Received: 31 May 2024 | Revised: 30 August 2024 | Accepted: 8 September 2024

DOI: 10.1002/alz.14293

Alzheimer's & Dementia®
THE JOURNAL OF THE ALZHEIMER'S ASSOCIATION

RESEARCH ARTICLE

Peripheral risk factors and their role in biomarker-based screening for dementia in the community

Yi-Han Hu[1] | Osorio Meirelles[1] | Eric J. Shiroma[1] | Claudia L. Satizabal[2] | Sudha Seshadri[3] | Russell P. Tracy[4] | Vilmundur Gudnason[5] | Lenore J. Launer[1]

[1]Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Baltimore, Maryland, USA

[2]Department of Population Health Sciences and Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases, University of Texas Health Science Center at San Antonio, San Antonio, Texas, USA

[3]Department of Neurology and Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases, University of Texas Health Science Center at San Antonio, San Antonio, Texas, USA

[4]Departments of Pathology & Laboratory Medicine, and Biochemistry, Larner College of Medicine, University of Vermont, Burlington, Vermont, USA

[5]Faculty of Medicine, University of Iceland, Reykjavík, Iceland

Correspondence
Lenore J. Launer, Laboratory of Epidemiology and Population Sciences, National Institute on Aging, 251 Bayview Blvd, Baltimore, MD 21224, USA.
Email: launerl@nia.nih.gov

Funding information
Icelandic Heart Association; Intramural Research Program at the National Institute on Aging, Grant/Award Numbers: N01-AG-12100, HHSN271201200022C; the Althingi (the Icelandic Parliament) and the Icelandic Centre for Research (RANNIS), Grant/Award Number: 141101–051; National Institute of Neurological Disorders and Stroke under the MarkVCID consortium, Grant/Award Numbers: UH3 NS100605, UF1 NS125513

Abstract

INTRODUCTION: Peripheral risk factors (PRFs) may correlate with dementia plasma biomarkers, potentially reflecting peripheral rather than brain health. This study explores the associations between PRFs and plasma biomarkers glial fibrillary acidic protein (GFAP), neurofilament light chain (NfL), and total-tau, and their role in predicting future dementia.

METHODS: Data from the Age, Gene/Environment Susceptibility–Reykjavik Study (2002–2015) included 4353 participants mean age of 76.6 years. A subsample of 910 participants tested their association with PRFs and plasma biomarkers' predictive performance. Sociodemographic, clinical, laboratory, sensory, and lifestyle variables ($n = 305$) were grouped into 34 clusters.

RESULTS: Besides age and estimated glomerular filtration rate (eGFR), significant associations were found between plasma biomarkers and clusters related to hemoglobin, red blood cell distribution, and inflammation. Incorporating these clusters into predictive models enhanced precision and sensitivity, though overall prediction improvement was modest (area under the precision-recall curve: GFAP 0.17 to 0.34, NfL 0.20 to 0.38).

DISCUSSION: PRFs are significantly associated with dementia plasma biomarkers; Considering these factors may enhance the predictive accuracy of dementia biomarkers.

KEYWORDS
cluster, incident dementia, peripheral health factors, plasma biomarker, predictive modeling

Highlights

- Machine learning identifies key peripheral factors influencing neurodegenerative biomarkers.
- Hemoglobin and red blood cell distribution cluster associates significantly with biomarker levels.
- Incorporating diverse peripheral factors modestly enhances incident dementia prediction accuracy in community settings.

1 | BACKGROUND

Alzheimer's disease and other dementias (AD/D) is a public health concern, characterized by multifactorial etiology, including but not limited to genetic and environmental factors.[1] With the aging population and the expected increase in dementia cases in the community, there's an urgent need for cost-effective diagnostic biomarkers. Current literature suggests blood-based biomarkers, such as plasma levels of different isoforms of phosphorylated tau (p-tau), glial fibrillary acidic protein (GFAP), and neurofilament light chain (NfL) may provide a signal of developing AD and dementia pathology.[1–10] Unlike cerebrospinal fluid assays or digital markers, these biomarkers can be measured through minimally invasive methods,[1] making them especially suitable for community-based settings where initial diagnosis commonly occurs.

The clinical or predictive utility of these plasma biomarkers is contingent on understanding what and how much peripheral factors influence their levels. For instance, evidence links kidney function, vital for protein clearance, with "AD" blood biomarkers levels, including p-tau 181 and p-tau 217.[8,11–13] However, the peripheral factors—external influences like cardiovascular risk factors, lifestyle choices, chronic diseases, and medications—that may affect biomarker levels and thus potentially lead to misclassification of dementia risk, are not well-understood. These factors contribute to the variability of biomarker measurements, influencing their accuracy in assessing dementia risk.[12,13]

The primary aim of this study is to explore the associations between peripheral risk factors (PRFs) and plasma biomarkers, specifically NfL and GFAP, and to assess their role in predicting future dementia. An increasing number of studies are demonstrating that NfL and GFAP are critically linked to neurodegenerative processes, underscoring their utility in predicting dementia and cognitive decline.[14–16] Peripheral health risk factors play a key role in diagnosing and managing diseases in older individuals with multiple morbidities, which often interrelate, making it hard to isolate single factors for targeted interventions. Clustering individual factors into domains[17] reduces the number of risk factors. For example, medical history could encompass chronic diseases and medications. Also, the multiple measures within a cluster could offer a range of similar factors that could be tested in different settings. It is important to base such analyses on community-based studies because ultimately, if tools are to be translated from the clinic to the community, there must be an understanding of the extent to which multimorbidity and heterogeneity affect levels of the biomarkers.[18]

To address these complex interrelations, we employ machine learning (ML), which identifies patterns in studies with multiple variables. ML can evaluate the role of various clusters of peripheral factors in a well-phenotyped, population-based prospective cohort. We also explore whether incorporating these peripheral clusters/factors improves the prediction power of these biomarkers for incident dementia in 10 years. Such analyses are essential for translating diagnostic tools from clinical settings to the community, where understanding the impact of multi-morbidity and heterogeneity on biomarker levels becomes crucial.

