaman H, Smithson C, Patrie JT. Telehealth-Guided Virtual Reality for Recovery of Upper Extremity Function Following Stroke. OTJR (Thorofare N J). 2023 Jul;43(3):446-456. doi: 10.1177/15394492231158375. Epub 2023 Mar 24. PMID: 36960762; PMCID: PMC10499117.

Gardening Program

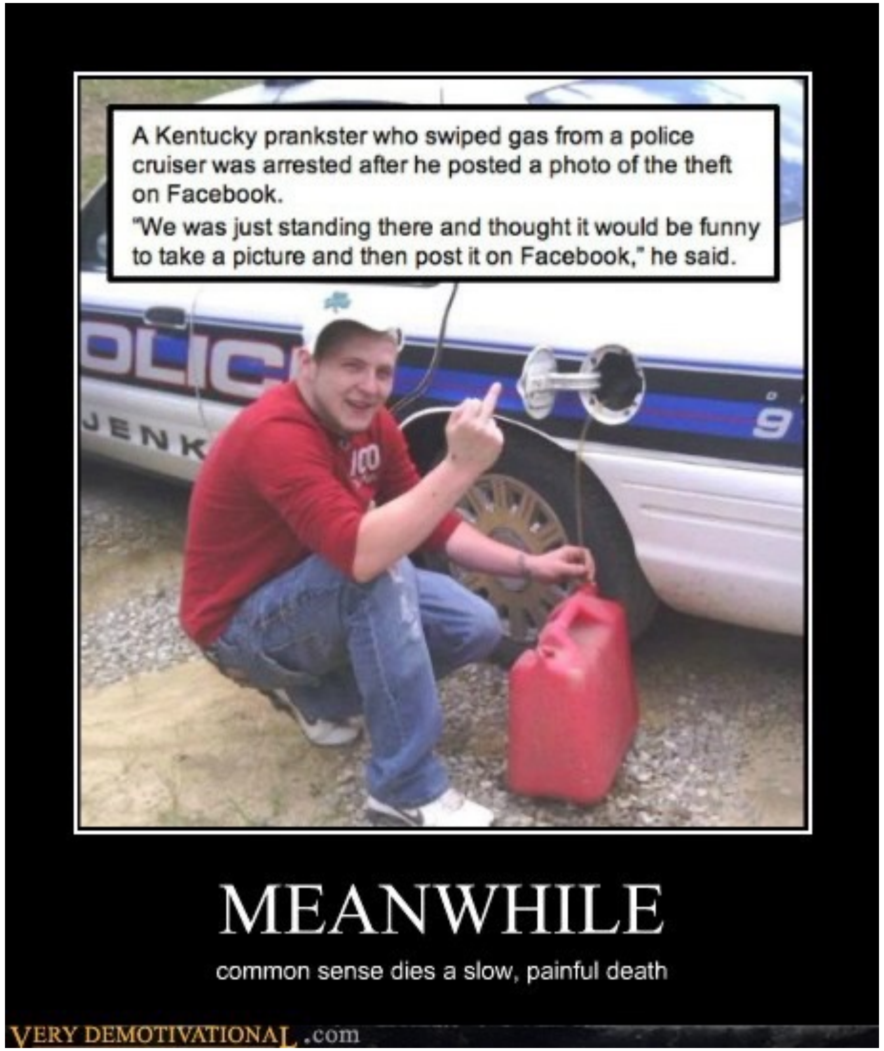
Adams RJ, Ellington AL, Kuccera KA, Leaman H, Smithson C, Patrie JT. Telehealth-Guided Virtual Reality for Recovery of Upper Extremity Function Following Stroke. OTJR (Thorofare N J). 2023 Jul;43(3):446-456. doi: 10.1177/15394492231158375. Epub 2023 Mar 24. PMID: 36960762; PMCID: PMC10499117.

