Accelerator for Clinical Transformation
Mass General Brigham

Generative Artificial Intelligence in Clinical Trials: A Driver of Efficiency and Democratization of Care

Alexander J. Blood, MD, MSc[1,2,3]
Associate Director, Accelerator for Clinical Transformation
Instructor of Medicine, Harvard Medical School

1 Accelerator for Clinical Transformation, Cardiovascular Division, Mass General Brigham, Somerville, MA, USA
2Brigham and Women's Hospital, 75 Francis St. Boston, MA, USA
3 Harvard Medical School, 25 Shattuck St., Boston, MA 02115

Disclosures

Grants: Boehringer Ingelheim, Novo Nordisk, Eli Lilly, General Electric Health, Merck

Consulting: Novo Nordisk, Corcept Therapeutics, Milestone Therapeutics, NODE Health, Alnylam Pharma, Color Health, Medscape, Nference Inc., Walgreens Health, Withings, HelloHeart

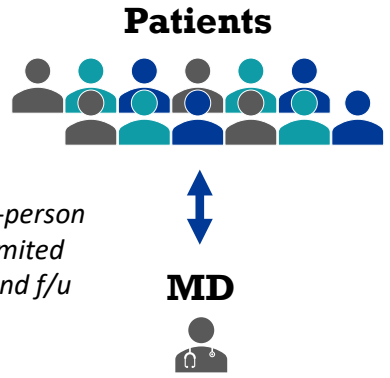
Equity: AIwithCare, Knownwell Health, Porter Health

Patents: Retrieval-augmented generation for medical data (pending)

Redesigned Care Model is Highly Scalable, Lower Cost

Traditional Care Model

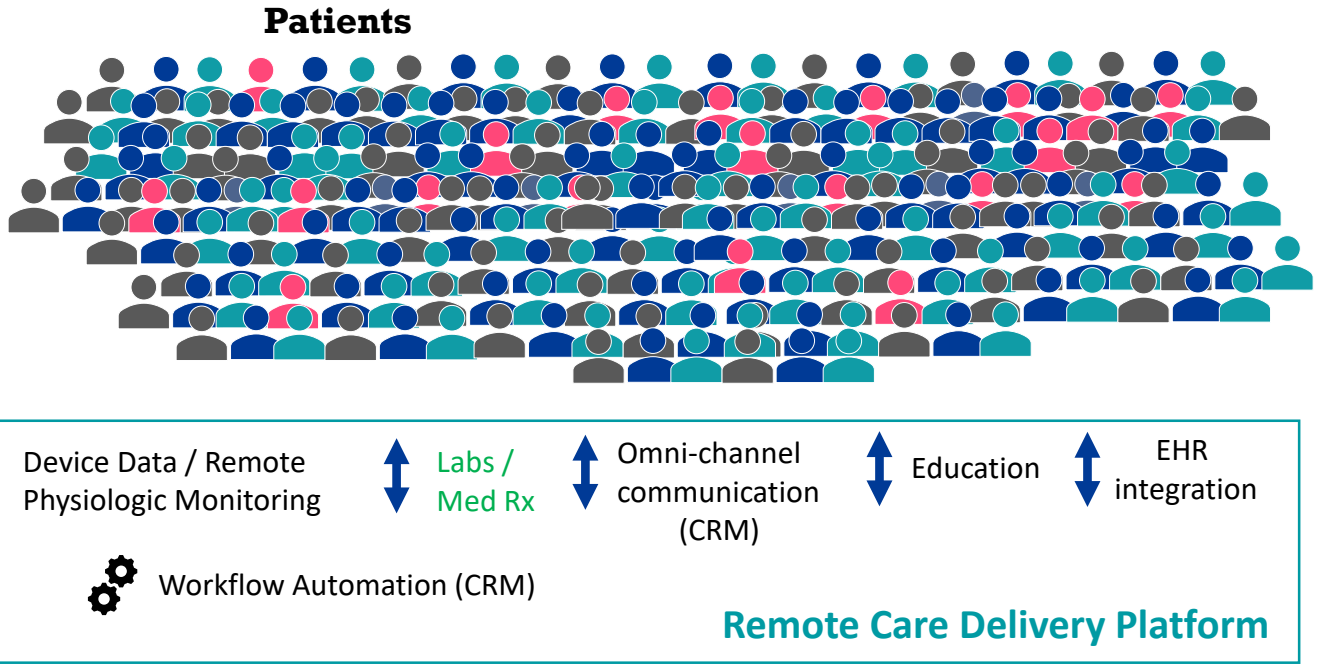
Expensive and unscalable



Limited to F2F (in-person or virtual) but limited time, cadence, and f/u

Redesigned Care Model for Remote Management

Highly scalable, cost-efficient, & enabled by technology at every step



THIEME OPEN ACCESS Special Section on Workflow Automation 1041

Workflow Automation for a Virtual Hypertension Management Program

Demonstrated Success

Research

JAMA Cardiology | Original Investigation

Results of a Remotely Delivered Hypertension and Lipid Program in More Than 10 000 Patients Across a Diverse Health Care Network

Alexander J. Blood, MD, MSc; Christopher P. Cannon, MD; William J. Gordon, MD; Charlotte Mailly, BS; Taylor MacLean, MS; Samantha Subramaniam, BA; Michela Tucci, MPH; Jennifer Crossen, PharmD; Hunter Nichols, PharmD; Kavishwar B. Wagholikar, MBBS, PhD; David Zelle, BA; Marian McPartlin, BA; Lina S. Matta, PharmD, MPH; Michael Oates; Samuel Aronson, ALM, MA; Shawn Murphy, MD, PhD; Adam Landman, MD; Naomi D. L. Fisher, MD; Thomas A. Gaziano, MD, MSc; Jorge Plutzky, MD; Benjamin M. Scirica, MD, MPH

Invited Commentary

CLINICAL STUDY DESIGN

CLINICAL CARDIOLOGY WILEY

A remote hypertension management program clinical algorithm

Hunter Nichols PharmD[1,2,3] | Christopher P. Cannon MD[1,3,4] | Benjamin M. Scirica MD, MPH[1,3,4] | Naomi D. L. Fisher MD[3,4,5]

Circulation

RESEARCH LETTER

Digital Care Transformation

Interim Report From the First 5000 Patients Enrolled in a Remote Algorithm-Based Cardiovascular Risk Management Program to Improve Lipid and Hypertension Control

JAMA Cardiology | Brief Report

Remote Optimization of Guideline-Directed Medical Therapy in Patients With Heart Failure With Reduced Ejection Fraction

Akshay S. Desai, MD; Taylor Maclean, MS; Alexander J. Blood, MD; Joshua Bosque-Hamilton, BA; Jacqueline Dunning, PharmD; Christina Fischer, MS; Liliana Fera, MSN; Katelyn V. Smith, PharmD; Kavishwar Wagholikar, MBBS, PhD; David Zelle, BA; Thomas Gaziano, MD; Jorge Plutzky, MD; Benjamin Scirica, MD; Calum A. MacRae, MD, PhD

Circulation

Circulation. 2024; [published online ahead of print]. DOI: 10.1161/CIRCULATIONAHA.124.069494

Randomized evaluation of a remote management program to improve guideline-directed medical therapy: The Diabetes Remote Intervention to improVe use of Evidence-based medications (DRIVE) Trial

Alexander J. Blood, MD, MSc; Lee-Shing Chang, MD; Shahzad Hassan, MBBS; Jacqueline Chasse, NP; Gretchen Stern, PharmD; Daniel Gabovitch, MBA; David Zelle, BA; Caitlin Colling, MD; Samuel J. Aronson, ALM, MA; Christian Figueroa, BS; Emma Collins, BS; Ryan Ruggiero, BA, AS; Emily Zacherle, MS, MBA; Joshua Noone, PhD; Carey Robar, MD; Jorge Plutzky, MD; Thomas A. Gaziano, MD, MSc; Christopher P. Cannon, MD; Deborah J. Wexler, MD, MSc; Benjamin M. Scirica, MD, MPH

Circulation
https://www.ahajournals.org/DOI: 10.1161/CIRCULATIONAHA.124.069494

Research Letter | AI in Medicine

JAMA

February 17, 2025

Manual vs AI-Assisted Prescreening for Trial Eligibility Using Large Language Models—A Randomized Clinical Trial

