AIME 2026

Causal Models with Tiny Data: The Case of Rural People Living with Dementia

Ranveer Singh, Saurabh Mathur, Kavimayil P. Komarasamy, Ameet Soni,
Cliff Whetung, Wayne Warry, Kristen Jacklin, Melissa Blind, Sriraam Natarajan

UT Dallas · TU Darmstadt · Swarthmore College · University of Minnesota Medical School, Duluth Campus

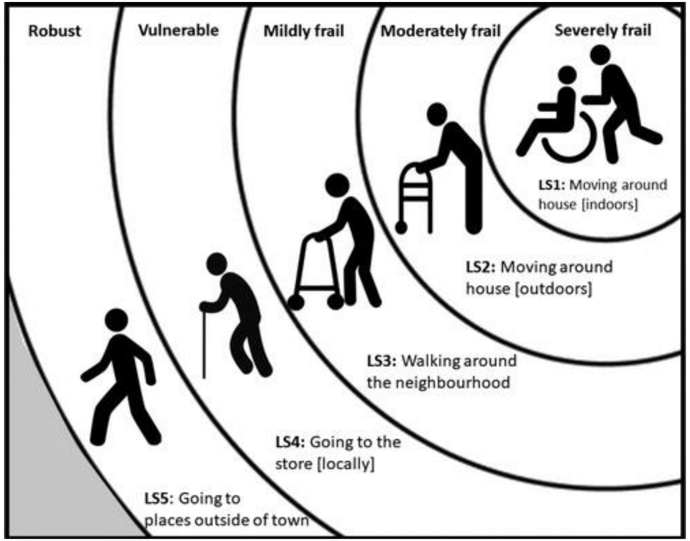
Supported by NIH awards R01NS133142 and 1R21AG072566

Speaker Declaration of Conflict of Interest

I do not have any relevant financial or professional affiliations with for-profit or not-for-profit organizations to disclose

Dementia: A Growing Health Crisis

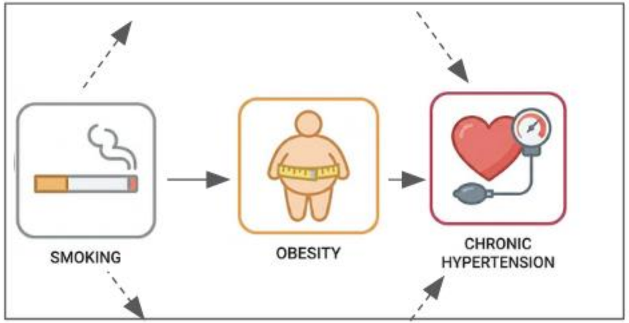
- Rising life expectancy and an aging population are driving up dementia risk
- Common assessment tool: Life-space assessment (LSA) survey to measure patient mobility
- Rural populations face higher risk of dementia than urban populations, yet they remain understudied
- Urgent need to identify interventions that improve outcomes in these populations



Weden et al (2018). Secular Trends in Dementia and Cognitive Impairment of U.S. Rural and Urban Older Adults. *Am. J. Prev. Med.*, *54*(2)
Hewston et al (2021). Frailty and Life-Space Mobility: Implications for Clinical Practice and Research. *Occupational therapy in health care*, *35*(1)

Identifying interventions by causal modeling

- Reasoning about interventions requires causal models.
- Problem: causal discovery requires large amount of data.
- What can tiny data reveal about causal relationships about the mobility of rural and Indigenous populations?



Causal Graph

Tiny Data: Nine boolean variables, n = 20

Source	Code	Variable	%
Life-Space Mobility	LS	Life-space score* (high)	45%
	TB	Total burden on caregiver (high)	40%
Caregiver diary	ND	Non-routine days (above median)	60%
	CD	Challenging days (above median)	45%
	CT	Community type (rural)	55%
	PS	Patient sex (female)	35%
Demographic	CS	Caregiver sex (female)	65%
	PE	Patient education (above threshold)	35%
	CE	Caregiver education (above threshold)	50%

*Adapted for rural populations

Four methods for causal graph construction

KNOWLEDGE SOURCE	METHOD	OUTPUT
Tiny dataset only 20 patients	FCI algorithm data-driven baseline	Data-driven graph (undirected assoc.)
Domain experts	Expert elicitation	Expert graph
Claude · Gemini · GPT	5 prompts each, pooled by weight	LLM consensus graph
Expert / LLM graph + tiny dataset	Subtractive refinement Min. Description Length	Refined graph (pruned)

Expert and LLM graphs additionally feed the Hybrid refinement step.

