Dynamic causal discovery in Alzheimer's disease through latent pseudotime modelling

Natalia Glazman
King's College London
natalia.glazman@kcl.ac.uk

Jyoti Mangal
King's College London
jyoti.mangal@kcl.ac.uk

Pedro Borges
Hologen & King's College London
pedro.borges@hologen.ai

Sebastien Ourselin
King's College London
sebastien.ourselin@kcl.ac.uk

M. Jorge Cardoso
King's College London
m.jorge.cardoso@kcl.ac.uk

Abstract

The application of causal discovery to diseases like Alzheimer's (AD) is limited by the static graph assumptions of most methods; such models cannot account for an evolving pathophysiology, modulated by a latent disease pseudotime. We propose to apply an existing latent variable model to real-world AD data, inferring a pseudotime that orders patients along a data-driven disease trajectory independent of chronological age, then learning how causal relationships evolve. Pseudotime outperformed age in predicting diagnosis (AUC 0.82 vs 0.59). Incorporating minimal, disease-agnostic background knowledge substantially improved graph accuracy and orientation. Our framework reveals dynamic interactions between novel (NfL, GFAP) and established AD markers, enabling practical causal discovery despite violated assumptions.

1 Introduction

With nearly $380 billion allocated annually to Alzheimer's disease (AD) research [1], clinical trials continue failing to halt disease progression [29]. This ongoing challenge arises from the complexity of the disease, involving thousands of pathways whose interactions remain poorly understood [3]. Causal inference offers a powerful framework for modelling these relationships. For example, causal graph discovery methods can recover causal graphs from observational data, enabling identification and estimation of causal effects. Mapping AD's underlying causal relationships could yield more effective treatments, facilitate biomarker discovery, and enable personalised treatment plans.

Applying causal discovery to healthcare data presents distinct challenges. First, the absence of empirical ground truth hinders development and validation of domain-specific tools. Second, fundamental assumptions of causal discovery, such as absence of cyclical graphs and unobserved confounders, are frequently violated or untestable. While recent methodological advances allow analysis of more realistic data by relaxing these assumptions [15; 38; 24], practical applications remain limited. Current AD studies predominantly focus on static consensus graphs, neglecting dynamic relationships and latent variables that characterise the disease [41]. Developing robust models of causal relationships in AD will advance both disease understanding and clinical application.

Accepted to the NeurIPS 2025 Workshop on CauScien: Uncovering Causality in Science.

Disease progression rates vary across AD patients [44], partly reflecting "cognitive reserve" - protective brain mechanisms arising from poorly characterised biological, lifestyle, and genetic factors [43; 34]. While approximating cognitive reserve through education and socioeconomic status partially explains individual variation, explicitly modelling latent disease pathway modulators remains crucial for comprehensive understanding of AD. The abstraction of pseudotime, which is a latent (unobserved) dimension measuring the progress of cell states through a transition is frequently applied to molecular-level disease modelling, and recent research has adopted this abstraction (both categorical and continuous) for modelling disease progression using electronic health records [22; 53].

Here, we apply causal discovery to real-world AD data through a latent pseudotime model introduced by Zhou et al. [58], investigating how causal interactions evolve along disease progression. We leverage known causal relationships from literature and validate our model through qualitative and quantitative evaluation against a consensus graph. Our analysis incorporates key AD biomarkers to characterise their interactions, including plasma NfL and GFAP, recently emerged as novel biomarkers for AD detection [51; 25; 7].

2 Methods

Data We analysed data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset [36], selecting 16 variables with established or putative causal roles in AD aetiology. These comprised demographic variables (years of education, sex, age, and APOE genotype), segmented brain region volumetric measurements (total intracranial volume (ICV), hippocampus, amygdala, temporal lobe), plasma biomarkers (Aβ40, Aβ42, pTau217, NfL, GFAP), and cognitive scores (Trail Making Test Part B (TRABSCORE), ADAS-Cog-13 (ADAS13), Rey Auditory Verbal Learning Test immediate (RAVLT)). The sample consisted of the following participant numbers, broken down by diagnosis: 48 AD, 117 MCI (mild cognitive impairment), and 215 CN (cognitively normal).

Pseudotime model of Alzheimer's disease We modelled AD progression using cross-sectional data from the ADNI dataset with Bayesian Networks with Latent Time Embedding (BN-LTE), a model proposed by Zhou et al. [58] to estimate causal graphs as a function of pseudotime. In disease modelling, pseudotime represents a latent variable that orders samples along a continuous trajectory of disease progression, capturing relative disease states rather than chronological time [21; 23]. This enables modelling of dynamic changes in biomarkers or molecular states when using irregularly sampled or cross-sectional data.

Theoretical identifiability of the pseudotime variable up to monotonic transformations is shown by Zhou et al. [58] under the condition that causal relationships change along the pseudotime axis.

We implemented the framework of Zhou et al. [58], described as follows. We assume faithfulness and causal sufficiency when incorporating the latent pseudotime variable. Let $\mathbf{X} = (X_1, ..., X_p) \in \mathbb{R}^p$ be a p-dimensional random variable vector, and let $G = (V, E)$ denote a directed acyclic graph (DAG) consisting of a node set $V = \{1, ..., p\}$ corresponding to elements of $\mathbf{X}$, and edge set $E \subset V \times V$ representing causal relationships between them. The distribution of $\mathbf{X}$ factorizes as

$$F(\mathbf{X}|Z, \Theta) = \prod_{j=1}^{p} F_j(X_j | X_{l \in pa_{G(Z)}}(j), Z, \Theta), \tag{1}$$

where $Z \in \mathbb{R}^n$ denotes pseudotime, $pa_{G(Z)}(j)$ denotes the parent nodes of j, and Θ represents model parameters. The conditional probability distribution of a single biomarker, F_j, is modelled as

$$X_j = a_j(Z) + \sum_{l=1}^{p} b_{jl}(Z) X_l + \epsilon_j, \epsilon_j \sim N(0, \sigma_j^2), \tag{2}$$

where $a_j(Z)$ represents the baseline function governing biomarker progression across pseudotime, $b_{jl}(Z)$ captures pseudotime-dependent causal effects, and ϵ_j denotes per-variable noise. Both functions are parametrised as cubic b-splines, following Zhou et al. [58]. Parameter posteriors were obtained using Markov chain Monte Carlo (MCMC). Details of the priors, sampling, posterior estimation, and convergence statistics are provided in the Supplementary Material.

