

A Literature-driven Framework For Population-level Causal Modeling Of Alzheimer's Disease

Yijiang Pang¹, Ramya Keerthi Majji², Hao Dai², Yu Huang², Jiang Bian², Jiayu Zhou³

¹ Michigan State University ² Indiana University ³ University of Michigan



AAIC 2026 Poster #6364 | Public Health / Innovative methods in epidemiology

A curated, evidence-traceable pipeline converts ADRD literature into causal triples and a population-level DAG, with validation against human annotations.

MOTIVATION AND CONTRIBUTION

- Population-level ADRD causal modeling requires explicit assumptions about risk factors, biomarkers, interventions, and cognitive outcomes.
- High-quality causal evidence is scattered across biomedical abstracts and is difficult to reuse in structured causal analyses.
- This work provides an auditable literature-to-DAG framework: every extracted edge can be traced back to source evidence and metadata.

The contribution is a reproducible path from biomedical evidence to causal assumptions that can be reviewed before downstream analysis.

METHOD: EVIDENCE TO CAUSAL TRIPLES

- Filter abstracts using ADRD-relevant risk-factor vocabularies.
- Use PICO-grounded prompting to extract Exposure -> causal relation -> Outcome triples.
- Preserve direction, effect type, dosage, effect size, confidence interval, p-value, and sample size when available.
- Score extracted triples by semantic alignment, then assemble normalized entities and directed edges into a DAG.

HOW THE DAG CAN BE USED

- Audit causal assumptions against PubMed evidence.
- Prioritize relationships for observational analyses.
- Identify multi-step pathways connecting risk factors.
- Export graph entities for downstream workflows.

Ensures interpretability and evidence traceability.

EVIDENCE BASE AND ANNOTATION

- PubMed abstracts were restricted to evidence types such as: randomized controlled trials, systematic reviews, and meta-analyses.
- Risk-factor vocabularies combine EHR-derived and literature-derived sources to guide filtering and normalization.
- Human-labeled validation and test sets provide a reference standard for evaluating open-selection causal extraction.

309

EHR risk-factor

764

Lit risk-factor

100

Abstracts val+test

696

Triples ground truth

Interpretable extraction metrics for valid evaluation scope.

VALIDATION RESULTS AND INTERPRETATION

0.887

F1 Score

0.865

Precision

0.910

Recall

0.797

Agreement

- Framework improves over PICO baseline F1 (0.796); design affects extraction quality.
- High recall is key for DAGs to ensure all plausible pathways are captured.
- Precision is sufficient for review, but expert human oversight is still required.

LIMITATIONS AND ACKNOWLEDGEMENTS

- DAG summarizes literature but does not identify effects.
- Evidence quality depends on abstracts and expert review.

Acknowledgements

This material is based in part upon work supported by the National Science Foundation under Grant IIS-2212174, National Institute on Aging (NIA) R01AG072449 and National Institute of General Medical Sciences (NIGMS) 1R01GM145700.



Intelligent Data Analytics Lab
UNIVERSITY OF MICHIGAN