Data-driven & hybrid discovery methods fail

Knowledge source	Method	Output
KNOWLEDGE SOURCE Tiny dataset only 20 patients	METHOD FCI algorithm data-driven baseline	OUTPUT Data-driven graph (undirected assoc.)
KNOWLEDGE SOURCE Domain experts	METHOD Expert elicitation	OUTPUT Expert graph
KNOWLEDGE SOURCE Claude · Gemini · GPT	METHOD 5 prompts each, pooled by weight	OUTPUT LLM consensus graph
KNOWLEDGE SOURCE Expert / LLM graph + tiny dataset	METHOD Subtractive refinement Min. Description Length	OUTPUT Refined graph (pruned)

Alternative: Is tiny data sufficient for evaluating causal hypotheses?

Research Questions

- Are the graphs obtained from experts and LLMs compatible with the statistical patterns present in the tiny dataset?
- Is there any consensus in causal relations identified by experts and LLMs?

Tiny Data for Conditional Independence Testing

Method:

- Full causal discovery isn't feasible with this little data; instead, test pairwise Conditional Independence (CI)
- G-test, conditioning set size 1, significance threshold 0.05
- Treat data-implied independencies as ground truth; compare against each graph's implied independencies
- Score each graph on precision and false positive rate (FPR)

Model	Total	Precision	FPR
Expert	113	0.73	0.53
LLM Consensus	31	0.68	0.48

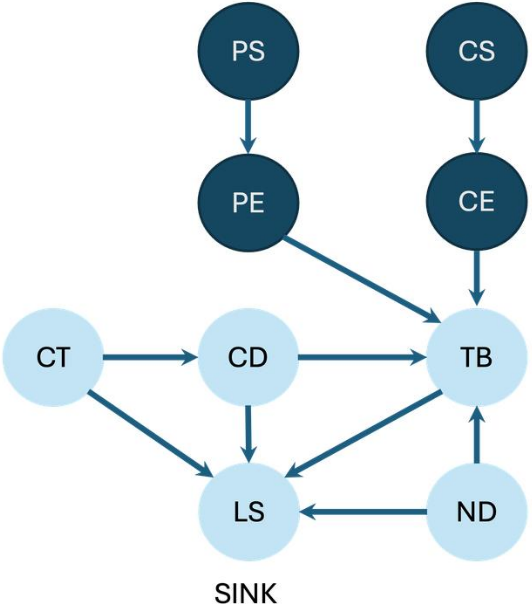
Finding:

- Expert and LLM graphs fail the CI test in different ways
- Experts have much higher number of independencies, a higher precision and FPR than LLMs

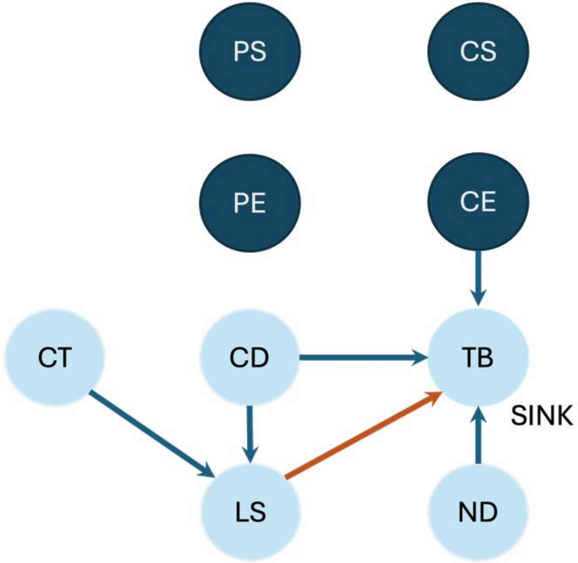
Expert vs. LLM causal graphs

Dark blue nodes = socioeconomic factors

Orange arrow = LLM-only



Expert Model



LLM Consensus

LLMs & Experts: Consensus and disagreement

Agreement across LLMs & experts

Non-routine days, community-type, and caregiver education drive total caregiver burden

ND → TB | CT → TB | CE → TB

Challenging days and community-type drive lifespace mobility

CD → LS | CT → LS

The disagreement

Experts

TB → LS

Life-space mobility (LS) is the final sink.
Experts include sex and education as causal drivers.

VS

LLMs

LS → TB

Caregiver burden (TB) is the ultimate sink.
LLMs exclude socioeconomic variables entirely.

Suggests a fundamental difference in modeling perspectives. May reflect a bidirectional relationship.

Conclusion and future work

CONCLUSION

Foundation for hybrid models

Combining expert and LLM causal knowledge supports more robust models.

Experts and LLMs diverge

They encode different assumptions about which variables matter most.

Implications for AI in Medicine

Understanding this gap is key for decision support in data-scarce settings.

FUTURE WORK

Is the TB-LS edge bidirectional?

May reflect a feedback loop, not a single direction.

Use disagreement to guide elicitation

Target expert feedback toward high-disagreement regions.

Rural vs. urban causal models

Implications for intervention design in underserved communities.