Mike Tyson Program



Adams RJ, Ellington AL, Kuccera KA, Leaman H, Smithson C, Patrie JT. Telehealth-Guided Virtual Reality for Recovery of Upper Extremity Function Following Stroke. OTJR (Thorofare N J). 2023 Jul;43(3):446-456. doi: 10.1177/15394492231158375. Epub 2023 Mar 24. PMID: 36960762; PMCID: PMC10499117.

The Common Sense Section



Papers that caught my eye but also seem like pretty common sense

Original Investigation | **Health Informatics**

Inpatient EHR User Experience and Hospital EHR Safety Performance

David C. Classen, MD, MS; Christopher A. Longhurst, MD, MS; Taylor Davis, MSStat, MBA; Julia Adler Milstein, PhD; David W. Bates, MD, MSc

The data gift that keeps on giving: Leapfrog and ARCH Collaborative

Does perceived EHR usability correlate with simulated safety scores?

Example Leapfrog

- DDI / DA / Therapeutic Dup

Example ARCH (1-5)

- This EHR is easy to use?
- This EHR enables high quality care?
- This EHR is integrated

"This means that a 1-point increase in the ARCH EHR Experience score (the difference between a clinician reporting that they agree vs strongly agree that the EHR was usable, efficient, integrated, and so forth) was associated with a 1.1 percentage point increase in overall Leapfrog Safety score "

Table 5. Models of the Association of the Component Leapfrog Scores (Dependent Variables) With the Overall KLAS Experience Average Score (Primary Independent Variable)

Dependent variable: Leapfrog electronic health record component score	β (95% CI)	*P* value
Drug-route	0.013 (0.006 to 0.020)	<.001
Drug-allergy	0.008 (0.002 to 0.014)	<.001
Therapeutic duplication	0.029 (0.014 to 0.045)	<.001
Drug-dose daily	0.018 (0.006 to 0.031)	<.001
Drug-diagnosis	−0.008 (−0.024 to 0.009)	.15
Drug-age	0.021 (0.005 to 0.036)	<.001
Drug-drug interaction	0.047 (0.031 to 0.062)	<.001
Drug-dose single	0.020 (0.009 to 0.032)	<.001
Drug-laboratory	−0.014 (−0.027 to −0.001)	.001
Drug-monitoring	−0.016 (−0.029 to −0.003)	<.001

JAMA Network Open. 2023;6(9):e2333152. doi:10.1001/jamanetworkopen.2023.33152

JAMA Network Open

Original Investigation | Health Informatics

Virtual Scribes and Physician Time Spent on Electronic Health Records

Lisa Rotenstein, MD, MBA, MSc; Edward R. Melnick, MD, MHS; Christine Iannaccone, MPH; Jianyi Zhang, PhD; Aqsa Mugal, BA; Stuart R. Lipsitz, PhD; Michael J. Healey, MD; Christopher Holland, MBA; Richard Snyder, MBA; Christine A. Sinsky, MD; David Ting, MD; David W. Bates, MD, MSc

