Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Personalised Neurodiagnostic Decision-Making for Alzheimer's Biomarker Evaluation: A Bayesian Utility Approach Using Plasma Neurofilament Light (NfL)

AFM Saiful Islam

afm.islam@ucl.ac.uk

University College London

Hossain M Pear

University of Hong Kong

Research Article

Keywords: Bayesian decision theory, Neurofilament light, Diagnostic utility, Alzheimer's disease, Personalised thresholding

Posted Date: May 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8641943/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Personalised Neurodiagnostic Decision-Making for Alzheimer's Biomarker Evaluation: A Bayesian Utility Approach Using Plasma Neurofilament Light (NfL)

AFM Saiful Islam[1] and M. Pear Hossain[2]

[1]Corresponding authour: Associate Professor, UCL Queen Square Institute of Neurology, 7 Queen Square, London WC1N 3AR, UK. Email: afm.islam@ucl.ac.uk

[2]Post doctoral research fellow, WHO Collaborating Centre for Infectious Disease Epidemiology and Control, School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong Special Administrative Region, China.

Abstract

Introduction

Plasma neurofilament light (NfL) is a blood biomarker of axonal injury and Alzheimer's

disease (AD) progression. Its interpretation is limited by age effects, overlap with other neurodegenerative conditions, and reliance on static cutoffs. We propose a Bayesian utility framework that integrates diagnostic threshold sensitivity (d_1) with decision intensity (d_2), enabling personalised, uncertainty-informed thresholding.

Methods

Log-transformed NfL was modelled with a Normal-Gamma prior and updated using synthetic data from published group statistics (controls, MCI-AD, ADD). Utility combined patient preference, clinician expertise, and clinical modifiers (MMSE, Tau, Aβ ratio, age). Expected utility was computed across thresholds using 5,000 posterior samples.

Results

Maximum expected utility rose with disease severity (ADD > MCI-AD > controls). Age, lower MMSE, clinician experience, and patient preference shifted thresholds upward, enhancing diagnostic confidence.

Discussion

This conjugate utility model embeds uncertainty and context into thresholding, supporting personalised AD diagnostics and reducing misclassification.

Keywords: Bayesian decision theory, Neurofilament light, Diagnostic utility, Alzheimer's disease, Personalised thresholding

1 Introduction

Bayesian decision theory provides a principled framework for clinical diagnostics by integrating prior knowledge, observed data, and uncertainty into a coherent decision-making process. The use of expected utility to guide optimal decisions under uncertainty lays the foundation

for probabilistic reasoning in medicine [1]. Earlier, De Groot introduced the concept of utility maximisation in statistical inference, which remains central to Bayesian clinical models today [2]. Recent efforts have focused on translating these theoretical principles into applied clinical research. Gunn-Sandell et al. offer a practical guide for implementing Bayesian analyses in clinical trials, emphasising how posterior distributions and utility functions can improve decision quality and transparency [4]. Frost et al. demonstrate the application of Bayes' theorem in diagnostic testing and trial design, showing how probabilistic reasoning enhances both sensitivity and specificity in clinical endpoints, focusing on evidence-based practice [5]. Utility functions have also been used to refine diagnostic accuracy by dynamically adjusting thresholds based on posterior evidence. Broemeling's review of Bayesian methods for medical test accuracy highlights how predictive values and ROC curves can be reinterpreted through a utility lens, improving classification performance in uncertain settings [6].

Chatzimichail and Hatjimihail developed a Bayesian software tool that estimates diagnostic confidence and predictive values while accounting for measurement uncertainty, demonstrating how utility-based reasoning can guide threshold selection in diabetes screening where this tool remains one of the few that visualise both diagnostic confidence and threshold sensitivity within a unified interface [7]. Similarly, Lancaster's Bayesian cutoff model proposes a utility-driven approach to biomarker thresholding, offering a flexible alternative to fixed clinical cutoffs [8]. Bayesian networks have further expanded the scope of utility-based diagnostics. In a scoping review of 60 studies, Polotskaya et al. report that Bayesian models consistently achieve high predictive accuracy, with average AUC values exceeding 75%, particularly when utility functions are embedded to rank diagnostic paths [9]. These models demonstrate the value of integrating probabilistic inference with decision-theoretic principles

to enhance diagnostic precision. Beyond accuracy, utility functions also support confident decision-making, though they are often implicit in clinical reasoning. Medow and Lucey argue that Bayesian models help overcome cognitive biases, such as anchoring and base-rate neglect, by formalising confidence through posterior updates, thereby improving diagnostic reliability [10]. Balci et al. extend this idea through Bayesian decision curve analysis, quantifying net clinical benefit across varying thresholds and confidence levels [11]. EditVerse's overview of Bayesian clinical decision-making emphasises the role of utility functions in capturing both uncertainty and readiness to act, especially in complex diagnostic scenarios [12]. Jaynes' philosophical treatise on probability theory reinforces the importance of utility in guiding rational action under uncertainty, providing a conceptual foundation for personalised diagnostics [13].

Plasma NfL has become a widely studied blood-based marker of neurodegeneration in Alzheimer's disease. As a structural axonal protein released after neuronal injury, its concentration in plasma reflects the burden of the disease, with higher levels associated with brain atrophy, cognitive decline, and progression from mild cognitive impairment to Alzheimer's dementia [17, 18]. Compared to cerebrospinal fluid (CSF) biomarkers, plasma NfL offers a less invasive and more scalable option for repeated assessments and population-level screening, making it attractive for early detection and stratification in clinical trials [15]. However, NfL lacks disease specificity, as elevations are also observed in conditions such as frontotemporal dementia, Lewy body dementia, and amyotrophic lateral sclerosis [19]. Its levels are further influenced by age and comorbidities, which complicates interpretation in heterogeneous patient populations [20]. These challenges highlight the need for integrative models that interpret NfL in combination with complementary biomarkers (Tau, $A\beta$ ratios) and cognitive measures. Therefore, a Bayesian utility framework proposed here aims to meet

this need by embedding NfL within a personalised decision-making structure.