Background knowledge Background or expert knowledge is commonly used to improve causal graph discovery performance when model assumptions are violated [42; 8; 11]. We introduced disease-

agnostic background knowledge to enhance model performance. To minimize bias, our background knowledge exclusively constrained causal and pseudotime relationships directed towards immutable variables (e.g., sex, genotype – considered root nodes) and prohibited cognitive score variables from forming outgoing causal effects (sink nodes). Although cognitive variables may influence lifestyle choices, such effects are unlikely to significantly impact other variables in the elderly ADNI population. Additional implementation details are found in the Supplementary Material.

3 Results

The patient distribution along the pseudotime axis corresponded to disease severity (Fig. 1), with cognitively normal participants clustering at early pseudotime, MCI patients concentrated in the intermediate region, and AD patients positioned towards late pseudotime. This ordering aligned with expected AD-associated biomarker trajectories, including decreased hippocampal volume, deteriorating cognitive scores, and elevated NfL and GFAP levels. To verify that pseudotime captured disease progression rather than chronological ageing, we compared predictive performance using a logistic generalized additive model, which captures complex, non-linear relationships. Pseudotime demonstrated superior predictive performance for diagnosis, achieving a mean AUC of 0.82 (95% CI: 0.81, 0.82) ($p < 0.001$) compared to 0.59 ($p < 0.01$) for age alone. We identified edges with the highest posterior inclusion probability throughout pseudotime for edges between variables with no background knowledge across different settings and validated them against literature evidence (Table 1). The model correctly identified several known disease-related pathways, including the causal effect of plasma NfL on hippocampal volume reduction and pTau-induced NfL elevation [16; 13].

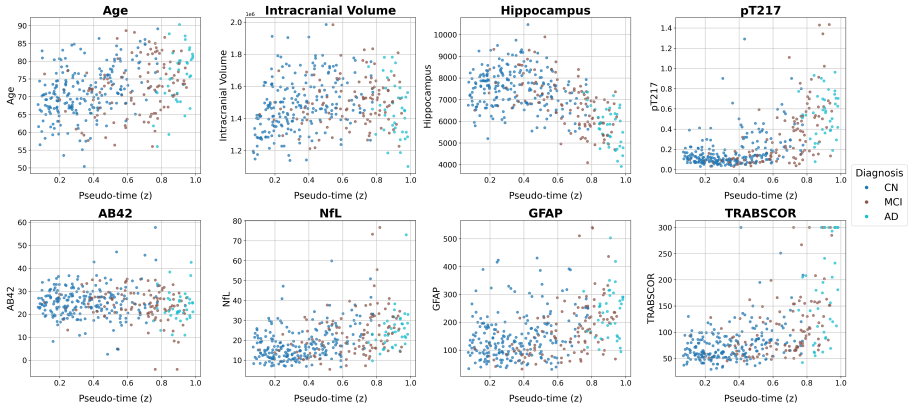

Figure 1: Demographic variables and biomarkers plotted as a function of the posterior mean of the inferred pseudotime values (z) of a single chain inferred by the BN-LTE model, with patients colour-coded according to their diagnosis.

Edge	PIP (No mod.)	PIP (RN)	PIP (RN + SN)	Lit. consensus
pTau217 → GFAP	0.64 (0.38)	0.93 (0.13)	0.80 (0.16)	Possible/unknown [13]
Aβ42 → Aβ40	0.75 (0.04)	0.75 (0.43)	0.75 (0.43)	Present [49]
pTau217 → NfL	0.35 (0.27)	0.61 (0.17)	0.57 (0.15)	Possible [13]
NfL → Hippocampus	0.25 (0.18)	0.63 (0.01)	0.53 (0.18)	Possible [27]
Aβ42 → NfL	0.60 (0.07)	0.49 (0.12)	0.46 (0.07)	Possible [56]

Table 1: Causal edges identified using BN-LTE with background knowledge, with the posterior probability of inclusion (PIP) junder different background knowledge settings. Values shown are chain-level mean and standard deviation for modifications of root nodes (RN) and sink nodes (SN).

To comprehensively evaluate model performance, we compared our estimated graphs against a consensus graph constructed from Alzheimer's literature (Fig. 2; further details listed in the Supplementary Material). The baseline model without modifications performed poorly across both edge detection and directionality metrics. Performance improved substantially with the incorporation of background knowledge and static features (Table 2). The improvement was particularly pronounced in orientation metrics and was reflected in the reconstruction accuracy.

After incorporating background knowledge for root and sink nodes, we evaluated performance improvements for variables without background knowledge constraints - specifically, biomarkers and brain volumetric measures. Given the induced orientation bias, we assessed this graph subset based only on edge presence rather than directionality. The model with background knowledge outperformed the unmodified model across various metrics (Table 2).

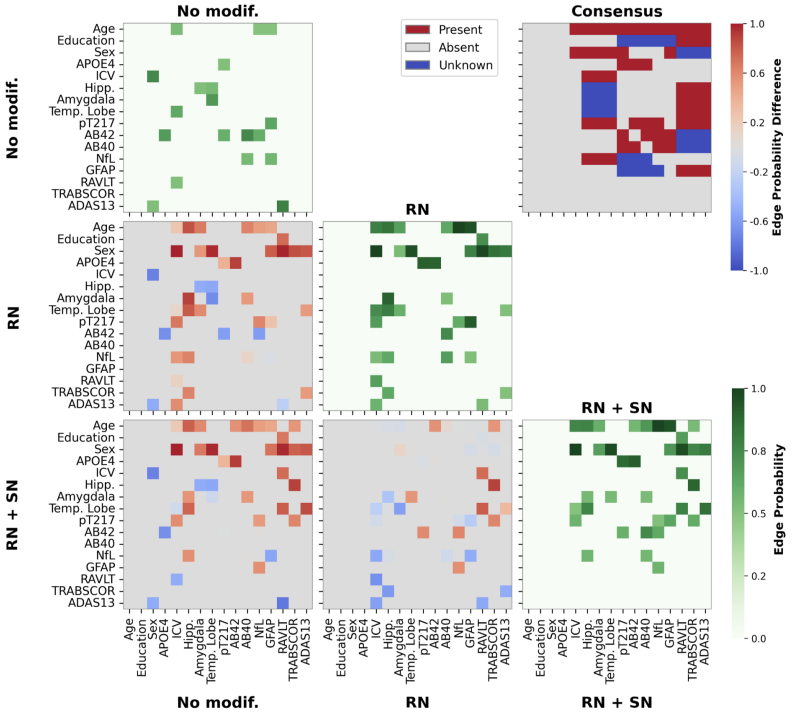

Figure 2: Established consensus graph compared to estimated causal graphs presented as adjacency matrices with edge indices averaged across pseudotime and over the four chains, resulting in an edge inclusion probability (only PIP > 0.5 shown). The matrices on the diagonal reflect the resulting matrices from different settings, while the strictly lower triangular matrices reflect differences between each pair of matrices.

