Adaptive AI for CV Adjudication:
Adjudication Across Different Definitions in the ODYSSEY OUTCOMES and EUCLID Trials

Sreek Vemulapalli, MD
Associate Professor of Medicine with Tenure
Duke University School of Medicine
Director, Cardiovascular Imaging, DUHS
Duke Clinical Research Institute
Duke-Margolis Institute for Health Policy

Ricardo Henao, PhD
Associate Professor of B&B, ECE, Surgery
Duke University School of Medicine
Associate Director, Clinical Trials AI
Duke Clinical Research Institute



Duke Clinical Research Institute | FROM THOUGHT LEADERSHIP TO CLINICAL PRACTICE

Disclosures

- Grants / Contracts: American College of Cardiology, American Heart Association, Food and Drug Administration, National Institutes of Health (R01 and UG3/UH3), Cytokinetics, Edwards Lifesciences, Abbott

- Advisory / Consulting: Edwards Lifesciences, Medtronic, Abbott, Astra Zeneca, Boehringer-Ingelheim, Veralox Therapeutics

- Grants / Contracts: American College of Cardiology, American Heart Association, Food and Drug Administration, National Institutes of Health (R01 and UG3/UH3), Cytokinetics, Edwards Lifesciences, Abbott

- Advisory / Consulting: Edwards Lifesciences, Medtronic, Abbott, Astra Zeneca, Boehringer-Ingelheim, Veralox Therapeutics

Outline

- Background
- History of NLP and AI for CEC efforts
- ADAPT-CEC
- Implications for Trials

Event adjudication is a trial bottleneck

High-value for rigor, but costly, slow, and inconsistently operationalized across adjudication groups

Clinical endpoint adjudication

The promise is standardization; the pain point is manual expert review at scale.

Cost pressure

Requires source document collection, dossier assembly, expert review, queries, reconciliation, and committee operations.

Operational drag

Events often wait for complete documents and multiple reviewers, creating turnaround delays before DSMB or final analyses.

Between-group variability

Event definitions, reviewer interpretation, source document completeness, and workflow mechanics can differ across adjudication groups.

Bottom line: adjudication improves endpoint credibility, but the current human-centered model is a major scalability constraint.

Expert review → Final endpoint



Evolution of Clinical Endpoint Committees

From independent adjudication in complex cardiovascular trials to standardized event definitions

CEC timeline

Key dates and shifts described in the text

Harmonization goal

Reduce classification variability

CEC adoption matured as trials became larger, geographically distributed, more data-intensive, and more dependent on harmonized endpoint definitions.



Core workflow: collect/validate endpoint data → assemble patient-event dossier → 2 physician reviews → reconcile disagreements → enter final adjudicated call

Note: "Recent" is shown as a relative milestone because the provided text did not specify a calendar year for FDA endpoint definitions.

The NLP Revolution: A 15-Year Timeline

2010-2013

Deep learning takes root in NLP. Models begin learning word meaning from data.

- RNNs for Language Modeling
- **Word2vec (2013):** Efficient Semantic Word Embeddings
- LSTMs Tackle Dependencies

2014-2016

Handling input/output sequences improves. Attention allows models to pinpoint relevant information.

- Sequence-to-Sequence (Seq2Seq) Architectures
- **Attention Mechanism (2014/15):** Models learn to focus, boosting Machine Translation

2017

A revolutionary architecture based solely on Attention enables parallel processing and better context handling, setting the stage for modern NLP.

- **"Attention Is All You Need":** Introduces the **Transformer** Architecture.
- ELMo: Contextual Word Embeddings

2018-2019

Large models pre-trained on massive web text become dominant. Fine-tuning adapts them for specific tasks with unprecedented performance.

- **BERT (2018):** Bidirectional context understanding, SOTA via Pre-train & Fine-tune
- **GPT & GPT-2 (2018/19):** Powerful Generative Models Emerge
- RoBERTa, T5, XLNet Refine Techniques

2020-2022

Model size explodes, unlocking new abilities. Conversational AI becomes highly capable, and models start understanding multiple data types (text, images).

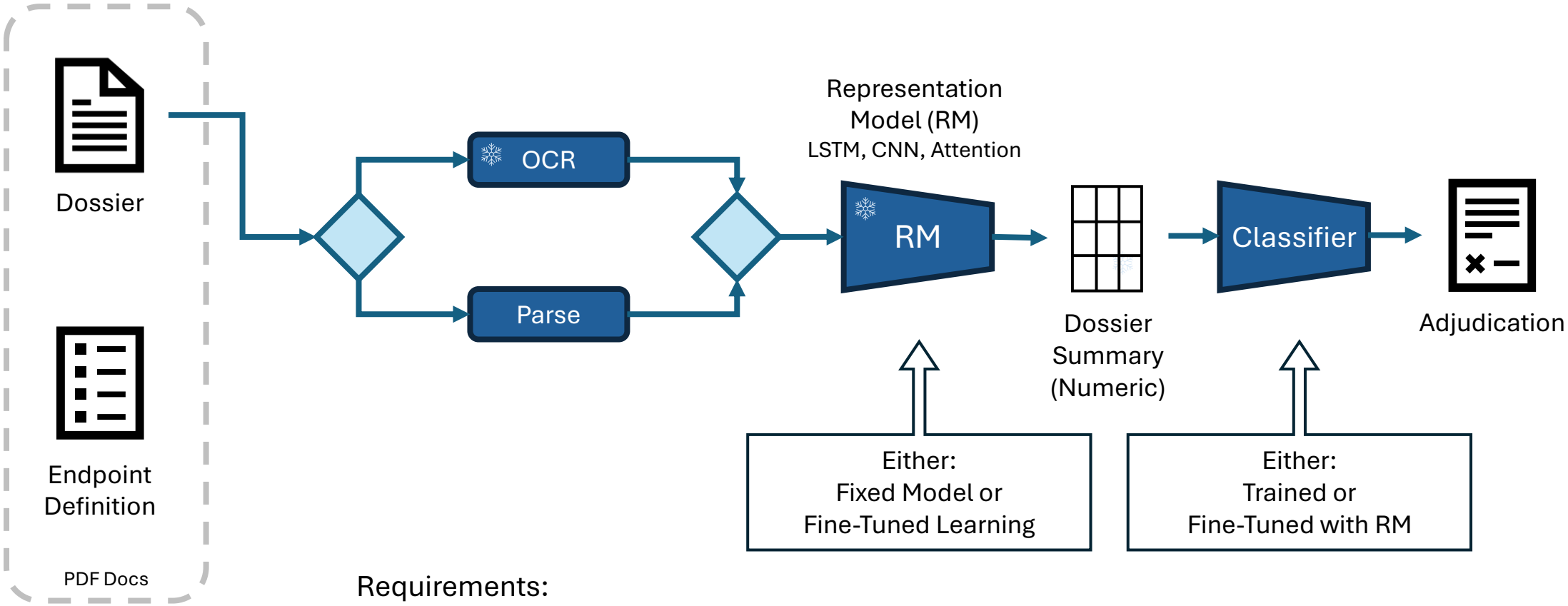
- **GPT-3 (2020):** Massive Scale (175B params), Impressive Few/Zero-Shot Learning
- CLIP & DALL-E (2021): Bridging Text and Images
- **ChatGPT (2022):** Conversational AI Goes Mainstream via RLHF

2023-Present

Focus shifts to reasoning, real-world interaction (agents), multimodality, and open-source alternatives alongside scaling and capability improvements.

- **GPT-4 (2023):** Enhanced Reasoning, Multimodal Input (Text + Image)
- **Open Models Thrive:** LLaMA, Mistral, Mixtral democratize access
- Rise of **AI Agents:** LLMs powering planning and tool use
- Focus on Efficiency (MoE), Safety, and Alternative Architectures (Mamba)

Early Approaches



Requirements:

- Large labeled datasets to train or fine-tune model
- Heavy preprocessing of dossiers to encourage generalization
- Endpoint informs the model *only* during training

Fixed Model

Challenges with Early Approaches (Pre-LLM)

Studies:

- (Discovery) EUCLID was a global, randomized, double-blind, active comparator clinical trial that included 13,885 patients with peripheral artery disease and compared monotherapy with ticagrelor and clopidogrel.
- (Validation) ODYSSEY OUTCOMES was a global, randomized, double-blind, placebo-controlled trial, that included 18,924 patients with a recent acute coronary syndrome and compared alirocumab with placebo.

Dataset	MI	Stroke	CV death
Validation (EUCLID)	0.90	0.73	0.81
Validation – No Adaptation (ODYSSEY)	0.60	0.43	0.59
Validation – Domain Adaptation (ODYSSEY)	0.54	0.49	0.58
Validation – Fine Tuning (ODYSSEY)	0.82	0.69	0.65

MI		Stroke		CV Death	
2,088		708		1,052	
Yes	No	Yes	No	Yes	No
853	1,235	485	223	496	556

- Processing: Textract or Tesseract
- Representation Model: GloVe
- Classification Model: CNN

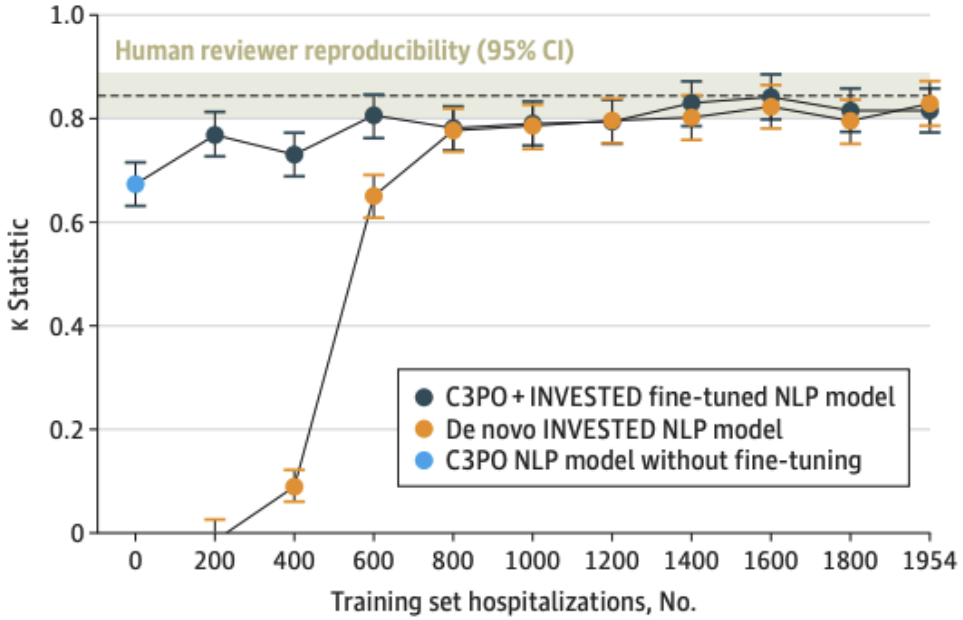
Challenges with Newer Approaches

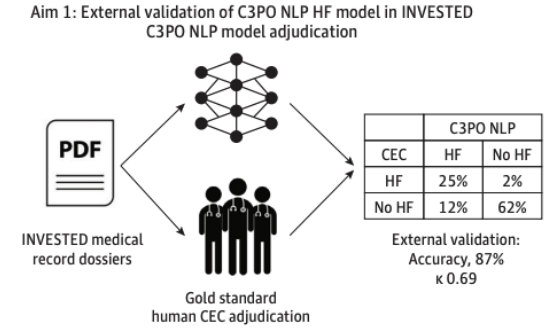
JAMA Cardiology | Original Investigation | AI IN CARDIOLOGY

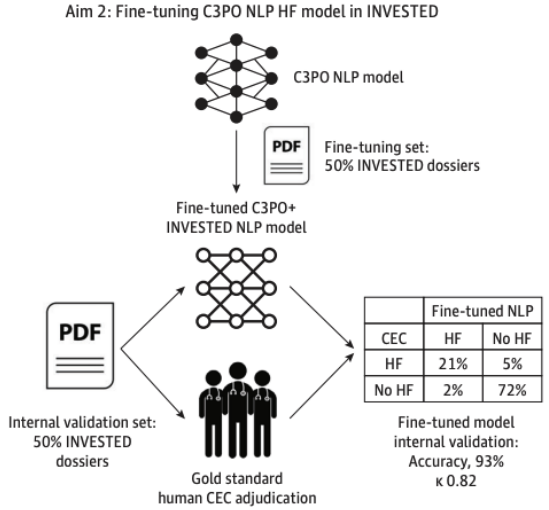
Natural Language Processing for Adjudication of Heart Failure in a Multicenter Clinical Trial

A Secondary Analysis of a Randomized Clinical Trial

Jonathan W. Cunningham, MD, MPH; Pulkit Singh, BA; Christopher Reeder, PhD; Brian Claggett, PhD; Pablo M. Marti-Castellote, MSc; Emily S. Lau, MD, MPH; Shaan Khurshid, MD, MPH; Puneet Batra, PhD; Steven A. Lubitz, MD, MPH; Mahnaz Maddah, PhD; Anthony Philippakis, MD, PhD; Akshay S. Desai, MD, MPH; Patrick T. Ellinor, MD, PhD; Orly Vardeny, PharmD, MS; Scott D. Solomon, MD; Jennifer E. Ho, MD

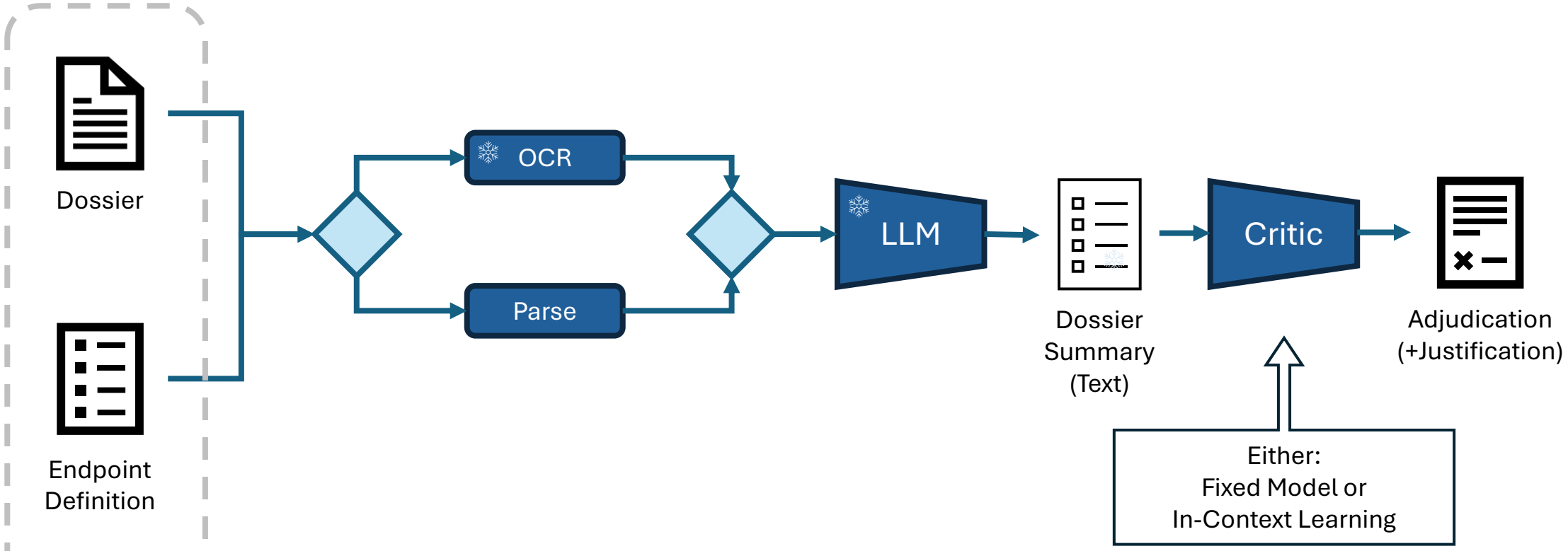






Retrospective analysis of the Influenza Vaccine to Effectively Stop Cardio Thoracic Events and Decompensated Heart Failure (INVESTED) trial, which compared 2 influenza vaccines in 5260 participants with cardiovascular disease at 157 sites in the US and Canada

New Approaches based on LLMs



Wants:

- Accept both parsable and non-parsable input documents
- Take endpoint definition as an input
- Require no (or very little) training (or refining)

Fixed Model

AI-assisted adjudication of MACE in clinical trials

Auto-MACE combines an LLM adjudicator with Clinical Longformer to separate confident events from cases needing human review

Marti-Castellote et al. | JACC 2026

Clinical trials + AI

WHY IT MATTERS

CEC adjudication is the standard but can be labor-intensive, slow, opaque, and imperfectly



Question: Can AI reproduce human CEC adjudication and preserve treatment-effect inference?

