Penn Medicine

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Department of Biostatistics, Epidemiology and Informatics

UNO: Neural Pruning Enhances AI Reliability of Observational Data-based Findings

Yue Wu
University of Pennsylvania

2026 OHDSI Symposium

CHASE
Center for Health AI & Synthesis of Evidence



ennCIL
A Computing · Inference · Learning lab at University of Pennsylvania

Motivation: Reliable AI for Observational Studies

1. **Observational data enable large-scale drug research.**

   They support evaluation of treatment effects across diverse populations.

2. **Systematic error can persist despite careful study design.**

   Even with target trial emulation, incompletely captured clinical factors can leave residual bias.

3. **Neural networks may learn spurious correlations from biased data.**

   These learned patterns can undermine the reliability of downstream findings.

> **How can we mitigate systematic error in AI models?**

Negative control outcomes (NCO): signals of potential systematic error

1 NCO definition

No causal effect of treatment.

Treatment ⊘→ **NCO**

May share unmeasured confounders with the primary outcome.

2 Estimate a signal

Compare predicted NCO outcomes.

Treated − **Control**

Same observed covariates

Differences can signal potential systematic error.

3 Guide refinement

Combine signals from multiple NCOs.

High-NCO-signal subset

Low-NCO-signal subset

Signal groups, not bias labels.

UNO uses NCO-derived signals to guide model refinement.

UNO: NCO-guided pruning and fine-tuning

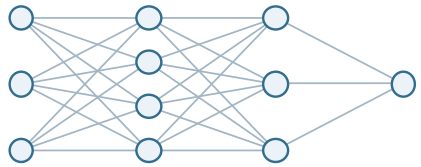
Here: a propensity-score neural network predicts treatment from baseline covariates.

NCO-derived subsets (from the previous slide)

High-NCO-signal

Low-NCO-signal

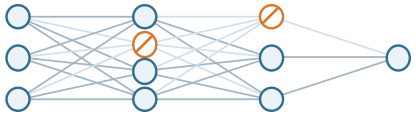
1 Base model



Predict treatment using covariates

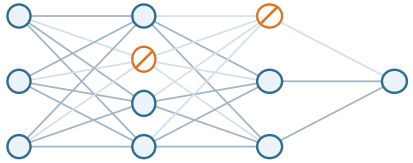
2 Score & prune

Mean activation² ratio (high / low)