RESEARCH IN CONTEXT

1. **Systematic review**: The literature was reviewed using conventional sources (e.g., PubMed and Google Scholar), focusing on the role of peripheral risk factors (PRFs) in influencing plasma biomarkers glial fibrillary acidic protein (GFAP), neurofilament light protein (NfL), and total tau (t-tau). While these biomarkers are promising for dementia risk stratification, there is significant evidence that they are associated with factors like kidney function, which might reflect peripheral health states rather than brain health.
2. **Interpretation**: Our findings highlight that PRFs, particularly hemoglobin and red blood cell distribution, inflammation, and other clusters significantly associate with dementia plasma biomarkers. Beyond age, sex, education, Apolipoprotein E*4, and estimated glomerular filtration rate (eGFR), and the biomarker self, incorporating these PRFs into predictive models for future dementia modestly enhances precision and sensitivity suggesting their potential utility in improving dementia risk predictions based on plasma biomarkers.
3. **Future directions**: Future research should aim at community based and longitudinal studies to further test the combination of these plasma biomarkers for improving the prediction of dementia, particularly focusing on expanding the study to diverse populations and integrating additional specific PRF, such as hematological factors. This will potentially enhance the predictive accuracy and utility of plasma biomarkers in both clinical and community settings, refining dementia screening and early diagnosis protocols.

2 | METHODS

2.1 | Study design

As previously described,[19] the Age, Gene/Environment Susceptibility–Reykjavik Study (AGES-RS; 2002–2015) is a longitudinal population-based cohort of 5764 Icelandic men and women aged 66–96, characterized by detailed evaluations of anatomy, physiology, lifestyle, and medical history. Written informed consent was obtained and the study was ethically approved.

This present study was conducted in two phases: in the first phase, we conducted a cluster analysis of risk factors, and then in the second

phase, we examined the predictive power for incident dementia of a biomarker model with and without the risk clusters (see Figure S1).

2.1.1 | Phase 1

Of 5764 participants, 1099 diagnosed with dementia at blood collection were excluded. This sample was used to define clusters of peripheral factors of interest. We further excluded 312 participants missing ≥ 50% of their data on all peripheral factors, giving 4353 participants for clustering analysis (hereafter referred to as "Cluster sample").

2.1.2 | Phase 2

As a part of the MarkVCID consortium to identify biomarkers for vascular cognitive impairment,[20] we selected 910 individuals who were non-demented and had available plasma samples (hereafter referred to as "Biomarker sample"). The comparison of characteristics between those included and those not included in this substudy is shown in Table S1.

2.2 | Measurement of plasma biomarker levels

We examined three biomarkers (GFAP, NfL, and total tau [t-tau]), selected for their known associations with dementia and MRI outcomes in the MarkVCID contributing cohorts.[14–16] Blood samples were collected at baseline during the initial examination phase of the AGES-RS, conducted between 2002 and 2006. These samples were assayed for biomarkers between November and December 2019 at the University of Vermont (R. Tracy) using standardized protocols on the Simoa Neurology 4-Plex Kit with a Simoa HD-1 Analyzer (Quanterix Corporation) to ensure quality control and consistency.[21]

2.3 | Incident dementia

Dementia was ascertained based on a three-step consensus adjudication following the *Diagnostic and Statistical Manual of Mental Disorders* (Fourth Edition, DSM-IV) criteria,[22] as previously described.[19,23] This involved initial screenings using cognitive assessments, followed by detailed evaluations for those screening positive, including proxy interviews and neurological examinations. The final diagnosis incorporated input from a consensus conference involving multiple specialists. Incident dementia cases were identified among participants following the baseline exam, up to the final follow-up in December 2015, with an average follow-up period of 10 years. These cases were ascertained either during the follow-up AGES-2 exam (2007–2011) or from subsequent nursing home records, medical records, or death certificates (2004–2015).[24]

2.4 | Statistical analysis

Our statistical analysis aimed to (1) identify clusters of peripheral factors in the Cluster sample for investigation with plasma biomarkers; (2) examine the associations of these clustered peripheral factors to the biomarkers in the Biomarker sample; and (3) assess if incorporating these clusters enhances the performance of the biomarker's power to predict incident dementia.

To reach these objectives, we built a five-step pipeline (Figure S2). Data preprocessing, clustering, and principal component analyses (PCA) to define the clusters were based on the Cluster sample. The clusters were then associated with plasma biomarker levels; and the significant clusters were added to a predictive biomarker model for incident dementia, described in detail below. R studio (version 4.4) was used for data preprocessing and clustering. PCA, cluster-biomarker association, and dementia prediction were performed in Python (version 3.7) with the Scikit-learn library (version 0.22.2).[25]

2.5 | Data preprocessing

For initial processing, we selected 424 peripheral factors from multiple domains that have been linked with age-related conditions.[19] These domains included sociodemographic, anthropometric, clinical chemistry (urine and blood), hearing evaluations (audiometry and questions), eye exams (retinal scans and vision testing), cardio-metabolic (biomarkers, imaging, and questionnaires), musculoskeletal (bone and muscle), and lifestyle (questionnaires) factors, as well as disease (questionnaires and medical records) and medication (vial presentation) information. The measurement of these PRFs in this cohort has been thoroughly described in the original AGES-RS publication.[19] A detailed list of these preselected variables is available from the corresponding author upon reasonable request. Cognitive measurements and MRI data were excluded as they provide most of the information used in the dementia diagnosis. All variables used in this study, except for the outcomes of interest, were collected during the AGES-RS baseline exam.

After we removed sex-specific variables (i.e., history of prostate disease drugs and age of menarche), those with low variance (< 0.001), or with >30% missing values, and retaining only one variable from variables set with high correlation ($|r| \geq 0.95$), we had 306 variables to impute missing values (Table S2) using Multiple Imputation by Chained Equation (MICE) in R. Skewed variables, including those with zeros, were adjusted by adding 1 before log-transformation and imputation.