AUTO-MACE WORKFLOW



PDF records

Plain text

Model outputs

o1-mini LLM primary adjudication

Clinical Longformer uncertainty flag

Confident

Reject

Uncertain

Uncertain cases are routed to human CEC review.

KEY FINDINGS

Agreement with human CEC among confident Auto-MACE events

Treatment-effect estimate preserved

Composite MACE HR: CEC 0.90 vs Auto-MACE 0.91

95% CIs overlapped closely: 0.77–1.05 vs 0.78–1.07

Take-home: AI can triage straightforward MACE events and focus human CEC effort on uncertainty.

Previous efforts at AI for CEC have:

– Derived separate models for each CV event type

– No demonstrated ability to adapt to new definitions or endpoints across trials

Visual abstract structure: concise paper teaser → question + methods + key outcome + implication; designed with minimal text and prominent visual hierarchy.

Auto-MACE: automated adjudication of CV death, non-fatal MI, and non-fatal stroke

Objectives

- Develop and train an AI algorithm to perform clinical endpoint adjudication that can:
  - Extend to new endpoints
  - Extend to new trials
- Compare algorithm performance to GPT 4.0 and hybrid adjudication (AI + human)
- Re-estimate a trial's primary endpoint using AI-CEC

Study Design

DEVELOPMENT PIPELINE

TRAINING DATA

ODYSSEY OUTCOMES

Source trial providing labeled clinical narratives for supervised learning

Trained on: MI Stroke

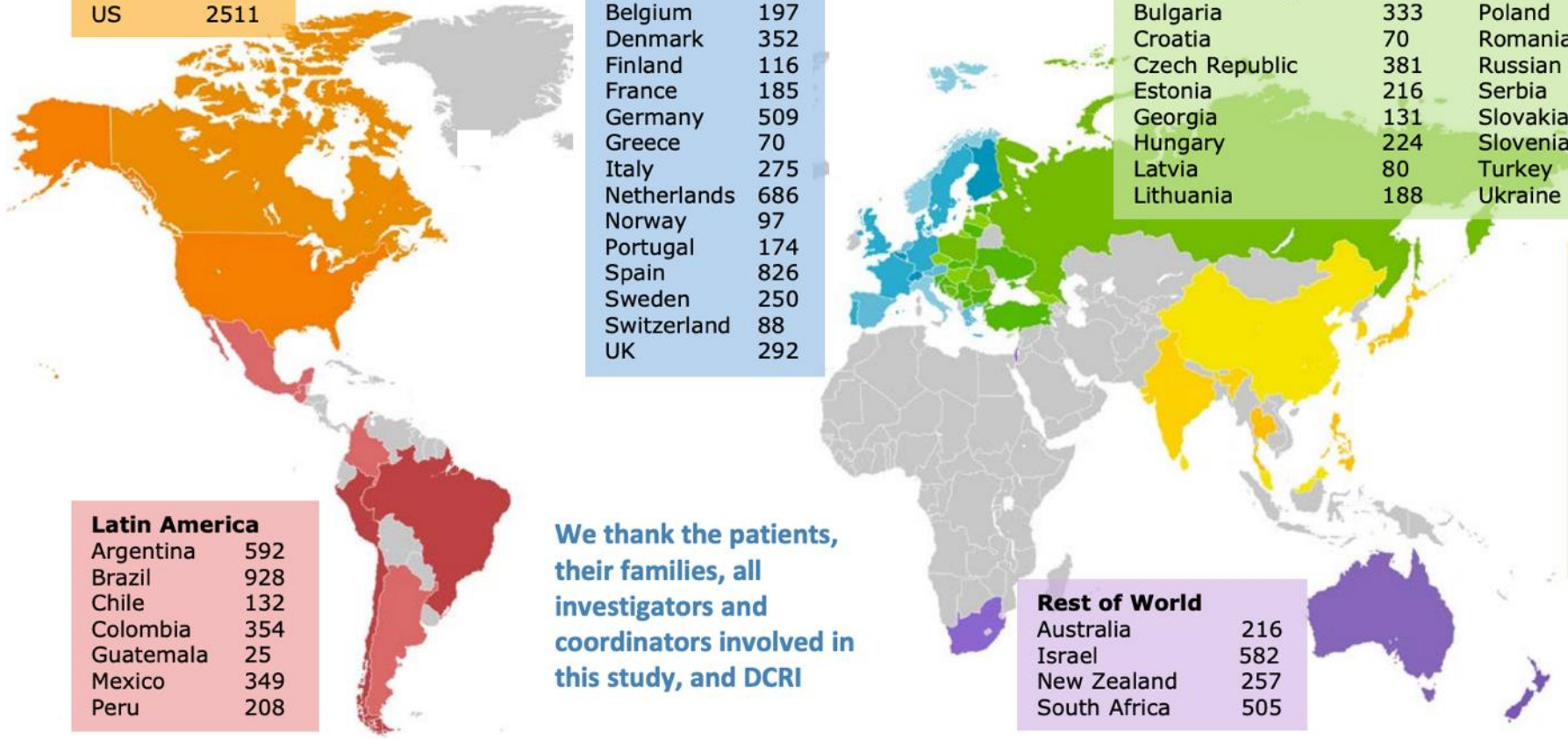
ALGORITHM

ADAPT-CEC Model

Model trained on labeled events to classify endpoints

ODYSSEY OUTCOMES

18,924 patients randomized at **1315 sites** in **57 countries**, Nov 2, 2012-Nov11,2017



Canada/USA	
Canada	361
US	2511

Western Europe	
Austria	58
Belgium	197
Denmark	352
Finland	116
France	185
Germany	509
Greece	70
Italy	275
Netherlands	686
Norway	97
Portugal	174
Spain	826
Sweden	250
Switzerland	88
UK	292

Central/Eastern Europe			
Bosnia–Herzegovina	156	Macedonia	132
Bulgaria	333	Poland	926
Croatia	70	Romania	145
Czech Republic	381	Russian Federation	1109
Estonia	216	Serbia	255
Georgia	131	Slovakia	340
Hungary	224	Slovenia	36
Latvia	80	Turkey	78
Lithuania	188	Ukraine	639

Asia	
China	614
Hong Kong	17
India	521
Japan	204
Korea	94
Malaysia	110
Philippines	116
Singapore	49
Sri Lanka	314
Taiwan	93
Thailand	161

Latin America	
Argentina	592
Brazil	928
Chile	132
Colombia	354
Guatemala	25
Mexico	349
Peru	208

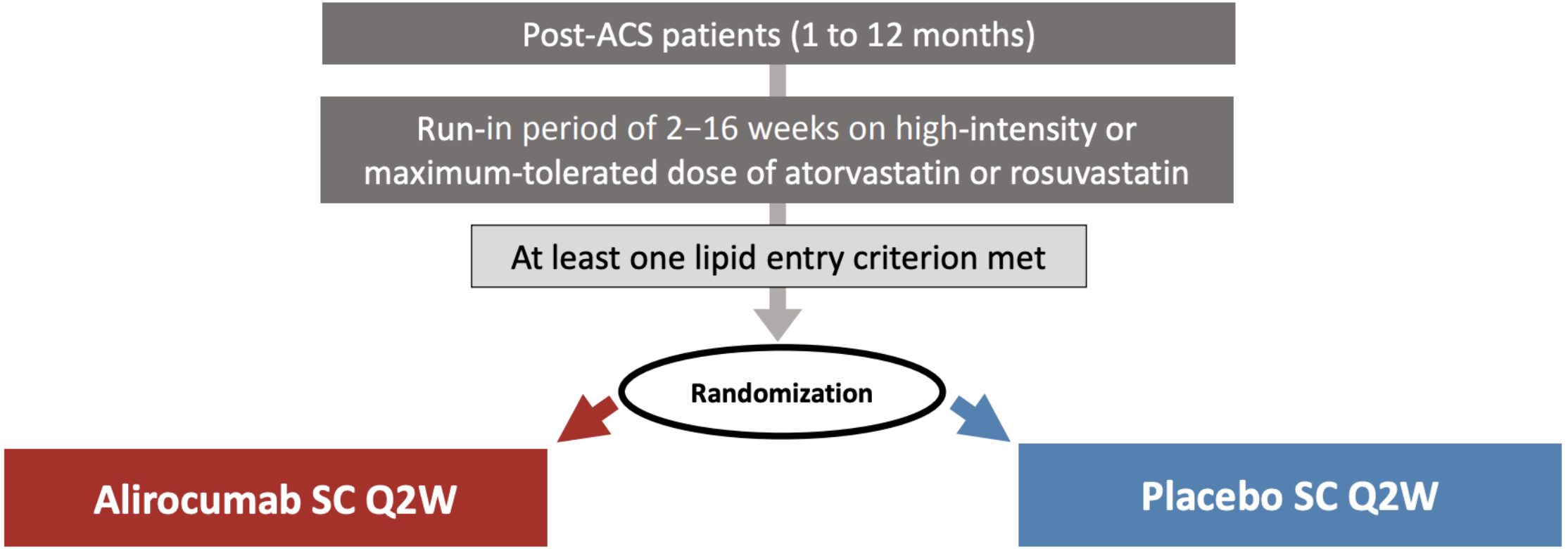
Rest of World	
Australia	216
Israel	582
New Zealand	257
South Africa	505

We thank the patients, their families, all investigators and coordinators involved in this study, and DCRI

ODYSSEY OUTCOMES

Steg et al. ACC 2018

Treatment Assignment

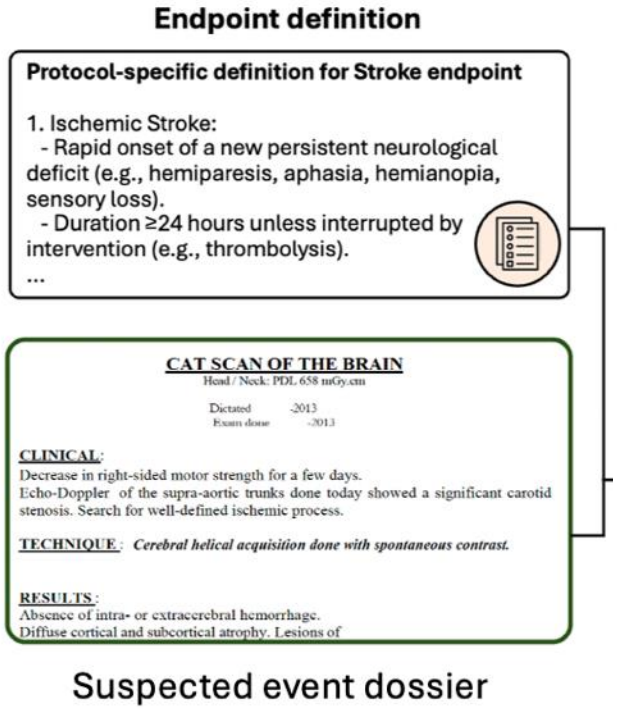


Patient and investigators remained blinded to treatment and lipid levels for the entire duration of the study

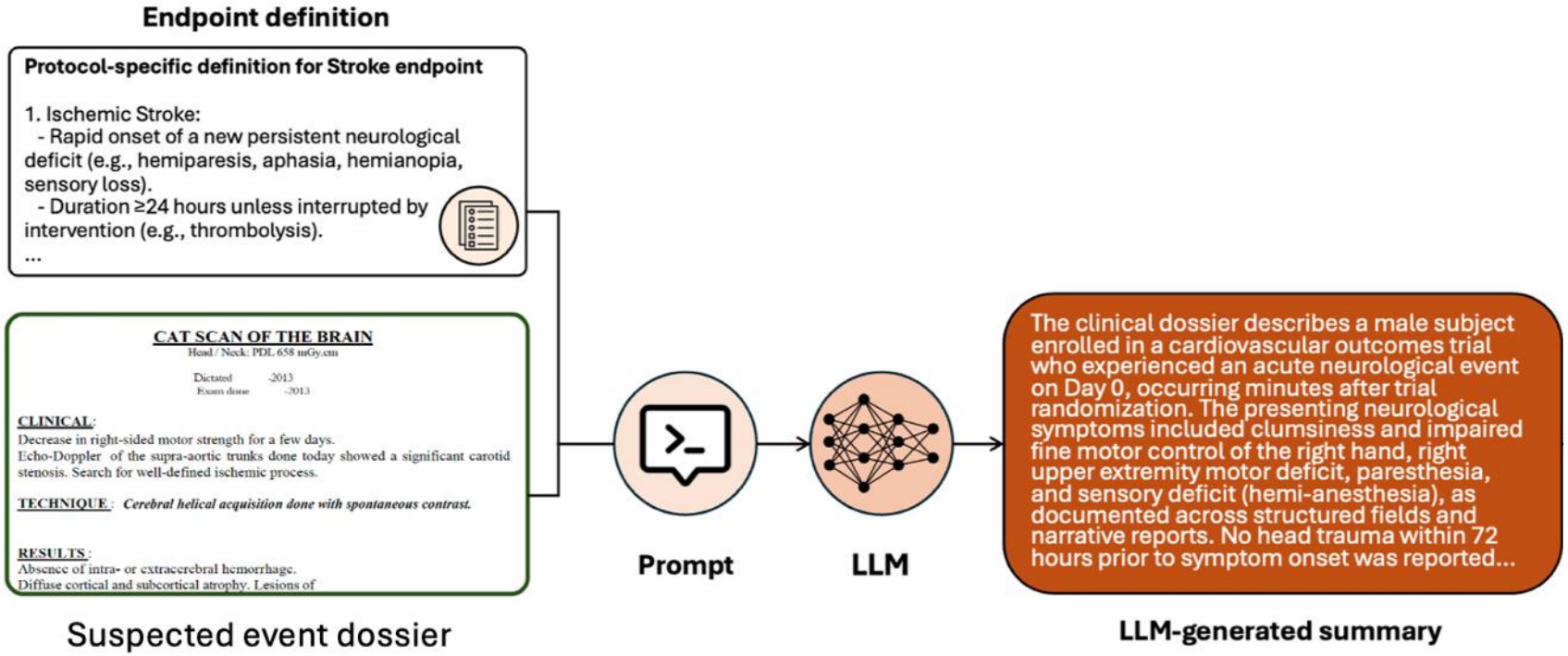
Schwartz GG, et al. Am Heart J 2014;168:682-689.e1.