Set top-scoring neurons to 0

3 Fine-tune



Low-signal training
Fixed pruning mask

4 Full cohort



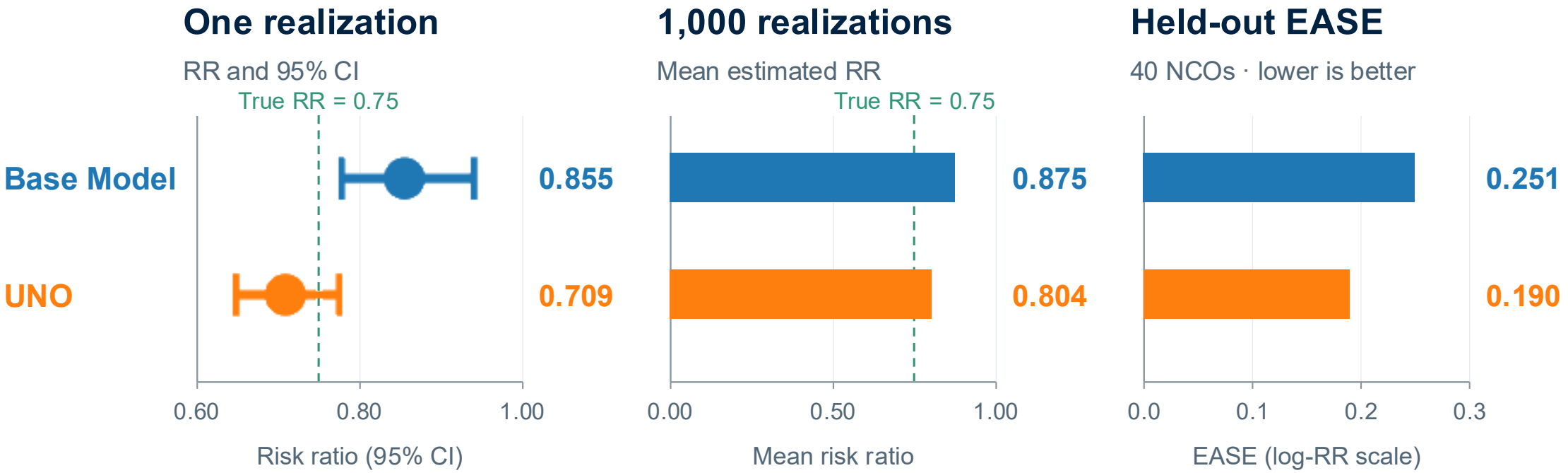
Propensity scores for every patient

Treatment-effect estimation uses the full cohort—both subsets retained.

An illustrative simulation example

10,000 individuals | True causal risk ratio (RR) = 0.75 | 40 held-out NCOs

NCOs and target outcome share unmeasured variables through different nonlinear functions.



Mean RR: 1,000 paired target-outcome realizations, with the cohort and fitted models held fixed.

In this example, UNO had a mean RR closer to 0.75 and lower held-out EASE than the Base Model.