2.6 | Clustering algorithm

We employed the agglomerative hierarchical clustering algorithm[26] to cluster all peripheral factors (except for age, sex, education, apolipoprotein E (APOE) e4 status, and estimated glomerular filtration rate [eGFR]) with moderate-to-high similarity based on their correlation-based distances. The Silhouette coefficient estimates intra

and intercorrelations among the variables, which guides the number of clusters to retain by assessing how similar each object is to its own cluster compared to other clusters;[27,28] we set a minimum of three variables for each cluster.

2.7 | Principal component analysis

We conducted separate PCA of the variables in each cluster.[29] In general, the first principal component (PC) explained at least 30% of the total variance within the cluster (Table S3). Additional details regarding the variables included in each cluster and their vector loadings on the first PC are included in Table S2.

2.8 | Identification of clusters associated with plasma biomarker levels

In the Biomarker sample, our analysis aimed to further reduce the clusters to those who had a good model fit when associated with each plasma biomarker. To do this, we employed three analytical methods: Univariate Linear Regression (ULR), Multiple Linear Regression (MLR), and Least Absolute Shrinkage and Selection Operator (LASSO) regression. For MLR, we developed six predictor sets, starting with age and progressively including sex, eGFR, ApoE4 status, education, and clusters' first PC scores, to evaluate model fit using the adjusted r-squared (higher is better) and normalized root mean squared error (NRMSE, lower is better). The beta coefficients of the all-inclusive model were visually represented using heatmaps. Our final selection of clusters was based on statistical significance in the MLR model, which required fewer variables and had the best balance of model fit and parsimony compared to LASSO regression. Additionally, because LASSO applies regularization to the coefficients, traditional *p*-values for hypothesis testing are not calculated. Detailed descriptions of the selected models, predictor sets, and the process of cluster selection are available in the Supplemental Methods.

2.9 | Evaluation of the contributions of the clusters to the predictive power of a biomarker model to predict incident dementia

We performed logistic regression to predict incident dementia with each plasma biomarker. Per biomarker, we tested models based on four predictor sets: (1) biomarker only; (2) plus age and sex; (3) a basic model including age, sex, education, eGFR, APOE e4 carriership; and (4) an expanded model incorporating the first PC scores of clusters significantly associated ($p < 0.05$) with the biomarker in the MLR model. The dementia prediction models were evaluated with five metrics: the average beta coefficients of plasma biomarkers, area under the precision-recall curve (AUPRC), precision, sensitivity (recall), and the F1 score. Sensitivity is particularly important in our study due to the low prevalence of incident dementia (6.7%). It measures the model's ability to correctly identify all actual cases of dementia, which is critical in clinical settings to ensure that no potential cases are missed. Precision is equally crucial as it aids in minimizing false positives in clinical diagnoses, ensuring that resources are appropriately allocated to those in need. Thus, we chose AUPRC as the primary metric because it integrates precision and sensitivity at different probability thresholds, providing a comprehensive view of model performance. This measure is more suitable for our data, where positive cases are relatively rare, unlike the area under the receiver operating characteristic (AUROC) curve: AUROC can be misleading in such contexts, often overestimating a model's performance when positive cases are rare.[30] ROC curves and AUROC values are provided in Figure S3 and should be interpreted with caution. The F1 score, harmonizing the mean of precision and sensitivity, is calculated at a fixed probability threshold of 0.5. It provides a balanced measure of the model's capabilities. A higher F1 score reflects a model that effectively identifies true dementia cases without overpredicting dementia where it does not exist. To enhance the robustness and reliability of our findings, we employed bootstrap resampling, executing this technique 5000 times to ensure comprehensive estimation and validation of our results. As t-tau was not associated with incident dementia, we provide the results in Figures S4, S5, and Tables S4 and S5.

2.10 | Sensitivity analysis

We conducted a sex-stratified analysis to investigate sex-specific variations in the associations between peripheral factors and biomarker levels. Briefly, we reran the ULR, MLR, and LASSO models with the predictor of age, eGFR, ApoE4, education, and each cluster's first PC scores for male and female subgroups separately (Figures S6 and S7).

3 | RESULTS

3.1 | Participant characteristics

Briefly, in the Cluster sample, the average age of participants was 75.7 years, 45.8% were male, and 22% had a primary school education. The average eGFR level across participants was 80.4, and 24.3% carried the APOE e4 allele. Among them, 93.3% did not develop dementia, with an average age of 75.3 years. In comparison, the 6.7% diagnosed with incident dementia were on average older, at 81.2 years.

In the Biomarker sample ($n = 910$), the mean levels of GFAP, NfL, and t-tau were 180.6, 24.3, and 2.8 pg/mL, respectively. Compared to the non-demented group the dementia group had higher levels of GFAP and Nfl (Figure 1A).

3.2 | Dimension reduction

The reduction of 306 variables into 34 clusters across nine health and lifestyle domains showed variance within clusters explained by the