Mass General 2020 – 2022

144 Docs using synchronous and asynchronous scribe services (Nuance / Speke /Scribble / Scribe America)

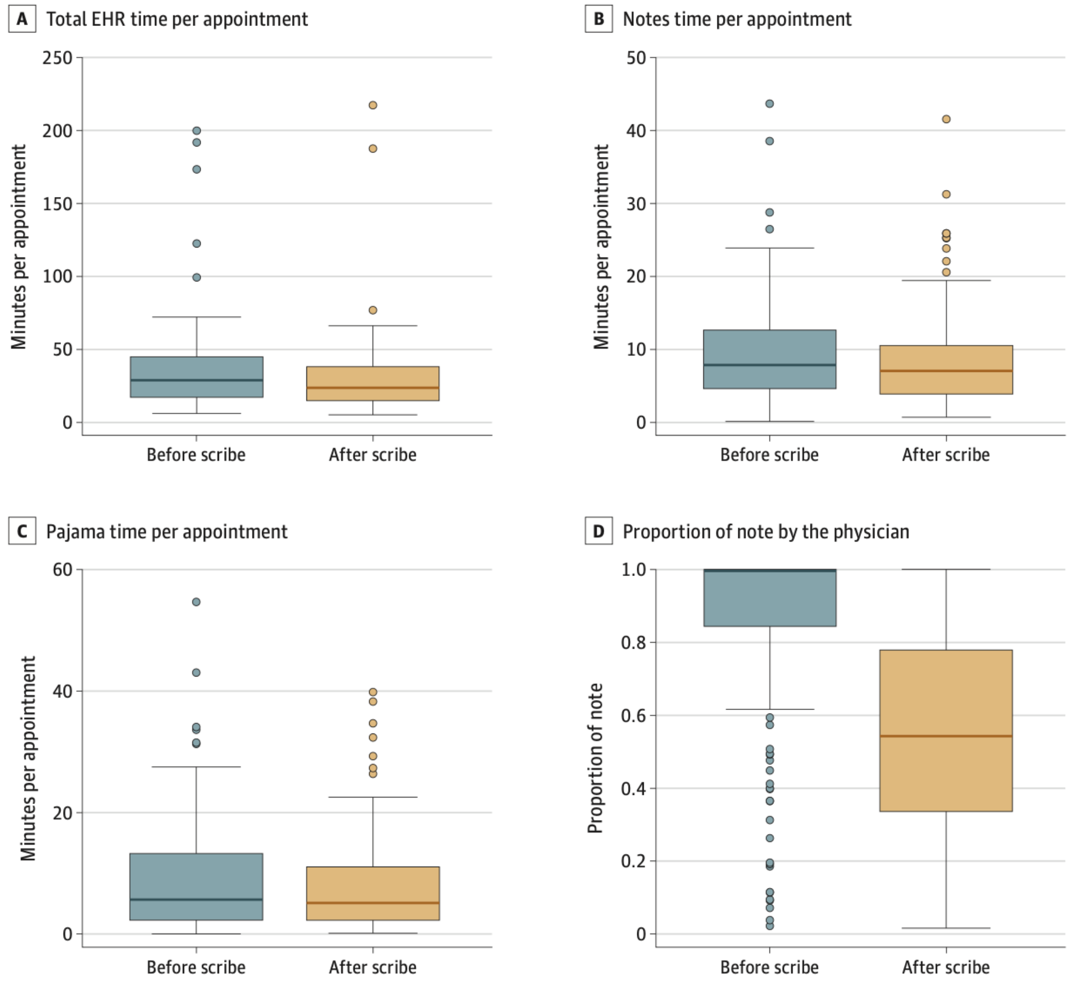
~ 50% PCP / 50 % Specialists

Impact on total EHR time? Time on Notes? Pajama Time? Proportion of note by physician? *YUP*

Impact on length of visit? *No*

Impact on proportion of orders placed by team? *NOPE*

Figure 1. Three-Month Change in Electronic Health Record (EHR) Metrics With Scribe Use for Overall Cohort



JAMA Network Open. 2024;7(5):e2413140. doi:10.1001/jamanetworkopen.2024.13140 (Reprinted)