Ozan Unlu, MD[1,2,3]; Matthew Varugheese, MS[1,4,5]; Jiyeon Shin, BS, ALM[1,4,5]; Samantha M. Subramaniam, BA[1,2]; David Walter Jacques Stein, PhD, MS[2,3]; John J. St Laurent, BSEE[1,4]; Charlotte J. Mailly, BA[1,4,5]; Marian J. McPartlin, BA[1,6]; Fei Wang, MSc[1,4,5]; Michael F. Oates[1,5,6]; Christopher P. Cannon, MD[1,2,3]; Benjamin M. Scirica, MD, MPH[1,2,3]; Kavishwar B. Wagholikar, MBBS, PhD[1,7]; Samuel J. Aronson, MA, ALM[1,4,5]; Alexander J. Blood, MD, MSc[1,2,3]

» Author Affiliations | Article Information

JAMA. Published online February 17, 2025. doi:10.1001/jama.2024.28047

NEJM AI

ORIGINAL ARTICLE

Retrieval-Augmented Generation–Enabled GPT-4 for Clinical Trial Screening

Ozan Unlu, M.D.,[1,2,3,4] Jiyeon Shin, A.L.M.,[1,5] Charlotte J. Mailly, B.A.,[1,5] Michael F. Oates,[1,5] Michela R. Tucci, M.P.H.,[1] Matthew Varugheese, M.S.,[1] Kavishwar Wagholikar, M.B.B.S., Ph.D.,[1,4] Fei Wang, M.Sc.,[1,5] Benjamin M. Scirica, M.D., M.P.H.,[1,2,4] Alexander J. Blood, M.D., M.Sc.,[1,2,4] and Samuel J. Aronson, A.L.M., M.A.[1,5]

American Heart Association. Scientific Sessions

Digital Care Transformation: Report from the First 10,000 Patients Enrolled in a Remote Algorithm-based Cardiovascular Risk Management Program to Improve Lipid and Hypertension Control

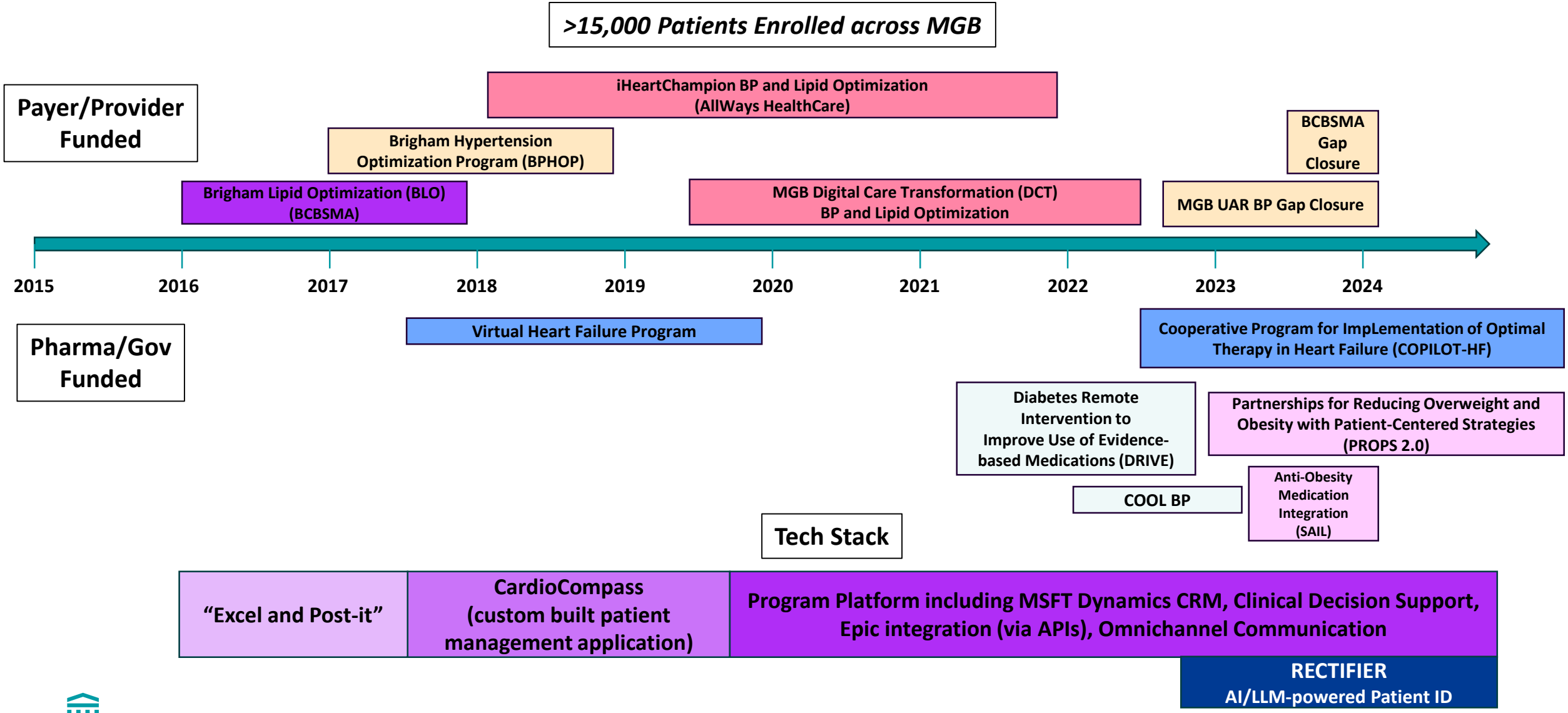
WILEY CLINICAL CARDIOLOGY

CLINICAL INVESTIGATIONS

Development of an entirely remote, non-physician led hypertension management program

Naomi D.L. Fisher | Liliana E. Fera | Jacqueline R. Dunning | Sonali Desai | Lina Matta | Victoria Liquori | Jaclyn Pagliaro | Erika Pabo | Mary Merriam | Calum A. MacRae | Benjamin M. Scirica

Accelerator for Clinical Transformation – the 1st Decade



Challenges

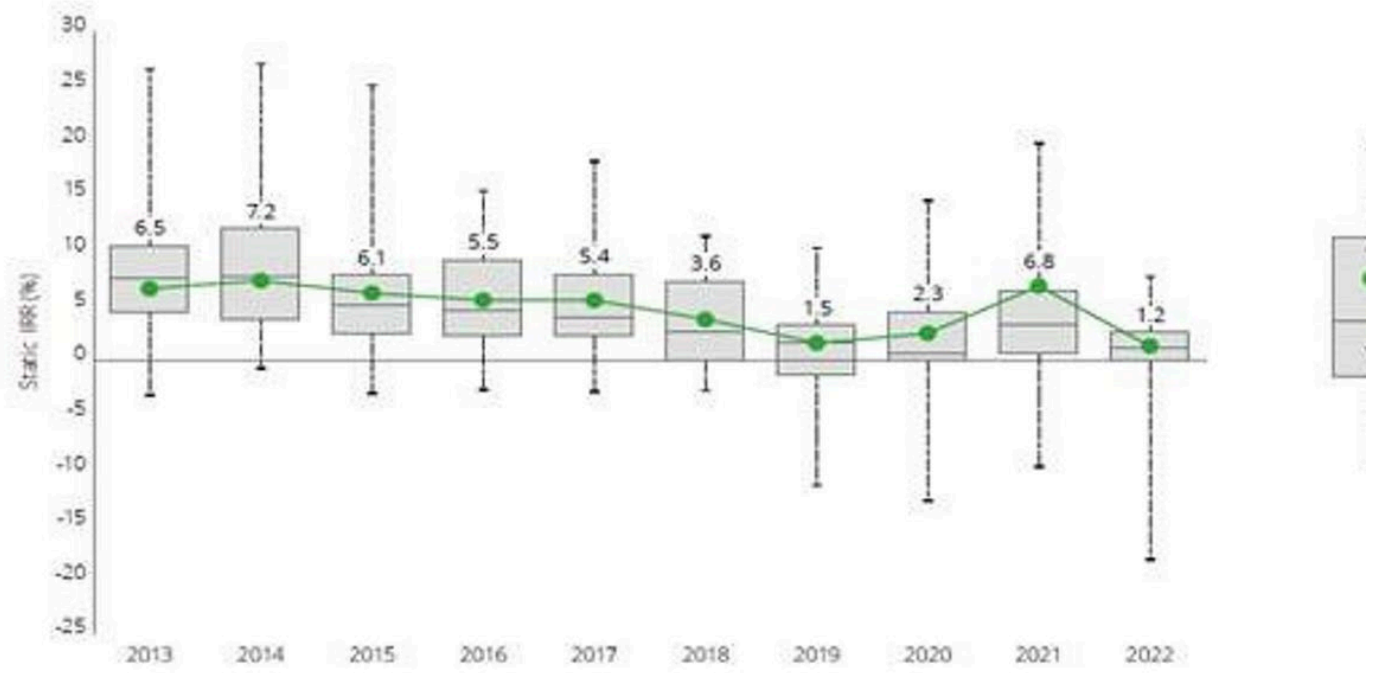
- **$2.3B** is the average cost to bring a new drug to market for the top 20 BioPharmas
- **80%** of clinical trials fail to complete on time
- **55%** of trials that fail to complete cite low accrual rate as the reason for termination
- **30%** of total trial costs are consumed by enrollment
- **20%** of sites cannot enroll any patients
- **$Millions** lost per day from reduced patent life and trial operational costs
- **$100'sM** annual budget reduction from Indirect cuts