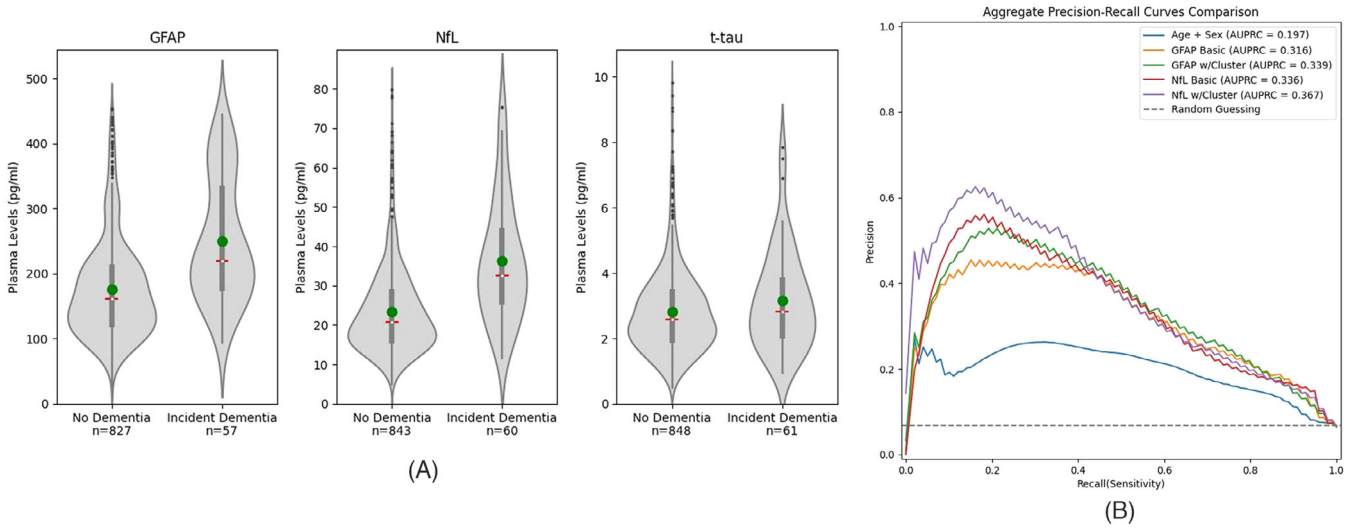

FIGURE 1 Plasma biomarkers and prediction performance in incident dementia. GFAP, glial fibrillary acidic protein; NfL, neurofilament light; t-tau, total tau; AUPRC, area under precision-recall curve. (A) presents a box-and-whisker plot illustrating the distribution of baseline plasma biomarkers by the dementia status at the follow-up. Each box represents the interquartile range (IQR), with the bottom and top edges indicating the first quartile (Q1) and third quartile (Q3), respectively. The red line within each box is the median value, while the green dot represents the mean. Whiskers extend to the lowest and highest values within 1.5 times the IQR from the quartiles. Dots located beyond the whiskers are those exceeding 1.5 times the IQR. (B) presents precision-recall curves for incident dementia prediction, with each curve's points corresponding to different thresholds. Each curve represents a different predictor set. The Basic model covers age, sex, education, APOE e4 status, and eGFR. The Clusters model incorporates the first principal component scores of associated clusters based on the MLR results. For GFAP, the prediction model includes four clusters: Metabolic Risk, Inflammation & WBC Distribution, Hemoglobin & RBC Distribution, and L1L2 Vertebral Bone Density. For NfL, five clusters are included: Hemoglobin & RBC Distribution, Physical Limitations, Thigh Muscle & Strength, Femur Cortical Structure and Density, and Coronary Calcification. Random Guessing indicates a baseline precision equal to the proportion of incident dementia (6.7%) in the Biomarker sub-sample. Precision-recall curves for total tau are provided in Figure S4.

first PC ranging from 30% to nearly 100%. Clusters varied significantly in their correlation with age, with audiometry showing a strong positive correlation, and some blood/urine test and medication clusters showing minimal correlation. (Table S3).

3.3 | Identification of clusters associated with plasma biomarker levels

Incorporating the first PC of clusters into basic models improved NRMSE and adjusted r-squared values for both NfL and GFAP. Similar improvements were observed with LASSO regression, indicating a consistent boost in model fit (Table S6).

3.4 | Factors associated with plasma NfL levels

Plasma NfL levels displayed a positive association with age and a negative association with eGFR (see the right set of columns in Figure 2). Among the clustered peripheral factors, C-Physical Limitations and C-Coronary Calcification demonstrated positive associations, while C-Thigh Muscle & Strength, C-Femur Cortical Structure & Density, and C-Hemoglobin & Red Blood Cell (RBC) Distribution exhibited negative associations.

In sex-stratified results (Figures S6 and S7), distinct patterns emerged. C-Physical Limitation (positive) and C-Inflammation & White blood cell (WBC) Distribution (positive), were only significant in females; in males, there were unique associations with C-Coronary Calcification and C-Urine Protein & Cell positively correlated, and C-Migraine negatively correlated in MLR.

3.5 | Factors associated with plasma GFAP levels

Six factors and clusters demonstrated significant associations with plasma GFAP across all three models (Figure 2). Age, APOE e4, and eGFR have a consistently positive relationship with GFAP levels. Among the clustered peripheral factors, C-L1L2 Vertebral Bone Density, C-Hemoglobin & RBC Distribution, and C-Metabolic Risk were negatively associated with GFAP levels, albeit with small effect sizes, across all models.

There were some differences by sex in which clusters were associated with the biomarkers (Figures S6 and S7). Two bone-related clusters [C-Hip & Femur Bone Density and C-L1L2 Vertebral Bone Density], C-Hemoglobin & RBC Distribution, and C-Metabolic Risk retained significance in the female subgroup; for the males, the C-Inflammation & WBC Distribution and C-Blood Pressure clusters emerged as significant in MLR.