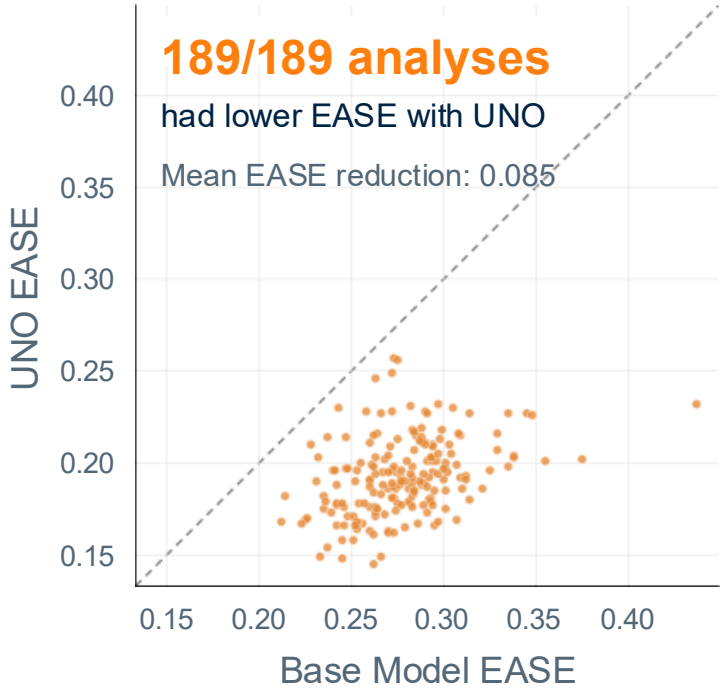
Result 1: GLP-1RAs vs DPP-4is

Pharmacologic outcome profiling | TriNetX, 2019–2024

303,580 treatment initiators with T2D | 189 outcome-specific analyses

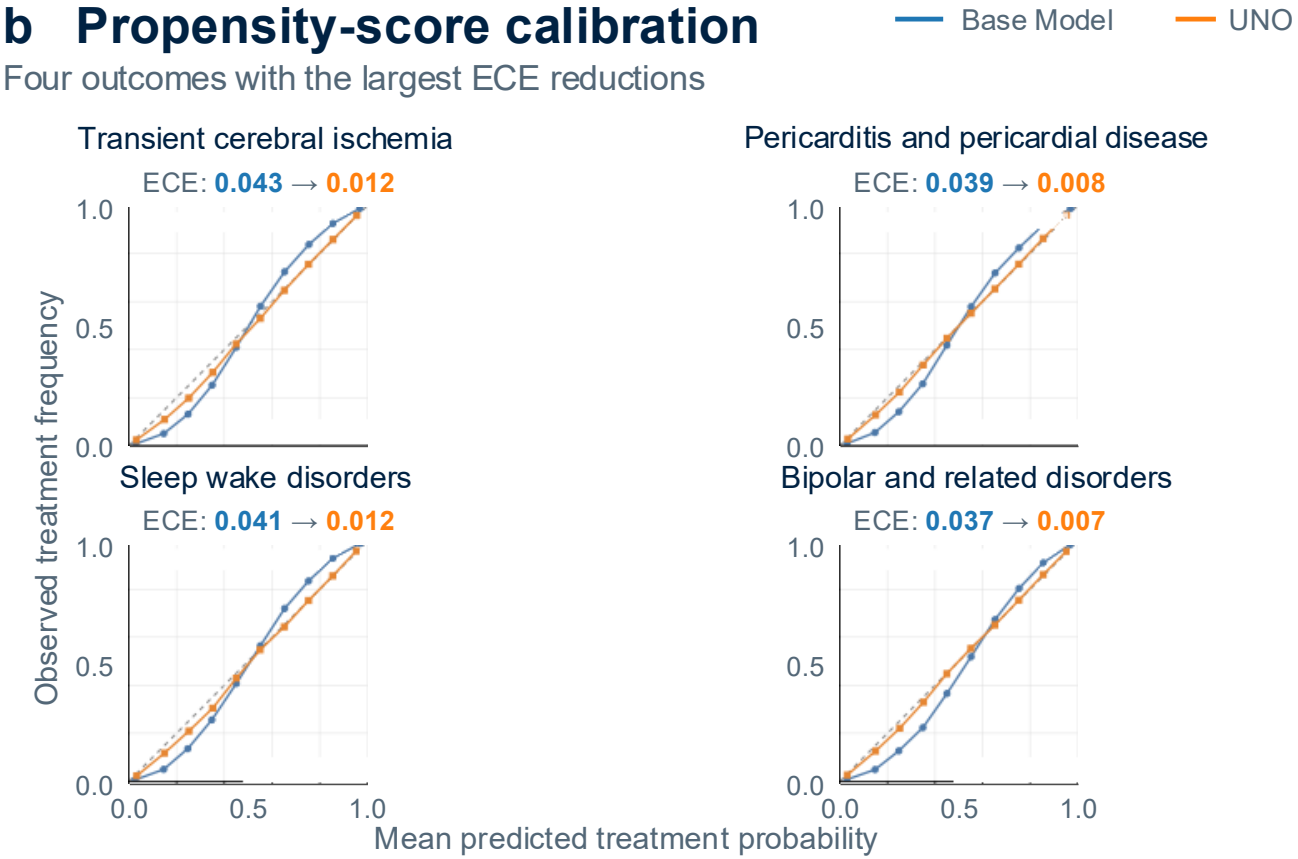
a Held-out EASE

50% of eligible NCOs reserved for evaluation



b Propensity-score calibration

Four outcomes with the largest ECE reductions



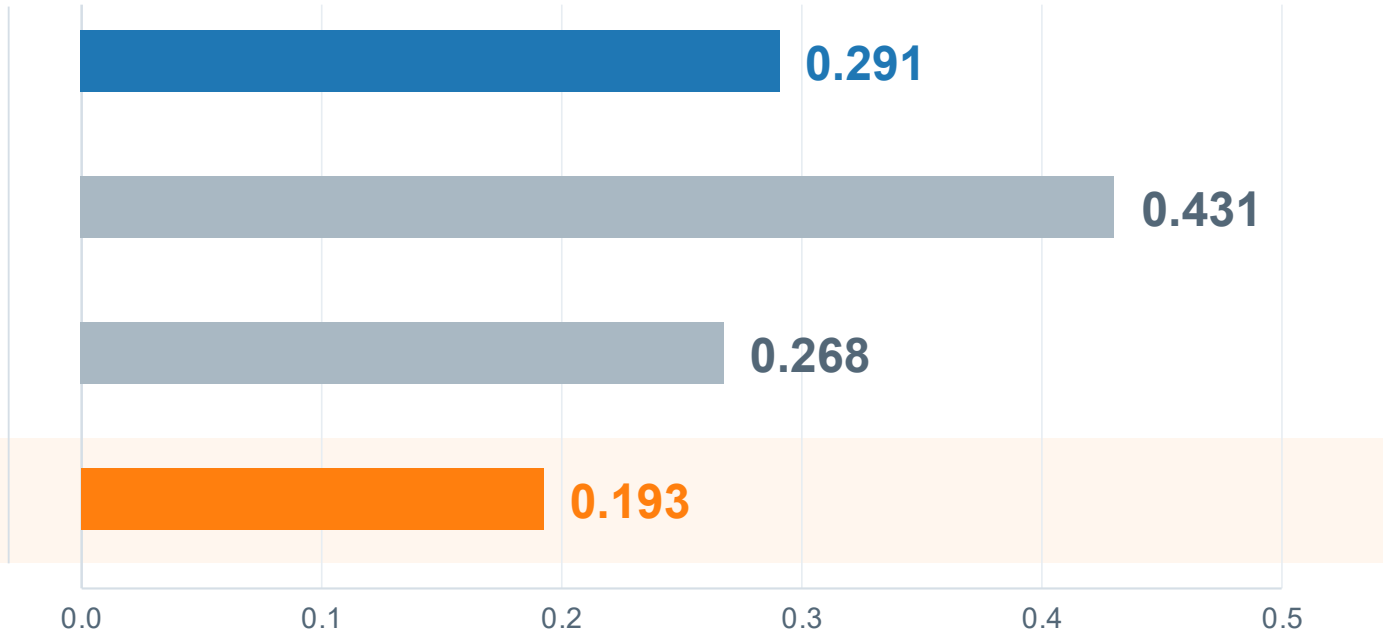
EASE: expected absolute systematic error. ECE: expected calibration error.