2 Methods

In our own work, we introduce a novel utility function based on the conjugate utility framework developed under the Bayesian paradigm by Islam [3]. This function explicitly combines diagnostic threshold sensitivity with decision confidence, and is defined as:

$$U(d_1, d_2) = \mathbb{E}_{\mu,\lambda}\left[\exp\left(-\frac{1}{2}(d_1-\mu)^2\right)\cdot(1-\exp(-\lambda d_2))\right],$$

where a Gaussian penalty for threshold deviation is combined with an exponential gain for decision intensity. This formulation is designed to capture both diagnostic accuracy and decision confidence within a Bayesian framework.

Here, d_1 represents the diagnostic threshold, typically a log-transformed biomarker value, at which a clinical decision is considered. Lower values of d_1 reflect earlier intervention, while higher values imply more conservative action. The parameter d_2 encodes decision intensity, a composite measure that integrates behavioural factors (such as patient preference and clinician experience) with clinical indicators associated with the biomarker, thereby quantifying the urgency or confidence to act. The terms μ and λ denote the posterior mean and precision of the biomarker signal, respectively, and are estimated from observed data using Bayesian inference.

The first component of the utility function, $\exp\left(-\frac{1}{2}(d_1-\mu)^2\right)$, penalises decisions that deviate from the true signal, promoting diagnostic accuracy. The second component, $1-\exp(-\lambda d_2)$, increases with both precision and decision intensity, reflecting greater confidence in acting. By taking the expectation over posterior samples of μ and λ, the function yields a

robust measure of expected utility that accounts for uncertainty in the biomarker distribution and the contextual readiness to intervene.

While no prior study has implemented this exact functional form, its design is conceptually consistent with the conjugate utility framework for exponential families, and builds directly on the principles of Bayesian expected utility maximisation [1, 2, 3]. The incorporation of posterior uncertainty in both mean and precision, alongside behavioural and clinical weighting in decision intensity, offers a flexible and interpretable model for clinical diagnostics.

By integrating statistical inference with contextual readiness, our model advances personalised decision-making in neurodiagnostic settings. In this framework, the diagnostic threshold d_1 represents the point at which a biomarker signal—such as plasma neurofilament light (NfL), is considered actionable. Rather than applying a fixed population-level cutoff, d_1 is evaluated dynamically against the patient's posterior biomarker distribution, allowing for nuanced interpretation tailored to individual context. The decision intensity d_2 complements this by incorporating behavioural and clinical modifiers, including patient preference, clinician experience, cognitive scores, and age. Together, these parameters define a utility curve that reflects the expected benefit of acting at each threshold for a specific patient. This enables clinicians to identify the most appropriate diagnostic action, grounded in both statistical evidence and personalised context—particularly valuable in conditions such as Alzheimer's disease, where early intervention is critical and diagnostic ambiguity is common [16].

Let $x_1, x_2, \ldots, x_n$ denote samples of plasma neurofilament light (NfL) concentrations, drawn from normal distributions with known means and standard deviations, separately defined for each diagnostic group. These raw samples are log-transformed to yield $y_i =$

$\log(x_i)$, which are assumed to follow a Gaussian distribution: $y_i \sim \mathcal{N}(\mu, \sigma^2)$, where both μ and $\lambda = 1/\sigma^2$ are unknown. The likelihood function of the log-transformed data $\mathbf{y} = (y_1, \ldots, y_n)$ is given by:

$$L(\mu, \lambda \mid \mathbf{y}) = \left(\frac{\lambda^{n/2}}{(2\pi)^{n/2}} \right) \exp\left(-\frac{\lambda}{2} \sum_{i=1}^{n} (y_i - \mu)^2 \right).$$

Assuming a conjugate Normal-Gamma prior for (μ, λ), we define:

$$\mu \mid \lambda \sim \mathcal{N}(\mu_0, (1/\kappa_0 \lambda)), \quad \lambda \sim \text{Gamma}(\alpha_0, \beta_0),$$

with fixed hyperparameters: $\mu_0 = 3.0, \quad \kappa_0 = 10, \quad \alpha_0 = 2.0, \quad \beta_0 = 1.0$. Given the observed log-transformed sample $\mathbf{y}$, the posterior distribution of (μ, λ) remains a Normal-Gamma distribution [25]:

$$\mu \mid \lambda, \mathbf{y} \sim \mathcal{N}(\mu_n, (1/\kappa_n \lambda)), \quad \lambda \mid \mathbf{y} \sim \text{Gamma}(\alpha_n, \beta_n),$$

with updated parameters:

$$\kappa_n = \kappa_0 + n, \quad \mu_n = \frac{\kappa_0 \mu_0 + n\bar{y}}{\kappa_n}, \quad \alpha_n = \alpha_0 + \frac{n}{2}, \quad \beta_n = \beta_0 + \frac{1}{2} \sum_{i=1}^{n} (y_i - \bar{y})^2 + \frac{\kappa_0 n (\bar{y} - \mu_0)^2}{2\kappa_n}.$$

Here, $\bar{y}$ denotes the sample mean of the log-transformed biomarker values. This posterior distribution captures uncertainty in both the mean and precision of the biomarker signal and forms the basis for evaluating expected utility $U(d_1, d_2)$ across diagnostic thresholds.