ODYSSEY OUTCOMES

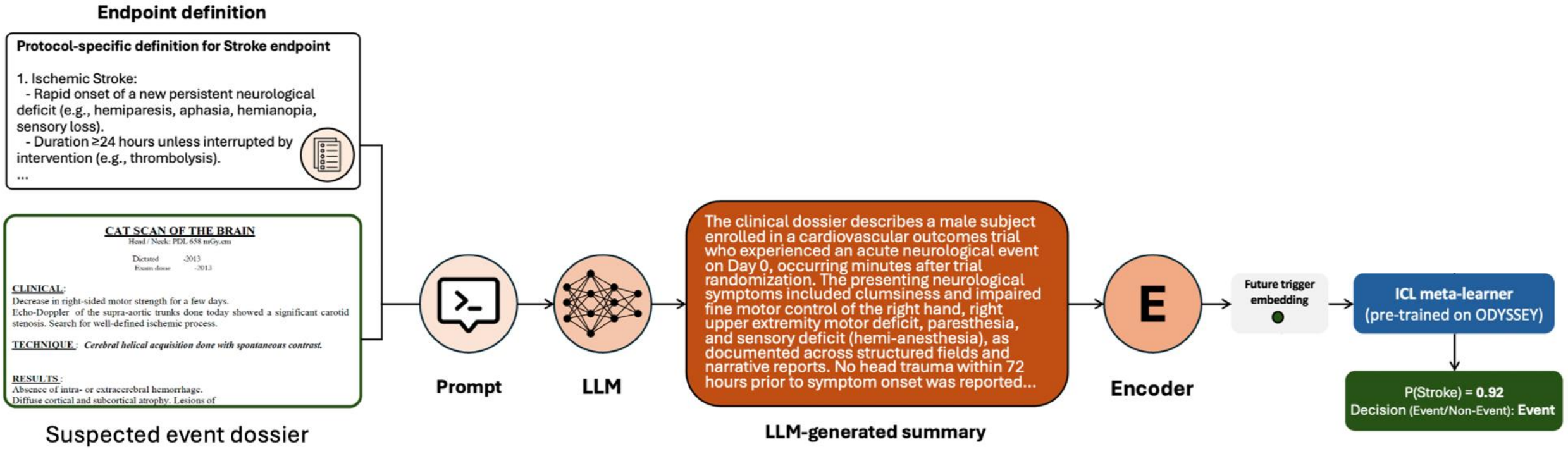
ADAPT-CEC Methods



ADAPT-CEC Methods



ADAPT-CEC Methods



Study Design

DEVELOPMENT PIPELINE

TRAINING DATA

ODYSSEY OUTCOMES

Source trial providing labeled clinical narratives for supervised learning

Trained on: MI | Stroke

ALGORITHM

ADAPT-CEC Model

Model trained on labeled events to classify endpoints

VALIDATION · EUCLID

Adjudication of New Events

Applied under new definitions to four endpoint categories

- MI (new definition)
- CV Death (new event)
- Stroke (new definition)
- Bleeding (new event)

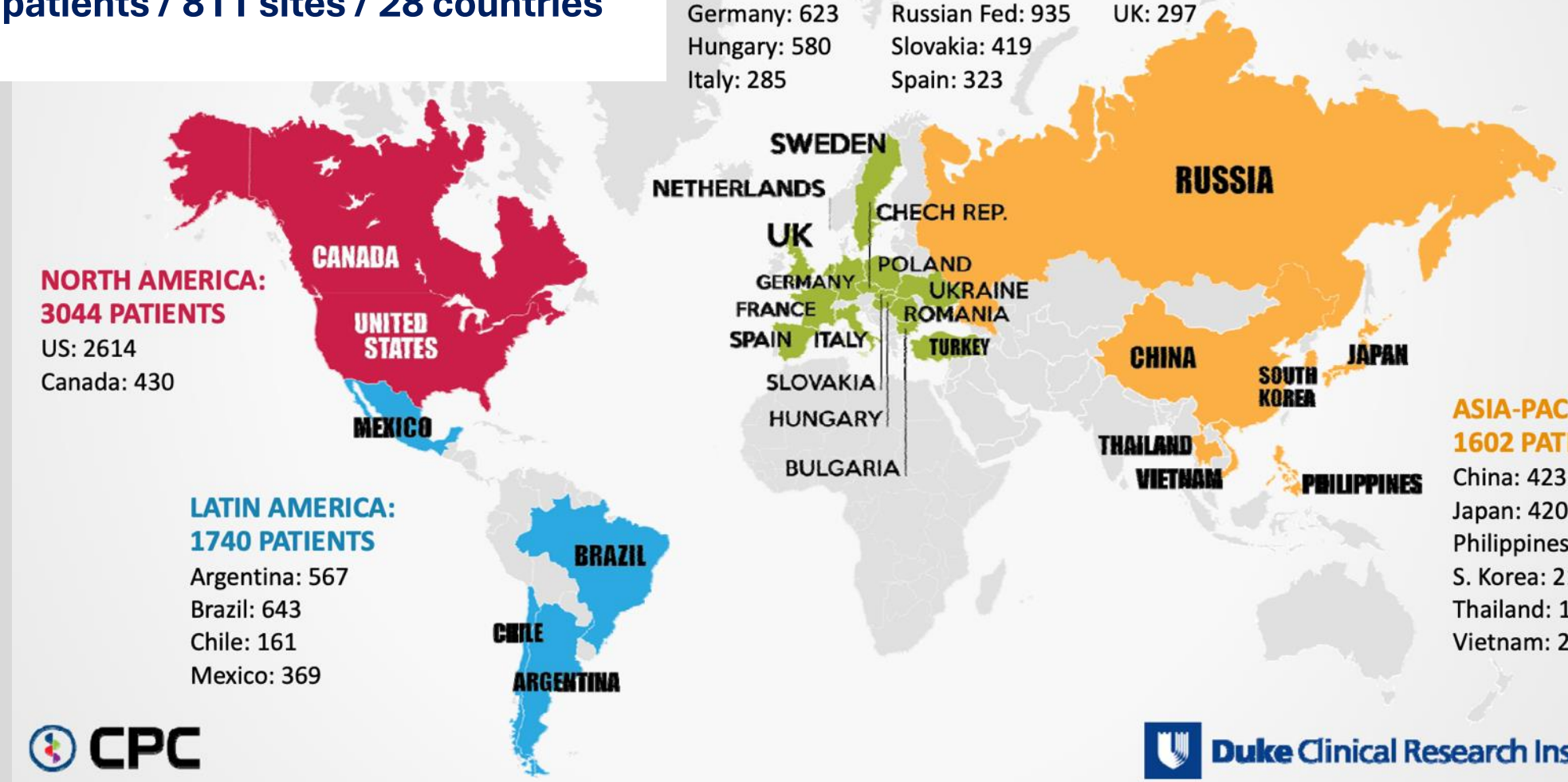
Step 1

Global Participation

13,855 patients / 811 sites / 28 countries

EUROPE: 7499 PATIENTS

Bulgaria: 679
Czech Rep: 723
France: 371
Germany: 623
Hungary: 580
Italy: 285
Netherlands: 234
Poland: 609
Romania: 508
Russian Fed: 935
Slovakia: 419
Spain: 323
Sweden: 240
Turkey: 137
Ukraine: 536
UK: 297



NORTH AMERICA: 3044 PATIENTS

US: 2614
Canada: 430

LATIN AMERICA: 1740 PATIENTS

Argentina: 567
Brazil: 643
Chile: 161
Mexico: 369

ASIA-PACIFIC: 1602 PATIENTS

China: 423
Japan: 420
Philippines: 128
S. Korea: 214
Thailand: 122
Vietnam: 295

CPC

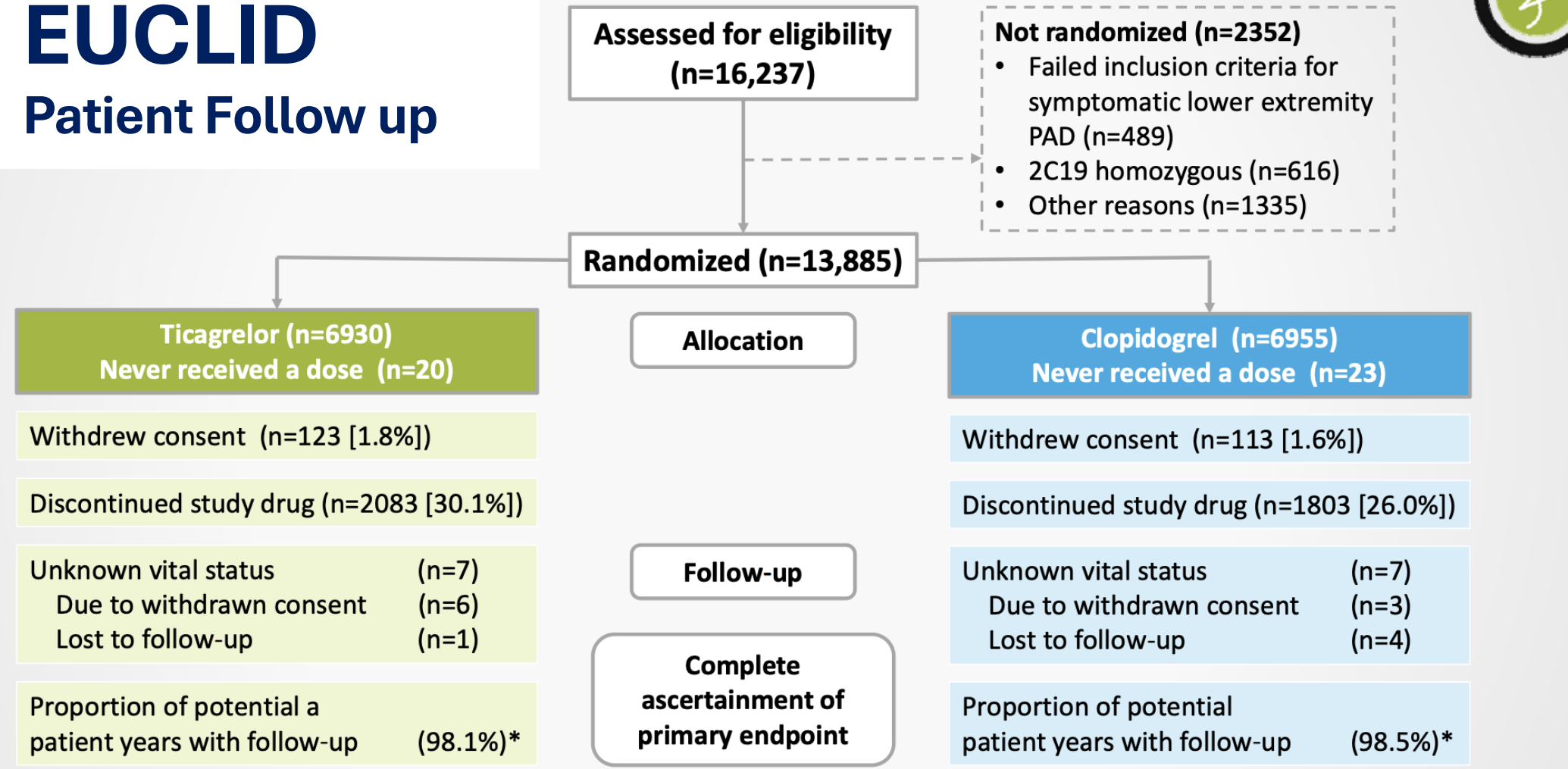
Duke Clinical Research Institute

ACC.26

Patel et al. AHA 2016

EUCLID

Patient Follow up



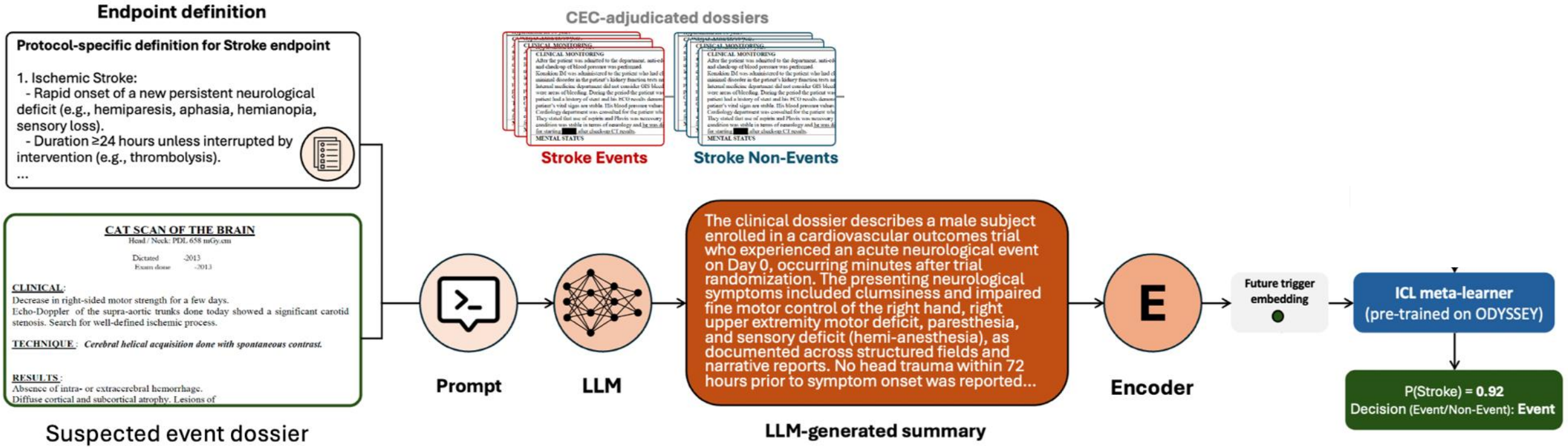
*Time from randomization until first primary event, censoring or death, divided by total time until first primary event, death or primary analysis censoring date.