Penn Medicine

Ablation: Contributions of pruning and fine-tuning

GLP-1RAs vs DPP-4is | 10 outcome-specific analyses

Variant	Pruning	Fine-tuning
Base Model	—	—
Pruning only	✓	—
Fine-tuning only	—	✓
Full UNO	✓	✓



Pruning: UNO-selected mask. Fine-tuning: low-NCO-signal subset. Means are equally weighted across analyses.

Full UNO had lower mean EASE than either step alone in these analyses.

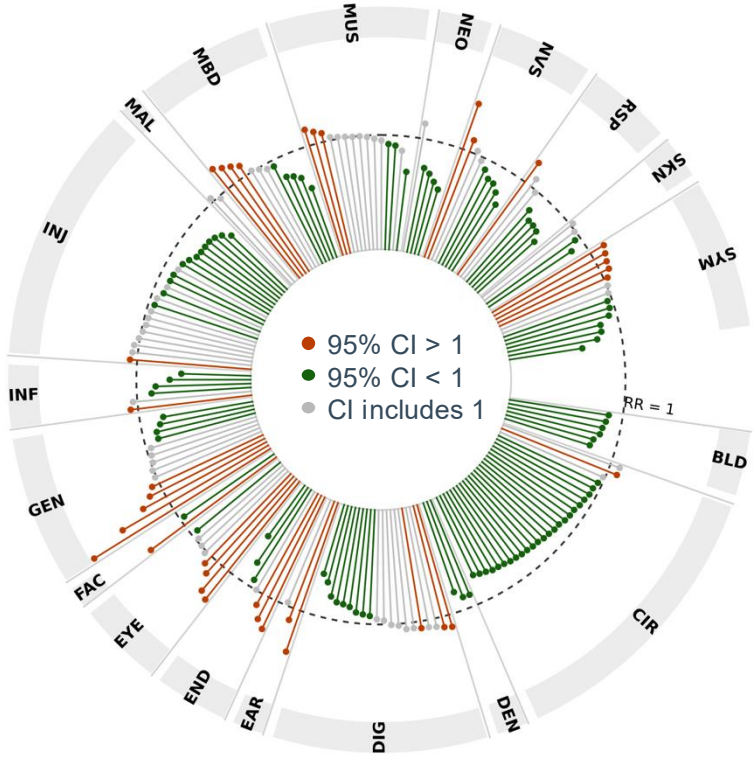
Same base-model checkpoint, evaluation cohort, and held-out NCO set within each analysis.