In this study, we considered results from a published investigation in which plasma concentrations of neurofilament light (NfL) were evaluated as a potential biomarker for screening neurodegeneration in Alzheimer's disease (AD) patients. The study compared NfL levels across three diagnostic groups: patients with mild cognitive impairment due to AD (MCI-AD), patients with early-stage AD dementia (ADD), and nondemented controls. The control

group exhibited a mean NfL concentration of 12.2 pg/mL with a standard deviation of 7.1; the MCI-AD group had a mean of 18.9 pg/mL with a standard deviation of 15.9; and the ADD group showed a mean of 27.8 pg/mL with a standard deviation of 18.4 [15].

To model these data, we generated synthetic samples of plasma NfL concentrations using the reported group-level means and standard deviations for each diagnostic category. These samples were log-transformed to approximate the distribution $x = \log(\text{NfL})$, assumed to follow a Gaussian form. The transformed data were then used to estimate the posterior distribution of the mean μ and precision λ under a conjugate Normal-Gamma prior. Specifically, we simulated a small sample of size $n = 22$ to reflect the observed distribution of raw NfL values reported in [15]. Using prior parameters $\mu_0 = 3.0$, $\kappa_0 = 10$, $\alpha_0 = 2.0$, and $\beta_0 = 1.0$, the corresponding posterior updates are:

$$\mu_n = \frac{\kappa_0 \mu_0 + n\bar{x}}{\kappa_0 + n}, \quad \alpha_n = \alpha_0 + \frac{n}{2}, \quad \beta_n = \beta_0 + \frac{1}{2}\sum (x_i - \bar{x})^2 + \frac{\kappa_0 n(\bar{x} - \mu_0)^2}{2(\kappa_0 + n)}.$$

2.1 Composite Decision Intensity d_2

In addition to NfL, patients' age(years), MMSE scores, CFS Tau (pg/ml) andCSF $A\beta_{1\text{–}42}/A\beta_{1\text{–}40}$ Ratio were considered from [15] to formulate the decision intensity term d_2, presented in Table 1.

While these variables were not used in their raw form, we constructed d_2 using a standardised composite derived from them. Each variable was first z-transformed within its respective group to ensure comparability, and then aggregated into a weighted score reflecting clinical severity and decision readiness. This standardisation allowed us to apply the same functional form of d_2 across all diagnostic groups, enabling consistent evaluation of

	Controls	**MCI-AD**	**ADD**
Age (years)	52.5 ± 13.1	69.2 ± 7.6	72.3 ± 8.3
MMSE Scores	29.3 ± 0.9	26.2 ± 2.1	21.5 ± 3.2
CSF Tau (pg/mL)	204 ± 101	412 ± 203	529 ± 248
CSF $A\beta_{1\text{–}42}/A\beta_{1\text{–}40}$ Ratio	0.10 ± 0.03	0.04 ± 0.01	0.04 ± 0.01

Table 1: Clinical and cognitive parameters across three diagnostic groups: patients with mild cognitive impairment due to Alzheimer's disease (MCI-AD), patients with early-stage Alzheimer's disease dementia (ADD), and nondemented controls summarised from Lewczuk et al. (2018). Values are reported as mean ± SD.

expected utility while preserving group-specific variation in clinical profiles. The resulting formulation captures both statistical evidence and contextual urgency, supporting personalised threshold selection in neurodiagnostic decision-making. The decision intensity term d_2 was formulated as a weighted sum of behavioural and clinical components, defined as $d_2 = w_b \cdot (u_1 \cdot \text{Patient Preference} + u_2 \cdot \text{Clinician Experience}) + w_c \cdot (v_1 \cdot \text{MMSE}_{\text{score}} + v_2 \cdot \text{Tau}_{\text{score}} + v_3 \cdot \text{A}\beta_{\text{score}} + v_4 \cdot \text{Age}_{\text{score}})$; subject to the constraints $w_b + w_c = 1$, $u_1 + u_2 = 1$, and $v_1 + v_2 + v_3 + v_4 = 1$. In this study, we adopted the following weights for all utility evaluations: behavioural weights $u_1 = 0.6$, $u_2 = 0.4$; global weights $w_b = 0.4$, $w_c = 0.6$; and clinical weights $v_1 = 0.25$, $v_2 = 0.35$, $v_3 = 0.25$, $v_4 = 0.15$. This formulation ensures that both subjective preferences and objective clinical indicators contribute proportionally to the decision intensity, enabling consistent and interpretable utility modelling across diagnostic scenarios.

The weighting scheme adopted for the decision intensity term d_2 reflects a deliberate balance between behavioural judgment and clinical evidence, grounded in both diagnostic

relevance and ethical AI principles. The global weights $w_b = 0.4$ and $w_c = 0.6$ prioritise clinical indicators slightly more than behavioural input, recognising the stronger predictive value of objective biomarkers in neurodegenerative conditions while ensuring that human values and contextual judgement remain integral to the decision process. Within the behavioural domain, weights $u_1 = 0.6$ and $u_2 = 0.4$ favour patient preference over clinician experience, consistent with ethical AI frameworks that promote autonomy, transparency, and human agency in decision-making. This approach aligns with the European Commission's guidelines for trustworthy AI, which emphasise respect for human autonomy, prevention of harm, fairness, and explicability [21]. The clinical weights: $v_1 = 0.25$ for MMSE, $v_2 = 0.35$ for Tau, $v_3 = 0.25$ for Aβ, and $v_4 = 0.15$ for age; were selected based on the relative diagnostic utility of each factor. Tau pathology was given the highest weight due to its strong correlation with disease progression, while MMSE and Aβ scores contribute equally as validated cognitive and biochemical markers. Age, though a known risk factor, was weighted lowest to avoid overgeneralisation and preserve individual specificity. Together, these weights ensure that decision intensity captures both the urgency and confidence to act, grounded in a multidimensional, interpretable, and ethically sound framework for personalised utility modelling in clinical diagnostics.