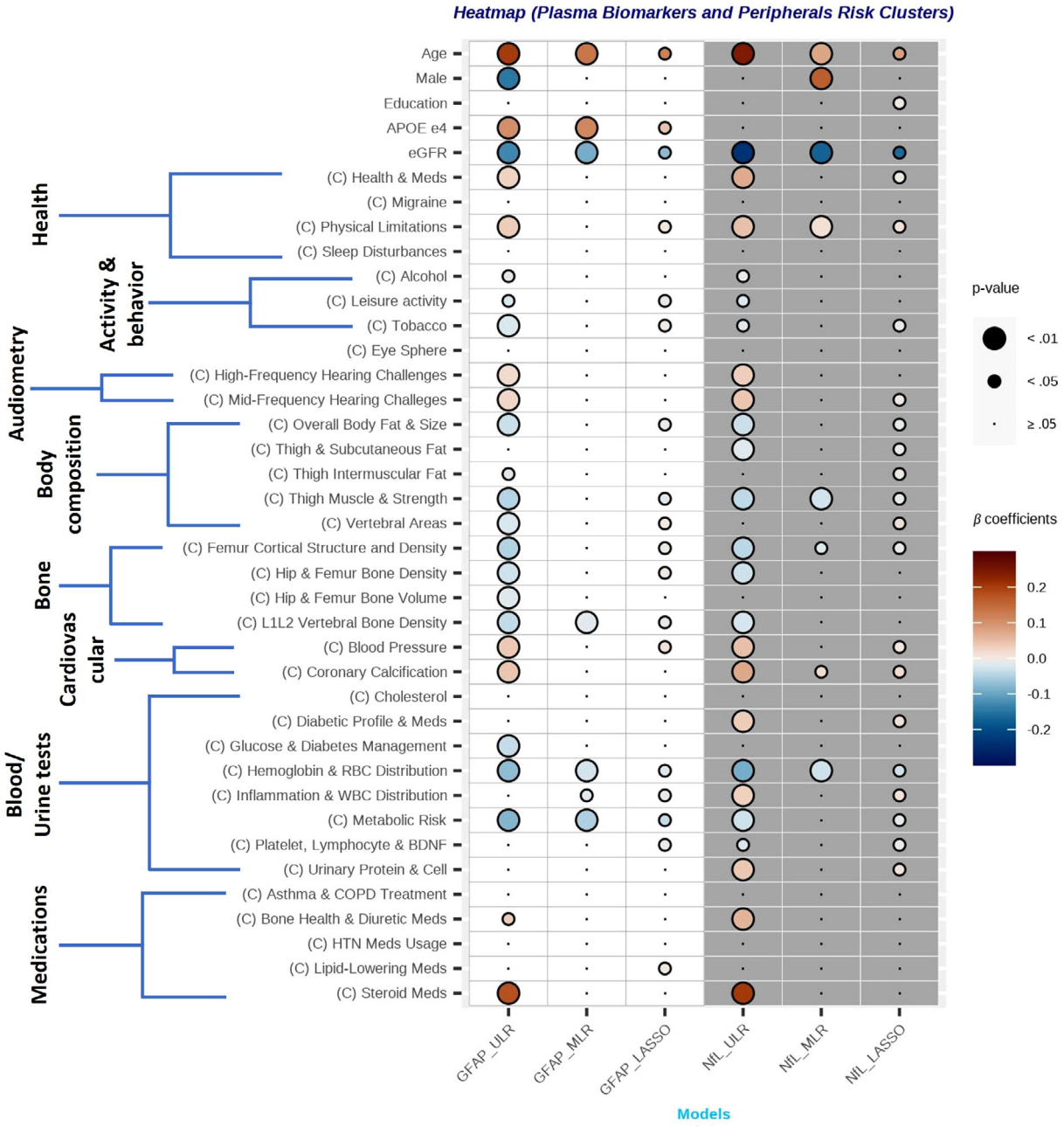

FIGURE 2 Peripheral factors/clusters associations with log-transformed GFAP and NfL in the AGES-RS cohort. GFAP, glial fibrillary acidic protein; NfL, neurofilament light; eGFR, estimated glomerular filtration rate; WBC, white blood cell; RBC, red blood cell; BDNF, brain-derived neurotrophic factor; COPD, chronic obstructive pulmonary disease; HTN, Hypertension; ULR, Univariate linear regression; MLR, Multiple linear regression; LASSO, least absolute shrinkage and selection operator regression. Total tau results are provided in Figure S5. The heatmap displays two sets of columns: the initial three pertain to the GFAP results ($n_{GFAP} = 884$) and the last trio corresponds to NfL ($n_{NfL} = 903$) results. On the Y-axis, labels beginning with '(C)' represent clusters of peripheral factors. The color gradients provide the strength and direction of these associations. The size of the dot indicates the *p*-value.

3.6 | Predicting incident dementia with plasma biomarkers and associated risk clusters

NfL models outperformed GFAP in dementia prediction, with or without cluster inclusion (Figure 1B and Table 1). The basic NfL model yielded an AUPRC of 0.34 and an F1 score of 0.13, while GFAP's basic model had an AUPRC of 0.32 and an F1 score of 0.06. Adding clusters like C-Hemoglobin & RBC Distribution to NfL improved its AUPRC to 0.37 and F1 to 0.26. GFAP had modest gains (AUPRC = 0.34, F1 = 0.09) with biomarker-associated clusters. NfL's effect size decreased from a $Beta_{basic}$ of 0.68 [0.68, 0.69] to 0.61 [0.61, 0.62] with clusters, suggesting that the effect of NfL was partially confounded or mediated by the added clusters. Despite the additions, NfL significantly predicts dementia with an effect size larger than any related clusters

TABLE 1 Evaluating logistic regression model performance to predict incident dementia with 5000 bootstraps.[a]