Penn Medicine

Result 1: Outcome-wide profiling results

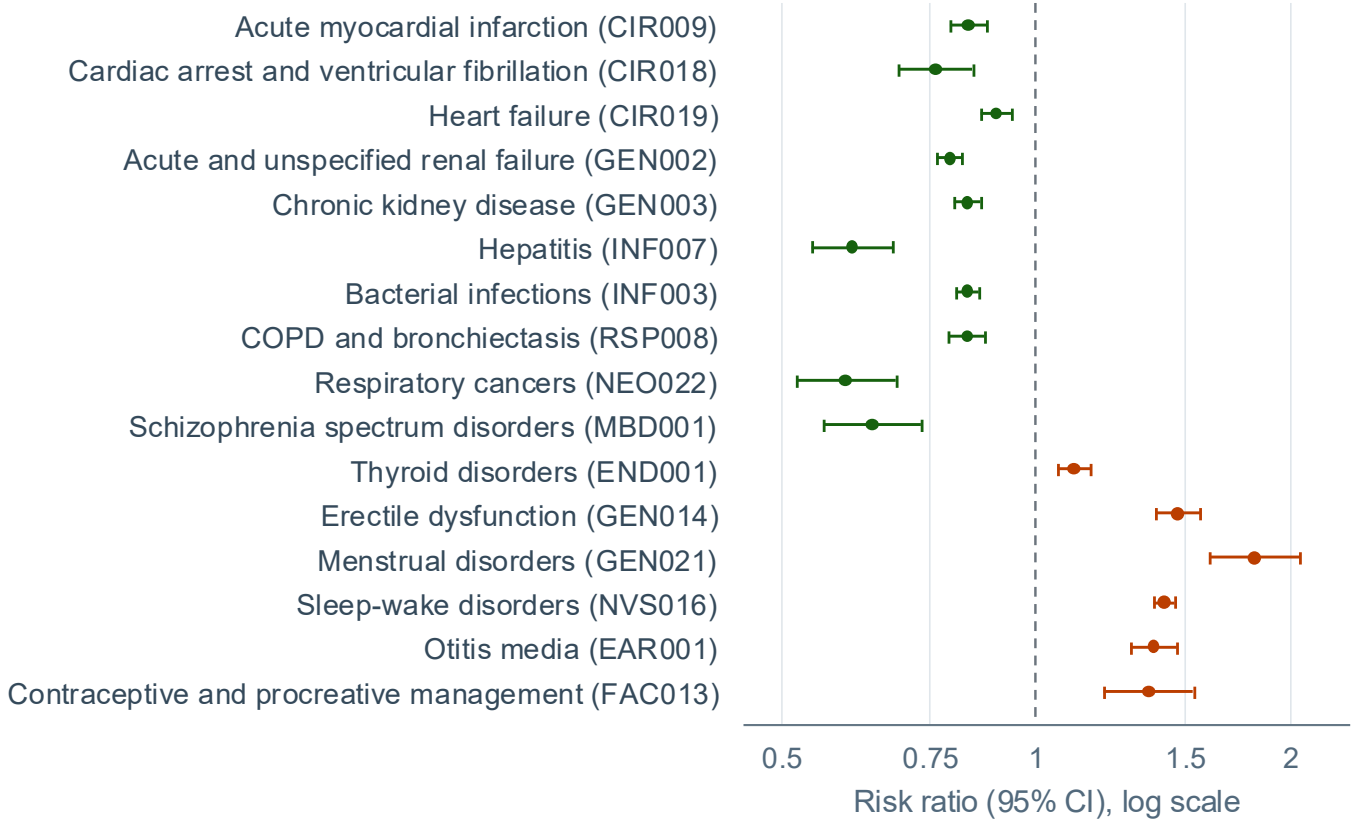
GLP-1RAs vs DPP-4is | 189 CCSR-defined clinical outcomes

a 189 outcomes by body system



Dashed circle: RR = 1

b 16 outcomes discussed



One treatment comparison → an outcome-wide association profile.