JAMA Network Open

Original Investigation | Health Informatics

Real-Time Electronic Patient Portal Use Among Emergency Department Patients

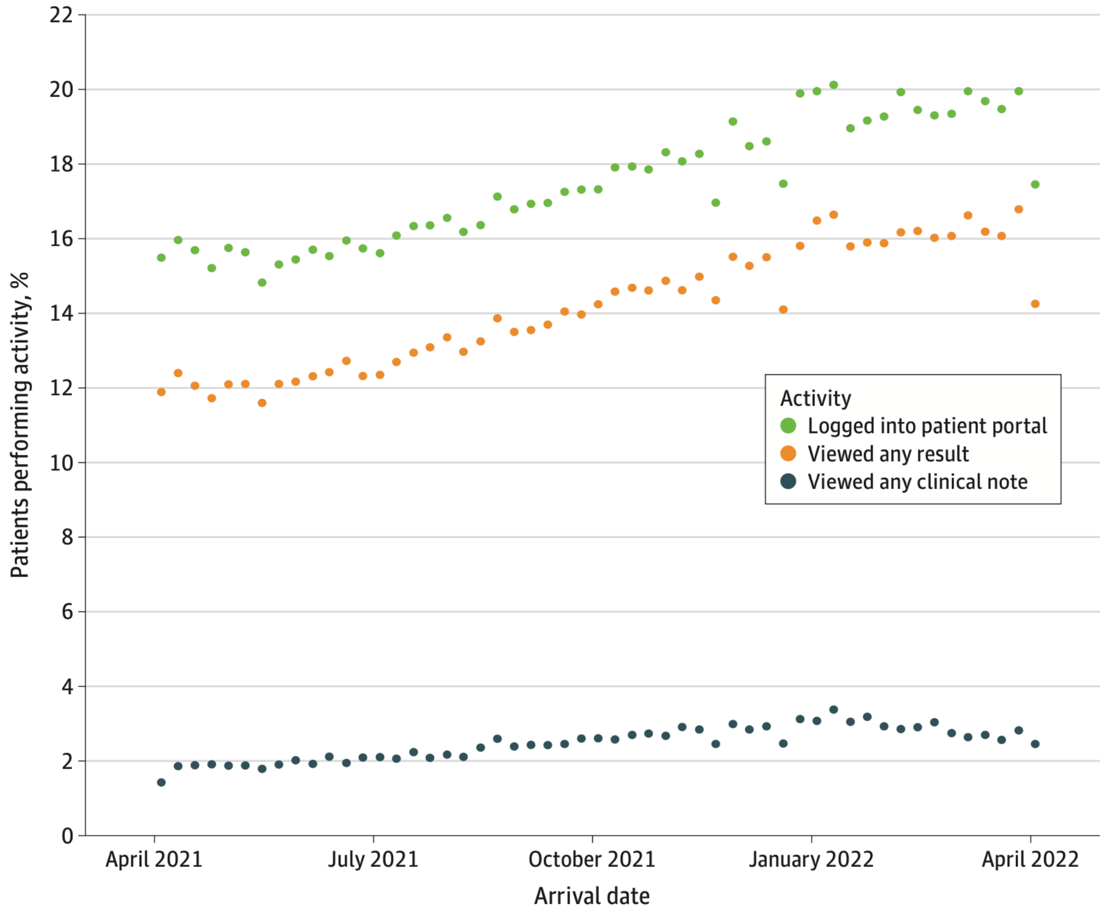
Robert W. Turer, MD; Samuel A. McDonald, MD; Christoph U. Lehmann, MD; Bhaskar Thakur, PhD; Sayon Dutta, MD, MPH; Richard A. Taylor, MD, MHS; Christian C. Rose, MD; Adam Frisch, MD; Kristian Feterik, MD; Craig Norquist, MD; Carrie K. Baker, DO; Jeffrey A. Nielson, MD; David Cha, MD; Brian Kwan, MD; Christian Dameff, MD; James P. Killeen, MD; Michael K. Hall, MD; Robert C. Doerning, MD; S. Trent Rosenbloom, MD; Casey Distaso, MD; Bryan D. Steitz, PhD

Are patients accessing their patient portal in real-time while being seen in the ED?