Modif.	Presence		Orientation		Subset (Presence)		MSE	SHD	Subset SHD
	P	R	P	R	P	R			
None	0.80 [0.80,82]	0.16 [0.16,16]	0.62 [0.62,0.63]	0.50 [0.49,0.50]	0.67 [0.65,0.71]	0.10 [0.10,0.10]	0.98 (0.045)	67 [67,67]	21 [21,21]
RN	0.72 [0.70,0.74]	0.35 [0.33,0.36]	0.89 [0.89,0.90]	0.84 [0.84,0.84]	0.74 [0.72,0.74]	0.29 [0.29,0.30]	0.95 (0.016)	53 [55,53]	23 [24,23]
RN + SN	0.88 [0.88,0.89]	0.45 [0.42,0.45]	0.96 [0.96,0.96]	0.88 [0.88,0.88]	0.79 [0.79,0.79]	0.39 [0.38,0.39]	0.93 (0.025)	41 [43,41]	21 [21,21]

Table 2: Statistics of causal graphs estimated with BN-LTE. Presence precision (P) and recall (R) measure edge presence using symmetric adjacency matrices. Orientation metrics evaluate directionality for edges present in both consensus and estimated graphs. Subset metrics exclude variables with background knowledge. Values presented as median, [95% CI] apart from MSE (SD).

The causal relationships between variables exhibited dynamic changes across pseudotime; for example, age showed a constant effect on GFAP, while pTau's influence on NfL manifested early in pseudotime progression (Fig 3)[18]. This pattern is consistent with AD pathophysiology, where the effect of pTau is known to plateau, while the effect of age on neuroinflammation most likely remains constant over the course of the disease [4].

We also identified edges that contradict established literature; for example, the proposed causal effect of pTau $\rightarrow$ GFAP and NfL $\rightarrow$ Aβ40 (Fig. 2) contradict studies that show that amyloid dysfunction precedes neurodegeneration, and GFAP elevation precedes changes in pTau [5].

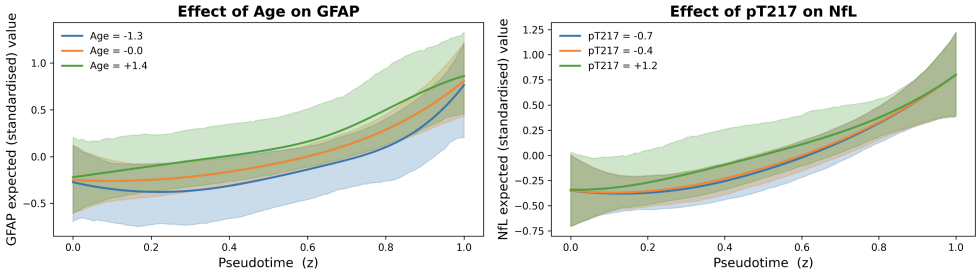

Figure 3: Biomarkers plotted against pseudotime, with solid lines representing the mean baseline trajectory of biomarkers and the corresponding effect of changing a specific parent variable. Parent variables are changed to the 5th, 50th, and 95th percentiles of the standardized values. The solid lines correspond to the mean of posteriors with shaded areas being 95% CI.

4 Discussion

Our findings demonstrate significant progress toward clinically applicable causal discovery in AD. Using BN-LTE, we estimated causal graphs as a function of pseudotime. Importantly, incorporating background knowledge proved essential for generating reliable causal graphs when analysing real-world data where model assumptions are violated. Notably, this knowledge remained disease-agnostic, constraining only demographic variables and cognitive outcomes, which significantly improved the recovered graphs. We identified both known and novel causal dynamics along the latent pseudotime through incorporation of emerging biomarkers (NfL and GFAP).

This pseudotime framework offers immediate translational potential. It can revolutionize clinical trial design by precisely stratifying patients for targeted interventions at optimal disease stages, potentially explaining heterogeneous treatment responses and identifying therapeutic windows. Furthermore, dynamic causal relationships across pseudotime suggest that combination therapies may require tailored sequencing based on a patient's disease stage.

AD progression arises from a complex interplay of unobserved factors such as cognitive reserve and lifestyle-driven neuroprotection, alongside other modulators of its pathological pathways [55]. Under the framework of [24], our approach treats pseudotime as a surrogate for latent modulating factors that drive non-stationary causal mechanisms throughout disease progression, ensuring identifiability of the causal graph. The inferred pseudotime separates patients by disease stage, predicts disease status better than age, and captures causal dynamics consistent with known AD pathology, demonstrating that it serves as a robust and clinically meaningful proxy for true disease progression. While this work demonstrates the value of a one-dimensional abstraction, future work could model these drivers as a multivariate construct to achieve a more granular understanding of AD's heterogeneity.

Several methodological advances would strengthen clinical applicability. First, model assumptions must be relaxed to accommodate biomedical systems with unobserved confounders. Second, our analysis was limited by dataset size and variable selection, preventing the inclusion of factors like ethnicity due to insufficient cohort heterogeneity. Expanding variables and sample size should enhance causal graph accuracy. Recent multi-dataset causal discovery advances [33] could enable more robust causal identification through cross-validation across independent cohorts. In addition, an evaluation of this model using longitudinal data could both validate the model and provide insights on AD progression through predictions grounded in causal reasoning.

In summary, our work establishes a foundation for integrating causal discovery into AD research and clinical practice. Combining pseudo-time modelling with disease-agnostic background knowledge provides a practical framework for balancing theoretical rigour with clinical applicability. As we refine these methods and expand to larger, diverse datasets, causal discovery can transform our understanding of AD heterogeneity and guide precision medicine development.

Acknowledgments Natalia Glazman is supported by the King's College London Digital Twins for Healthcare CDT and cofunded by Siemens Healthineers. Jyoti Mangal is supported by NHS England.

References

[1] RePORT: Funding for Various Research, Condition, and Disease Categories (RCDC), June 2025. URL https://report.nih.gov/funding/categorical-spending#/.

[2] F. L. Andronie-Cioara, A. I. Ardelean, C. D. Nistor-Cseppento, A. Jurcau, M. C. Jurcau, N. Pascalau, and F. Marcu. Molecular Mechanisms of Neuroinflammation in Aging and Alzheimer's Disease Progression. *International Journal of Molecular Sciences*, 24(3):1869, Jan. 2023. ISSN 1422-0067. doi: 10.3390/ijms24031869. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9915182/.