CPC

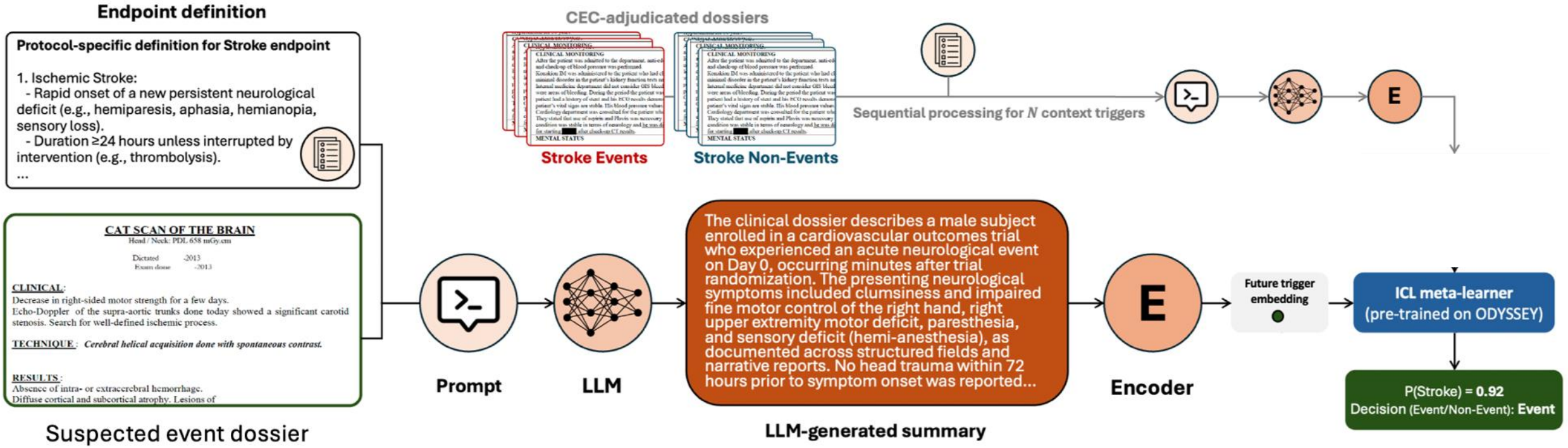
Duke Clinical Research Institute

Patel et al. AHA 2016

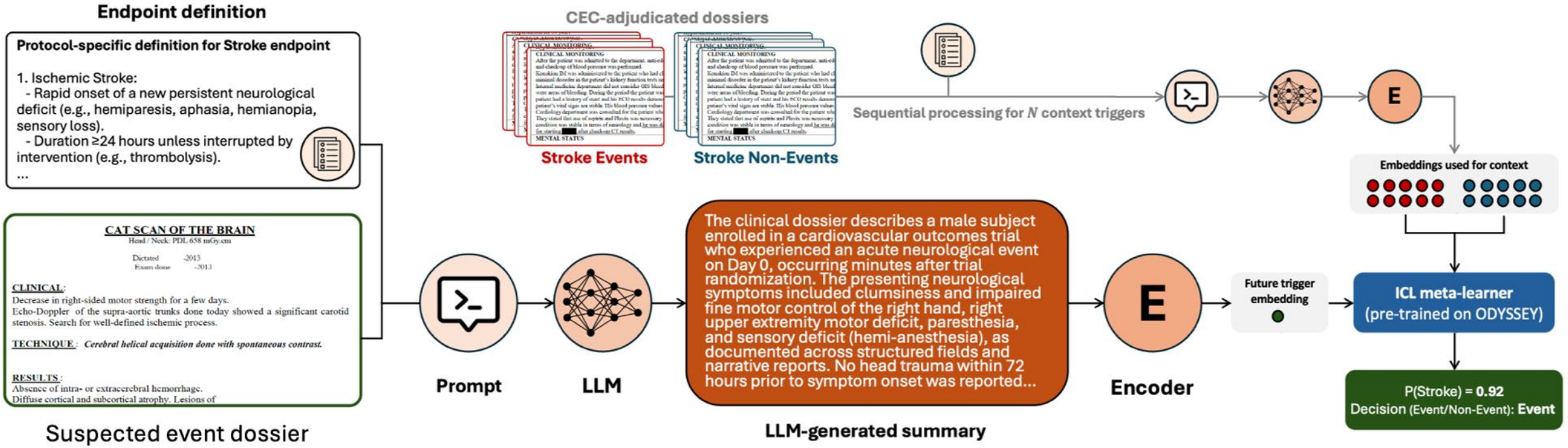
Adaptation to New Definitions / Endpoints



Adaptation to New Definitions / Endpoints



Adaptation to New Definitions / Endpoints



Study Design

DEVELOPMENT PIPELINE

TRAINING DATA

ODYSSEY OUTCOMES

Source trial providing labeled clinical narratives for supervised learning

Trained on: MI | Stroke

ALGORITHM

ADAPT-CEC Model

Model trained on labeled events to classify endpoints

VALIDATION · EUCLID

Adjudication of New Events

Applied under new definitions to four endpoint categories

- MI (new definition)
- CV Death (new event)
- Stroke (new definition)
- Bleeding (new event)

STRATEGY COMPARISON

ADAPT-CEC

Automated AI adjudication: no human review in the loop

VS

Hybrid ADAPT-CEC + Human

AI pre-screening with expert review on uncertain cases

VS

GPT-4

Large language model zero-shot adjudication

Step 1 | Step 2

Step 2: ADAPT-CEC in New Definitions in EUCLID

Endpoint	Metric	ADAPT-CEC Adaptive Training Size=20 (95% CI)	Hybrid ADAPT-CEC – Human Adjudication 30% Filtered (95% CI)	GPT4o (95% CI)	p-value (ADAPT vs. GPT)
MI	F1	0.731 (0.709–0.752)	0.795 (0.771-0.818)	0.697 (0.669–0.725)	0.031
	PPV	0.635 (0.607–0.663)	0.701 (0.669-0.733)	0.821 (0.791–0.851)	0.000
	TPR	0.859 (0.836–0.883)	0.917 (0.894 – 0.983)	0.606 (0.572–0.639)	0.000
	AUC	0.825 (0.807–0.842)			
	AUPRC	0.715 (0.678–0.750)			
	ECE	0.066 (0.050–0.082)			
Stroke	F1	0.780 (0.745–0.813)	0.821 (0.783 – 0.857)	0.812 (0.780–0.842)	0.048
	PPV	0.754 (0.709–0.796)	0.829 (0.780 – 0.877)	0.787 (0.745–0.828)	0.035
	TPR	0.809 (0.765–0.849)	0.812 (0.762 – 0.860)	0.838 (0.798–0.876)	0.231
	AUC	0.800 (0.764–0.834)			
	AUPRC	0.793 (0.741–0.843)			
	ECE	0.083 (0.057–0.111)			

ADAPT-CEC and Type I vs. Type II MI

Endpoint	Metric	ALL MI TYPES vs. No Event	TYPE 1 MI vs. TYPE 2 MI	Difference ALL MI TYPES vs. No Event - TYPE1 vs TYPE2	p-value
		ADAPT-CEC Ctxt=20 (95% CI)	ADAPT-CEC Ctxt=20 (95% CI)		
MI	Accuracy	0.760 (0.742–0.778)	0.739 (0.708–0.770)	+0.021	0.241
	F1	0.731 (0.709–0.752)	0.759 (0.725–0.789)	-0.028	0.161
	PPV	0.635 (0.607–0.663)	0.819 (0.780–0.856)	-0.184	0.000
	TPR	0.859 (0.836–0.883)	0.706 (0.665–0.747)	+0.153	0.000
	AUC	0.825 (0.807–0.842)	0.786 (0.752–0.818)	+0.039	0.038
	AUPRC	0.715 (0.678–0.750)	0.848 (0.813–0.880)	-0.135	0.000
	ECE	0.066 (0.050–0.082)	0.085 (0.067–0.119)	-0.023	0.091

Step 2: ADAPT-CEC in New Endpoints in EUCLID

Endpoint	Metric	ADAPT-CEC Adaptive Training Size=20 (95% CI)	Hybrid ADAPT-CEC – Human Adjudication 30% Filtered (95% CI)	GPT4o (95% CI)	p-value (ADAPT vs. GPT)
CV Death	F1	0.873 (0.854–0.891)	0.940 (0.925 – 0.955)	0.862 (0.842–0.882)	0.295
	PPV	0.891 (0.867–0.913)	0.931 (0.909 – 0.953)	0.954 (0.936–0.970)	0.000
	TPR	0.857 (0.830–0.882)	0.949 (0.929 – 0.967)	0.787 (0.757–0.817)	0.000
	AUC	0.919 (0.903–0.934)			
	AUPRC	0.937 (0.920–0.952)			
	ECE	0.088 (0.071–0.106)			
Bleeding	F1	0.780 (0.745–0.813)	0.833 (0.819 – 0.846)	0.564 (0.546–0.581)	0.000
	PPV	0.754 (0.709–0.796)	0.951 (0.939 – 0.961)	0.707 (0.686–0.727)	0.000
	TPR	0.809 (0.765–0.849)	0.741 (0.721 – 0.760)	0.468 (0.449–0.487)	0.000
	AUC	0.800 (0.764–0.834)			
	AUPRC	0.793 (0.741–0.843)			
	ECE	0.083 (0.057–0.111)			

Study Design

DEVELOPMENT PIPELINE

TRAINING DATA

ODYSSEY OUTCOMES

Source trial providing labeled clinical narratives for supervised learning

Trained on: MI | Stroke

ALGORITHM

ADAPT-CEC Model

Model trained on labeled events to classify endpoints

VALIDATION · EUCLID

Adjudication of New Events

Applied under new definitions to four endpoint categories

- MI (new definition)
- CV Death (new event)
- Stroke (new definition)
- Bleeding (new event)