2.2 Normalisation of Clinical and behavioural Inputs

To ensure comparability across heterogeneous scales, each clinical and demographic input contributing to d_2 was normalised to a bounded range, typically between 0 and 1. This included inverted linear scaling for MMSE scores—where lower cognitive performance corresponds to higher diagnostic confidence, defined as $\text{MMSE}_{\text{score}} = 1 - \frac{\text{MMSE}-20}{9}$, following conventions used in cognitive decline modelling [22]. Tau levels were scaled using stan-

dard min-max normalisation over the clinical range: $\text{Tau}_{\text{score}} = \frac{\text{Tau}-244}{542-244}$, based on observed distributions in Alzheimer's disease cohorts [15]. Aβ ratios were transformed using inverted min-max scaling to reflect greater pathology at lower values, expressed as $\text{A}\beta_{\text{score}} = 1 - \frac{\text{A}\beta - 0.06}{0.04}$, consistent with biomarker interpretation in amyloid burden studies [23]. Age was scaled linearly with clipping to capture increasing diagnostic urgency beyond midlife, using $\text{Age}_{\text{score}} = \text{clip}\left(\frac{\text{Age}-50}{40}, 0, 1\right)$, reflecting age-related risk thresholds in neurodegeneration [24]. These transformations enabled coherent integration of diverse patient-specific modifiers into the composite decision intensity, supporting personalised threshold selection in the presence of diagnostic ambiguity.

2.3 Expected utility evaluation

For each threshold value $d_1 \in [2.5, 4.0]$, representing the log-transformed plasma neurofilament light (NfL) level at which a diagnostic decision is considered, we computed the expected utility $U(d_1, d_2)$ using 5,000 posterior samples drawn from $\mu \sim \mathcal{N}(\mu_n, (\kappa_n \lambda)^{-1})$ and $\lambda \sim \text{Gamma}(\alpha_n, \beta_n^{-1})$. The parameter d_1 governs the sensitivity of the decision to the biomarker signal—lower values reflect earlier intervention, while higher values imply more conservative action. The decision intensity d_2, a composite measure of readiness to act, incorporates behavioural factors (patient preference, clinician experience) and clinical indicators (MMSE, Tau, Aβ ratio, and age).

We evaluated expected utility under four distinct scenarios. First, in a single-patient setting, d_2 was held fixed based on known inputs, while d_1 was varied to assess its influence on personalised decision-making. Second, we examined how age interacts with NfL by varying age while keeping other components of d_2 constant. Third, we explored the effect of cognitive performance by varying MMSE scores under similar constraints. Fourth, we assessed the role

of clinician experience by adjusting this component within d_2. In each case, we generated a three-dimensional utility surface to understand how these factors shape optimal diagnostic action. Comparisons were made across three diagnostic groups: patients with mild cognitive impairment due to Alzheimer's disease (MCI-AD), those with early-stage Alzheimer's dementia (ADD), and nondemented healthy controls. The x-axis represents the diagnostic threshold d_1, which corresponds to the log-transformed plasma NfL value ($d_1 = \log(\text{NfL})$). The y-axis indicates the expected utility. Each curve represents the group-specific expected utility computed using posterior samples of the biomarker mean μ and precision λ, integrated with the decision intensity term d_2.

3 Results

3.1 Maximum expected utility by diagnostic group for a single patient

In a single-patient scenario, we considered a clinical and behavioural profile defined by the following parameters: MMSE score of 24, Tau level of 500, $A\beta$ ratio of 0.05, age 72, patient preference of 0.7, and clinician experience of 0.8. These values reflect an older individual with moderate cognitive impairment who actively seeks referral and is being assessed by a highly experienced clinician. The resulting curve represents the expected utility of acting at each diagnostic threshold, tailored to this fixed patient-specific context. It illustrates how personalised decision-making can be guided by both biomarker evidence and contextual modifiers within a Bayesian framework.

List of figures

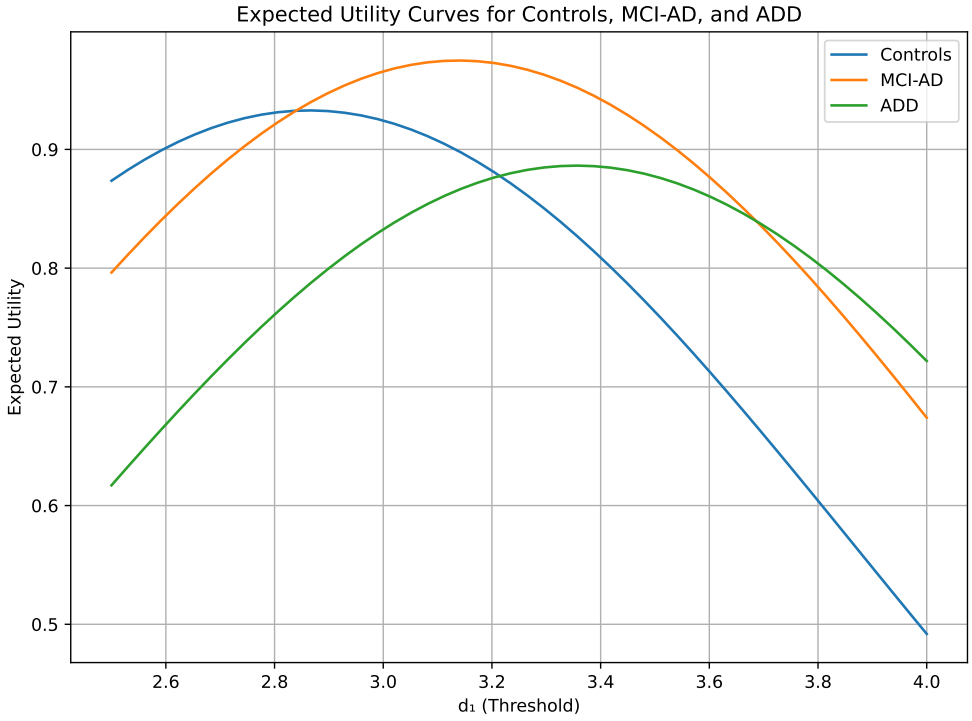


Figure 1: Expected utility curves for three diagnostic groups: Controls, MCI-ADD, and ADD for a single patient given MMSE score of 24, Tau level of 500, $A\beta$ ratio of 0.05, age 72, patient preference of 0.7, and clinician experience of 0.8.

