Predicting progression from mild cognitive impairment to Alzheimer's disease using survival analysis

Anonymous Authors[1]

Abstract

Identifying which individuals with mild cognitive impairment (MCI) will progress to Alzheimer's disease (AD) is critical for therapeutic planning and clinical trial enrollment. We used Random Survival Forest models to evaluate the prognostic value of structured blood biomarkers (p-Tau217, Aβ42/40, NfL, GFAP), amyloid-β PET imaging, APOE ε4 genotype, and demographic data to predict MCI-to-AD conversion in the Alzheimer's Disease Neuroimaging Initiative (n=201). At near 4 years, all biomarker-based models outperformed a demographics-and-APOE ε4 baseline, with a multi-blood biomarker panel achieving the highest AUROC (0.818, 95% CI [0.748–0.888]). However, p-Tau217 alone (AUROC = 0.787 [0.705–0.868]) performed comparably to both amyloid-β PET (0.758 [0.669–0.848]) and the multi-marker panel. Partial dependence analysis revealed a critical window in which increasing p-Tau217 levels from 0.08 to 0.56 pg/mL corresponded to a rise in predicted 4-year conversion probability from approximately 10% to 45%. A two-cutoff thresholding strategy to stratify high (20%), intermediate (31.4%), and low (48.6%) risk groups based on 90% specificity and 90% sensitivity had a negative predictive value of 83.9% and positive predictive value of 44.4%. These findings support p-Tau217 as a minimally invasive and scalable prognostic marker and demonstrate the value of survival modeling in structured clinical data for risk stratification in prodromal AD.

1. Introduction

Alzheimer's disease (AD) affects over 35 million people globally, with mild cognitive impairment (MCI) as a common prodromal stage (Alzheimer's Association, 2025; World Health Organization, 2025). However, because only 10–15% of MCI patients progress to AD dementia, reliable prognostic tools are essential for identifying those at highest risk, particularly as disease-modifying therapies appear to offer the greatest benefit when administered early (Liss et al., 2021; Sims et al., 2023).

Amyloid-β PET is the clinical gold standard for detecting AD pathology but its high cost and limited accessibility constrain broader use(Alzheimer's Association, 2025). Plasma phosphorylated tau 217 (p-Tau217) has emerged as a minimally invasive and scalable alternative. P-Tau217 is a blood-based biomarker strongly associated with amyloid and tau pathology, the core pathological hallmarks of AD (Khalafi et al., 2025; Rudolph et al., 2025). Notably, the p-Tau217/Aβ42 ratio is FDA approved as an aid in AD diagnosis (Ayalew et al., 2026). However, while p-Tau217 has shown strong performance in cross-sectional diagnostic classification, its prognostic value for predicting *when* patients with MCI will progress to AD remains less well understood. It also remains unclear whether p-Tau217 alone can match amyloid-β PET for prognosis, and whether multi-marker blood panels provide meaningful added predictive value.

We address these gaps using Random Survival Forest (RSF) models built on structured data spanning demographics, genetics, and blood-based biomarkers with three objectives: (1) evaluate whether biomarkers significantly improve MCI-to-AD progression prediction beyond demographics and APOE genotype; (2) compare p-Tau217 head-to-head with amyloid-β PET; and (3) assess whether additional blood markers (Aβ42/40, NfL, GFAP) improve upon p-Tau217 alone. We show that p-Tau217 alone demonstrates comparable prognostic performance to amyloid-β PET and a multi-blood biomarker panel. Furthermore, we derive clinically actionable p-Tau217 thresholds for AD risk stratification within a critical range associated with elevated conversion risk.

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

2. Methods

2.1. ADNI Tabular Data

We analyzed structured tabular data from patients clinically diagnosed with MCI in the Alzheimer's Disease Neuroimaging Initiative (ADNI) who had both amyloid-β PET and blood biomarker measurements near time of diagnosis. ADNI is a multi-center longitudinal study collecting structured clinical, genetic, neuroimaging, and blood biomarker data from over 2,500 participants across North America (Aisen et al., 2024). Additional details on ADNI, tabular PET (Centiloids), and blood biomarker data can be found in the Supplementary Methods B.1-B.3.

2.2. Survival Analysis

Survival time was defined as the number of days from MCI diagnosis to AD diagnosis. Patients who did not develop AD were right-censored at their last visit, and all features were measured at a single baseline time point near MCI diagnosis. For the prediction window, we took the 5th and 81st percentiles of follow-up times across cross-validation folds and generated 21 uniformly spaced time points. Primary results are reported at 4.07 years (1,486 days), the grid point closest to 4 years, which is a clinically relevant horizon as MCI-to-AD conversion commonly occurs within 3–5 years(Lombardo et al., 2026). For simplicity, throughout the remainder of the paper, reported year labels correspond to the nearest time points on the evaluation grid, with 2.98, 4.07, and 4.985 years referred to as 3, 4, and 5 years, respectively.