STRATEGY COMPARISON

ADAPT-CEC
Automated AI adjudication: no human review in the loop

VS

Hybrid ADAPT-CEC + Human
AI pre-screening with expert review on uncertain cases

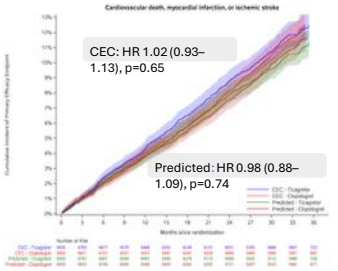
VS

GPT-4
Large language model zero-shot adjudication

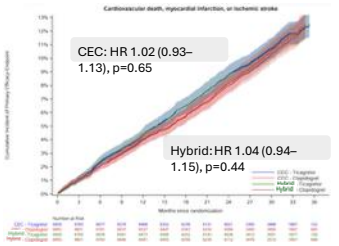
EUCLID PRIMARY ENDPOINT REPLICATION

Kaplan–Meier cumulative incidence: CV Death, MI, or Ischemic Stroke

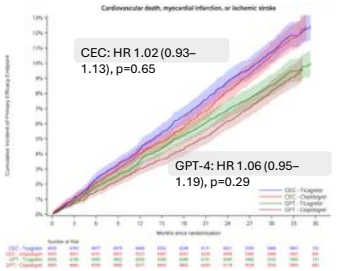
ADAPT-CEC



Hybrid ADAPT-CEC + Human

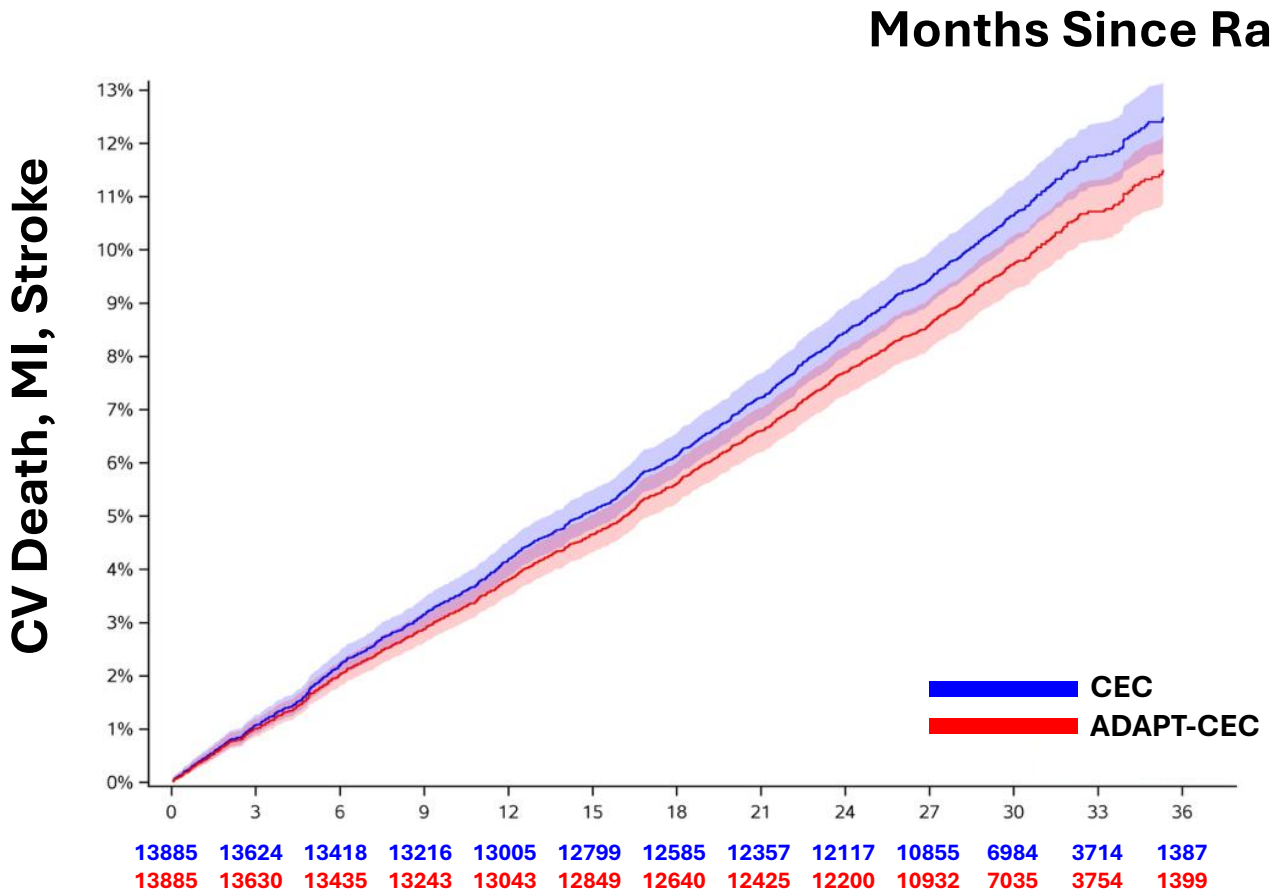


GPT-4

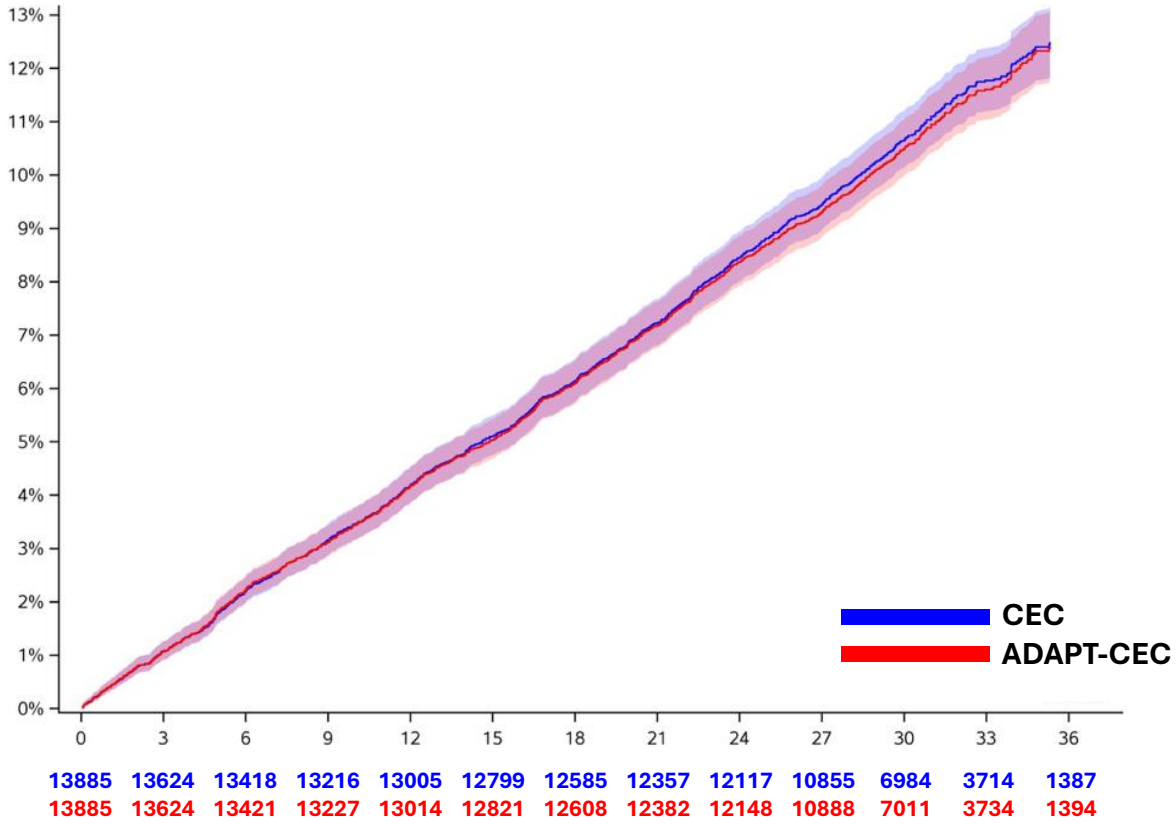


Step 1 | **Step 2** | **Step 3**

Step 3: EUCLID 1º Endpoint Cumulative Incidence

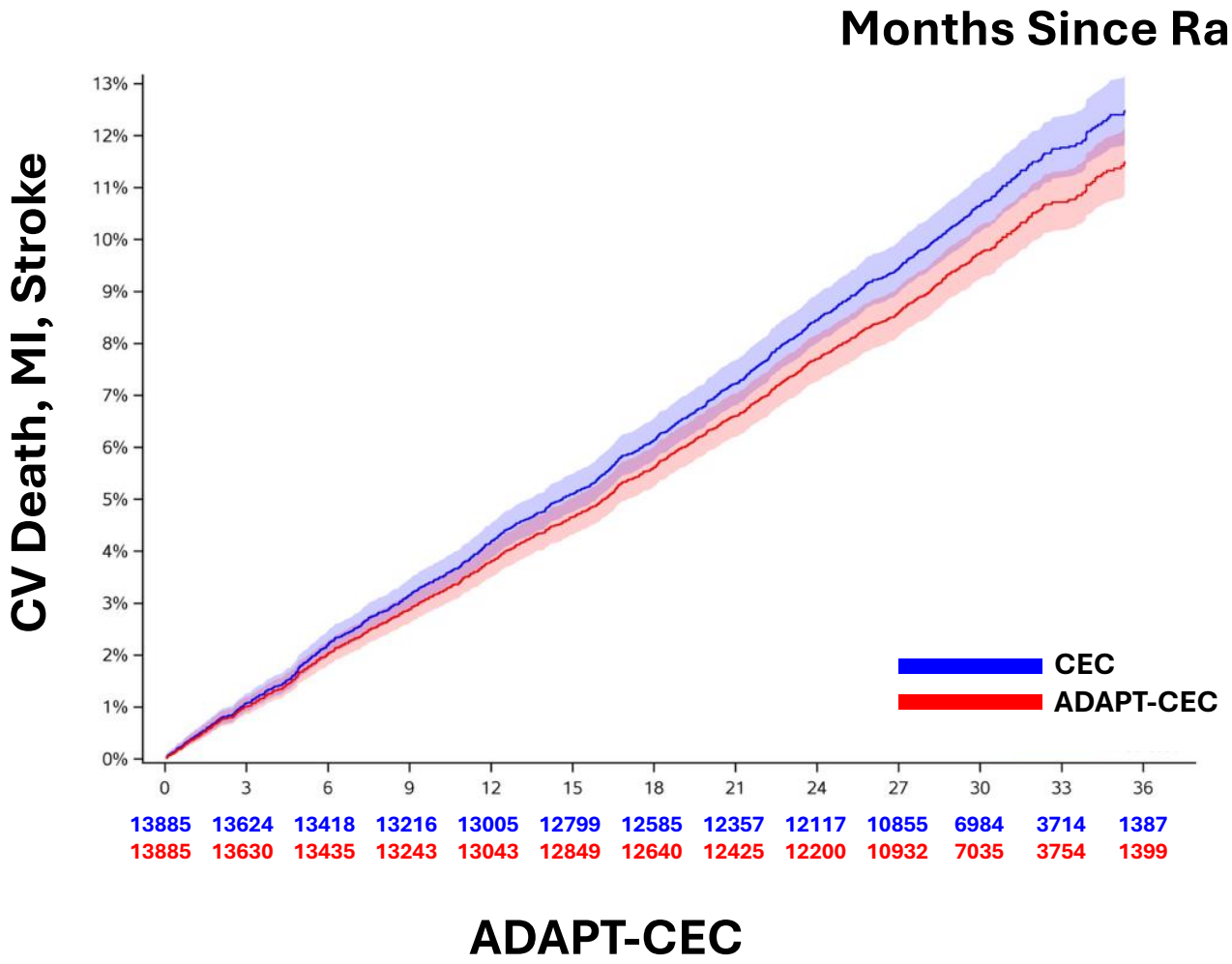


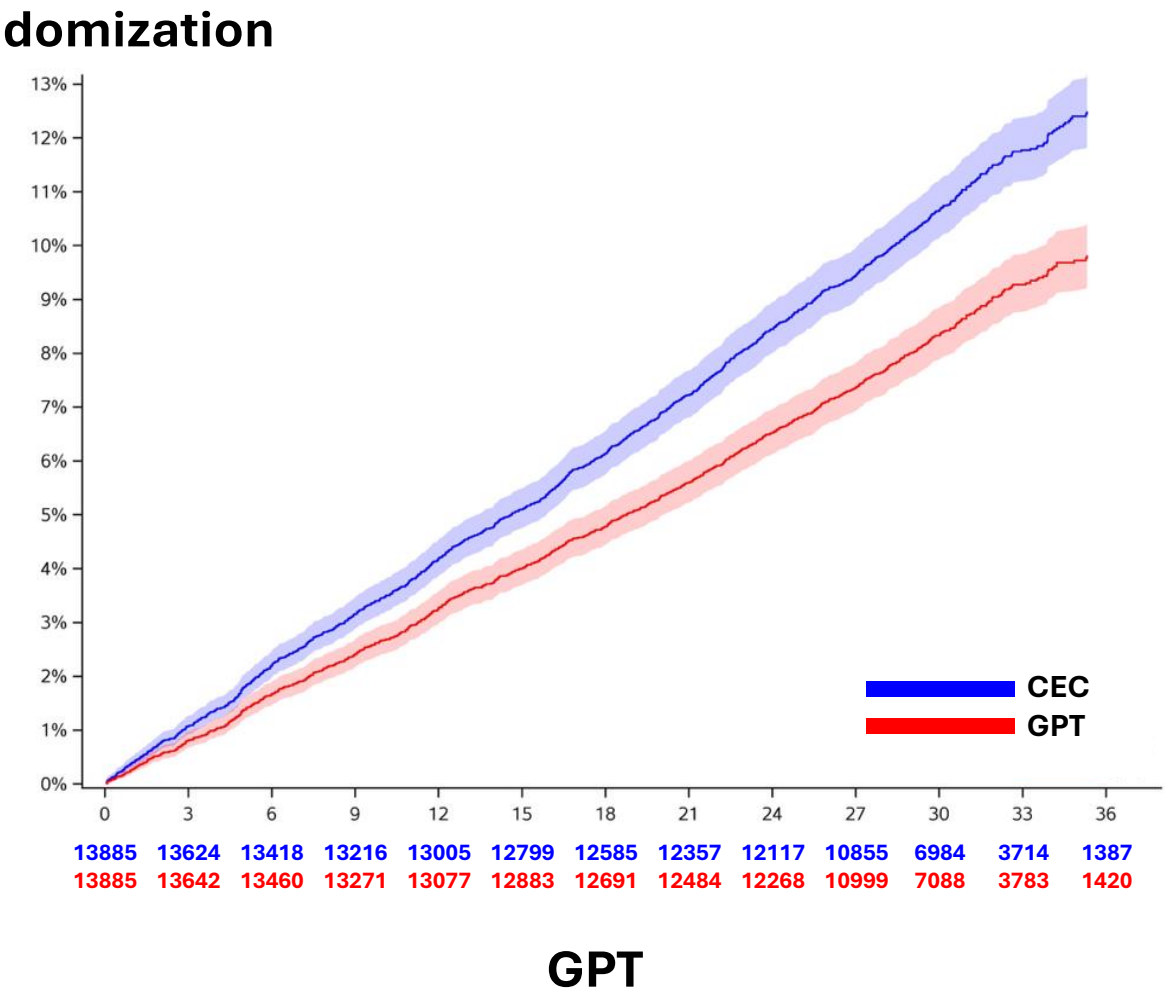
ADAPT-CEC



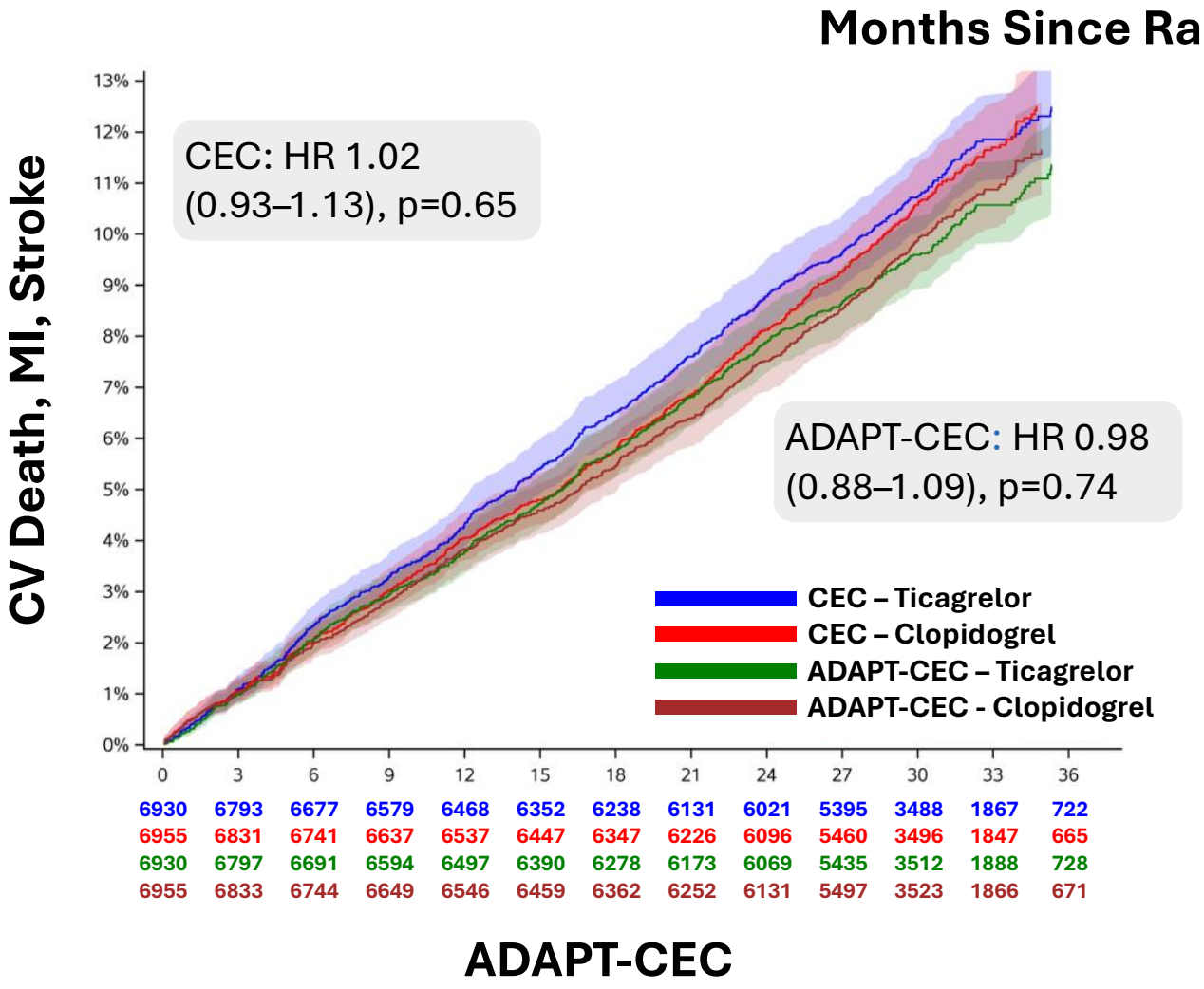
HYBRID

Step 3: EUCLID 1º Endpoint Cumulative Incidence

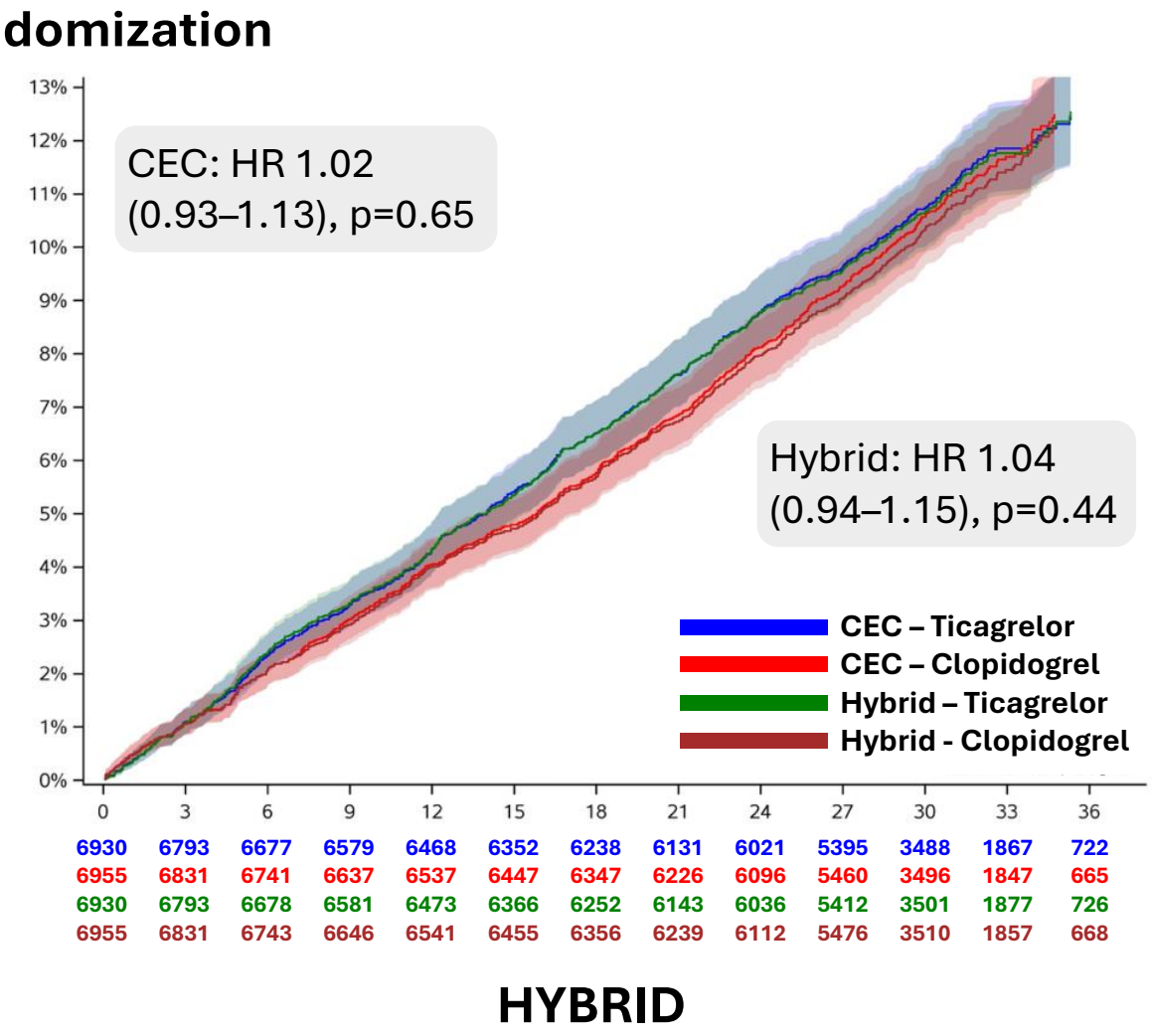




Step 3: EUCLID 1º Endpoint Treatment Effect

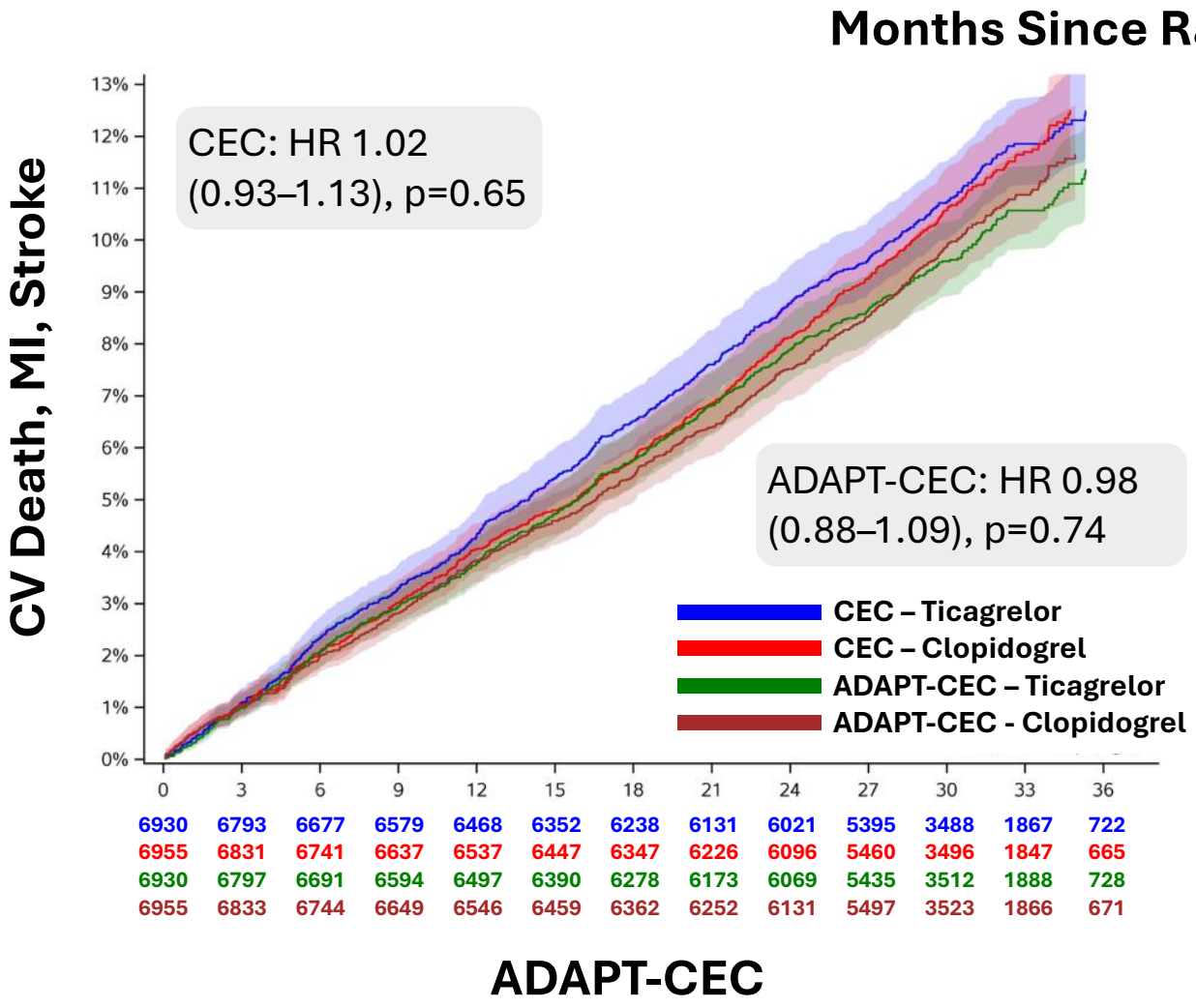


ADAPT-CEC

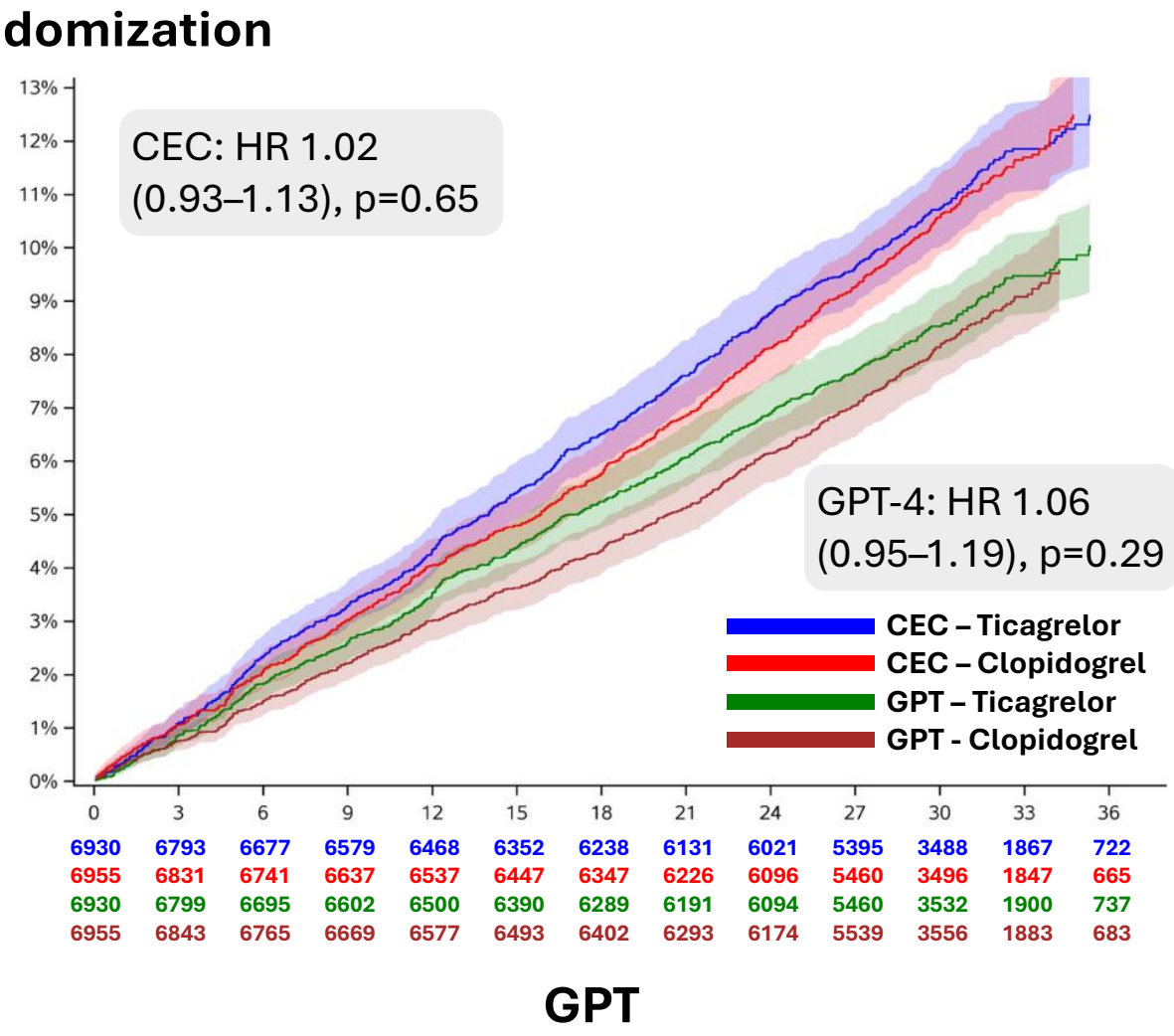


HYBRID

EUCLID 1º Endpoint Treatment Effect



ADAPT-CEC



GPT

Limitations

1. Further work will be needed to validate ADAPT-CEC on other endpoints