[3] S. E. Arnold, B. T. Hyman, R. A. Betensky, and H. H. Dodge. Pathways to personalized medicine—Embracing heterogeneity for progress in clinical therapeutics research in Alzheimer's disease. *Alzheimer's & Dementia*, 20(10):7384–7394, 2024. ISSN 1552-5279. doi: 10.1002/alz.14063. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/alz.14063. _eprint: https://alz-journals.onlinelibrary.wiley.com/doi/pdf/10.1002/alz.14063.

[4] M. Bertsch, B. Franchi, M. C. Tesi, and V. Tora. The role of A$\beta $ and Tau proteins in Alzheimer's disease: a mathematical model on graphs. *Journal of Mathematical Biology*, 87(3):49, 2023. ISSN 0303-6812. doi: 10.1007/s00285-023-01985-7. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10468937/.

[5] M. Bilgel, Y. An, K. A. Walker, A. R. Moghekar, N. J. Ashton, P. R. Kac, T. K. Karikari, K. Blennow, H. Zetterberg, B. M. Jedynak, M. Thambisetty, L. Ferrucci, and S. M. Resnick. Longitudinal changes in Alzheimer's-related plasma biomarkers and brain amyloid. *Alzheimer's & dementia : the journal of the Alzheimer's Association*, 19(10):4335–4345, Oct. 2023. ISSN 1552-5260. doi: 10.1002/alz.13157. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10592628/.

[6] G. S. Bloom. Amyloid- and Tau: The Trigger and Bullet in Alzheimer Disease Pathogenesis. *JAMA Neurology*, 71(4):505–508, Apr. 2014. ISSN 2168-6149. doi: 10.1001/jamaneurol.2013.5847. URL https://doi.org/10.1001/jamaneurol.2013.5847.

[7] K. Bolsewig, A. A. van Unnik, E. R. Blujdea, M. C. Gonzalez, N. J. Ashton, D. Aarsland, H. Zetterberg, A. Padovani, L. Bonanni, B. Mollenhauer, S. Schade, R. Vandenberghe, K. Poesen, M. G. Kramberger, C. Paquet, O. Bousiges, B. Cretin, E. A. Willemse, C. E. Teunissen, A. W. Lemstra, and for the European-Dementia With Lewy Bodies (E-DLB) Consortium. Association of Plasma Amyloid, P-Tau, GFAP, and NfL With CSF, Clinical, and Cognitive Features in Patients With Dementia With Lewy Bodies. *Neurology*, 102(12):e209418, June 2024. doi: 10.1212/WNL.0000000000209418. URL https://www.neurology.org/doi/10.1212/WNL.0000000000209418. Publisher: Wolters Kluwer.

[8] G. Borboudakis and I. Tsamardinos. Incorporating Causal Prior Knowledge as Path-Constraints in Bayesian Networks and Maximal Ancestral Graphs, June 2012. URL http://arxiv.org/abs/1206.6390. arXiv:1206.6390 [cs].

[9] M. Botdorf, K. L. Canada, and T. Riggins. A meta-analysis of the relation between hippocampal volume and memory ability in typically developing children and adolescents. *Hippocampus*, 32(5):386–400, May 2022. ISSN 1050-9631. doi: 10.1002/hipo.23414. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9313816/.

[10] Y. Caspi, R. M. Brouwer, H. G. Schnack, M. E. van de Nieuwenhuijzen, W. Cahn, R. S. Kahn, W. J. Niessen, A. van der Lugt, and H. H. Pol. Changes in the intracranial volume from early adulthood to the sixth decade of life: A longitudinal study. *NeuroImage*, 220:116842, Oct. 2020. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2020.116842. URL https://www.sciencedirect.com/science/article/pii/S1053811920303293.

[11] R. Castelo and A. Siebes. Priors on network structures. Biasing the search for Bayesian networks. *International Journal of Approximate Reasoning*, 24(1):39–57, Apr. 2000. ISSN 0888-613X. doi: 10.1016/S0888-613X(99)00041-9. URL https://www.sciencedirect.com/science/article/pii/S0888613X99000419.

[12] S. Chatterjee, M. Sealey, E. Ruiz, C. M. Pegasiou, K. Brookes, S. Green, A. Crisford, M. Duque-Vasquez, E. Luckett, R. Robertson, P. Richardson, G. Vajramani, P. Grundy, D. Bulters, C. Proud, M. Vargas-Caballero, and A. Mudher. Age-related changes in tau and autophagy in human brain in the absence of neurodegeneration. *PLOS ONE*, 18(1):e0262792, Jan. 2023. ISSN 1932-6203. doi: 10.1371/journal.pone.0262792. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9879510/.

[13] A. Didonna. Tau at the interface between neurodegeneration and neuroinflammation. *Genes & Immunity*, 21(5):288–300, Nov. 2020. ISSN 1476-5470. doi: 10.1038/s41435-020-00113-5. URL https://www.nature.com/articles/s41435-020-00113-5. Publisher: Nature Publishing Group.

[14] J. P. Ferrari-Souza, F. Z. Lussier, D. T. Leffa, J. Therriault, C. Tissot, B. Bellaver, P. C. L. Ferreira, M. Malpetti, Y.-T. Wang, G. Povala, A. L. Benedet, N. J. Ashton, M. Chamoun, S. Servaes, G. Bezgin, M. S. Kang, J. Stevenson, N. Rahmouni, V. Pallen, N. M. Poltronetti, J. T. O'Brien, J. B. Rowe, A. D. Cohen, O. L. Lopez, D. L. Tudorascu, T. K. Karikari, W. E. Klunk, V. L. Villemagne, J.-P. Soucy, S. Gauthier, D. O. Souza, H. Zetterberg, K. Blennow, E. R. Zimmer, P. Rosa-Neto, and T. A. Pascoal. APOE4 associates with microglial activation independently of A plaques and tau tangles. *Science Advances*, 9(14):eade1474, Apr. 2023. ISSN 2375-2548. doi: 10.1126/sciadv.ade1474.

[15] P. Forré and J. M. Mooij. Constraint-based Causal Discovery for Non-Linear Structural Causal Models with Cycles and Latent Confounders. July 2018. URL `https://www.auai.org/uai2018/proceedings/papers/117.pdf`.