We trained Random Survival Forest (RSF) models using `scikit-survival`, selecting RSF for its ability to handle non-linear relationships and multi-collinearity (Pölsterl, 2020). Furthermore, prior comparative studies have shown it outperforms Cox proportional hazards, gradient boosting, and survival trees for AD progression (Sarica et al., 2023). Six configurations were compared, each adding biomarker features to a shared baseline of demographics (age, sex, education) and APOE ε4: (1) baseline only, (2) +amyloid-β PET, (3) +p-Tau217, (4) +p-Tau217/Aβ42 ratio, (5) +amyloid-β PET and p-Tau217, and (6) +p-Tau217, Aβ42/40, NfL, and GFAP. Hyperparameters were optimized via nested stratified 5-fold cross-validation.

Model discrimination was evaluated using time-dependent area under the receiver operating characteristic curve (AUROC) (Uno et al., 2007; Hung & Chiang, 2010; Lambert & Chevret, 2016), inverse probability of censoring weighted (IPCW) concordance index (Uno et al., 2011), and Brier score for calibration (Graf et al., 1999). Additional modeling metric details are provided in Supplementary Methods B.4 and B.5.

2.3. Deriving Biomarker Thresholds

To translate continuous risk estimates into actionable categories, we used partial dependence plots (PDPs) to characterize how predicted conversion probability varies as a function of individual biomarker values (Molnar, 2020).

We then applied a two-cutoff thresholding strategy to p-Tau217, setting a lower threshold at 90% sensitivity and an upper threshold at 90% specificity, adapted from prior work on blood biomarker-based detection of AD pathology and consistent with recommendations from the Global CEO Initiative on Alzheimer's Disease (Giacomucci et al., 2025; Schindler et al., 2024). Patients below the lower cutoff were classified as low risk, those above the upper cutoff as high risk, and the remainder as intermediate risk.

3. Results

3.1. Cohort Characteristics

The cohort included 157 stable MCI (sMCI) and 44 progressive MCI (pMCI) participants with a mean follow-up of 4.34 years. The two groups were similar in age (74.4 vs. 75.3 years), sex (61.1% vs. 61.4% female), and education (16.1 vs. 16.3 years), while APOE ε4 carrier frequency was significantly higher in pMCI (50.0% vs. 39.5%, $p = 0.037$). Individuals who progressed had significantly higher baseline p-Tau217, NfL, and GFAP levels than those who remained stable ($p < 0.01$). Descriptive statistics and boxplots of baseline biomarker comparisons are provided in Supplementary Table S1 and Supplementary Figure S1, respectively.

3.2. Biomarker Models Outperform Demographics + APOE ε4

At 4 years, all biomarker-based RSF models significantly outperformed the demographics-and-APOE ε4 baseline (AUROC = 0.432; Table 1; Figure 1). The baseline model's near-chance discrimination (C-index = 0.580 ± 0.189) reflects the absence of significant demographic differences between sMCI and pMCI in this cohort (Supplementary Table S1). The combined blood biomarker panel achieved the highest AUROC (0.818 [0.748–0.888]), followed by PET+p-Tau217 (0.808 [0.733-0.883]), p-Tau217 alone (0.787 [0.705-0.868]), and amyloid-β PET (0.758 [0.669-0.848]).

3.3. p-Tau217 Matches Amyloid-β PET and the Clinically Approved Ratio

Plasma p-Tau217 achieved numerically higher AUROCs than amyloid-β PET at all timepoints, though no differences reached significance (4-year: 0.787 vs. 0.758, $p_{\text{adj}} = 0.850$; Table 2). Adding PET to p-Tau217 yielded only a small, non-

Table 1. Time-dependent AUROCs from biomarker and baseline (demographics + APOE $\varepsilon 4$) models at 4.07-year window.

Model	AUROC [95% CI]	p_{adj}
Baseline	0.432 [.331–.534]	—
Aβ PET	0.758 [.669–.848]	3.41e-7
p-Tau217	0.787 [.705–.868]	1.30e-9
PET+p-Tau217	0.808 [.733–.883]	5.33e-11
Blood panel	0.818 [.748–.888]	6.19e-12

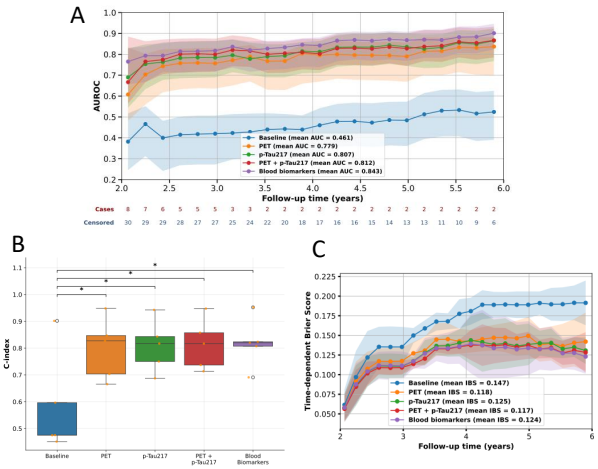

Figure 1. Performance of RSF models across feature configurations. **(A)** Time-dependent area under the receiver operating characteristic curve for predicting conversion to Alzheimer's dementia from mild cognitive impairment. Models tested include 1) Baseline (age, sex, education, and APOE $\varepsilon 4$), 2) PET, 3) p-Tau217, 4) PET+p-tau217, and 5) a combined blood-based biomarker panel (p-Tau217, NfL, GFAP, Aβ42/40). All non-baseline models also include the baseline variables. **(B)** C-index box plots; asterisks indicate significant improvement over baseline ($p < 0.05$). **(C)** Time-dependent Brier scores, where lower score indicates better calibration. A) and C) are represented as averages across 5 cross-validation folds, with 95% CIs for A) and standard deviations for C).

significant gain (0.808 vs. 0.787; $p_{adj} = 0.595$), suggesting limited incremental value from imaging beyond what p-Tau217 already captures.