2. ADAPT-CEC uses an LLM to summarize an event dossier, making it susceptible to changes in the underlying LLM

3. Performance across event types was not uniform

4. Training and Validation performed on adjudications at a single ARO

5. Analyses were retrospective and prospective assessment of AI-CEC will be needed

Conclusions

- A single AI algorithm trained on a single trial can, after brief adaptation:
  - identify the same events with a new definition
  - Identify new events

- Validated in the EUCLID Trial

- ADAPT-CEC alone or as a hybrid strategy with human adjudication reproduced the EUCLID primary endpoint and effect size

- Further prospective studies are needed to determine how adaptive AI adjudication can be implemented in future trials.

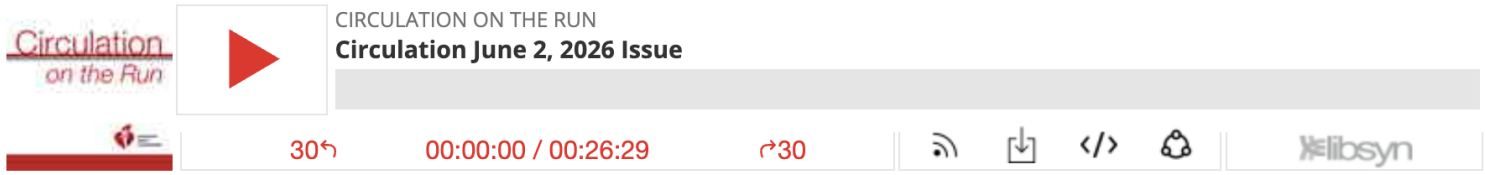
Circulation

Ad
Ev
OU
Sr
W
Rc
G.

CIRCULATION ON THE RUN | ORIGINALLY PUBLISHED JUNE 1, 2026

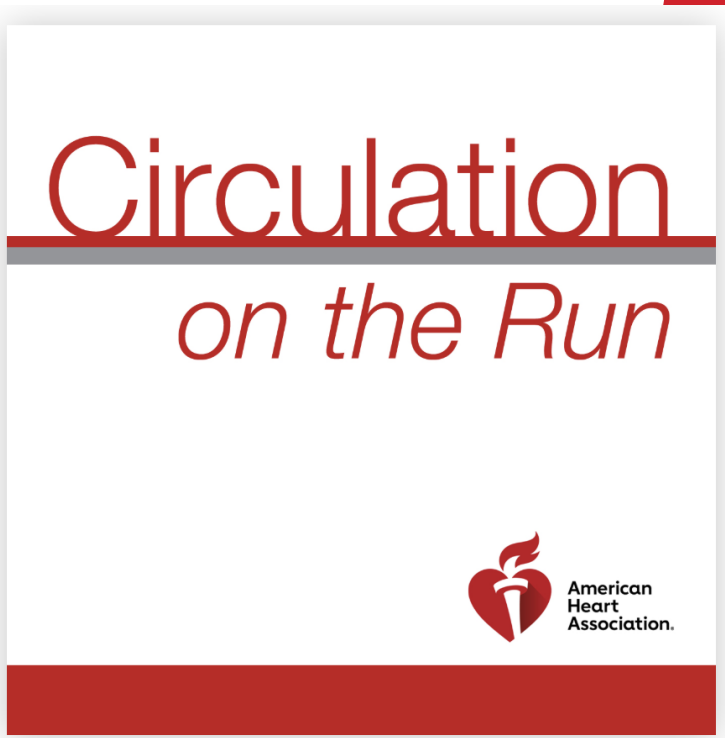
Circulation on the Run: June 2, 2026

W. Gregory Hundley, MD | Peder L. Myhre, MD, PhD | Shirin Doroudgar, PhD



This week, please join author Sreekanth Vemulapalli as he discusses the article "Adaptive AI for Cardiovascular Event Adjudication: Cardiovascular Event Adjudication Across Different Definitions in the ODYSSEY OUTCOMES and EUCLID Trials."