	Incident dementia				
Predictor set[a]	**B(95% CI)[d]**	**AUPRC[e]**	**Sensitivity (Recall)**	**Precision**	**F1**
Age + Sex only (n = 910; 6.7% incident dementia)	–	0.197 (0.196, 0.197)	0.002 (0.002, 0.002)	0.030 (0.027, 0.033)	0.003 (0.003, 0.004)
GFAP + (n = 884; 6.4% incident dementia)					
Biomarker only	0.935 (0.935, 0.935)	0.173 (0.172, 0.173)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
Age + Sex	0.600 (0.598, 0.602)	0.228 (0.227, 0.228)	0.009 (0.008, 0.009)	0.114 (0.109, 0.118)	0.016 (0.015, 0.016)
Basic model[b]	0.585 (0.582, 0.588)	0.316 (0.315, 0.317)	0.035 (0.034, 0.036)	0.236 (0.232, 0.240)	0.061 (0.059, 0.062)
Basic model[b] + 4 associated clusters[c]	0.637 (0.635, 0.640)	0.339 (0.339, 0.340)	0.051 (0.050, 0.051)	0.304 (0.300, 0.308)	0.086 (0.085, 0.088)
NfL + (n = 903; 6.6% incident dementia)					
Biomarker only	0.982 (0.982, 0.984)	0.201 (0.200, 0.201)	0.001 (0.001, 0.001)	0.014 (0.012, 0.017)	0.002 (0.002, 0.002)
Age + Sex	0.673 (0.671, 0.675)	0.243 (0.242, 0.244)	0.021 (0.020, 0.021)	0.256 (0.250, 0.262)	0.037 (0.037, 0.038)
Basic model[b]	0.682 (0.680, 0.685)	0.336 (0.336, 0.337)	0.077 (0.076, 0.078)	0.413 (0.409, 0.417)	0.130 (0.128, 0.131)
Basic model[b] + 5 associated clusters[c]	0.614 (0.611, 0.617)	0.367 (0.367, 0.368)	0.169 (0.168, 0.170)	0.614 (0.612, 0.616)	0.264 (0.262, 0.265)

Note: The Age, Gene/Environment Susceptibility–Reykjavik Study (AGES-RS), 2002–2015.
Abbreviations: AUPRC, area under precision-recall curve; GFAP, glial fibrillary acidic protein; NfL, neurofilament light.
[a]The 95% confidence interval for beta coefficients and each metric were calculated across the 5000 bootstraps.
[b]Basic model includes age, sex, education, APOE e4 status (yes/no), and eGFR.
[c]The first principal component scores of associated clusters based on the MLR between clusters and biomarkers. For GFAP, four clusters were included in the prediction model: Metabolic Risk, Inflammation & WBC Distribution, Hemoglobin & RBC Distribution, and L1L2 Vertebral Bone Density. For NfL, five clusters, including Hemoglobin & RBC Distribution, Physical Limitations, Thigh Muscle & Strength, Femur Cortical Structure and Density, and Coronary Calcification, were included.
[d]Average beta coefficients of plasma biomarkers in logistic regression model predicting incident dementia across 5000 bootstrapped results are reported. Results are per unit increase in the standardized natural log of plasma biomarkers.
[e]AUPRC is derived by integrating the area beneath the curve of precision against recall (sensitivity) across various thresholds. It is considered more appropriate than the area under the receiver operating characteristic (AUROC) in scenarios where the prevalence of positive cases is low.

(Table 2). GFAP showed improved predictive power with clusters ($Beta_{basic}$ = 0.59 [0.58, 0.59] to 0.64 [0.64, 0.64]). While t-tau displayed prediction performance metrics similar to GFAP, its beta estimate was relatively small (Tables S4 and S5).

4 | DISCUSSION

In a population-based cohort, we explored how peripheral factors affect neurodegenerative biomarker levels (GFAP, NfL, t-tau) and their predictive capacity for dementia over 10 years. Age, eGFR, and C-Hemoglobin & RBC distribution significantly influenced all three biomarkers, with GFAP and NfL levels positively associated with age but negatively with eGFR and C-Hemoglobin & RBC. Integrating these identified peripheral clusters into predictive models showcased the potential to enhance performance metrics, especially for NfL when predicting incident dementia. This finding has implications for interpreting levels of neurodegenerative markers as a reflection only of brain health. The biomarker levels may also partly reflect levels of risk factors associated with dementia.

Our study also highlights other challenges in utilizing these biomarkers for dementia prediction in a community setting. Despite incorporating age, sex, education, and APOE e4 status to minimize confounding, the predictive models had limited effectiveness. The addition of specific biomarker-associated clusters leads to some improvement, but the overall performance remained modest. Specifically, in the GFAP model with all four associated clusters, only about five out of every 100 incident dementia cases were correctly identified. The NfL model performed better, predicting around 17% of cases with a 61% precision rate, which means when the model predicts a participant as an incident dementia case, it is correct only about 61% of the time.

In contrast to longitudinal studies like Beyer et al., which reported AUROC of 0.74 for GFAP and 0.68 for NfL alongside age and sex in a nested case-control sample from a community-based setting (68 AD cases and 240 controls),[9] our results reflect the real-world challenges in identifying dementia cases in a general population. Another case-control study using the Alzheimer's Disease Cardiff Cohort, with 1439 AD cases and 508 controls, reported AUROCs of 0.64 for GFAP and 0.63 for NfL, adjusted for age and sex.[10] These studies present a different context compared to our approach, which captures a wide variation in cognitive functions and health. This variation leads to significant overlap in biomarker distributions, as shown in Figure 1A, suggesting biomarker studies in communities will vary from those in curated case-control studies.

Our models, particularly the NfL model with clusters, show modest precision in identifying dementia within a diverse population. However,

TABLE 2 Logistic regression models predicting incident dementia with 5000 bootstrapped results.[a]

	GFAP		NfL	
Parameter	**ORs**	**95% CI**	**ORs**	**95% CI**
Plasma biomarker levels (GFAP/NfL)[b]	1.891	(1.886, 1.896)	1.848	(1.842, 1.854)
Age	2.369	(2.362, 2.375)	1.774	(1.769, 1.779)
Male	1.543	(1.535, 1.551)	2.389	(2.377, 2.401)
Education	0.475	(0.474, 0.477)	0.533	(0.531, 0.534)
APOE e4	2.554	(2.543, 2.564)	2.669	(2.657, 2.680)
eGFR	0.908	(0.906, 0.910)	0.965	(0.963, 0.967)
(C) Hemoglobin & RBC Distribution	0.960	(0.959, 0.961)	1.138	(1.137, 1.14)
(C) Inflammation & WBC Distribution	1.150	(1.148, 1.151)	N/A	N/A
(C) Metabolic Risk	1.170	(1.167, 1.172)	N/A	N/A
(C) L1L2 Vertebral Bone Density	0.979	(0.978, 0.979)	N/A	N/A
(C) Thigh Muscle & Strength	N/A	N/A	0.919	(0.917, 0.92)
(C) Cardiovascular Calcification	N/A	N/A	0.992	(0.991, 0.994)
(C) Femur Cortical Structure and Density	N/A	N/A	0.957	(0.955, 0.959)
(C) Physical Limitations	N/A	N/A	1.134	(1.133, 1.135)

Note: The Age, Gene/Environment Susceptibility–Reykjavik Study (AGES-RS), 2002–2015.