Importantly, p-Tau217 alone performed comparably to the FDA-cleared p-Tau217/Aβ42 ratio at all timepoints (Figure 2), with near-identical AUROC curves and C-index distributions. At 4 years, the difference was insignificant (0.787 vs. 0.777; $p_{adj} = 0.874$), indicating that the simpler single-analyte measure may be sufficient for prognostic risk stratification.

Table 2. Pairwise AUROC comparisons across follow-up windows, with 4 years as the primary timepoint of evaluation.

Model A	Model B	Yr	ΔAUC	p_{adj}
Aβ PET	p-Tau217	3	+.048	0.566
		4	+.029	0.850
		5	+.058	0.686
p-Tau217	PET+pTau	4	+.021	0.595
p-Tau217	pTau/Aβ42	4	–.010	0.874

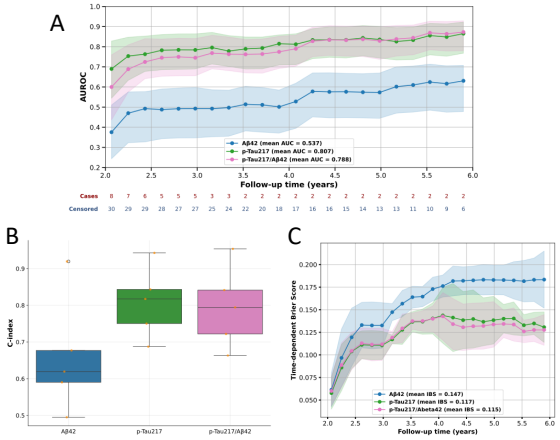

Figure 2. Comparison of p-Tau217, Aβ42, and the p-Tau217/Aβ42 ratio for predicting MCI-to-AD conversion. **(A)** Time-dependent AUROC. **(B)** C-index boxplots. **(C)** Brier scores. p-Tau217 and the clinically approved ratio show closely overlapping performance.

3.4. Multi-Marker Panel Does Not Significantly Outperform p-Tau217

The combined panel (p-Tau217+Aβ42/40+NfL+GFAP) achieved a 4-year AUROC of 0.818 vs. 0.787 for p-Tau217 alone (ΔAUROC = +0.031, $p_{adj} = 0.310$). Among individual biomarkers, p-Tau217 demonstrated the strongest standalone performance (C-index = 0.808 ± 0.097), followed by Aβ42/40 (0.743 ± 0.104), GFAP (0.735 ± 0.116), and NfL (0.632 ± 0.156). SHAP analysis reaffirmed p-Tau217 as the dominant predictor (mean $|\text{SHAP}| = 3.71$), far exceeding Aβ42/40 (1.42), GFAP (0.42), NfL (0.38), and all demographic variables (≤ 0.27) (Supplementary Figure S2).

3.5. Biomarker Thresholds and Partial Dependence for Risk Stratification

Partial dependence analysis revealed a critical window of increased conversion risk: as p-Tau217 increased from 0.08 to 0.56 pg/mL, the predicted 4-year conversion probability rose from approximately 10% to 45%(Figure 3A). This signal is not merely a proxy for age, as the pMCI group

exhibited consistently higher p-Tau217 levels than the sMCI group across the full observed age range (Figure 3B).

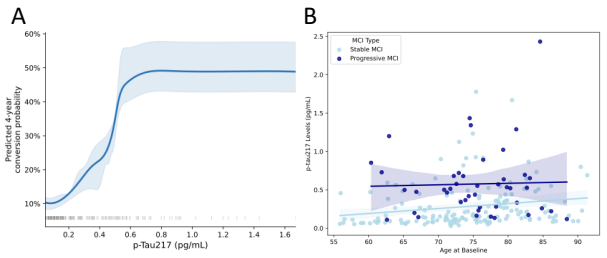

Figure 3. Critical window of increased conversion risk within p-Tau217 levels of 0.08-0.56 pg/mL. **(A)** Partial dependence plot shows the independent relationship between plasma p-Tau217 and the predicted probability of conversion from MCI to AD within 4 years. **(B)** Scatterplot illustrating the association of p-Tau217 levels with baseline age in stable MCI and progressive MCI groups. Trend lines represent ordinary least squares linear regression fits, with shaded regions indicating 95% confidence intervals.