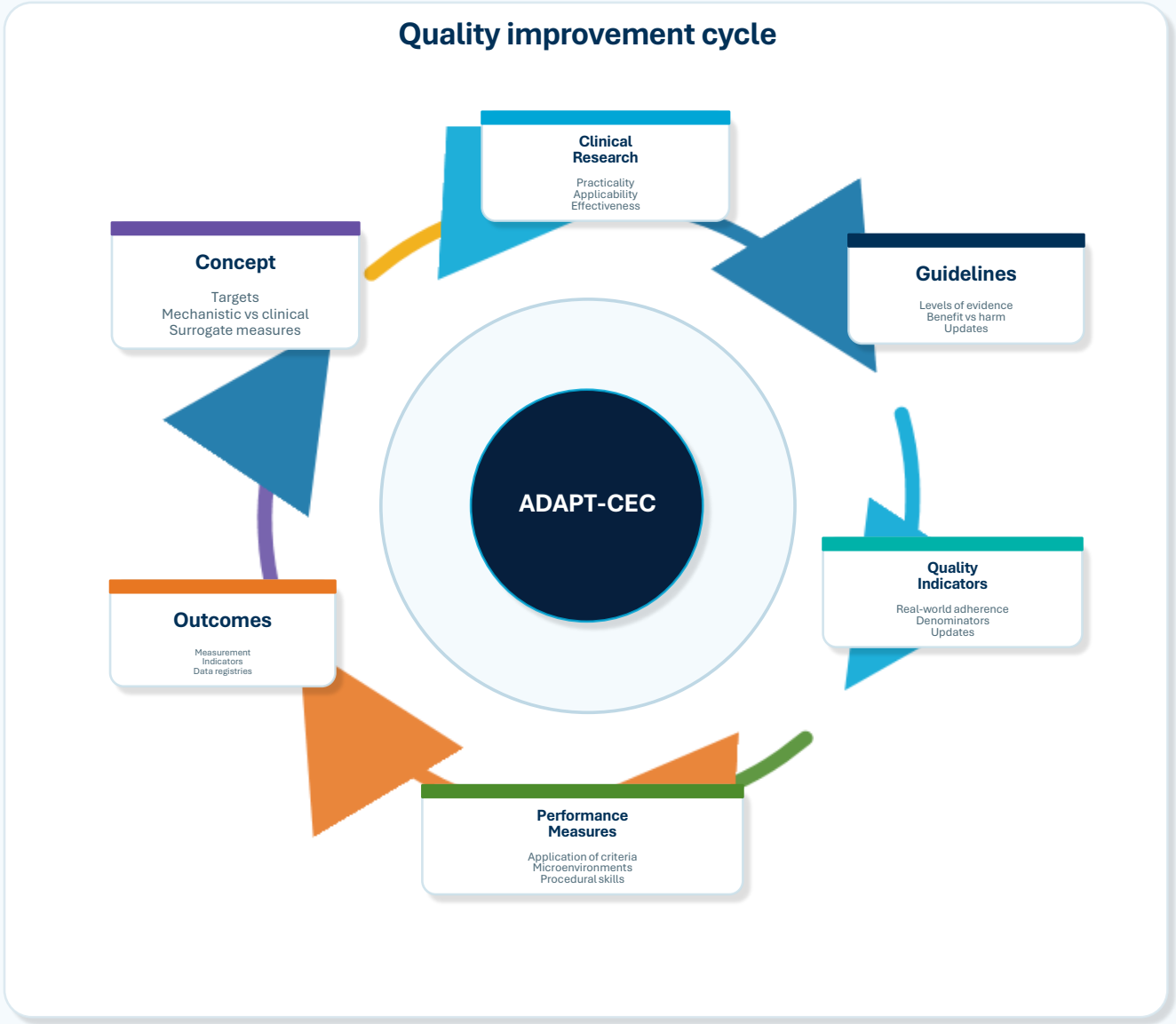
Ricardo Henao



CIRCULATION. 2026; [PUBLISHED ONLINE AHEAD OF PRINT]. DOI: 10.1161/CIRCULATIONAHA.126.080072

Implications

ADAPT-CEC in the Quality Improvement Cycle



Adapted from Califf RM et al. JACC 2002

ADAPT-CEC in the Quality Improvement Cycle

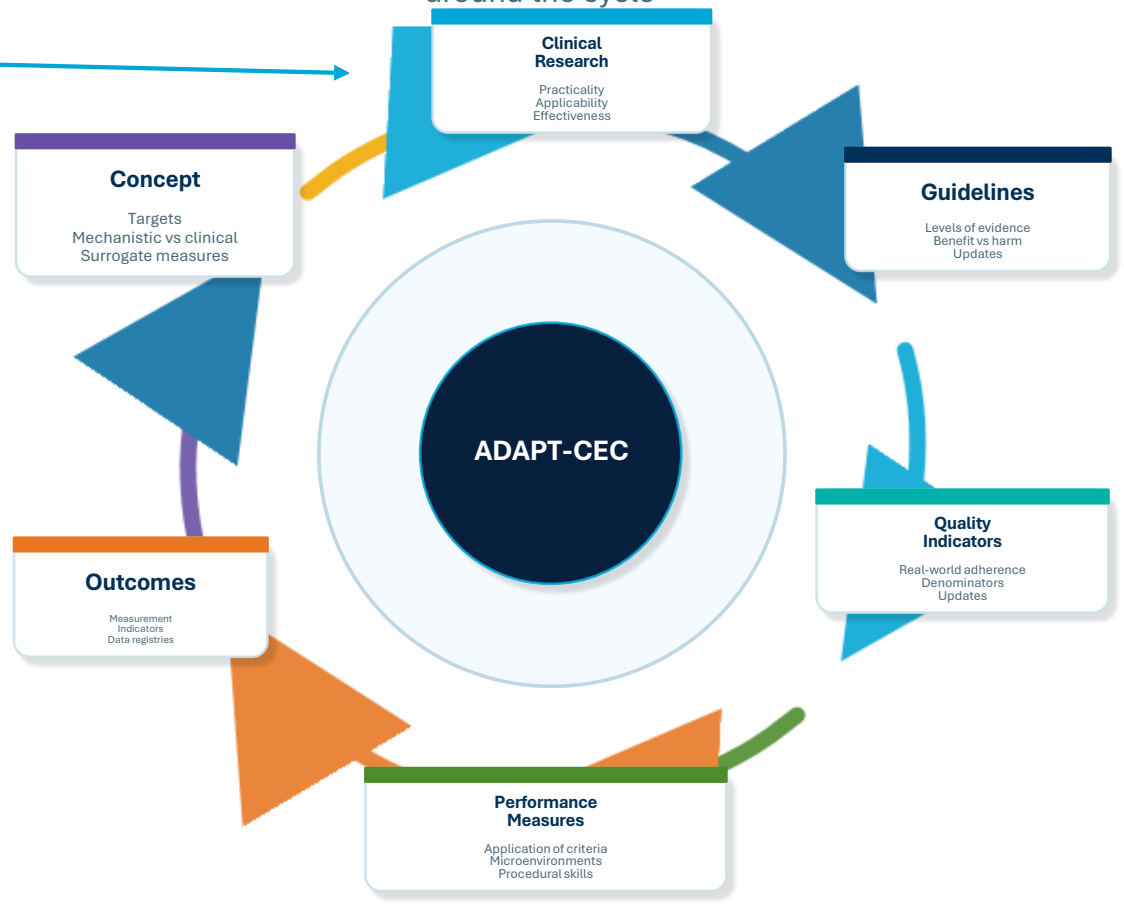
Trial Evidence Generation

1 Bespoke clinical trials

Event dossiers + CRFs are adjudicated against trial-specific endpoint definitions. ADAPT-CEC can triage high-certainty events while routing uncertain cases to the human CEC.

Quality improvement cycle

ADAPT-CEC supplies adjudicated, auditable event outcomes that can generate evidence around the cycle



Adapted from Califf RM et al. JACC 2002

ADAPT-CEC in the Quality Improvement Cycle

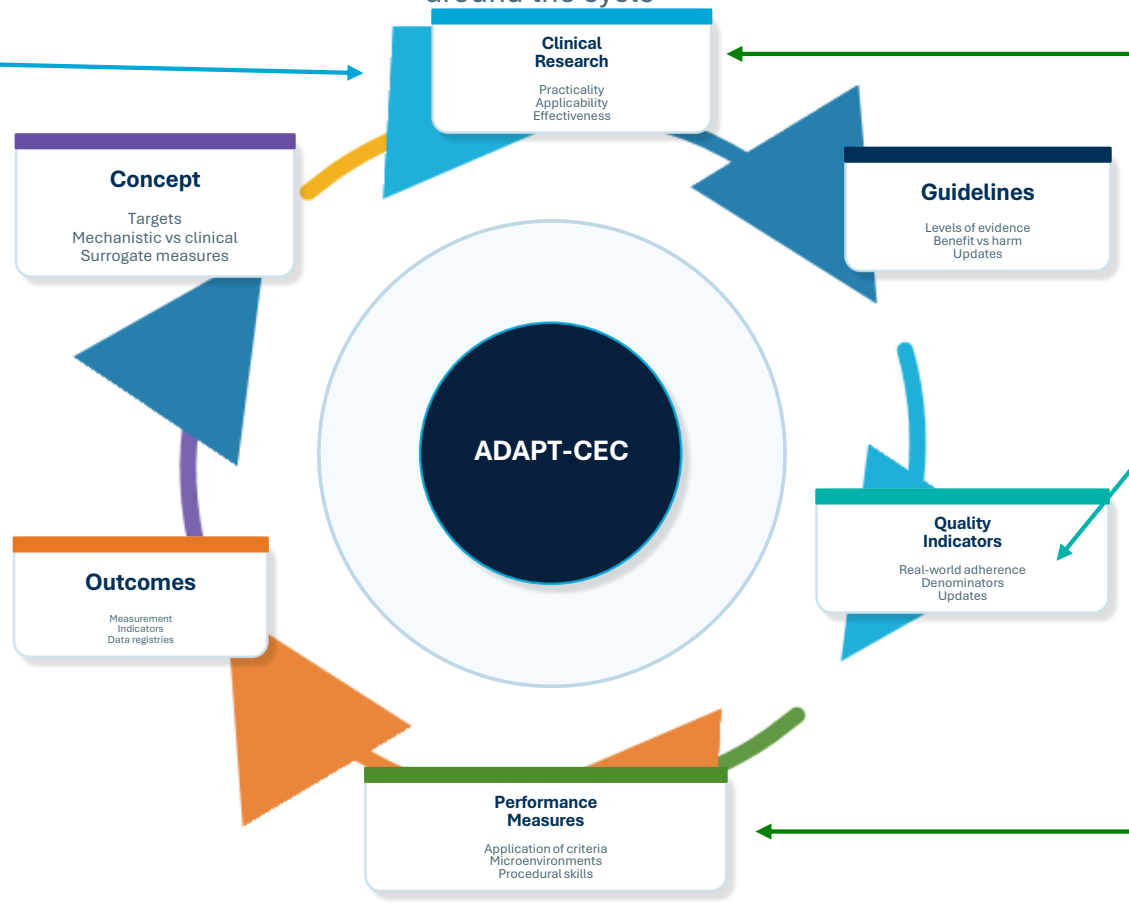
Trial Evidence Generation

1 Bespoke clinical trials

Event dossiers + CRFs are adjudicated against trial-specific endpoint definitions. ADAPT-CEC can triage high-certainty events while routing uncertain cases to the human CEC.

Quality improvement cycle

ADAPT-CEC supplies adjudicated, auditable event outcomes that can generate evidence around the cycle



Practice-Embedded Measurement

2 Quality Assessment and Pragmatic EHR-embedded trials

ADAPT-CEC can be embedded into EHR-linked pragmatic workflows so endpoint definitions are applied directly to clinical notes, labs, imaging reports, and procedures.

Adapted from Califf RM et al. JACC 2002

ADAPT-CEC in the Quality Improvement Cycle

Trial Evidence Generation

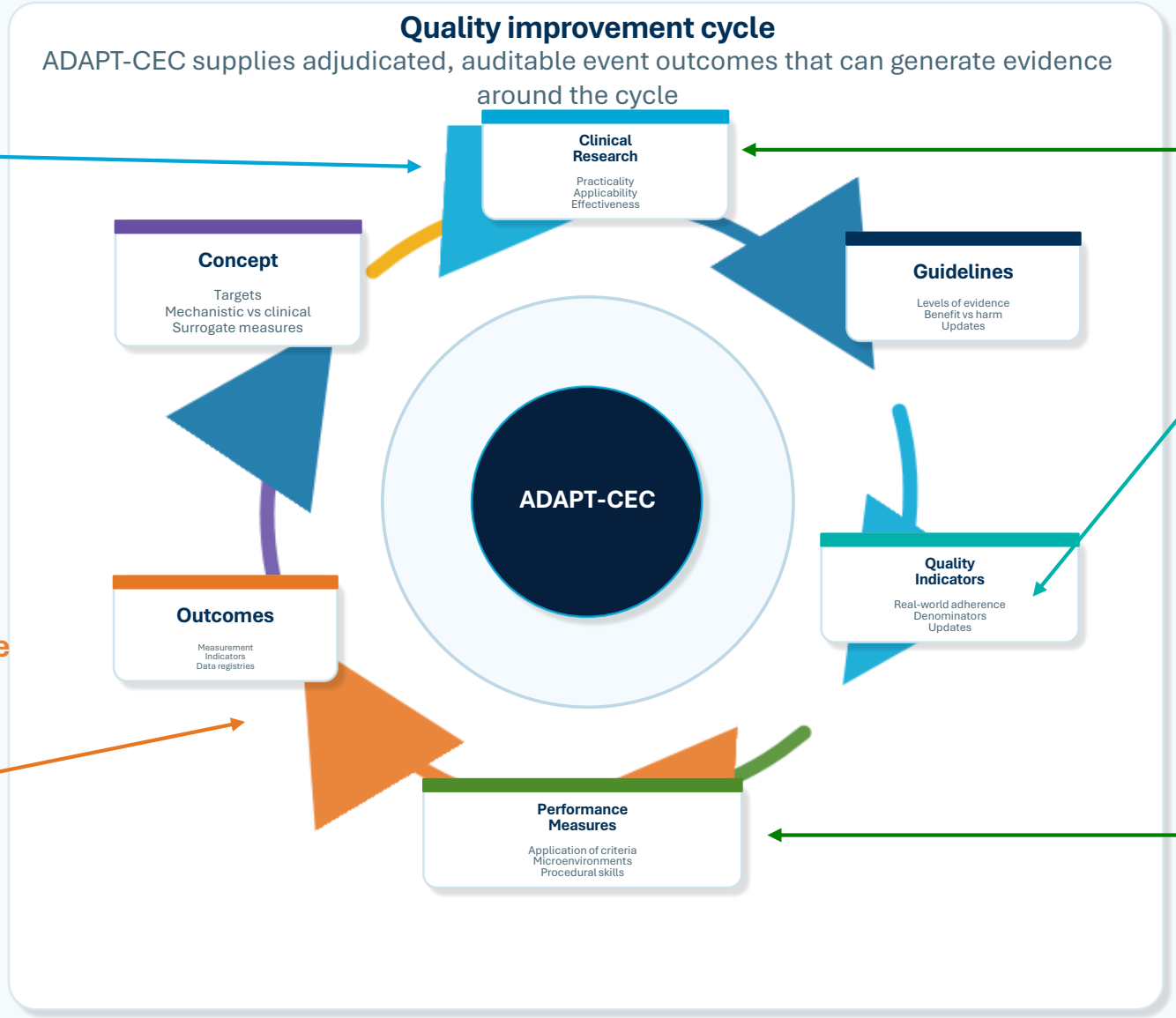
1 Bespoke clinical trials

Event dossiers + CRFs are adjudicated against trial-specific endpoint definitions. ADAPT-CEC can triage high-certainty events while routing uncertain cases to the human CEC.

Outcomes Feedback + Surveillance

3 Post-market RWE surveillance

Longitudinal EHR data can be adjudicated to support safety and effectiveness surveillance, with auditable outputs feeding outcomes, quality indicators, and future study concepts.