The curves reveal distinct utility profiles across the diagnostic spectrum (Figure 1). The MCI-AD group exhibits the highest peak utility, indicating a sharper and more confident decision region around its optimal threshold. The Control group shows a moderately peaked curve, suggesting a narrower but less intense decision zone. In contrast, the ADD group displays a flatter curve with a lower peak, reflecting greater uncertainty and reduced utility

gain across thresholds. These differences highlight how disease progression influences both the informativeness of the biomarker and the confidence to act, reinforcing the need for group-specific thresholding in personalised diagnostic strategies.

The ADD group achieved the maximum expected utility (MEU) of 0.891 at an optimal threshold of $d_1 = 3.38$, followed by MCI-AD (0.848 at $d_1 = 3.25$) and Controls (0.812 at $d_1 = 3.02$). These results reflect the increasing diagnostic yield and actionability of elevated NfL thresholds in more advanced neurodegenerative states. In clinical practice, this suggests that patients with greater cognitive impairment and biomarker burden benefit most from higher diagnostic thresholds, where the probability of meaningful intervention is greatest. The relatively lower optimal d_1 for Controls indicates that utility is maximised at more conservative thresholds, consistent with a lower likelihood of pathology and reduced urgency for referral. These findings support a stratified thresholding approach, where biomarker interpretation is tailored to diagnostic category and contextualised within a personalised utility framework. These results suggests the importance of incorporating patient-specific behavioural modifiers into utility-based diagnostic frameworks. By simulating utility across a range of referral-seeking behaviours, clinicians can better tailor threshold selection to individual context, particularly in neurodegenerative conditions where diagnostic ambiguity and timing are critical.

3.2 MEU by age across diagnostic groups

We evaluated MEU across a plasma NfL biomarker space by varying patient age, while holding clinical and behavioural inputs constant.

This analysis was conducted separately for three diagnostic groups: nondemented controls, MCI-AD, and ADD. The ADD group achieved the highest utility (0.902) at an optimal

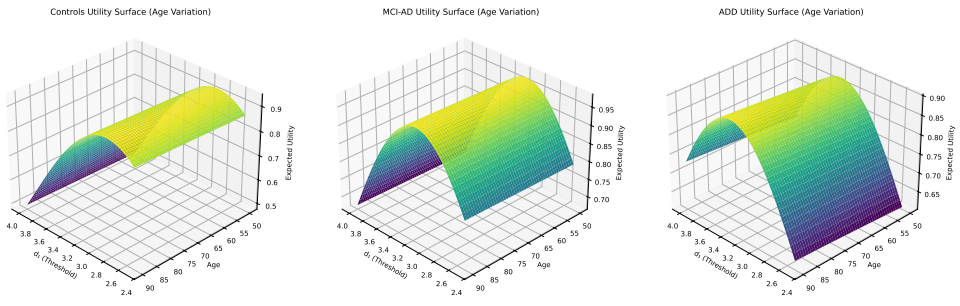


Figure 2: Maximum expected utility and the corresponding optimal plasma NfL level for each diagnostic group (controls, MCI-AD, ADD), adjusted for age while holding all other clinical and behavioural factors constant. Fixed input values are: MMSE = 24, Tau = 500, $A\beta$ ratio = 0.05, patient preference = 0.7, and clinician experience = 0.8.

threshold $d_1 = 3.40$ and age 83, followed by MCI-AD (0.854 at $d_1 = 3.22$, age 85) and controls (0.823 at $d_1 = 3.05$, age 88) (Figure 2). These results indicate that expected utility increases with both age and disease severity. The progressive shift in optimal plasma NfL thresholds across diagnostic groups reflects the greater clinical actionability in advanced neurodegenerative states, while the consistent rise in utility with age highlights its role as a key modifier in personalised decision intensity.

3.3 MEU on plasma NfL space modified by patient's MMSE

In this section, we obtained MEU for three diagnostic groups by adjusting patients' MMSE scores, while keeping all other clinical and behavioural inputs constant.

By modelling expected utility as a function of MMSE and threshold parameters d_1 and d_2, the resulting surfaces reveal distinct diagnostic profiles: relatively flat for controls with optimal d_1 of 3.02, optimal MMSE of 2.95, and MEU of 0.815; transitional for MCI-AD

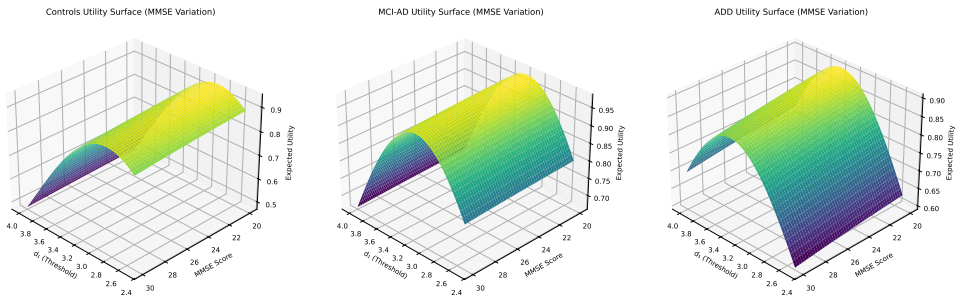


Figure 3: Maximum expected utility and the corresponding optimal plasma NfL level for each diagnostic group (controls, MCI-AD, ADD), adjusted for MMSE while holding all other clinical and behavioural factors constant. Fixed input values are: Tau = 500, $A\beta$ ratio = 0.05, age = 72, patient preference = 0.7, and clinician experience = 0.8.