To translate these continuous risk estimates into actionable clinical categories, we applied a two-cutoff p-Tau217 framework (Figure 4). Using thresholds at 90% sensitivity (0.199 pg/mL) and 90% specificity (0.553 pg/mL), patients were classified into low risk ($n = 68$; NPV = 83.9%), intermediate ($n = 44$; 31.4% of cohort), and high risk ($n = 28$; PPV = 44.4%). Under this approach, 68.6% of patients could be classified with high confidence using p-Tau217 alone, potentially avoiding confirmatory PET imaging. The cutoff values fall within the steepest portion of the PDP curve, reinforcing that the region of greatest prognostic information corresponds to a biologically meaningful transition in disease risk.

4. Discussion

We demonstrate that structured tabular data can be used within a survival analysis framework to predict MCI-to-AD conversion. Plasma p-Tau217 showed prognostic performance comparable to amyloid-β PET and the FDA-cleared p-Tau217/Aβ42 ratio. The combined blood biomarker panel did not significantly outperform p-Tau217 alone, consistent with findings that p-Tau217 captures the dominant AD-specific prognostic signal at the MCI stage (Palmqvist et al., 2021; Grande et al., 2025).

Although prior survival-based ML studies have reported higher C-indices of 0.87 and 0.89, they incorporated cognitive variables such as the Functional Activities Questionnaire and Clinical Dementia Rating Scale collected within 4 years of AD onset (Song et al., 2023; Sarica et al., 2023). Because clinical diagnosis is itself based largely on cognitive symptoms, such measures may inflate prognostic performance. In contrast, the resulting C-index of 0.808

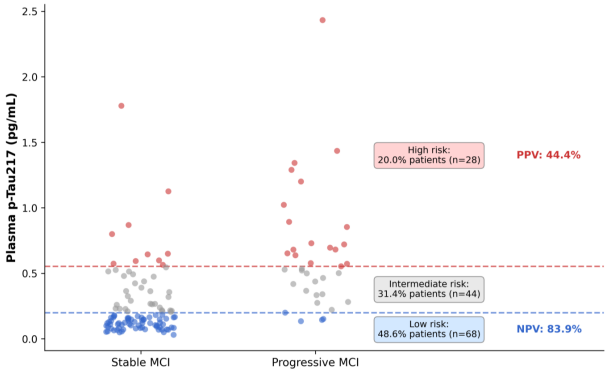

Figure 4. Two-cutoff p-Tau217 thresholds for AD risk stratification. Patients were classified into three risk groups based on baseline plasma p-Tau217 levels using thresholds set at 90% sensitivity (lower: 0.199 pg/mL) and 90% specificity (upper: 0.553 pg/mL). Low-risk = below 0.199 pg/mL (blue zone); High-risk = above 0.553 pg/mL (red zone); Intermediate = between thresholds (gray zone).

and AUROC of 0.787 for our p-Tau217 + baseline model was achieved without cognitive testing or brain imaging, supporting feasibility for blood-based risk stratification in primary care settings (Palmqvist et al., 2019). This prognostic application is distinct from the FDA-cleared diagnostic use of the p-Tau217/Aβ42 ratio: our work addresses not whether a patient has AD pathology, but whether and *when* they are likely to progress to AD dementia.

Partial dependence curves and two-cutoff thresholds suggest that a narrow range of p-Tau217 concentrations captures the critical transition from low to high conversion risk. The cutoffs of 0.199 and 0.553 pg/mL fall within a similar range to what other studies have reported on the same type of assay for detecting patients with AD pathology (Giacomucci et al., 2025; Kang et al., 2026). Patients below the lower threshold (NPV = 83.9%) could be monitored routinely, while those above could be prioritized for treatment evaluation or clinical trial enrollment, with the intermediate group referred for confirmatory PET.

Limitations and Future Work. The ADNI cohort is modest in size and predominantly White and highly educated and therefore may not represent the general AD population. Furthermore, we utilized single biomarker measurements; longitudinal measurements may improve prediction and the understanding of MCI development on a biological level. Future work should validate the prediction model and two-cutoff thresholds in diverse cohorts and explore time-varying features.

Impact Statement

This paper presents work at the intersection of machine learning and clinical risk stratification. Our goal is to improve prognostic modeling for progression from mild cognitive impairment to Alzheimer's disease using structured biomarker data. If validated more broadly, such approaches may support earlier and more scalable risk stratification.

References

Aisen, P. S., Donohue, M. C., Raman, R., Rafii, M. S., Petersen, R. C., and Initiative, A. D. N. The alzheimer's disease neuroimaging initiative clinical core. *Alzheimer's & Dementia*, 20(10):7361–7368, 2024.

Alzheimer's Association. 2025 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, 21(4):e70235, 2025.

Ayalew, B. D., Alemayehu, Z. G., Bonger, T. D., Sime, B. L., Zewdie, Y. A., Tewodros, Y. T., and Bano, S. Fda approval of the lumipulse g ptau217/β-amyloid 1-42 plasma ratio test: a new era in accessible alzheimer's diagnosis, 2026.

Chatterjee, P., Pedrini, S., Doecke, J. D., Thota, R., Villemagne, V. L., Doré, V., Singh, A. K., Wang, P., Rainey-Smith, S., Fowler, C., et al. Plasma aβ42/40 ratio, p-tau181, gfap, and nfl across the alzheimer's disease continuum: a cross-sectional and longitudinal study in the aibl cohort. *Alzheimer's & Dementia*, 19(4):1117–1134, 2023.