YUP

Looking at labs and occasionally notes

Figure 1. Temporal Trends of Real-Time Emergency Department Patient Portal Use



National Trends in Billing Secure Messages as E-Visits

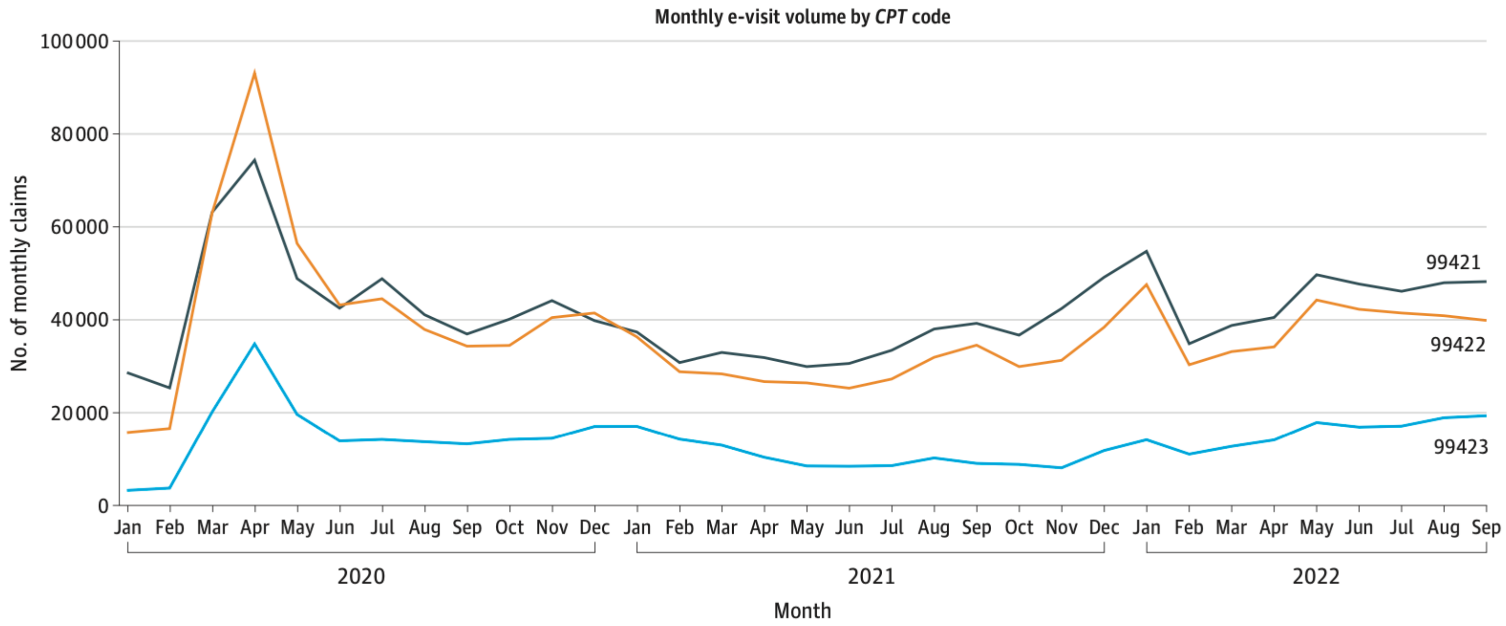
UCSF study

Utilizing the all payers claims database

Hot topic for AMDIS since COVID-19

Previous papers regarding patient and provider perception of billing for secure messaging

Figure. E-Visit Volume by *Current Procedural Terminology* Code and Number of Care Delivery Organizations Billing



- Billing peaked at the onset of the pandemic, then fell, and is now rebounding slowly
- Most common CPT codes? Exactly what you'd expect
- ? Organizations embracing this as potential long-term revenue generation

Table. Top 10 Diagnosis Codes Associated With Billed E-Visits, by *Current Procedural Terminology* Code (N = 3 068 367)