[16] L. Gaetani, K. Blennow, P. Calabresi, M. D. Filippo, L. Parnetti, and H. Zetterberg. Neurofilament light chain as a biomarker in neurological disorders. *Journal of Neurology, Neurosurgery & Psychiatry*, 90(8): 870–881, Aug. 2019. ISSN 0022-3050, 1468-330X. doi: 10.1136/jnnp-2018-320106. URL `https://jnnp.bmj.com/content/90/8/870`. Publisher: BMJ Publishing Group Ltd Section: Neurodegeneration.

[17] N. A. Gillespie, M. C. Neale, M. S. Panizzon, R. E. McKenzie, X. M. Tu, H. Xian, C. A. Reynolds, M. J. Lyons, R. A. Rissman, J. A. Elman, C. Franz, and W. S. Kremen. Testing the causal impact of plasma amyloid on total Tau using a genetically informative sample of adult male twins. *Aging Brain*, 7:100139, Jan. 2025. ISSN 2589-9589. doi: 10.1016/j.nbas.2025.100139. URL `https://www.sciencedirect.com/science/article/pii/S2589958925000052`.

[18] H. Hampel, J. Cummings, K. Blennow, P. Gao, C. R. Jack, and A. Vergallo. Developing the ATX(N) classification for use across the Alzheimer disease continuum. *Nature Reviews Neurology*, 17(9):580–589, Sept. 2021. ISSN 1759-4766. doi: 10.1038/s41582-021-00520-w. URL `https://www.nature.com/articles/s41582-021-00520-w`. Publisher: Nature Publishing Group.

[19] H. Hampel, J. Hardy, K. Blennow, C. Chen, G. Perry, S. H. Kim, V. L. Villemagne, P. Aisen, M. Vendruscolo, T. Iwatsubo, C. L. Masters, M. Cho, L. Lannfelt, J. L. Cummings, and A. Vergallo. The Amyloid- Pathway in Alzheimer's Disease. *Molecular Psychiatry*, 26(10):5481–5503, Oct. 2021. ISSN 1476-5578. doi: 10.1038/s41380-021-01249-0. URL `https://www.nature.com/articles/s41380-021-01249-0`. Publisher: Nature Publishing Group.

[20] K. Hasegawa, I. Yamaguchi, S. Omata, F. Gejyo, and H. Naiki. Interaction between A(142) and A(140) in Alzheimer's -Amyloid Fibril Formation in Vitro. *Biochemistry*, 38(47):15514–15521, Nov. 1999. ISSN 0006-2960. doi: 10.1021/bi991161m. URL `https://doi.org/10.1021/bi991161m`. Publisher: American Chemical Society.

[21] B. He, S. Zhang, S. L. Risacher, A. J. Saykin, and J. Yan. Multi-modal Imaging-based Pseudotime Analysis of Alzheimer progression. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 30: 664–674, 2025. ISSN 2335-6928. doi: 10.1142/9789819807024_0047. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC12044618/`.

[22] L. Heumos, P. Ehmele, T. Treis, J. Upmeier zu Belzen, E. Roellin, L. May, A. Namsaraeva, N. Horlava, V. A. Shitov, X. Zhang, L. Zappia, R. Knoll, N. J. Lang, L. Hetzel, I. Virshup, L. Sikkema, F. Curion, R. Eils, H. B. Schiller, A. Hilgendorff, and F. J. Theis. An open-source framework for end-to-end analysis of electronic health record data. *Nature Medicine*, 30(11):3369–3380, Nov. 2024. ISSN 1546-170X. doi: 10.1038/s41591-024-03214-0. URL `https://www.nature.com/articles/s41591-024-03214-0`. Publisher: Nature Publishing Group.

[23] J. Hong, S. K. Kang, I. Alberts, J. Lu, R. Sznitman, J. S. Lee, A. Rominger, H. Choi, K. Shi, and the Alzheimer's Disease Neuroimaging Initiative. Image-level trajectory inference of tau pathology using variational autoencoder for Flortaucipir PET. *European Journal of Nuclear Medicine and Molecular Imaging*, 49(9):3061–3072, July 2022. ISSN 1619-7089. doi: 10.1007/s00259-021-05662-z. URL `https://doi.org/10.1007/s00259-021-05662-z`.

[24] B. Huang, K. Zhang, J. Zhang, J. Ramsey, R. Sanchez-Romero, C. Glymour, and B. Schölkopf. Causal Discovery from Heterogeneous/Nonstationary Data. *Journal of Machine Learning Research*, 21(89):1–53, 2020. ISSN 1533-7928. URL `http://jmlr.org/papers/v21/19-232.html`.

[25] A. Ingannato, S. Bagnoli, S. Mazzeo, G. Giacomucci, V. Bessi, C. Ferrari, S. Sorbi, and B. Nacmias. Plasma GFAP, NfL and pTau 181 detect preclinical stages of dementia. *Frontiers in Endocrinology*, 15: 1375302, Apr. 2024. ISSN 1664-2392. doi: 10.3389/fendo.2024.1375302. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11035722/`.

[26] C. S. Inman, J. R. Manns, K. R. Bijanki, D. I. Bass, S. Hamann, D. L. Drane, R. E. Fasano, C. K. Kovach, R. E. Gross, and J. T. Willie. Direct electrical stimulation of the amygdala enhances declarative memory in humans. *Proceedings of the National Academy of Sciences*, 115(1):98–103, Jan. 2018. doi: 10.1073/pnas.1714058114. URL https://www.pnas.org/doi/10.1073/pnas.1714058114. Publisher: Proceedings of the National Academy of Sciences.

[27] Y. Jung and J. S. Damoiseaux. The potential of blood neurofilament light as a marker of neurodegeneration for Alzheimer's disease. *Brain*, 147(1):12–25, Aug. 2023. ISSN 0006-8950. doi: 10.1093/brain/awad267. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11484517/.

[28] T. Kanekiyo, H. Xu, and G. Bu. ApoE and A in Alzheimer's disease: accidental encounters or partners? *Neuron*, 81(4):740–754, Feb. 2014. ISSN 0896-6273. doi: 10.1016/j.neuron.2014.01.045. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3983361/.

[29] C. K. Kim, Y. R. Lee, L. Ong, M. Gold, A. Kalali, and J. Sarkar. Alzheimer's Disease: Key Insights from Two Decades of Clinical Trial Failures. *Journal of Alzheimer's Disease*, 87(1):83–100, May 2022. ISSN 1387-2877. doi: 10.3233/JAD-215699. URL https://journals.sagepub.com/action/showAbstract. Publisher: SAGE Publications.