Abbreviations: CI, confidence interval; eGFR, estimated glomerular filtration rate; GFAP, glial fibrillary acidic protein; NfL, neurofilament light; ORs, odds ratios; RBC, red blood cell; WBC, white blood cell.

[a]The first principal component scores of associated clusters are based on the MLR between clusters and biomarkers. For GFAP, four clusters were included in the prediction model: Metabolic Risk, Inflammation & WBC Distribution, Hemoglobin & RBC Distribution, and L1L2 Vertebral Bone Density. For NfL, five clusters, including Hemoglobin & RBC Distribution, Physical Limitations, Thigh Muscle & Strength, Femur Cortical Structure and Density, and Coronary Calcification, were included.

[b]Average ORs across 5000 bootstrapped results are reported. For biomarkers' results, per unit increase in the standardized natural log of plasma biomarkers.

the observed improvement with the inclusion of additional clusters associated with biomarker levels suggests avenues for enhancing their utility in diverse community settings. Future research should aim to identify and integrate more specific and sensitive markers or factors that may be affecting biomarker levels, thus, potentially enhancing predictive accuracy and aiding in interpreting biomarker levels about brain health.

We confirmed previous findings of biomarkers' positive association with age and negative with eGFR. Interestingly, including the ApoE allele did not enhance dementia prediction. Our data also indicated sex differences in how clusters associate with biomarkers, pointing to the need for further research on sex-specific clusters and models for more accurate dementia prediction.

Of interest, we found several hematological factors associated with plasma biomarkers. For example, high C-Hemoglobin (Hb) and RBC were associated with lower GFAP and NFL levels. These factors are crucial for oxygen transport to tissues including the brain. Additionally, stroke and small vessel disease, often associated with alterations in hematological characteristics[31,32] may also influence NfL's association with dementia risk.[16] We also found a significantly negative association between C-Inflammation & WBC Distribution and GFAP, which is consistent with GFAP as an indicator of neuroinflammation.[33]

Our study has several strengths. A key strength is our comprehensive dataset covering health metrics, behavioral indicators, clinical exams, and medication usage, enhancing our understanding of factors affecting neurodegenerative biomarkers. The employment of ML for dimension reduction is another strength, which allows the discovery of complex patterns from a wide array of potential factors and hypotheses beyond traditional analysis. Furthermore, our findings can provide a basis for other studies with fewer variables to explore individual variables within identified clusters to potentially replicate or extend our results.

While our study provides valuable findings, it is important to acknowledge its limitations. Our cohort predominantly consists of individuals of European descent, potentially limiting the generalizability of our findings to other populations. Additionally, the relatively small size of the Biomarker sample may affect the robustness of our conclusions. Also, the small sample size prevents reliable stratification by cognitive status (e.g., mild cognitive impairment vs. cognitively unimpaired), limiting the granularity of our subgroup analyses. Furthermore, p-tau was not available in this study, as our focus was on general dementia; therefore, its potential contributions to dementia prediction could not be evaluated.

5 | CONCLUSION

Our study offers a comprehensive exploration of how peripheral factors influence neurodegenerative biomarkers in a population-based cohort, highlighting Hemoglobin & RBC Distribution as key

influences. These findings underscore the importance of understanding the broader physiological context when interpreting biomarker levels. While our results suggest that integrating diverse peripheral factors can aid in understanding biomarker variations, they also reveal the limitations of these biomarkers in predicting dementia within a general population.

ACKNOWLEDGMENTS

The AGES–Reykjavik Study was supported by the Icelandic Heart Association, the Intramural Research Program at the National Institute on Aging (grant numbers N01-AG-12100 and HHSN271201200022C), the Althingi (the Icelandic Parliament), and the Icelandic Centre for Research (RANNIS) (grant number 141101–051). The measurement of fluid markers was funded by the National Institute of Neurological Disorders and Stroke under the MarkVCID consortium (UH3 NS100605, UF1 NS125513).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest related to this article. Author disclosures are available in the Supporting Information.

CONSENT STATEMENT

Written informed consent was obtained from all participants.