Practice-Embedded Measurement

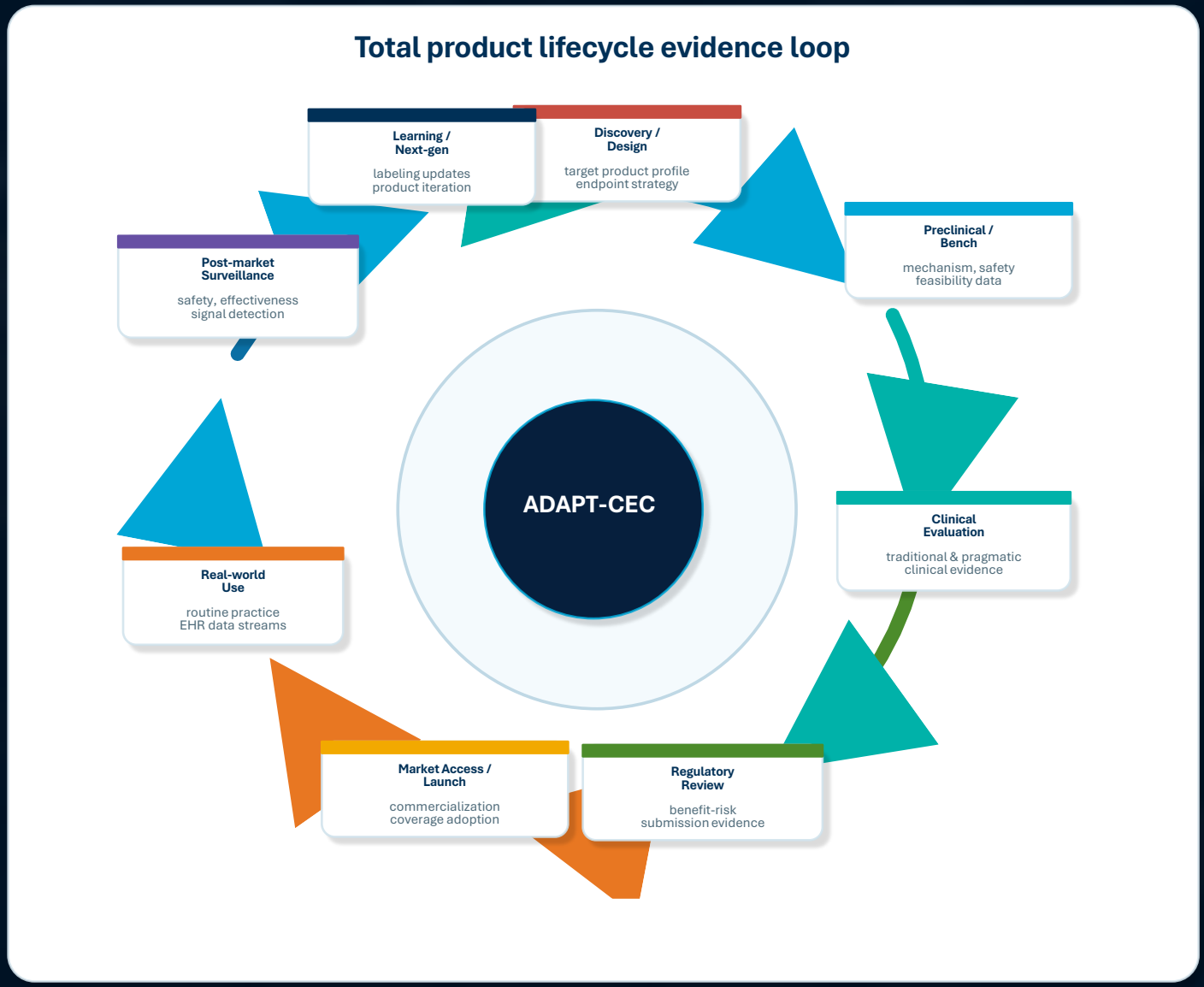
2 Quality Assessment and Pragmatic EHR-embedded trials

ADAPT-CEC can be embedded into EHR-linked pragmatic workflows so endpoint definitions are applied directly to clinical notes, labs, imaging reports, and procedures.

Adapted from Califf RM et al. JACC 2002

ADAPT-CEC across the Total Product Lifecycle

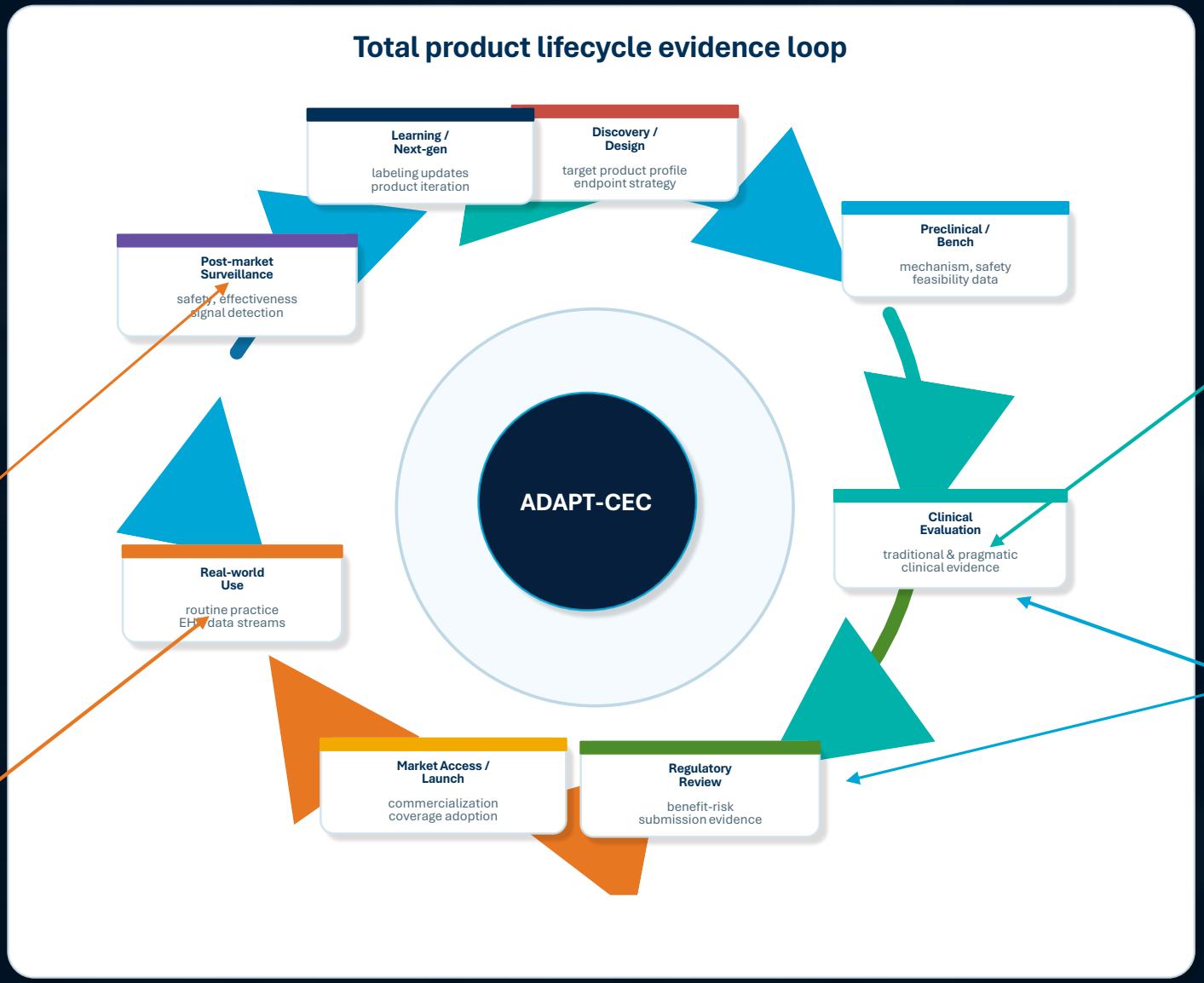
Event adjudication as a reusable evidence-generation layer from pivotal studies to real-world surveillance



ADAPT-CEC across the Total Product Lifecycle

Event adjudication as a reusable evidence-generation layer from pivotal studies to real-world surveillance

Total product lifecycle evidence loop



Practice-Embedded Clinical Evidence

2 EHR-embedded pragmatic trials

Applies trial endpoint definitions directly to routine-care EHR data streams to support lower-burden pragmatic trial adjudication during clinical evaluation and real-world use.

Premarket / Pivotal Evidence

1 Bespoke pivotal trials

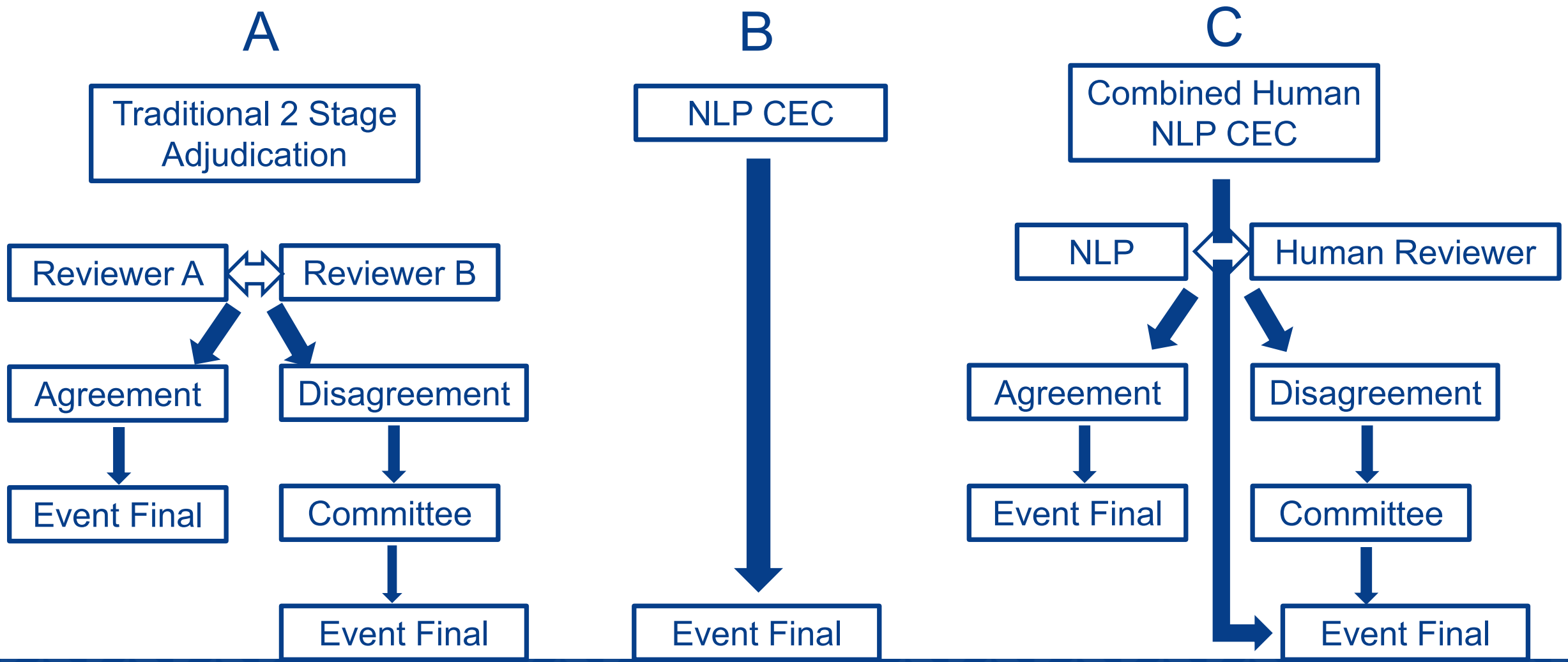
Facilitates endpoint adjudication for traditional sponsor/CRO-style studies using event dossiers, CRFs, and protocol-defined endpoints; maps mainly to clinical evaluation and regulatory review.

Postmarket Feedback Loop

3 Post-market RWE surveillance

Provides adjudication of longitudinal EHR data to generate real-world evidence for safety, effectiveness, signal evaluation, labeling updates, and next-generation development.

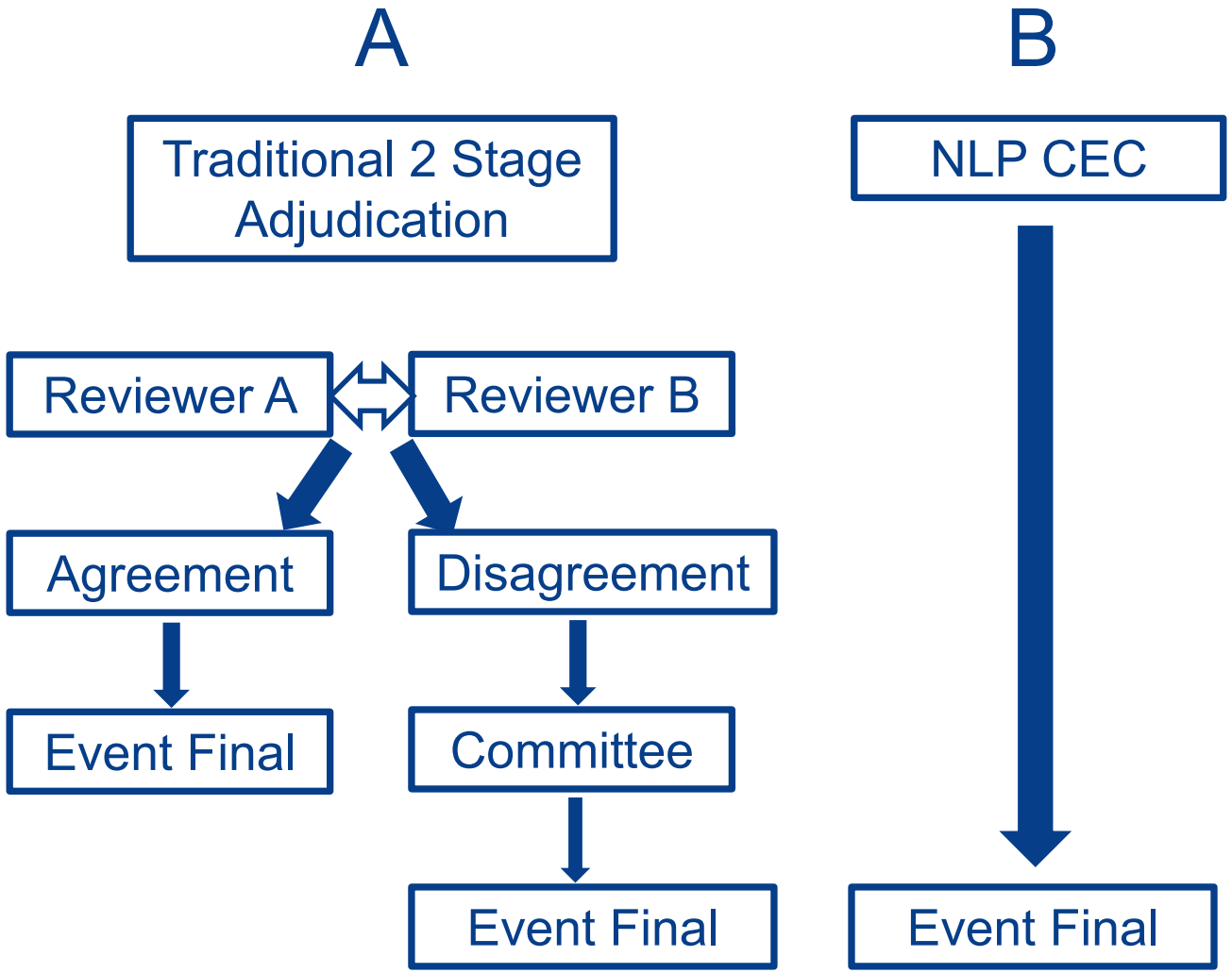
Prospectively Adjudicate a Trial



Implementation in an EHR

Rationale:

- EMR-based clinical trials



Conclusions

- A single AI algorithm trained on a single trial can, after brief adaptation:
  - identify the same events with a new definition
  - Identify new events
- ADAPT-CEC alone or as a hybrid strategy with human adjudication reproduced the EUCLID primary endpoint and effect size
- Further prospective studies are needed to determine how adaptive AI adjudication can be implemented in future trials.
- The potential impact of AI-enabled automated adjudication of outcomes on developing evidence for quality improvement and regulatory evidence is substantial