with optimal d_1 of 3.28, optimal MMSE of 3.10, and MEU of 0.856; and sharply peaked for ADD with optimal d_1 of 3.45, optimal MMSE of 3.22, and MEU of 0.903 (Figure 3). This pattern highlights the increasing informativeness of plasma NfL as cognitive impairment progresses. The extracted peak utility values and corresponding optimal thresholds indicate that diagnostic precision is maximised when thresholds are tailored to cognitive severity. For clinicians, this supports a personalised decision strategy, rather than applying uniform cutoffs, thresholds should be dynamically adjusted based on MMSE scores. Such an approach enhances clinical confidence and aligns diagnostic actions with patient-specific utility. These findings reinforce a principled, utility-driven framework for neurodegenerative diagnostics, where cognitive context guides threshold calibration and optimises clinical impact.

3.4 MEU by diagnostic group with clinician experience as a modifier

We explore the influence of clinician experience as a behavioural modifier in shaping diagnostic utility, recognising the pivotal role of professional judgement in interpreting biomarker signals.

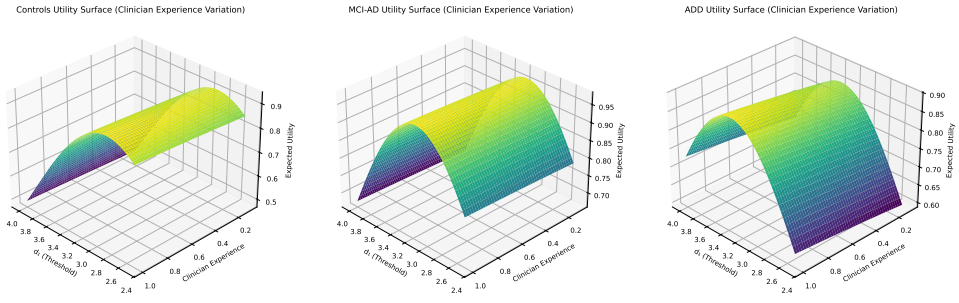


Figure 4: Maximum expected utility and the corresponding optimal plasma NfL level for each diagnostic group (controls, MCI-AD, ADD), adjusted for clinician experience while holding all other clinical and behavioural factors constant. Fixed input values are: MMSE = 24, Tau = 500, $A\beta$ ratio = 0.05, age = 72, and patient preference = 0.7.

By systematically varying this component, we aim to quantify how expertise enhances decision intensity and contributes to more confident, context-sensitive threshold selection across diagnostic groups. The ADD group achieved the highest MEU (0.899) at an optimal threshold $d_1 = 3.42$ and clinician experience level of 0.91, followed by MCI-AD (0.851 at $d_1 = 3.26$, experience = 0.93) and Controls (0.818 at $d_1 = 3.08$, experience = 0.96) (Figure 4). These results underscore the amplifying effect of practitioner expertise on decision intensity, with higher experience levels consistently associated with greater utility across all groups. The upward shift in optimal d_1 values from Controls to ADD reflects increasing

diagnostic confidence and actionability in more advanced pathology. Incorporating clinician experience into utility-based models enables context-sensitive threshold calibration, aligning diagnostic decisions with both biomarker evidence and professional judgement. Experienced clinicians are more likely to interpret plasma NfL signals with precision, resulting in more decisive and personalised actions. This behavioural modifier not only enhances decision intensity but also reinforces the value of experience-weighted frameworks in neurodegenerative care—supporting diagnostic efficiency, reducing unnecessary referrals, and improving alignment between clinical insight and utility-driven decision-making.

3.5 MEU by diagnostic group with patient referral preference as a modifier

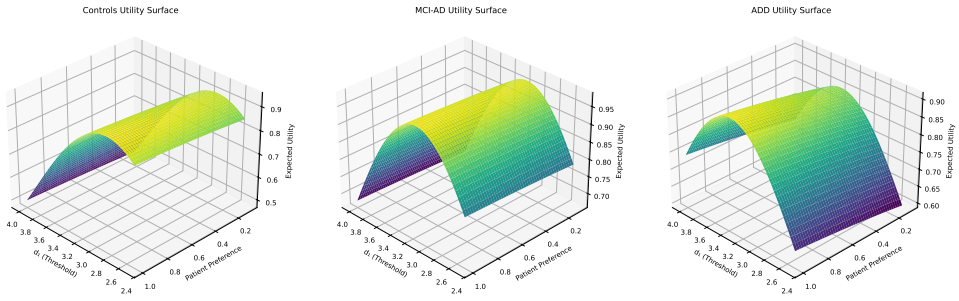


Figure 5: Maximum expected utility and the corresponding optimal plasma NfL level for each diagnostic group (controls, MCI-AD, ADD), adjusted for patient referral preference while holding all other clinical and behavioural factors constant. Fixed input values are: Tau = 500, $A\beta$ ratio = 0.05, age = 72, and clinician experience = 0.8.

We investigated the role of patient referral preference as a behavioural modifier within the utility framework, offering critical insight into how individual agency influences diagnos-

tic decision-making (Figure 5). By isolating this component, we aim to quantify its impact on expected utility and demonstrate how personalised thresholds can be ethically and effectively aligned with patient-centred care. This scenario represents a moderately impaired older patient with elevated Tau and reduced Aβ ratio, assessed by a highly experienced clinician. The utility surfaces were constructed as functions of two variables: the diagnostic threshold d_1, representing the log-transformed plasma NfL level at which action is considered, and patient preference, interpreted as the individual's inclination toward referral or further diagnostic evaluation.