Current Procedural Terminology Code					
99421 (5-10 min)		**99422 (11-20 min)**		**99423 (≥21 min)**	
Diagnosis	**No. (%)**	**Diagnosis**	**No. (%)**	**Diagnosis**	**No. (%)**
Acute sinusitis	97 530 (7.1)	Acute respiratory infection	52 674 (4.2)	Essential (primary) hypertension	81 506 (18.0)
Urinary tract infection	96 931 (7.0)	Acute sinusitis	50 940 (4.1)	Encounter for general adult medical examination without abnormal findings	58 646 (13.0)
Acute respiratory infection	61 225 (4.5)	Essential (primary) hypertension	47 459 (3.8)	Contact with and (suspected) exposure to other viral communicable diseases	15 562 (3.4)
Essential (primary) hypertension	40 487 (2.9)	COVID-19	43 420 (3.5)	Contact with and (suspected) exposure to COVID-19	11 494 (2.5)
COVID-19	36 350 (2.6)	Acute pharyngitis	28 030 (2.3)	Encounter for observation for suspected exposure to other biological agents ruled out	11 482 (2.5)
Contact with and (suspected) exposure to other viral communicable diseases	33 770 (2.5)	Urinary tract infection	26 151 (2.1)	Alcohol dependence, uncomplicated	10 209 (2.3)
Contact with and (suspected) exposure to COVID-19	32 907 (2.4)	Acute cystitis without hematuria	22 281 (1.8)	COVID-19	7236 (1.6)
Candidiasis of vulva and vagina	29 982 (2.2)	Cough	21 213 (1.7)	Type 2 diabetes with hyperglycemia	7064 (1.6)
Acute pharyngitis	25 667 (1.9)	Rash and other nonspecific skin eruption	20 206 (1.6)	Type 2 diabetes without complications	5456 (1.2)
Cough	23 887 (1.7)	Viral infection, unspecified	17 434 (1.4)	Chronic pain syndrome	4007 (0.9)

Research Letter | Medical Education

Accuracy of Chatbots in Citing Journal Articles

Anjun Chen, PhD; Drake O. Chen, BS

Hey ChatGPT 3.5 and 4 give me some citations on these Learning Health System topics

Oops….

3.5 = 98% fake

4 – 20% fake

Don't do it folks, not yet

Table 1. Examples of a Sequence of Prompts to Engage GPT Chatbots for Discussing LHS Topics

Topic[a]	Order[b]	Prompt[c]
LHS	1	LHS vision will transform our health care systems. What is LHS? Provide some journal articles for LHS as reference.
Clinical study	2	LHS embeds clinical research in care delivery. Provide some journal articles for embedded clinical studies.
Clinical study	3	In LHS, I can conduct observational studies. Give me 10 journal articles on observational studies.
Data	4	LHS uses terminology standards for patient data. Provide 10 journal articles on medical terminology standards.
Data	5	UMLS integrates all standard vocabularies. Please provide 10 journal articles on UMLS standard.
ML	6	ML can build risk prediction models. Provide 10 journal articles for machine learning risk prediction models.
ML	7	XGBoost is a common ML algorithm. Provide 10 journal articles for XGBoost risk prediction models.
ML	8	ML can use EHR data. Provide 8 journal articles for stroke risk prediction models using EHR data.
ML	9	Deploying ML model is challenging, right? Give me some journal articles about deployment of risk prediction ML models.
Regulation	10	Deploying ML models in health care is regulated. Do you have some journal articles about regulation of using risk prediction ML or AI models in clinical settings.

Table 2. Fake Journal Article References Cited by ChatGPT

	ChatGPT model	
Fact-checking	GPT-4	GPT-3.5
Total No. of articles checked	257	162
No. of fake articles	53	159
Error rate (95% CI), %[a]	20.6 (15.8-26.1)	98.1 (94.7-99.6)
Example of fake journal articles[b]	Kesselheim AS, Cresswell K. Implementing learning health systems in the UK NHS. *BMJ*. 2017; 357: j2449. doi:10.1136/bmj.j2449	Rubin JC et al. Building a learning health system: challenges and opportunities. *J Am Med Inform Assoc*. 2015.
	Niska R, Hane CA, Castillo RC. Development and validation of the XGBoost prediction model for stroke risk: a large-scale electronic health record-based cohort study. *J Stroke Cerebrovasc Dis*. 2018;27(9):2413-2422. doi:10.1016/j.jstrokecerebrovasdis.2018.04.010	Chen et al. Integrating patient graphs and knowledge graphs for lung cancer risk factor identification. *J Biomed Inform*. 2022.

The END ?

billg@uic.edu

@colin_banas
cbanas@drfirst.com