Deloitte. (2023)
Sertkaya, A. *JAMA network open.* (2024)
Global Data. *Pharm Tech*. (2018)
Khan, M.S. *JAMA Cardiology*. (2025)
Butler J. *J Am Coll Cardiol*. (2013)
Boston Globe (2025)

Pharma Economics

- The time it takes for a new drug to move from starting a clinical trial to approval - increased to 7.1 years in 2022 from 6.9 years in 2021
- As a result of the above and other pressures, the estimated average cost of developing a drug, including the cost of failure, increased from $2 Billion in 2021 to $2.3 Billion in 2022.
- In total, the 20 companies analyzed spent $139 billion on R&D in 2022, a decrease of two per cent compared to 2021 ($141 billion).

Return on late-stage pipeline, 2013-2022



Source: Deloitte analysis, 2022.

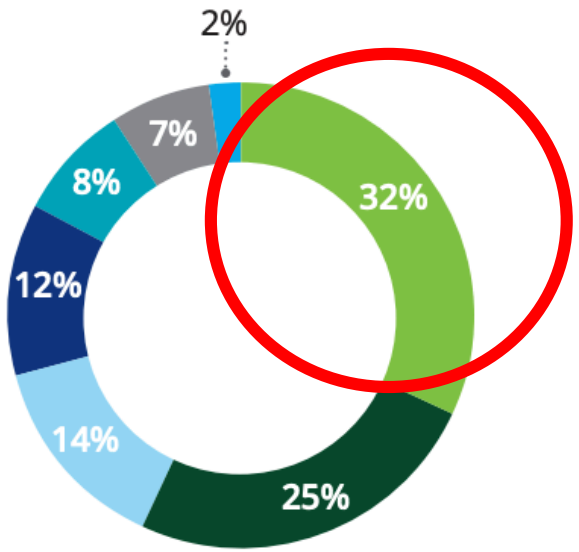
Please note: 2013-2019 data includes the 15 companies of the combined cohort; 2020-2022 data includes the results of the to 2019 R&D spend. See appendix 1 for the data of each cohort. Compared to last year's report 2020 and 2021 figures have bee the top 20 companies by R&D spend as of 2020.

Deloitte. (2023)

Applications of AI-enabled technology in clinical trials

Cost drivers in clinical trials

- Patient recruitment
- Outsourcing costs
- Site recruitment
- Clinical trial management system and other technology
- Site retention
- Data management and validation
- Patient retention



	TRIAL DESIGN	TRIAL STARTUP	TRIAL CONDUCT	STUDY CLOSEOUT
Advanced data analytics and AI automation	Assess feasibility of protocol design for patient recruitment using RWD. Assess site performance (e.g. enrolment and dropout rates) with real-time monitoring. Analyse and interpret unstructured and structured data from previous trials and scientific literature.	Mine EHRs and publicly available content, including trial databases and social media, to help match patients with trials, by using NLP and ML. Create drafts of investigator and site contracts and confidentiality agreements by smart automation.	Assess site performance (e.g. enrolment and dropout rates) with real-time monitoring. Analyse digital biomarkers on disease progression, and other quality-of-life indicators. Automate sharing of data across multiple systems.	Complete sections of the final clinical trial report for submission by using NLP. Data cleaning by ML methods.
AI-enhanced mobile applications, wearables, biosensors and connected devices		Expedite recruitment and create a more representative study cohort through cloud-based applications. Simplify and accelerate the informed consent process using eConsent.	Enhance adherence through smartphone alerts and reminders. eTracking of medication using smart pillboxes, and tools for visual confirmation of treatment compliance. eTracking of missed clinic visits, and trigger non-adherence alerts.	

Source: Deloitte analysis.

Trial performance in the United States lagging

Table 2. Site-Specific Enrollment Characteristics for Trials Reporting Patient-per-Site Data

	Patients enrolled			No. (%)			Enrollment rate, median (IQR), No.	
					Sites enrolling <10 patients			
Trial	Total No.	US, No. (%)	Total sites, No.	US sites	Overall	US[a]	Overall[b]	US
Total	89 172	17 705 (19.9)	4388	1133 (25.8)	2014 (45.8)	659 (58.1)	0.3 (0.1-0.7)	0.2 (0.1-0.5)
ISCHEMIA-CKD	777	159 (20.5)	118	36 (30.5)	95 (80.5)	31 (86.1)	0.1 (0-0.2)	0 (0-0.1)
COP-AF	3209	355 (11.1)	45	8 (17.8)	12 (26.7)	3 (37.5)	0.4 (0.2-1.5)	0.4 (0.1-0.8)
THEMIS	19 220	2266 (11.8)	1297	307 (23.7)	668 (51.5)	228 (74.3)	0.4 (0.1-0.7)	0.2 (0.1-0.4)
ILUMIEN IV: OPTIMAL PCI	2487	909 (36.6)	80	35 (43.8)	23 (28.7)	14 (40.0)	0.6 (0.2-1.1)	0.4 (0.1-0.7)
PARADISE-MI	5702	454 (8.0)	494	82 (16.6)	284 (57.5)	73 (89.0)	0.2 (0.1-0.4)	0.1 (0.1-0.2)
REPRIEVE	7769	3787 (48.7)	145	100 (69.0)	23 (15.9)	18 (18.0)	0.7 (0.3-1.3)	0.7 (0.2-1.1)
SELECT	17 604	3652 (20.7)	804	201 (25.0)	220 (27.4)	58 (28.9)	0.7 (0.3-1.0)	0.5 (0.3-0.8)
ISCHEMIA	5179	853 (16.5)	319	109 (34.2)	207 (64.9)	88 (80.7)	0.1 (0-0.2)	0.1 (0-0.1)
AEGIS-II	18 219	1993 (10.9)	899	196 (21.8)	450 (50.1)	138 (70.4)	0.2 (0.1-0.4)	0.1 (0-0.2)
TWILIGHT	9006	3277 (36.4)	187	59 (31.6)	32 (17.1)	8 (13.6)	0.1 (0.5-2.0)	1.2 (0.5-2.5)

Khan MS, Jamil A, Shakoor M, et al. Patient Enrollment for Cardiovascular Clinical Trials in the United States. *JAMA Cardiol.* 2025;10(3):298–300. doi:10.1001/jamacardio.2024.5537

Background

- Traditional Clinical Trial Screening:
  - Burdensome, labor intensive, and time-consuming.
  - Relies on diagnosis codes, problem lists, billing data to create a screening pool.
  - Relies on manual reviews for unstructured clinical data like notes or radiology reports.

- Natural Language Processing (NLP) can improve screening and automate extraction and analyses; however, traditional NLP has limitations, particularly in handling complex, unstructured data (e.g. EHR data)

- Large language models (LLMs), such as (GPT-4o), significantly advanced state-of-the-art performance on many NLP tasks

- LLMs create opportunity to modernize patient identification and screening for clinical trials