Giacomucci, G., Mazzeo, S., Bagnoli, S., Ingannato, A., Leccese, D., Berti, V., Padiglioni, S., Galdo, G., Ferrari, C., Sorbi, S., et al. Plasma neurofilament light chain as a biomarker of alzheimer's disease in subjective cognitive decline and mild cognitive impairment. *Journal of neurology*, 269(8):4270–4280, 2022.

Giacomucci, G., Crucitti, C., Ingannato, A., Moschini, V., Bagnoli, S., Pozzi, F. E., Marcantelli, E., Padiglioni, S., Morinelli, C., Mazzeo, S., et al. The two cut-offs approach for plasma p-tau217 in detecting alzheimer's disease in subjective cognitive decline and mild cognitive impairment. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 17(2):e70116, 2025.

Graf, E., Schmoor, C., Sauerbrei, W., and Schumacher, M. Assessment and comparison of prognostic classification schemes for survival data. *Statistics in medicine*, 18(17-18):2529–2545, 1999.

Grande, G., Valletta, M., Rizzuto, D., Xia, X., Qiu, C., Orsini, N., Dale, M., Andersson, S., Fredolini, C., Winblad, B., et al. Blood-based biomarkers of alzheimer's disease and incident dementia in the community. *Nature Medicine*, 31(6):2027–2035, 2025.

Harrell, F. E., Lee, K. L., Califf, R. M., Pryor, D. B., and Rosati, R. A. Regression modelling strategies for improved prognostic prediction. *Statistics in medicine*, 3(2): 143–152, 1984.

Hung, H. and Chiang, C.-T. Estimation methods for time-dependent auc models with survival data. *Canadian Journal of Statistics*, 38(1):8–26, 2010.

Kang, H., Gu, Y., Lee, J., Yoon, S., Zetterberg, H., Blennow, K., Gonzalez-Ortiz, F., Ashton, N. J., Day, T. A., Weiner, M. W., et al. Double cutoff strategies for plasma ptau217 to predict tau pet positivity across multiple assay platforms: Tau-enriched and tau-scarce cohorts for cost-effective clinical use. *Alzheimer's Research & Therapy*, 2026.

Khalafi, M., Dartora, W. J., McIntire, L. B. J., Butler, T. A., Wartchow, K. M., Hojjati, S. H., Razlighi, Q. R., Shirbandi, K., Zhou, L., Chen, K., et al. Diagnostic accuracy of phosphorylated tau217 in detecting alzheimer's disease pathology among cognitively impaired and unimpaired: A systematic review and meta-analysis. *Alzheimer's & Dementia*, 21(2):e14458, 2025.

Kim, K. Y., Shin, K. Y., and Chang, K.-A. Gfap as a potential biomarker for alzheimer's disease: a systematic review and meta-analysis. *Cells*, 12(9):1309, 2023.

Klunk, W. E., Koeppe, R. A., Price, J. C., Benzinger, T. L., Devous Sr, M. D., Jagust, W. J., Johnson, K. A., Mathis, C. A., Minhas, D., Pontecorvo, M. J., et al. The centiloid project: standardizing quantitative amyloid plaque estimation by pet. *Alzheimer's & dementia*, 11(1):1–15, 2015.

Lambert, J. and Chevret, S. Summary measure of discrimination in survival models based on cumulative/dynamic time-dependent roc curves. *Statistical methods in medical research*, 25(5):2088–2102, 2016.

Liss, J., Seleri Assunção S. Cummings, J. A., Geldmacher, A., et al. Practical recommendations for timely, accurate diagnosis of symptomatic alzheimer's disease (mci and dementia) in primary care: a review and synthesis. *Journal of internal medicine*, 290(2):310–334, 2021.

Lombardo, F. L., Caraglia, N., Lorenzini, P., et al. Mild cognitive impairment-to-alzheimer's dementia progression risk: the contribution of the interceptor project. *Alzheimer's & Dementia*, 22(4):e71204, 2026.

Molnar, C. *Interpretable machine learning*. Lulu. com, 2020.

Palmqvist, S., Janelidze, S., Stomrud, E., Zetterberg, H., Karl, J., Zink, K., Bittner, T., Mattsson, N., Eichenlaub, U., Blennow, K., et al. Performance of fully automated plasma assays as screening tests for alzheimer disease–related β-amyloid status. *JAMA neurology*, 76(9):1060–1069, 2019.

Palmqvist, S., Tideman, P., Cullen, N., Zetterberg, H., Blennow, K., Initiative, A. D. N., Dage, J. L., Stomrud, E., Janelidze, S., Mattsson-Carlgren, N., et al. Prediction of future alzheimer's disease dementia using plasma phospho-tau combined with other accessible measures. *Nature medicine*, 27(6):1034–1042, 2021.

Pérez-Grijalba, V., Romero, J., Pesini, P., Sarasa, L., Monleón, I., San-José, I., Arbizu, J., Martínez-Lage, P., Munuera, J., Ruiz, A., et al. Plasma aβ42/40 ratio detects early stages of alzheimer's disease and correlates with csf and neuroimaging biomarkers in the ab255 study. *The journal of prevention of Alzheimer's disease*, 6(1):34–41, 2019.