ORCID

Yi-Han Hu https://orcid.org/0000-0003-0882-6563

REFERENCES

1. Hampel H, Hu Y, Cummings J, et al. Blood-based biomarkers for Alzheimer's disease: current state and future use in a transformed global healthcare landscape. *Neuron*. 2023;111(18):2781-2799.
2. Moscoso A, Grothe MJ, Ashton NJ, et al. Time course of phosphorylated-tau181 in blood across the Alzheimer's disease spectrum. *Brain*. 2021;144(1):325-339.
3. Karikari TK, Pascoal TA, Ashton NJ, et al. Blood phosphorylated tau 181 as a biomarker for Alzheimer's disease: a diagnostic performance and prediction modelling study using data from four prospective cohorts. *Lancet Neurol*. 2020;19(5):422-433.
4. Blennow K, Zetterberg H. Biomarkers for Alzheimer's disease: current status and prospects for the future. *J Intern Med*. 2018;284(6):643-663.
5. Palmqvist S, Janelidze S, Quiroz YT, et al. Discriminative accuracy of plasma phospho-tau217 for Alzheimer disease vs. other neurodegenerative disorders. *JAMA*. 2020;324(8):772-781.
6. Giacomucci G, Mazzeo S, Bagnoli S, et al. Plasma neurofilament light chain as a biomarker of Alzheimer's disease in subjective cognitive decline and mild cognitive impairment. *J Neurol*. 2022;269(8):4270-4280.
7. Khalil M, Teunissen CE, Otto M, et al. Neurofilaments as biomarkers in neurological disorders. *Nat Rev Neurol*. 2018;14(10):577-589.
8. Kim KY, Shin KY, Chang KA. GFAP as a potential biomarker for Alzheimer's disease: a systematic review and meta-analysis. *Cells*. 2023;12(9):1309.
9. Beyer L, Stocker H, Rujescu D, et al. Amyloid-beta misfolding and GFAP predict risk of clinical Alzheimer's disease diagnosis within 17 years. *Alzheimers Dement*. 2023;19(3):1020-1028.
10. Stevenson-Hoare J, Heslegrave A, Leonenko G, et al. Plasma biomarkers and genetics in the diagnosis and prediction of Alzheimer's disease. *Brain*. 2023;146(2):690-699.
11. Stocker H, Beyer L, Trares K, et al. Association of kidney function with development of Alzheimer disease and other dementias and dementia-related blood biomarkers. *JAMA Netw Open*. 2023;6(1):e2252387.
12. Pichet Binette A, Janelidze S, Cullen N, et al. Confounding factors of Alzheimer's disease plasma biomarkers and their impact on clinical performance. *Alzheimers Dement*. 2023;19(4):1403-1414.
13. Syrjanen JA, Campbell MR, Algeciras-Schimnich A, et al. Associations of amyloid and neurodegeneration plasma biomarkers with comorbidities. *Alzheimers Dement*. 2022;18(6):1128-1140.
14. Cronjé HT, Liu X, Odden MC, et al. Serum NfL and GFAP are associated with incident dementia and dementia mortality in older adults: the cardiovascular health study. *Alzheimers Dement*. 2023;19(12):5672-5680.
15. Sapkota S, Erickson K, Harvey D, et al. Plasma biomarkers predict cognitive trajectories in an ethnoracially and clinically diverse cohort: mediation with hippocampal volume. *Alzheimers Dement*. 2022;14(1):e12349.
16. van Gennip ACE, Satizabal CL, Tracy RP, et al. Associations of plasma NfL, GFAP, and t-tau with cerebral small vessel disease and incident dementia: longitudinal data of the AGES–Reykjavik Study. *GeroScience*. 2024;46(1):505-516.
17. Kim J, Park Y, Park S, et al. Prediction of tau accumulation in prodromal Alzheimer's disease using an ensemble machine learning approach. *Sci Rep*. 2021;11(1):5706.
18. Mielke MM, Dage JL, Frank RD, et al. Performance of plasma phosphorylated tau 181 and 217 in the community. *Nat Med*. 2022;28(7):1398-1405.
19. Harris TB, Launer LJ, Eiriksdottir G, et al. Age, gene/environment susceptibility–Reykjavik Study: multidisciplinary applied phenomics. *Am J Epidemiol*. 2007;165(9):1076-1087.
20. Wilcock D, Jicha G, Blacker D, et al. MarkVCID cerebral small vessel consortium: I. enrollment, clinical, fluid protocols. *Alzheimers Dement*. 2021;17(4):704-715.
21. O'Bryant SE, Gupta V, Henriksen K, et al. Guidelines for the standardization of preanalytic variables for blood-based biomarker studies in Alzheimer's disease research. *Alzheimers Dement*. 2015;11(5):549-560.
22. American Psychiatric Association A, Association AP. *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV*. vol 4. American psychiatric association;1994.
23. Moonen JE, Sabayan B, Sigurdsson S, et al. Contributions of cerebral blood flow to associations between blood pressure levels and cognition: the age, gene/environment susceptibility–Reykjavik Study. *Hypertension*. 2021;77(6):2075-2083.
24. Jørgensen L, el Kholy K, Damkjaer K, Deis A, Schroll M. " RAI"–an international system for assessment of nursing home residents. *Ugeskr Laeger*. 1997;159(43):6371-6376.
25. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Machine Learning Res*. 2011;12:2825-2830.
26. Liu F, Zhou Z, Cai M, Wen Y, Zhang J. AGNEP: an agglomerative nesting clustering algorithm for phenotypic dimension reduction in a joint analysis of multiple phenotypes. *Front Genet*. 2021;12:648831.
27. Sasirekha K, Baby P. Agglomerative hierarchical clustering algorithm-a review. *Int J Scientific Res Publications*. 2013;3:2250-3153.
28. Zhou HB, Gao JT. Automatic method for determining cluster number based on silhouette coefficient. *AMR*. 2014;951:227-230.
29. Wold S, Esbensen K, Geladi P. Principal component analysis. *Chemom Intell Lab Syst*. 1987;2(1-3):37-52.
30. Ozenne B, Subtil F, Maucort-Boulch D. The precision–recall curve overcame the optimism of the receiver operating characteristic curve in rare diseases. *J Clin Epidemiol*. 2015;68(8):855-859.

31. Heo J, Youk TM, Seo KD. Anemia is a risk factor for the development of ischemic stroke and post-stroke mortality. *J Clin Med*. 2021;10(12):2556.
32. Wang H, Li H, Wang Y, et al. Hematological parameters and early-onset coronary artery disease: a retrospective case–control study based on 3366 participants. *Ther Adv Chronic Dis*. 2023;14:20406223221142670.
33. Abdelhak A, Foschi M, Abu-Rumeileh S, et al. Blood GFAP as an emerging biomarker in brain and spinal cord disorders. *Nat Rev Neurol*. 2022;18(3):158-172.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Hu Y-H, Meirelles O, Shiroma EJ, et al. Peripheral risk factors and their role in biomarker-based screening for dementia in the community. *Alzheimer's Dement*. 2024;20:8556–8565. https://doi.org/10.1002/alz.14293