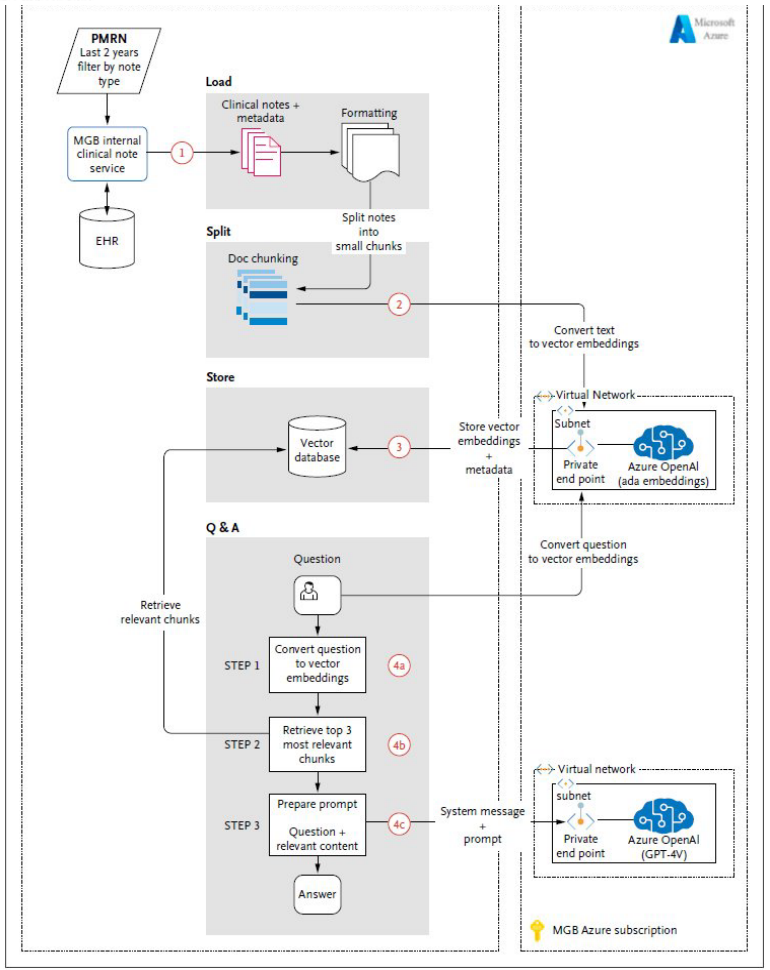
Kim, J. *Health Tech Inform.* (2022)
Elm, J.J. *Clin Trials* (2014)
Idnay, B. *JAMIA* (2021)
Perlis, R.H. *JAMA Net Open* (2023)

In a pilot study, LLM (RECTIFIER) was highly accurate to determine eligibility criteria with a potential to streamline screening

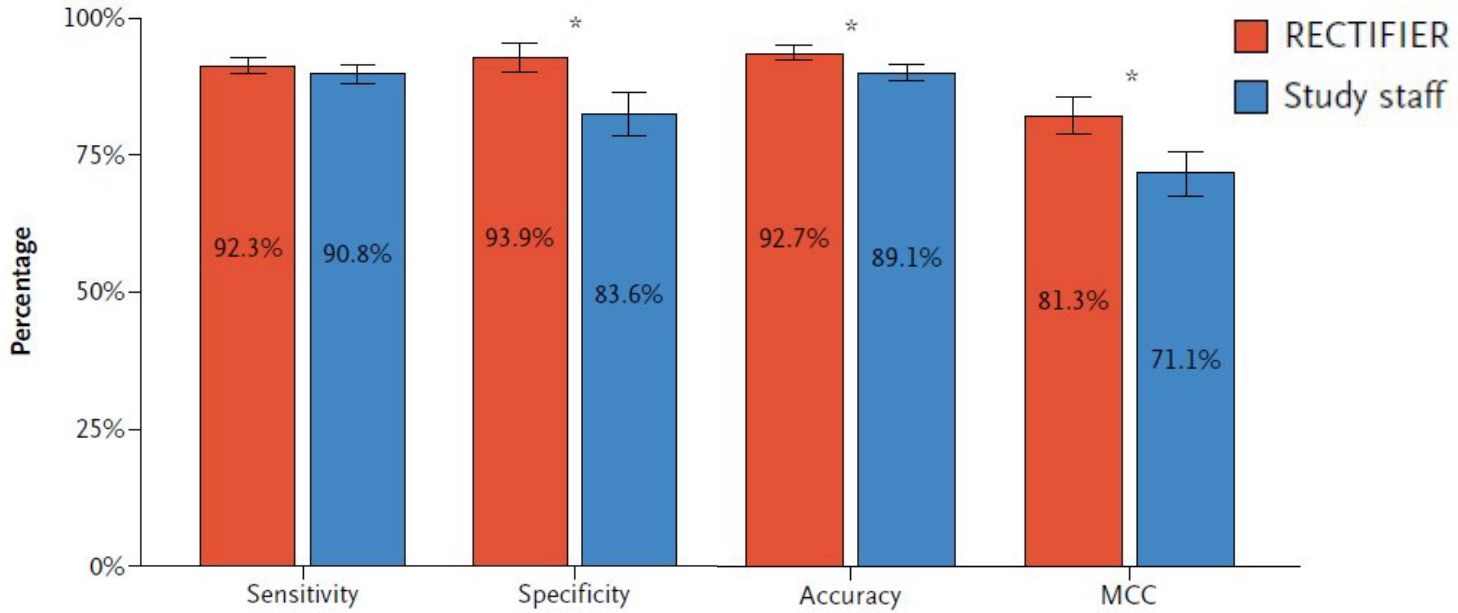
ORIGINAL ARTICLE

Retrieval-Augmented Generation–Enabled GPT-4 for Clinical Trial Screening

Ozan Unlu, M.D.,[1,2,3,4] Jiyeon Shin, A.L.M.,[1,5] Charlotte J. Mailly, B.A.,[1,5] Michael F. Oates,[1,5] Michela R. Tucci, M.P.H.,[1] Matthew Varugheese, M.S.,[1] Kavishwar Wagholikar, M.B.B.S., Ph.D.,[1,4] Fei Wang, M.Sc.,[1,5] Benjamin M. Scirica, M.D., M.P.H.,[1,2,4] Alexander J. Blood, M.D., M.Sc.,[1,2,4] and Samuel J. Aronson, A.L.M., M.A.[1,5]



- Embedded LLM project within existing active clinical trial of patients with heart failure
- Compared screening performance metrics in determining patient eligibility for each criterion between traditional Study Staff screening vs RECTIFIER with expert review as reference standard
- RAG model important to enable use and limit costs



Unlu, O. *NEJM AI.* (2024)

Limitations of Large Language Models by themselves

Table S4. Two Years of Clinical Notes Exceeding the Context Window for GPT-4 and GPT-3.5 Processing

We calculated the number and percentage of patients in the test set with the past two years of clinical notes exceeding the context window for GPT-4 and GPT-3.5 processing. These patients require additional processing if RAG is not used, where their notes would need to be segmented and a mechanism would need to be developed to combine the model's response for each question.

Token Limit	Number of Patients (n=1894)	Percentage of Patients
> 128K	493	26.03%
> 16K	1695	89.49%
> 8K	1832	96.73%

Supplementary Table 10 B.

	Embedding	Q & A		Total Cost		Cost per Patient	
	Ada-002	GPT-4 Turbo-1106	GPT 3.5 Turbo-1106	GPT-4 Turbo-1106	GPT 3.5 Turbo-1106	GPT-4 Turbo-1106	GPT 3.5 Turbo-1106
RAG	$21.46	$195.38	$19.51	$216.85	$40.98	$0.11	$0.02
No RAG	$0.00	$30,067.40	$3,006.71	$30,067.40	$3,006.71	$15.88	$1.59