Petersen, K. K., Milà-Alomà, M., Li, Y., Du, L., Xiong, C., Tosun, D., Saef, B., Saad, Z. S., Du-Cuny, L., Coomaraswamy, J., et al. Predicting onset of symptomatic alzheimers disease with plasma p-tau217 clocks. *Nature Medicine*, pp. 1–10, 2026.

Pölsterl, S. scikit-survival: A library for time-to-event analysis built on top of scikit-learn. *Journal of Machine Learning Research*, 21(212):1–6, 2020.

Rudolph, M. D., Sutphen, C. L., Register, T. C., Lockhart, S. N., Rundle, M. M., Hughes, T. M., Bateman, J. R., Sai, K. K. S., Whitlow, C. T., Craft, S., et al. Evaluation of plasma p-tau217 for detecting amyloid pathology in a heterogeneous community-based cohort. *Alzheimer's & Dementia*, 21(7):e70426, 2025.

Sarica, A., Aracri, F., Bianco, M. G., Vaccaro, M. G., Quattrone, A., and Quattrone, A. Conversion from mild cognitive impairment to alzheimer's disease: A comparison of tree-based machine learning algorithms for survival analysis. In *International conference on brain informatics*, pp. 179–190. Springer, 2023.

Schindler, S. E., Galasko, D., Pereira, A. C., Rabinovici, G. D., Salloway, S., Suárez-Calvet, M., Khachaturian, A. S., Mielke, M. M., Udeh-Momoh, C., Weiss, J., et al. Acceptable performance of blood biomarker tests of amyloid pathology—recommendations from the global ceo initiative on alzheimer's disease. *Nature Reviews Neurology*, 20(7):426–439, 2024.

Sims, J. R., Zimmer, J. A., Evans, C. D., Lu, M., Ardayfio, P., Sparks, J., Wessels, A. M., Shcherbinin, S., Wang, H., Monkul Nery, E. S., et al. Donanemab in early symptomatic alzheimer disease: the trailblazer-alz 2 randomized clinical trial. *Jama*, 330(6):512–527, 2023.

Song, S., Asken, B., Armstrong, M. J., Yang, Y., Li, Z., and Initiative, A. D. N. Predicting progression to clinical alzheimer's disease dementia using the random survival forest. *Journal of Alzheimer's Disease*, 95(2):535–548, 2023.

Stocker, H., Beyer, L., Perna, L., Rujescu, D., Holleczek, B., Beyreuther, K., Stockmann, J., Schöttker, B., Gerwert,

K., and Brenner, H. Association of plasma biomarkers, p-tau181, glial fibrillary acidic protein, and neurofilament light, with intermediate and long-term clinical alzheimer's disease risk: results from a prospective cohort followed over 17 years. *Alzheimer's & Dementia*, 19 (1):25–35, 2023.

Uno, H., Cai, T., Tian, L., and Wei, L.-J. Evaluating prediction rules for t-year survivors with censored regression models. *Journal of the American Statistical Association*, 102(478):527–537, 2007.

Uno, H., Cai, T., Pencina, M. J., D'Agostino, R. B., and Wei, L.-J. On the c-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Statistics in medicine*, 30(10):1105–1117, 2011.

World Health Organization. Dementia, March 31 2025. Fact sheet. Accessed April 28, 2026.

A. Supplementary Data.

A.1. Cohort Characteristics at Baseline

Table S1. Baseline characteristics of the study cohort stratified by MCI conversion status.

Characteristic	sMCI (n=157)	pMCI (n=44)	p-value	Overall (n=201)
Age, years, mean (SD)	74.3 ± 8.2	75.3 ± 6.8	0.501	74.5 ± 7.9
Sex, n (%)			1.000	
Female	96 (61.1%)	27 (61.4%)		123 (61.2%)
Male	61 (38.9%)	17 (38.6%)		78 (38.8%)
Education, years, mean (SD)	16.1 ± 2.7	16.3 ± 2.5	0.785	16.1 ± 2.6
APOE ε4, n (%)			0.037*	
0 copies	95 (60.5%)	22 (50.0%)		117 (58.2%)
1 copy	46 (29.3%)	19 (43.2%)		65 (32.3%)
2 copies	16 (10.2%)	3 (6.8%)		19 (9.5%)

Values are mean $\pm$ SD for continuous variables and n (%) for categorical variables.

P-values were calculated using two-sample t-tests for continuous variables, Fisher's exact test for sex, and chi-squared test for APOE ε4 status

sMCI = stable mild cognitive impairment; pMCI = progressive mild cognitive impairment. * $p < 0.05$.