[30] R. La Joie, A. V. Visani, O. H. Lesman-Segev, S. L. Baker, L. Edwards, L. Iaccarino, D. N. Soleimani-Meigooni, T. Mellinger, M. Janabi, Z. A. Miller, D. C. Perry, J. Pham, A. Strom, M. L. Gorno-Tempini, H. J. Rosen, B. L. Miller, W. J. Jagust, and G. D. Rabinovici. Association of APOE4 and Clinical Variability in Alzheimer Disease With the Pattern of Tau- and Amyloid-PET. *Neurology*, 96(5):e650–e661, Feb. 2021. ISSN 0028-3878. doi: 10.1212/WNL.0000000000011270. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7884991/.

[31] S. Langella, P. J. Mucha, K. S. Giovanello, and E. Dayan. The association between hippocampal volume and memory in pathological aging is mediated by functional redundancy. *Neurobiology of aging*, 108:179–188, Dec. 2021. ISSN 0197-4580. doi: 10.1016/j.neurobiolaging.2021.09.002. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8616845/.

[32] J. McGaugh, L. Cahill, and B. Roozendaal. Involvement of the amygdala in memory storage: Interaction with other brainsystems. *Proceedings of the National Academy of Sciences*, 93(24):13508–13514, Nov. 1996. doi: 10.1073/pnas.93.24.13508. URL https://www.pnas.org/doi/10.1073/pnas.93.24.13508. Publisher: Proceedings of the National Academy of Sciences.

[33] J. M. Mooij, S. Magliacane, and T. Claassen. Joint causal inference from multiple contexts. *J. Mach. Learn. Res.*, 21(1):99:3919–99:4026, Jan. 2020. ISSN 1532-4435.

[34] M. E. Nelson, D. J. Jester, A. J. Petkus, and R. Andel. Cognitive Reserve, Alzheimer's Neuropathology, and Risk of Dementia: A Systematic Review and Meta-Analysis. *Neuropsychology Review*, 31(2):233–250, June 2021. ISSN 1573-6660. doi: 10.1007/s11065-021-09478-4. URL https://doi.org/10.1007/s11065-021-09478-4.

[35] J. B. Pereira, S. Janelidze, E. Stomrud, S. Palmqvist, D. van Westen, J. L. Dage, N. Mattsson-Carlgren, and O. Hansson. Plasma markers predict changes in amyloid, tau, atrophy and cognition in non-demented subjects. *Brain*, 144(9):2826–2836, June 2021. ISSN 0006-8950. doi: 10.1093/brain/awab163. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8557344/.

[36] R. C. Petersen, P. S. Aisen, L. A. Beckett, M. C. Donohue, A. C. Gamst, D. J. Harvey, C. R. Jack, W. J. Jagust, L. M. Shaw, A. W. Toga, J. Q. Trojanowski, and M. W. Weiner. Alzheimer's Disease Neuroimaging Initiative (ADNI). *Neurology*, 74(3):201–209, Jan. 2010. doi: 10.1212/WNL.0b013e3181cb3e25. URL https://www.neurology.org/doi/10.1212/WNL.0b013e3181cb3e25. Publisher: Wolters Kluwer.

[37] G. Pobric, E. Jefferies, and M. A. L. Ralph. Anterior temporal lobes mediate semantic representation: mimicking semantic dementia by using rTMS in normal participants. *Proceedings of the National Academy of Sciences of the United States of America*, 104(50):20137–20141, Dec. 2007. ISSN 1091-6490. doi: 10.1073/pnas.0707383104.

[38] T. S. Richardson. A Discovery Algorithm for Directed Cyclic Graphs, Feb. 2013. URL http://arxiv.org/abs/1302.3599. arXiv:1302.3599 [cs].

[39] S. J. Ritchie, S. R. Cox, X. Shen, M. V. Lombardo, L. M. Reus, C. Alloza, M. A. Harris, H. L. Alderson, S. Hunter, E. Neilson, D. C. M. Liewald, B. Auyeung, H. C. Whalley, S. M. Lawrie, C. R. Gale, M. E. Bastin, A. M. McIntosh, and I. J. Deary. Sex Differences in the Adult Human Brain: Evidence from 5216 UK Biobank Participants. *Cerebral Cortex (New York, NY)*, 28(8):2959–2975, Aug. 2018. ISSN 1047-3211. doi: 10.1093/cercor/bhy109. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6041980/.

[40] L. A. Ross, D. McCoy, H. B. Coslett, I. R. Olson, and D. A. Wolk. Improved proper name recall in aging after electrical stimulation of the anterior temporal lobes. *Frontiers in Aging Neuroscience*, 3:16, 2011. ISSN 1663-4365. doi: 10.3389/fnagi.2011.00016.

[41] X. Shen, S. Ma, P. Vemuri, and G. Simon. Challenges and Opportunities with Causal Discovery Algorithms: Application to Alzheimer's Pathophysiology. *Scientific Reports*, 10(1):2975, Feb. 2020. ISSN 2045-2322. doi: 10.1038/s41598-020-59669-x. URL https://www.nature.com/articles/s41598-020-59669-x. Publisher: Nature Publishing Group.

[42] M. Sinha and S. A. Ramsey. Using a General Prior Knowledge Graph to Improve Data-Driven Causal Network Learning. *AAAI Spring Symposium: Combining Machine Learning with Knowledge Engineering*, 2021.

[43] Y. Stern. Cognitive reserve in ageing and Alzheimer's disease. *The Lancet Neurology*, 11(11):1006–1012, Nov. 2012. ISSN 1474-4422, 1474-4465. doi: 10.1016/S1474-4422(12)70191-6. URL https://www.thelancet.com/journals/laneur/article/PIIS1474-4422(12)70191-6/fulltext. Publisher: Elsevier.

[44] Y. Stern, B. Gurland, T. K. Tatemichi, M. X. Tang, D. Wilder, and R. Mayeux. Influence of Education and Occupation on the Incidence of Alzheimer's Disease. *JAMA*, 271(13):1004–1010, Apr. 1994. ISSN 0098-7484. doi: 10.1001/jama.1994.03510370056032. URL https://doi.org/10.1001/jama.1994.03510370056032.

[45] N. H. Stricker, T. J. Christianson, E. S. Lundt, E. C. Alden, M. M. Machulda, J. A. Fields, W. K. Kremers, C. R. Jack, D. S. Knopman, M. M. Mielke, and R. C. Petersen. Mayo Normative Studies: Regression-Based Normative Data for the Auditory Verbal Learning Test for Ages 30-91 Years and the Importance of Adjusting for Sex. *Journal of the International Neuropsychological Society : JINS*, 27(3):211–226, Mar. 2021. ISSN 1355-6177. doi: 10.1017/S1355617720000752. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7895855/.