~150x more expensive without Retrieval Augmented Generation

Fast

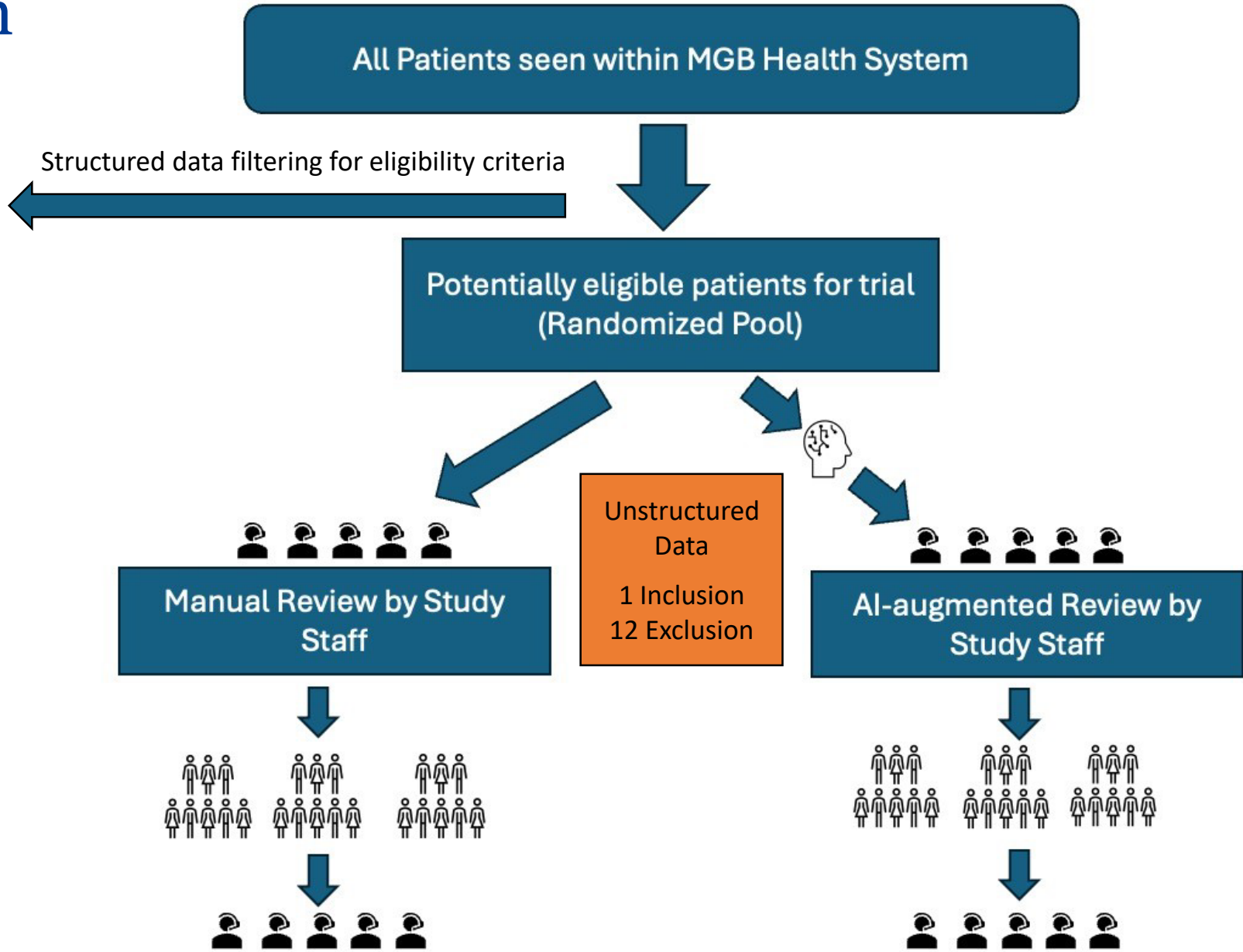
Accurate

Cheap

MAPS-LLM Trial Design

Inclusion
- Age 18-90
- Documented diagnosis of heart failure
- Most recent left ventricular ejection fraction (LVEF) assessed within the past 24 months
- Seen MGB provider within the last 24 months
- Speaks English or Spanish

Exclusion
- LVEF <50% currently prescribed or intolerant to an evidence-based beta-blocker, ARNI, MRA, and SGLT2i at least 50% goal dose
- LVEF>50% currently prescribed or intolerant to SGLT2i
- Systolic blood pressure (SBP) <90 mmHg at last measure



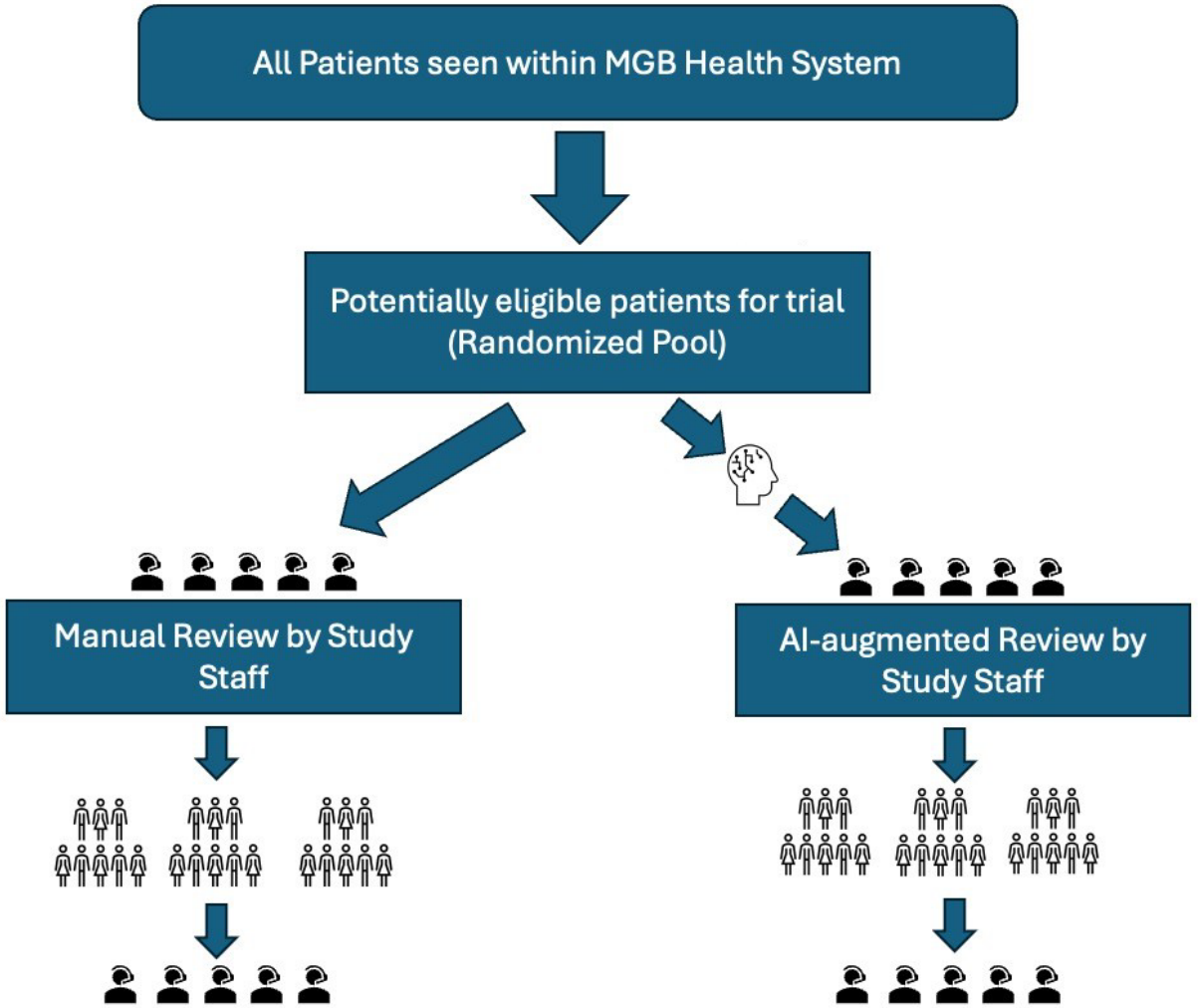
MAPS-LLM Trial Design

- Study staff are allocated an equal number of screening hours each week for each group.

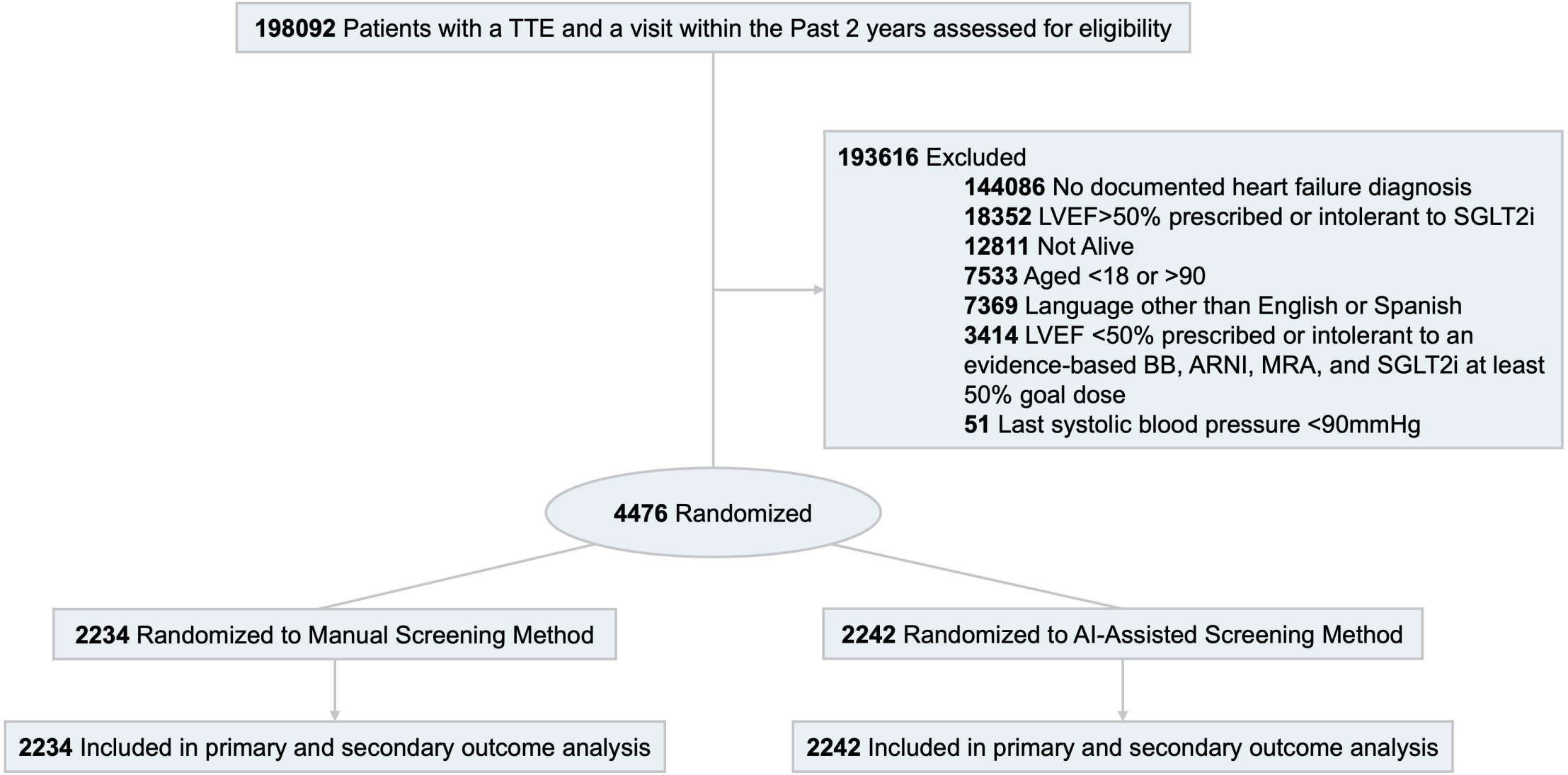
- Screened patients are placed in a pool and assigned to a study staff member, different from the one who conducted their screening, for enrollment calls.