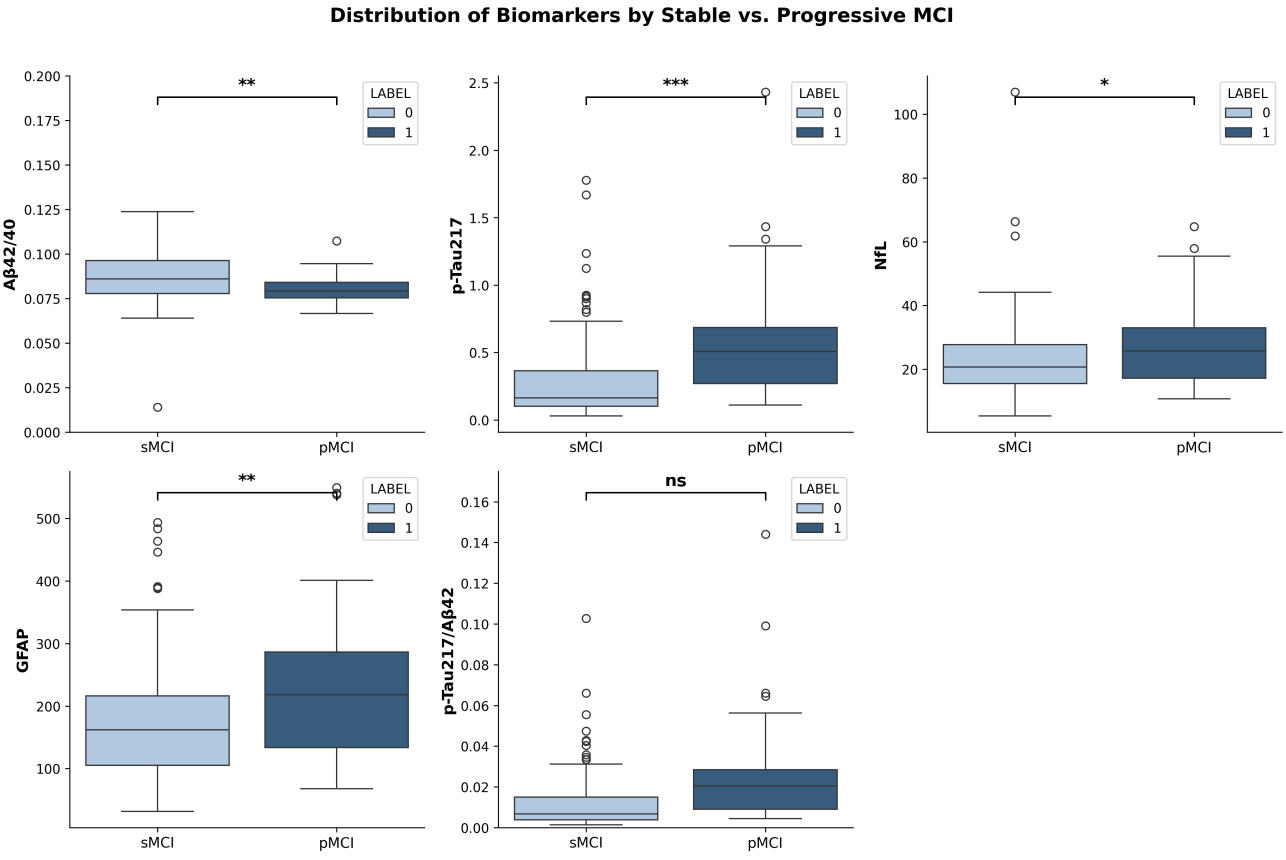

Figure S1. Baseline biomarker levels by stable and progressive MCI. Box plots show the median (center line) and interquartile range (bounds of box) as well as the 2.5th and 97.5th percentiles (whiskers). Horizontal bars indicate significant differences between the two groups. P-values were derived from ordinary least squares regression with group (sMCI vs. pMCI) as the predictor of interest, controlling for age at MCI diagnosis, sex, and APOE ε4 carrier status. ***p <0.001, **p <0.01, *p <0.05.