[46] E. E. Sundermann, M. S. Panizzon, X. Chen, M. Andrews, D. Galasko, and S. J. Banks. Sex differences in Alzheimer's-related Tau biomarkers and a mediating effect of testosterone. *Biology of Sex Differences*, 11: 33, June 2020. ISSN 2042-6410. doi: 10.1186/s13293-020-00310-x. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7304096/.

[47] S. Tan, W. Chen, G. Kong, L. Wei, and Y. Xie. Peripheral inflammation and neurocognitive impairment: correlations, underlying mechanisms, and therapeutic implications. *Frontiers in Aging Neuroscience*, 15, Nov. 2023. ISSN 1663-4365. doi: 10.3389/fnagi.2023.1305790. URL https://www.frontiersin.org/journals/aging-neuroscience/articles/10.3389/fnagi.2023.1305790/full. Publisher: Frontiers.

[48] J. Therriault, A. L. Benedet, T. A. Pascoal, S. Mathotaarachchi, M. Chamoun, M. Savard, E. Thomas, M. S. Kang, F. Lussier, C. Tissot, M. Parsons, M. N. I. Qureshi, P. Vitali, G. Massarweh, J.-P. Soucy, S. Rej, P. Saha-Chaudhuri, S. Gauthier, and P. Rosa-Neto. Association of Apolipoprotein E 4 With Medial Temporal Tau Independent of Amyloid-. *JAMA Neurology*, 77(4):470–479, Apr. 2020. ISSN 2168-6149. doi: 10.1001/jamaneurol.2019.4421. URL https://doi.org/10.1001/jamaneurol.2019.4421.

[49] J. Tran, D. Chang, F. Hsu, H. Wang, and Z. Guo. Cross-seeding between AB40 and AB42 in Alzheimer's disease. *FEBS Letters*, 591(1):177–185, 2017. ISSN 1873-3468. doi: 10.1002/1873-3468.12526. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/1873-3468.12526. _eprint: https://febs.onlinelibrary.wiley.com/doi/pdf/10.1002/1873-3468.12526.

[50] K. B. Walhovd, L. T. Westlye, I. Amlien, T. Espeseth, I. Reinvang, N. Raz, I. Agartz, D. H. Salat, D. N. Greve, B. Fischl, A. M. Dale, and A. M. Fjell. Consistent neuroanatomical age-related volume differences across multiple samples. *Neurobiology of aging*, 32(5):916–932, May 2011. ISSN 0197-4580. doi: 10.1016/j.neurobiolaging.2009.05.013. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4040218/.

[51] X. Wang, Z. Shi, Y. Qiu, D. Sun, and H. Zhou. Peripheral GFAP and NfL as early biomarkers for dementia: longitudinal insights from the UK Biobank. *BMC Medicine*, 22(1):192, May 2024. ISSN 1741-7015. doi: 10.1186/s12916-024-03418-8. URL https://doi.org/10.1186/s12916-024-03418-8.

[52] H.-Y. Wu, P.-C. Kuo, Y.-T. Wang, H.-T. Lin, A. D. Roe, B. Y. Wang, C.-L. Han, B. T. Hyman, Y.-J. Chen, and H.-C. Tai. -Amyloid Induces Pathology-Related Patterns of Tau Hyperphosphorylation at Synaptic Terminals. *Journal of Neuropathology & Experimental Neurology*, 77(9):814–826, Sept. 2018. ISSN 0022-3069. doi: 10.1093/jnen/nly059. URL https://doi.org/10.1093/jnen/nly059.

[53] A. L. Young, R. V. Marinescu, N. P. Oxtoby, M. Bocchetta, K. Yong, N. C. Firth, D. M. Cash, D. L. Thomas, K. M. Dick, J. Cardoso, J. van Swieten, B. Borroni, D. Galimberti, M. Masellis, M. C. Tartaglia, J. B. Rowe, C. Graff, F. Tagliavini, G. B. Frisoni, R. Laforce, E. Finger, A. de Mendonça, S. Sorbi, J. D. Warren, S. Crutch, N. C. Fox, S. Ourselin, J. M. Schott, J. D. Rohrer, and D. C. Alexander. Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with Subtype and Stage Inference. *Nature Communications*, 9(1):4273, Oct. 2018. ISSN 2041-1723. doi: 10.1038/s41467-018-05892-0. URL https://www.nature.com/articles/s41467-018-05892-0. Publisher: Nature Publishing Group.

[54] C. B. Young, E. Johns, G. Kennedy, M. E. Belloy, P. S. Insel, M. D. Greicius, R. A. Sperling, K. A. Johnson, K. L. Poston, E. C. Mormino, for the Alzheimer's Disease Neuroimaging Initiative, and the A4 Study Team. APOE effects on regional tau in preclinical Alzheimer's disease. *Molecular Neurodegeneration*, 18(1):1, Jan. 2023. ISSN 1750-1326. doi: 10.1186/s13024-022-00590-4. URL https://doi.org/10.1186/s13024-022-00590-4.

[55] J. Zhang, Y. Zhang, J. Wang, Y. Xia, J. Zhang, and L. Chen. Recent advances in Alzheimer's disease: mechanisms, clinical trials and new drug development strategies. *Signal Transduction and Targeted Therapy*, 9(1):211, Aug. 2024. ISSN 2059-3635. doi: 10.1038/s41392-024-01911-3. URL https://www.nature.com/articles/s41392-024-01911-3. Publisher: Nature Publishing Group.

[56] Y. Zhang, H. Chen, R. Li, K. Sterling, and W. Song. Amyloid -based therapy for Alzheimer's disease: challenges, successes and future. *Signal Transduction and Targeted Therapy*, 8(1):248, June 2023. ISSN 2059-3635. doi: 10.1038/s41392-023-01484-7. URL https://www.nature.com/articles/s41392-023-01484-7. Publisher: Nature Publishing Group.

[57] Y. Zheng and X. Hu. Healthcare predictive analytics for disease progression: a longitudinal data fusion approach. *Journal of Intelligent Information Systems*, 55(2):351–369, Oct. 2020. ISSN 1573-7675. doi: 10.1007/s10844-020-00606-9. URL https://doi.org/10.1007/s10844-020-00606-9.

[58] F. Zhou, K. He, and Y. Ni. Individualized causal discovery with latent trajectory embedded Bayesian networks. *Biometrics*, 79(4):3191–3202, 2023. ISSN 1541-0420. doi: 10.1111/biom.13843. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/biom.13843. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/biom.13843.