Study End Points:

Primary Endpoint	*Eligibility Determination*
Secondary Endpoint	Hierarchical win ratio prioritizing enrollments over eligibility determinations
Exploratory Outcomes	- Enrollments - Equity of eligibility determination



MAPS-LLM Consort Diagram

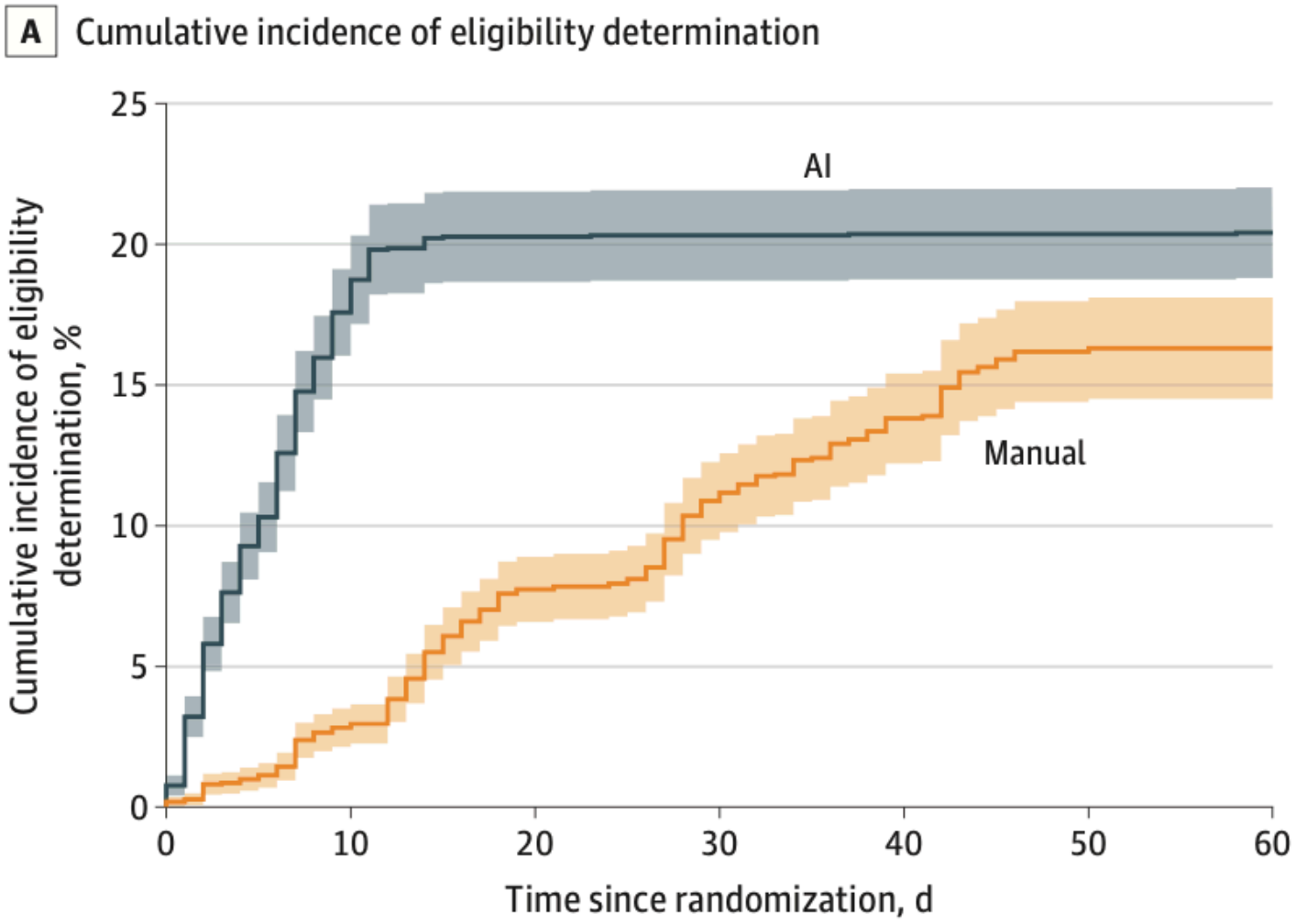


Baseline Characteristics

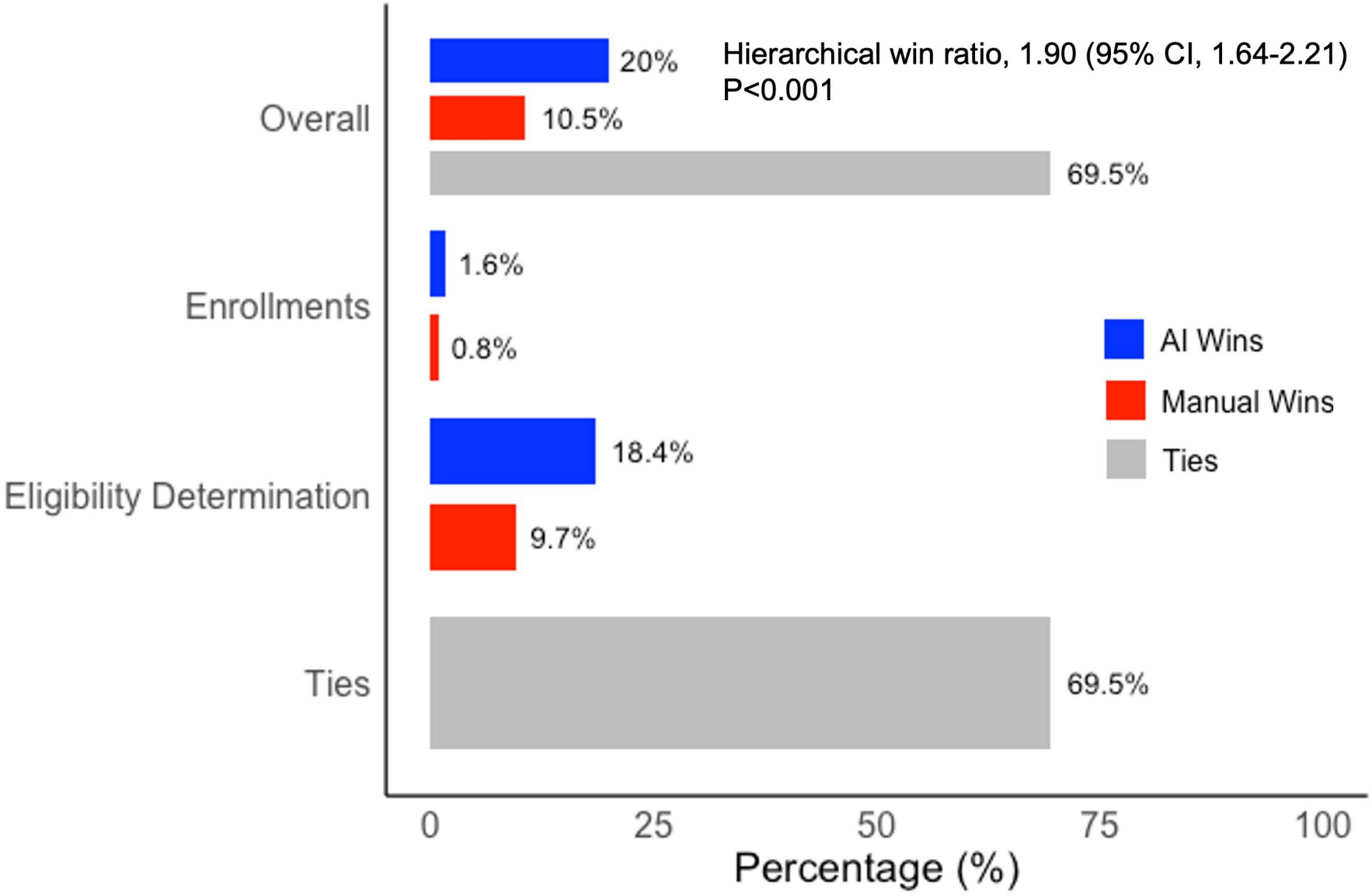
Characteristic	Manual Screening (n=2,234)	AI-Assisted Screening (n=2,242)	Total (N=4,476)
Age, mean (SD), y	74.7 (12.1)	75.1 (11.2)	74.9 (11.6)
Female, No. (%)	1,085 (48.6)	1,166 (52)	2,251 (50.3)
Race, No. (%)			
White	1985 (44.3)	1965 (43.9)	3950(88.2)
Black	111 (2.5)	139 (3.1)	250(5.6)
Asian	21 (0.5)	18 (0.4)	139(3.1)
Other	70 (1.6)	69 (1.5)	39(0.9)
Declined sharing	47 (1.1)	51 (1.1)	98(2.2)