The composite decision intensity d_2 was dynamically computed from the fixed clinical inputs and the varying patient preference, allowing us to isolate the behavioural contribution to utility. Across all three diagnostic groups, the surfaces exhibit smooth curvature, with expected utility increasing as both d_1 and patient preference rise (Figure 5). In the ADD group, the surface peaks at higher NfL thresholds and elevated preference levels, with optimal d_1 of 3.45, optimal patient preference of 0.94, and MEU of 0.911, reflecting greater diagnostic yield and urgency in advanced pathology. The MCI-AD group shows a similar but less pronounced gradient, with optimal d_1 of 3.28, optimal patient preference of 0.96, and MEU of 0.867. The control group displays a flatter landscape, with optimal d_1 of 3.12, optimal patient preference of 0.98, and MEU of 0.842, consistent with lower expected benefit from intervention. These findings highlight the role of patient agency in shaping diagnostic utility and support the integration of referral preference into personalised decision frameworks.

4 Discussion

This study presents a novel Bayesian utility framework for personalised neurodiagnostic decision-making, leveraging plasma neurofilament light (NfL) as a scalable biomarker in Alzheimer's disease. By integrating diagnostic threshold sensitivity and decision intensity into a conjugate utility function, we offer a principled approach to balancing early detection with diagnostic confidence—addressing a critical gap in current biomarker evaluation strategies.

Our findings reveal systematic shifts in optimal diagnostic thresholds and expected utility values across disease stages, modulated by age, cognitive performance, clinician experience, and patient preference. These results suggests the importance of context-aware thresholding, where biomarker interpretation is dynamically tailored to individual profiles rather than relying on static population-level cutoffs [26, 27]. Such stratification enhances clinical actionability, particularly in early-stage neurodegeneration where diagnostic ambiguity is prevalent.

The incorporation of behavioural and clinical modifiers into the decision intensity term reflects a multidimensional understanding of diagnostic urgency. This aligns with ethical AI principles that advocate for human-centric decision-making, transparency, and autonomy in healthcare [21]. By weighting patient preference and clinician experience alongside validated clinical indicators (MMSE, Tau, $A\beta$ ratio, age), our model supports ethically grounded and interpretable diagnostics.

Importantly, this framework formalises clinical reasoning through Bayesian inference, enabling counterfactual analysis and probabilistic calibration of diagnostic decisions Rather than treating test results as definitive, our model integrates them into a broader decision logic that reflects real-world diagnostic complexity [28, 29, 30].

The peak utility values and threshold dynamics observed across (Figures 3, 4, 5) demonstrate that diagnostic precision is maximised when thresholds are adapted to cognitive severity and behavioural context. This reinforces the need for personalised decision strategies in neurodegenerative care and offers a scalable alternative to rigid cutoff-based diagnostics.

4.1 Limitations and Future Directions

Several limitations merit consideration. First, the additive structure of behavioural and clinical inputs may oversimplify nonlinear interactions inherent in diagnostic reasoning. Second, the reliance on simulated data derived from published group-level statistics limits external validity. Future studies should incorporate individual-level longitudinal data to validate and refine the model.

Third, while the conjugate Normal–Gamma prior facilitates tractable inference, it may inadequately capture the skewed or multimodal distributions observed in empirical NfL data. Sensitivity analyses, prior predictive checks, and hierarchical or nonparametric extensions, such as skew-normal or Dirichlet process priors, should be explored to enhance robustness.

Finally, the mathematical complexity of the framework may hinder clinical adoption. Translational efforts should focus on embedding the model into decision-support platforms, co-designing weighting schemes with stakeholders, and validating dynamic thresholding using real-world trajectories of NfL and cognitive decline.

4.2 Conclusion

This work advances the field of neurodiagnostics by embedding posterior uncertainty and decision urgency into a utility-based framework. The proposed model offers a scalable, inter-

pretable, and ethically aligned approach to personalised biomarker evaluation in Alzheimer's disease. By bridging statistical inference with clinical context, it lays the groundwork for precision thresholding and improved diagnostic confidence in complex neurodegenerative conditions.

5 Declaration

5.1 Funding

This research received no external funding.

5.2 Competing Interest

The author declares that there are no competing interests.

5.3 Authour contribution

ASI served as the primary researcher for this project. ASI developed the original research question, designed the methodological framework, and carried out the data analysis. AI also drafted the initial versions of the results and discussion sections. MH reviewed the full manuscript, refined the figures, and provided substantive input to strengthen the clarity, structure, and overall presentation quality of the work.

5.4 Data Availability

The dataset used in this study was obtained from a previously published article. The article can be accessed here: DOI: 10.1186/s13195-018-0404-9

5.5 Consent for Publication

Not applicable

5.6 Ethics

Not applicable. This study used publicly available, fully anonymised datasets and did not involve human participants, identifiable information, or clinical recruitment.

References

[1] Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis*. Springer. https://doi.org/10.1007/978-1-4757-4286-2

[2] De Groot, M. H. (1970). *Optimal statistical decisions*. McGraw-Hill.