A.2. Model Interpretability Analysis

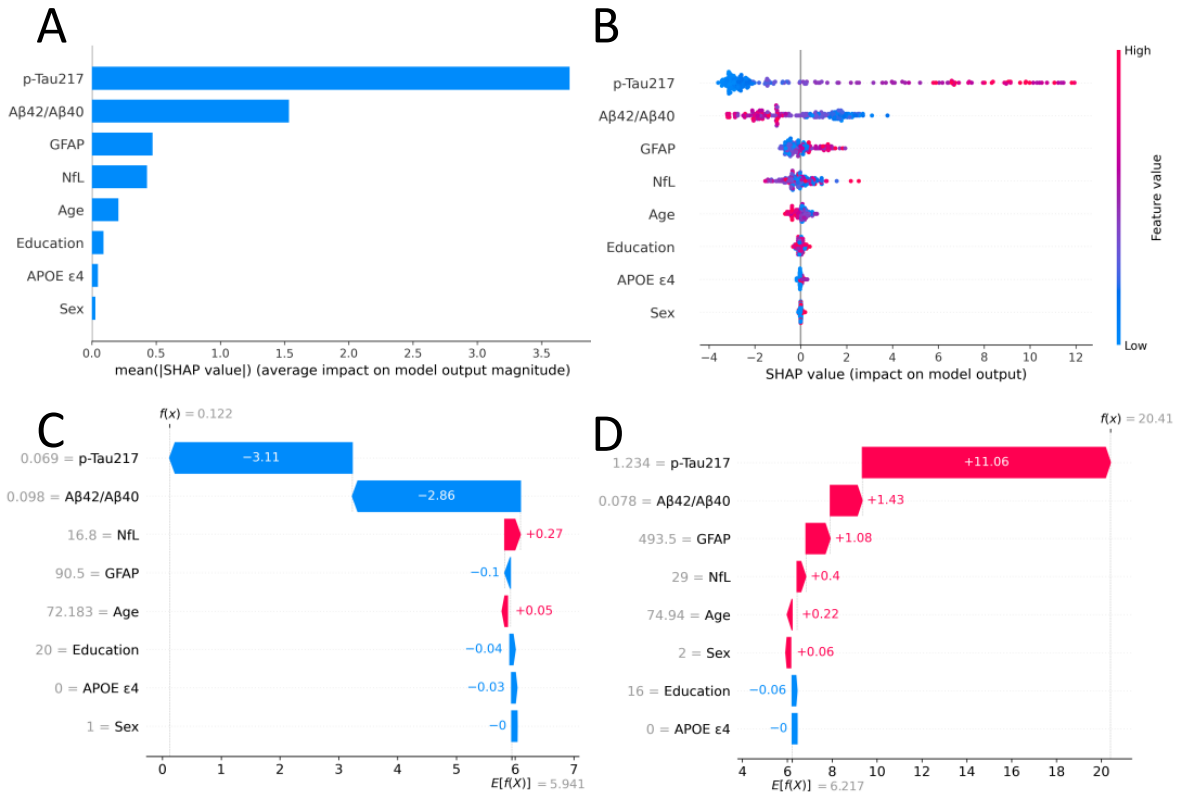

Figure S2. SHAP analysis for MCI-to-AD progression risk prediction using RSF model with plasma biomarker panel, age, sex, education and APOE as predictors.**(A)** Mean absolute SHAP importance values. **(B)** Summary plot of SHAP values showing risk-increasing (red) and risk-decreasing (blue) effects, with features ranked by mean absolute SHAP importance. **(C)** Individual SHAP contributions for lowest-risk patient (predicted risk = 0.122). **(D)** Individual SHAP contributions for highest-risk patient (predicted risk = 20.410).

B. Supplementary Methods

B.1. ADNI Dataset

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD).

B.2. Centiloids

Centiloids are units for a standardized 100-point scale across radiotracers, scanners, and clinical sites. Radiotracers include ^{18}F-Fluorodeoxyglucose, ^{11}C-labeled Pittsburgh compound B, ^{18}F-florbetaben, ^{18}F-florbetapir, and ^{18}F-flutafuranol in ADNI (Klunk et al., 2015).

B.3. Blood-based Biomarkers

The blood-based biomarkers incorporated in this work include plasma phosphorylated Tau 217 (p-Tau217), neurofilament light chain (NfL), glial fibrillary acidic protein (GFAP), and the amyloid-β42/40 ratio, all of which have been validated in prior studies as markers of Alzheimer's pathology or neurodegeneration (Chatterjee et al., 2023; Stocker et al., 2023; Pérez-Grijalba et al., 2019). Plasma p-Tau217 has been shown to correlate strongly with both amyloid and tau pathology and to track disease progression over time (Petersen et al., 2026). NfL is a component of the neuronal cytoskeleton and is released from neurons into the plasma following neurodegeneration. This makes it a valuable marker of neuronal damage, with studies finding elevated NfL levels in patients with MCI and AD (Giacomucci et al., 2022). GFAP is an intermediate filament protein expressed primarily by astrocytes. GFAP has been found to be elevated in the blood of patients with AD (Kim et al., 2023). Lastly, amyloid-β42 and amyloid-β40 are peptides derived from amyloid precursor protein. Amyloid-β42 is more aggregation-prone and more strongly linked to plaque formation, whereas amyloid-β40 is more abundant in plasma and is found in only a subset of plaques. Accordingly, the plasma amyloid-β42/40 ratio is used as a surrogate marker of amyloid burden, with lower values generally reflecting greater plaque accumulation (Pérez-Grijalba et al., 2019).

B.4. Hyperparameter optimization

Grid search (GridSearchCV) with fivefold cross-validation was implemented to determine the optimal values of hyperparameters that maximized the training set performance. Hyperparameters for RSF were number of trees (n_estimators), minimum number of samples to split a node (min_samples_split), minimum samples in a leaf node (min_samples_leaf), maximum depth (max_depth), and number of features to consider when looking for the best split (max_features). For CoxPH, grid search was conducted for the optimal regularization parameter for ridge regression penalty (alpha).

B.5. Evaluation Metrics

We assessed how well the model distinguishes subjects who experience an event across time points with a time-dependent AUC (Uno et al., 2007; Hung & Chiang, 2010; Lambert & Chevret, 2016). In addition, we calculated the concordance index (C-index) with inverse probability of censoring weighted (IPCW), an alternative to Harrell's C-index that provides a less biased estimate, especially with high censoring. The C-index is a generalization of the ROC-AUC that can assess a model's ability to rank survival times based on individual risk scores (Harrell et al., 1984; Uno et al., 2011). Additionally, the brier score (BS) was calculated to assess model calibration, with lower values indicating better calibration (Graf et al., 1999). Finally, the integrated brier score (IBS) was used to evaluate the model's performance across all available time points.