5 Supplementary material

5.1 MCMC procedure and convergence statistics

In order to confirm the stability of the MCMC results, we ran 4 chains across 5000 iterations with 1000 iterations for burn-in and evaluate their convergence statistics. We used 5 knots for both causal and baseline cubic b-splines. The rest of the hyperparameters were set according to Zhou et al. [58]. The $\hat{R}$ values and ESS values are presented in the table below.

Variable	No mod.		SF		SF + BK	
	$\hat{R}$	ESS	$\hat{R}$	ESS	$\hat{R}$	ESS
Per-variable noise variance	1.02	1269.58	1.03	364.62	1.01	400.10
Causal spline coefficients	1.00	4041.07	1.00	730.3	1.00	723.14
Baseline spline coefficients	1.03	346.7	1.01	432.10	1.01	217.42
Roughness parameter	1.00	1504.03	1.00	711.52	1.01	250.85
Variance of spline coefficients	1.01	233.74	1.01	219.91	1.01	199.33

Table S1: Convergence statistics on the MCMC results

We followed the procedure outlined in the supplementary material of Zhou et al. [58] in order to sample the parameter posteriors. To obtain the PIPs of the edges, we averaged the edge inclusion indicator parameters (r_{jk}) across the four chains, and used a threshold of $P \geq 0.5$ to construct the final causal graphs, which were then used to calculate statistics in Table 2.

In addition, we remove the Coulomb prior as described by Zhou et al. [58] from the model, which serves to ensure a distance between samples and prevent clustering, as we do not make an assumption that there exists a uniform distribution of patients across disease pseudotime due to the uneven distribution of patient diagnoses, as outlined in the Methods section.

All experiments were performed on a MacBook Pro with an Apple M4 Pro CPU and 24GB of RAM. A single run of the MCMC algorithm (4 chains) required approximately 30 minutes to complete.

Implementation of background knowledge involved specifying the variables that were root nodes and sink nodes. Root nodes were then skipped during sampling of baseline and incoming causal trajectory variables, and remained set at 0. For sink nodes, only outgoing causal trajectories were set to 0. When calculating residuals in the sampling, root nodes were excluded.

For settings with background knowledge, MSE was calculated over only the variables that were non-root nodes.

5.2 Consensus and Estimated Causal Graph

The consensus causal graph was created using a literature search of consensus mechanisms of AD progression. Where variables are grouped together, the citation shows evidence of the effect across all the variables. Below are the identified causal relationships:

Age → ICV[10], [Hippocampus, Amygdala, Temporal lobe][50], pTau [12], [Aβ42, Aβ40] [57], GFAP [2], [RAVLT, TRABSCORE, ADAS13]

Education → [RAVLT, TRABSCORE, ADAS13] [45]

Sex → [ICV, Hippocampus, Amygdala, Temporal lobe] [39], pTau [46], RAVLT, TRABSCORE, ADAS13 [45]

APOE → pTau[14; 54; 30; 48], [Aβ42, Aβ40] [28]

ICV → Hippocampus, Amygdala, Temporal lobe

Hippocampus → [RAVLT, TRABSCORE, ADAS13] [31; 9]

Amygdala → [RAVLT, TRABSCORE, ADAS13] [32; 26]

Temporal lobe → [RAVLT, TRABSCORE, ADAS13] [40; 37]

pTau → [Aβ42, Aβ40] [6], NfL [13], [RAVLT, TRABSCORE, ADAS13] [35]

Aβ42 → Aβ40 [20], pTau217 [17; 52; 6] NfL [56], GFAP [19]

Aβ40 → Aβ42 [20], pTau217 [17]

NfL → [Hippocampus, Amygdala, Temporal lobe] [16]

GFAP → [RAVLT, TRABSCORE, ADAS13] [47]

The rest of the edges were marked as being absent, aside from the following, which were left out of the analysis due to a lack of causal understanding in literature:

Education → [pTau217, Aβ40, Aβ42, NfL, GFAP]

Hippocampus - [Temporal Lobe, Amygdala]

[Aβ42, Aβ42] → [RAVLT, TRABSCORE, ADAS13]

[NfL, GFAP] → [pTau217, Aβ40, Aβ442]

GFAP → NfL

These edges were not included when calculating statistics such as precision, recall, and SHD. It is important to note that the consensus graph contains three feedback systems, namely between pTau217 and AB42 and AB40. This must be taken into account when considering the accuracy values obtained for graph recovery, as the model restricts any cyclical relationships.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction accurately describe the novel application of a latent pseudotime model for dynamic causal discovery in Alzheimer's disease. The claims align with the results presented, which demonstrate the model's ability to infer evolving causal relationships the improvement to the model via the addition of background knowledge.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The discussion section explicitly addresses the limitations of the study. The authors acknowledge that their analysis was constrained by the dataset size and the selection of variables, as well as theoretical identifiability. We also mention that important factors such as ethnicity could not be included due to a lack of cohort heterogeneity.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

   Justification: The paper does not introduce new theoretical results that would require formal proofs. Instead, it applies an existing framework (BN-LTE) to a new domain. The key assumptions of the model are stated in the Methods section.

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The paper provides a good overview of the methodology and the data used (ADNI dataset). The supplementary material further details the MCMC procedure and the construction of the consensus causal graph, which should be sufficient for other researchers to replicate the study.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [No]

   Justification: The code has not been provided as it is not a novel algorithm, although it will be made public upon publication.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: The Methods section and the supplementary material provide the necessary details about the experimental setup. This includes the number of MCMC chains and iterations, the hyperparameters used, and the specifics of the BN-LTE model implementation

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: The paper reports 95% confidence intervals for the performance metrics in Table 2 and the p-values for the AUC comparison. This provides a clear indication of the statistical significance of the findings.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: We provide the computational hardware details (Apple M4 Pro CPU, 24GB RAM) and approximate execution time in the Supplementary Material. This information is sufficient for others to gauge the resources required for reproduction.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The authors ackowledge that the paper conforms to the NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [Yes]

Justification: The discussion section highlights the potential positive societal impacts, such as revolutionizing clinical trial design and enabling personalized treatment plans for Alzheimer's patients. The paper does not explicitly discuss negative societal impacts, as this work should not have a negative societal impact.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: The paper does not introduce any models that have a high risk of misuse

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: The ADNI dataset was used in this work, and a citation for the original dataset was included to credit the data source.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: No new assets have been introduced in the paper

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: The usage of LLMs is not a component of this paper

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.