Hispanic individuals, No. (%)	101(4.5)	123(5.5)	224 (5.0)
ADI State Ranks* , No. (%)			
1st (1–20)	496 (22.2)	515 (23.0)	1011 (22.6)
2nd (21–40)	496 (22.2)	480 (21.4)	976 (21.8)
3rd (41–60)	485 (21.7)	520 (23.2)	1005 (22.5)
4th (61–80)	456 (20.4)	500 (22.3)	956 (21.4)
5th (81–100)	205 (9.2)	195 (8.7)	400 (8.9)
Cannot be Calculated	96 (4.3)	32 (1.43)	128 (2.9)
HFrEF (LVEF <50%), No. (%)	662 (29.6)	633(28.2)	1295 (28.9)
HFpEF (LVEF ≥ 50%), No. (%)	1572 (70.4)	1609 (71.8)	3181(71.1)

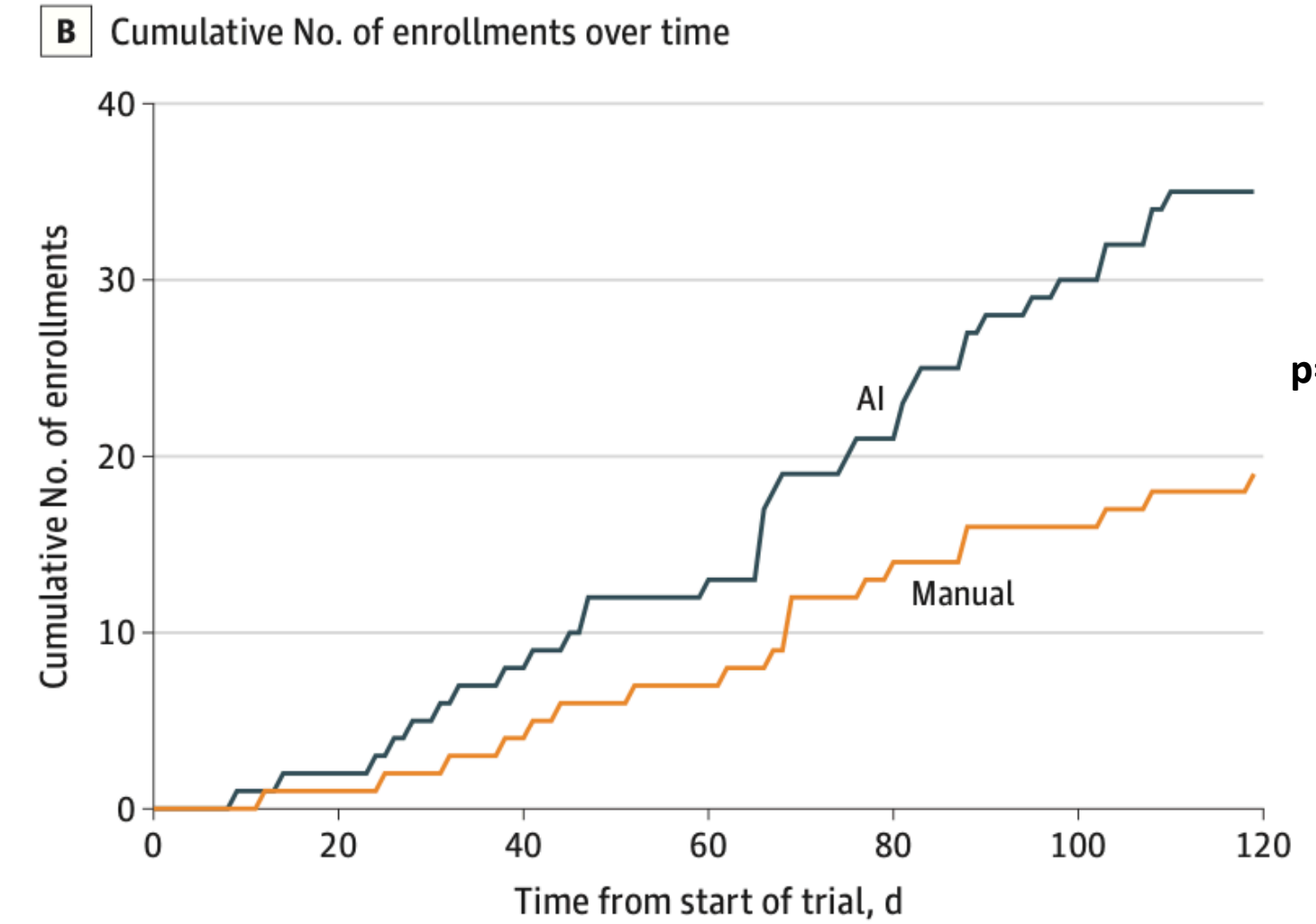
Primary Outcome: Eligibility Determination

Figure 2. Cumulative Incidence of Eligibility Determination



Secondary Outcome: Hierarchical Win Ratio




B
Cumulative No. of enrollments over time
Cumulative No. of enrollments
40
30
20
10
0
0
20
40
60
80
100
120
Time from start of trial, d
AI
Manual