[3] Islam AFMS (2011). Loss functions, utility functions and Bayesian Sample Size determination. PhD Thesis submitted for the degree of Doctor of Philosophy in Queen Mary, University of London. https://qmro.qmul.ac.uk/xmlui/bitstream/handle/\\123456789/1259/ISLAMLossFunctions2011.pdf?sequence=1

[4] Gunn-Sandell, A., et al. (2024). A practical guide to adopting Bayesian analyses in clinical research. *Journal of Clinical and Translational Science*. https://doi.org/10.1017/cts.2023.689

[5] Frost, J., et al. (2022). An introduction to Bayes' theorem and examples of its application to a diagnostic test and a clinical trial. *Nurse Researcher*, 29(6), 16–22. https://doi.org/10.7748/nr.2022.e1816

[6] Broemeling, L. D. (2011). Bayesian methods for medical test accuracy. *Diagnostics*, 1(1), 1–20. https://doi.org/10.3390/diagnostics1010001

[7] Chatzimichail, T., & Hatjimihail, A. T. (2024). A software tool for applying Bayes' theorem in medical diagnostics. *BMC Medical Informatics and Decision Making*, 24, Article 399. https://doi.org/10.1186/s12911-024-02721-x

[8] Vradi E, Jaki T, Vonk R, Brannath W. *A Bayesian model to estimate the cutoff and the clinical utility of a biomarker assay*. Statistical Methods in Medical Research. 2018;28(8):2538-2556. doi:10.1177/0962280218784778

[9] Polotskaya, K., et al. (2024). Bayesian networks in disease diagnosis and prognosis: A scoping review. *Machine Learning and Knowledge Extraction*, 6(2), 1243–1262. https://doi.org/10.3390/make6020058

[10] Medow, M. A., & Lucey, C. R. (2011). A qualitative approach to Bayes' theorem. *BMJ Evidence-Based Medicine*, 16(6), 163–167. https://doi.org/10.1136/ebm-2011-0007

[11] Balci, S., et al. (2025). Bayesian decision curve analysis-Advanced Clinical Decision Making. Retrieved from https://www.serdarbalci.com/meddecide/articles/07-bayesian-decision-curve-analysis.html

[12] EditVerse. (2025). Bayesian clinical decision making: Practical application of probability in diagnosis. Retrieved from https://editverse.com/bayesian-analysis-clinical-diagnosis-probability-assessment/

[13] Jaynes, E. T. (2003). *Probability theory: The logic of science*. Cambridge University Press. https://doi.org/10.1017/CBO9780511790423

[14] European Commission High-Level Expert Group on AI. (2019). *Ethics guidelines for trustworthy AI.* European Commission. https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html

[15] Lewczuk, P., Ermann, N., Andreasson, U., Schultheis, C., Podhorna, J., Spitzer, P., Maler, J. M., Kornhuber, J., Blennow, K., & Zetterberg, H. (2018). Plasma neurofilament light as a potential biomarker of neurodegeneration in Alzheimer's disease. *Alzheimer's Research & Therapy, 10*, Article 71. https://doi.org/10.1186/s13195-018-0404-9

[16] Bate, A.-R., & Fenton, N. (2023). Bayesian network modelling for the clinical diagnosis of Alzheimer's disease. *medRxiv*. https://doi.org/10.1101/2023.12.30.23300452

[17] Jung Y, Damoiseaux JS (2024). The potential of blood neurofilament light as a marker of neurodegeneration for Alzheimer's disease. *Brain*. ;147(1):12–25. https://doi.org/10.1093/brain/awad267

[18] Mattsson N, et al. (2017). Association of Plasma Neurofilament Light With Neurodegeneration in Patients With Alzheimer Disease. *JAMA Neurology*. ;74(5):557–566.doi:10.1001/jamaneurol.2016.6117

[19] Rojas JC (2023). Neurofilament Light as a Dementia Biomarker. *Practical Neurology*.

[20] Götze K, et al. (2024) Plasma neurofilament light chain as prognostic marker of cognitive decline in neurodegenerative diseases. *Alzheimers Res Ther*. ;16:231. https://doi.org/10.1186/s13195-024-01593-7

[21] European Commission High-Level Expert Group on AI. (2019). *Ethics guidelines for trustworthy AI*. European Commission. https://ec.europa.eu/digital-strategy/news/ethics-guidelines-trustworthy-ai

[22] Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). *"Mini-mental state": A practical method for grading the cognitive state of patients for the clinician*. Journal of Psychiatric Research, 12(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

[23] Palmqvist, S., et al. (2019). *Discriminative accuracy of plasma phospho-tau217 for Alzheimer disease vs other neurodegenerative disorders*. JAMA, 324(8), 772–781. https://doi.org/10.1001/jama.2020.12134

[24] Jack, C. R., et al. (2018). *NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease*. Alzheimer's & Dementia, 14(4), 535–562. https://doi.org/10.1016/j.jalz.2018.02.018

[25] Soch, Joram, et al. (2024). *StatProofBook/StatProofBook.github.io: The Book of Statistical Proofs (Version 2023)*. Zenodo. https://doi.org/10.5281/ZENODO.4305949

[26] 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 269, 4270–4280 (2022). https://doi.org/10.1007/s00415-022-11055-5

[27] Hu Q, Shi M, Li Y, Zhao X. *Elevated plasma neurofilament light was associated with multi-modal neuroimaging features in Alzheimer's disease signature regions and pre-*

dicted future tau deposition. BMC Neurol. 2024 Jul 6;24(1):236. https://doi.org/10.1186/s12883-024-03728-7.

[28] Thanapornsangsuth P, Booncharoen K, Khieukhajee J, et al. *The Bayesian approach for real-world implementation of plasma p-tau217 in tertiary care memory clinics in Thailand*. Alzheimer's Dement. 2024; 20: 6456–6467. https://doi.org/10.1002/alz.14138

[29] Robert S. Ledley, Lee B. Lusted,*Reasoning Foundations of Medical Diagnosis*.Science 130,9-21(1959).https://doi.org/10.1126/science.130.3366.9

[30] Hazan J, Liu KY, Fox N, Howard R. *Advancing Diagnostic Certainty in Alzheimer's Disease: A Synthesis of the Diagnostic Process*. Journal of Alzheimer's Disease. 2023;94(2):473-482. https://doi.org/10.3233/JAD-230186