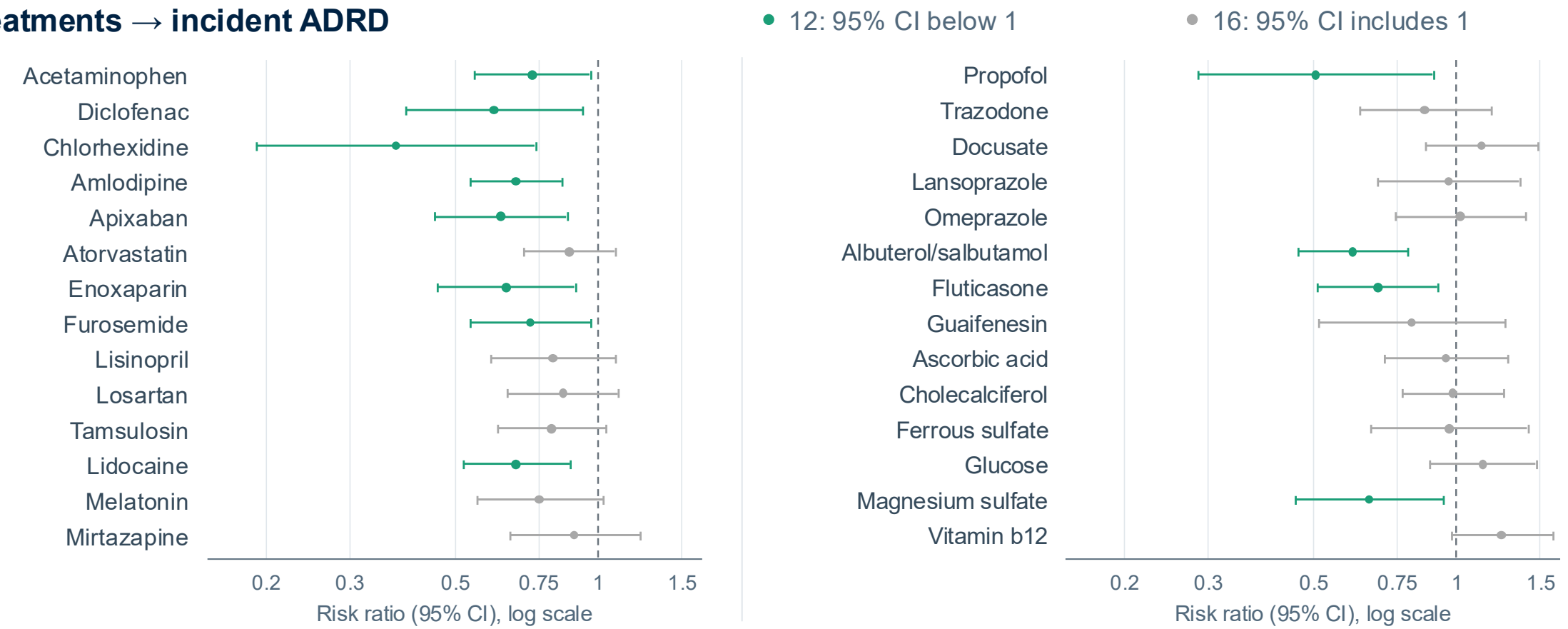
all 189 outcomes in panel a and all 16 main-text outcomes in panel b. CCSR: Clinical Classifications Software Refined.

Penn Medicine

Result 2: ADRD drug repurposing

Penn Medicine, 2012–2020 | 16,184 patients with MCI | 28 candidate medications

28 treatments → incident ADRD



Prioritize candidates for further study.

MCI: mild cognitive impairment. ADRD: Alzheimer's disease and related dementias.

Penn Medicine

Summary

1 NCO-guided model refinement

UNO uses negative control outcomes to guide neural pruning and fine-tuning.

2 Improved reliability

Lower held-out EASE and better propensity-score calibration than the Base Model.
Performance depends on relevant NCOs. Residual confounding may remain.

3 Compatible with existing analysis pipelines and large-scale studies

Applicable across tasks and neural network architectures.

Acknowledgments

*† Co-first authors * Senior authors*

- **Yiwen Lu†** University of Pennsylvania
- Huiyuan Wang University of Pennsylvania
- Dazheng Zhang University of Pennsylvania
- Huilin Tang University of Pennsylvania
- Bingyu Zhang University of Pennsylvania
- Lu Li University of Pennsylvania
- Wenjie Hu University of Pennsylvania
- Jiajie Chen University of Pennsylvania
- Tingyin Wang University of Pennsylvania
- Han Wu The Harker School
- Ting Zhou University of Pennsylvania
- Yuqing Lei University of Pennsylvania
- Xinyao Jian University of Pennsylvania
- Tsai Hor Chan University of Pennsylvania
- Siqi Chen University of Pennsylvania
- Yong Lin Eli Lilly and Company
- Demissie Alemayehu Pfizer Inc.
- **James Zou*** Stanford University
- **David Asch*** University of Pennsylvania
- **Yong Chen*** University of Pennsylvania

github.com/Penncil/UNO

Thank you!