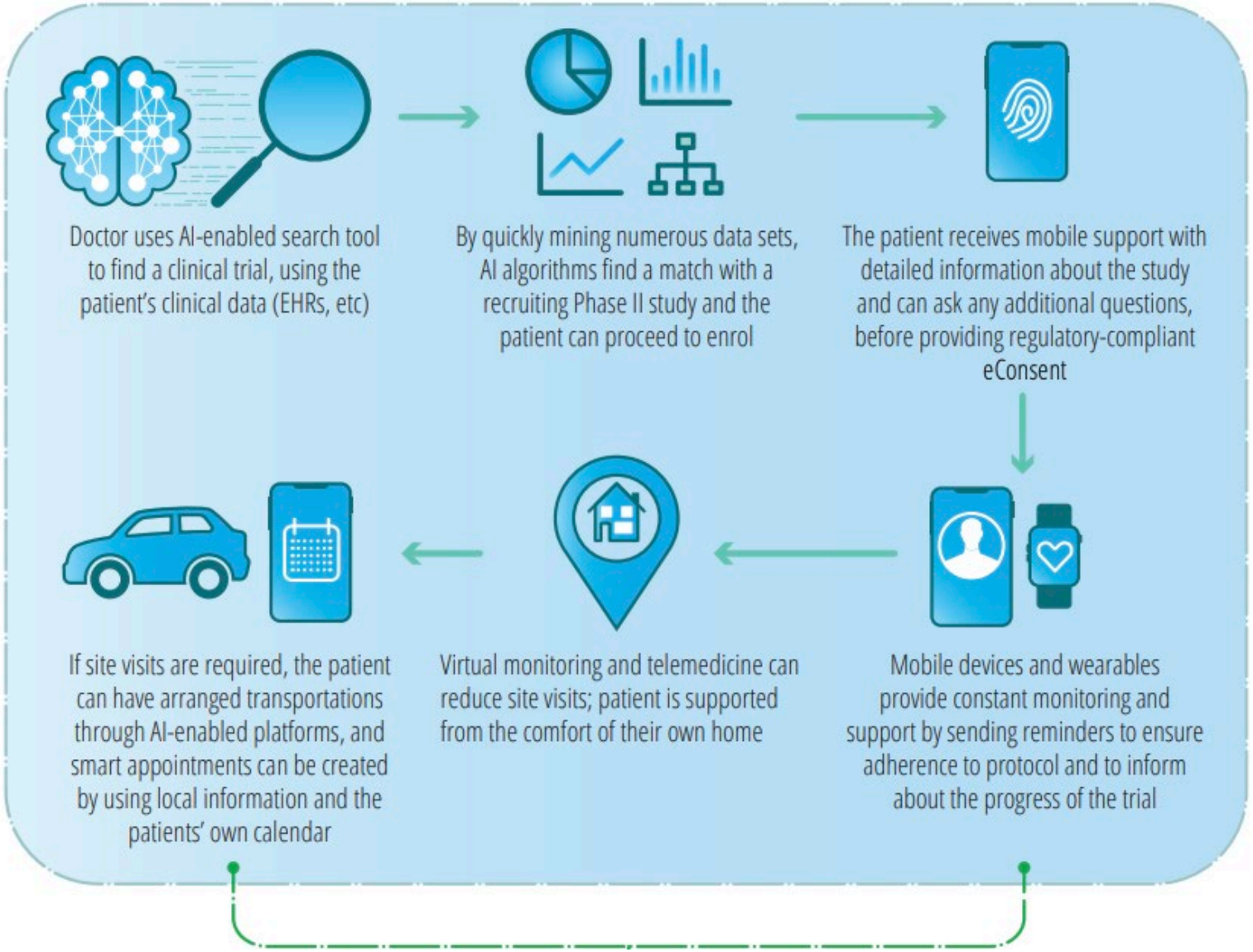
p= 0.041

Exploratory Outcomes

Equity

Demographic Group	Eligibility Rate		P Value
	Manual Screening	AI-Assisted Screening	
Race			
White	21.6%	22.4%	0.629
Black	14.3%	17.1%	0.612
Asian	27.3%	23.5%	0.999
Other	12.5%	13.8%	0.844
Ethnicity			
Hispanic	9.4%	18.3%	0.137
Non-Hispanic	21.7%	21.9%	0.890
Sex			
Female	21.7%	20.7%	0.621
Male	21.6%	21.4%	0.921
ADI State Ranks			
1st (1–20)	22.1%	24.4%	0.428
2nd (21–40)	18.5%	20.3%	0.535
3rd (41–60)	21.2%	18.8%	0.406
4th (61–80)	20.2%	21.1%	0.785
5th (81–100)	24.2%	23.7%	0.918

Patient journey through an AI-enabled Clinical Trial



Deloitte. (2023)

Application beyond Clinical Trials

NEJM AI

Large Language Models for More Efficient Reporting of Hospital Quality Measures

Authors: Aaron Boussina, Ph.D. , Rishivardhan Krishnamoorthy, M.S. , Kimberly Quintero, R.N., M.S. , Shreyansh Joshi , Gabriel Wardi, M.D. , Hayden Pour, M.S. , Nicholas Hilbert, R.N., M.S.N. , +6 , and Shamim Nemati, Ph.D. Author Info & Affiliations

Published October 21, 2024 | NEJM AI 2024;1(11) | DOI: 10.1056/AIcs2400420 | VOL. 1 NO. 11

- Hospital quality measures are a vital component of a learning health system, yet they can be costly to report, statistically underpowered, and inconsistent due to poor interrater reliability.
- Authors assessed a large language model's ability to complete Severe Sepsis and Septic Shock Management Bundle (SEP-1) abstraction.
- The LLM system achieved agreement with manual abstractors on the measure category assignment in 90 of the abstractions (90%; κ=0.82; 95% confidence interval, 0.71 to 0.92). Expert review of the 10 discordant cases identified four that were mistakes introduced by manual abstraction.

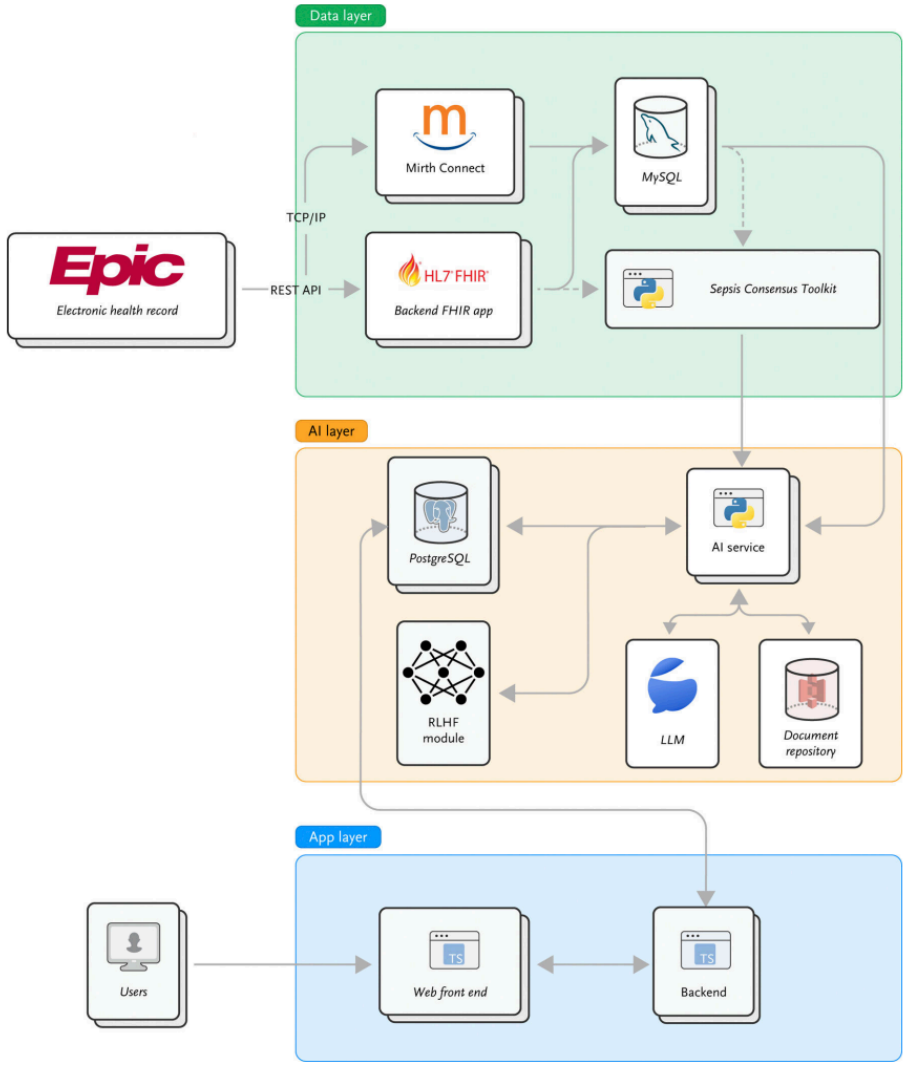


Figure 1. System Architecture for Automation of Hospital Quality Measures.
The data layer (green) enables the collection of electronic health record data through Fast Healthcare Interoperability Resources (FHIR) and the computation of clinical criteria. Mirth Connect stores all encounters from admission-discharge-transfer messages. The backend FHIR application then queries encounter data and stores it in MySQL. The Sepsis Consensus Toolkit applies standard rule-based criteria to the structured data to identify systemic inflammatory response syndrome and organ failure events. The artificial intelligence layer (orange) manages the large language model for abstraction. The app layer

Conclusion

- AI-assisted patient screening using the RECTIFIER system significantly improved eligibility determination and enrollment in a heart failure clinical trial compared with manual screening.

- Implementing AI-assisted tools like RECTIFIER can enhance clinical trial efficiency, reduce resource utilization, and promote equitable recruitment, potentially leading to faster trial completion and earlier patient access to novel therapies.

- Generative AI is likely to play a significant role in the future of Clinical Trials

Thank you for your time and attention

ABlood@BWH.Harvard.Edu