Online Supplemental Material: Full Analytic Description

Discrete-Time Hazard Model and Prediction

In the discrete-time hazard model, the logged odds of death for individual $i$in interval t are modeled as

$$\text{logit}\{P(Y_{it}=1 \mid \text{Race}_i, \text{Age}_{it}, X_{it}, u_i)\} = \beta_0 + \beta_{\text{race}}\text{Race}_i + \beta_{\text{age}}\text{Age}_{it} + \beta^{\top}x_{it} + u_i \quad \text{(Eq.S1)}$$

where $u_i \sim N(0, \sigma_u^2)$ captures unobserved individual heterogeneity. This multilevel specification represents the discrete-time hazard as a linear predictor on the log-odds scale, with coefficients associated with age, race, and other time-varying covariates collected in survival records. Bio-physical and psychometric measures are not included in this specification because, within this additive regression framework, their inclusion would be expected to attenuate the race coefficient by redistributing variance across domains of risk.

The corresponding predicted probability is given by

$$\hat{p}_{it} = \frac{1}{1+\exp(-\text{logit}(h_{it}))} = P(Y_{it}=1 \mid \text{Race}_i, \text{Age}_{it}, X_{it}, u_i) \quad \text{(Eq.S2)}$$

Regression coefficients are estimated by maximizing the log-likelihood, optimizing probabilistic fit without encoding any explicit classification rule. In rare-event settings such as mortality, this estimation procedure tends to prioritize correct prediction of the majority outcome—survival—leading to elevated false-negative rates for deaths when predictions are translated into classifications (Huang et al., 2022; Krieger et al., 2024).

Because Eq. 4 yields continuous predicted probabilities rather than discrete outcomes, classification requires an externally imposed decision threshold τ. For evaluation purposes, we select τsuch that the predicted positive rate aligns with the observed mortality prevalence:

$P(\hat{p}_{it} \geq \tau) \approx \pi$, where $\pi = \frac{1}{N}\sum_{it} Y_{it}$ (Eq.S3)

For example, if deaths occur in 2% of person–intervals, the threshold is set at the 98th percentile of the predicted probability distribution. This prevalence-calibrated thresholding is used solely for evaluation and does not alter the underlying regression estimates. Performance under this rule is compared with that of the neural-network models using their respective prediction outputs.

NN-Hazard Model and Prediction

Architecture

In contrast to regression-based hazard models, the NN-based hazard model is trained using a classification-oriented loss function that places greater emphasis on correctly identifying death events, while still producing probabilistic predictions on the log-odds scale.

We employ a feedforward multilayer perceptron (MLP)—a sequence of fully connected layers that sequentially learn representations of information across multiple data branches. All continuous covariates are standardized to z-scores prior to training, except for age variables, which are log-transformed and retained on their original scale; binary indicators are left unstandardized. Inputs are organized into domain-specific branches: static covariates (race, sex, education, cohort, and parental education) and time-varying self-reported health measures (self-rated health, depressive symptoms, mobility limitations, activities of daily living, and diagnosed conditions).

Age variables are not standardized because they define the discrete-time hazard scale and enter explicitly into the recency encoding used to align biomarker and psychometric measurements with the current interval. Standardization of non-age continuous variables serves

two purposes: first, to prevent differences in measurement scale from disproportionately influencing learned representations; and second, to encode covariates relative to their population-level distributions per interval/Wave rather than in raw units. Together with the log-transformation of age, these preprocessing steps ensure that chronological age structures population-level hazard patterns while individual heterogeneity is represented relative to the population distribution. Furthermore, bio-physical and psychometric domains are incorporated by allowing group disparities to emerge through the joint structure of predicted hazards.

Branch-level representations and base-data projection

Within each data branch, inputs are transformed through two successive linear layers with nonlinear activation. This depth provides a minimal architecture capable of learning nonlinear representations beyond simple linear reweighting of the original inputs within each branch. Each branch therefore produces a fixed-length embedding summarizing domain-specific information, denoted $h_i^{(\text{static})}$ for static covariates and $h_{it}^{(\text{tv})}$ for time-varying information.

These branch-level embeddings are concatenated and passed through an additional nonlinear mixing layer, allowing interactions across static and time-varying domains. The base log-odds component is then defined as

$$z_{it} = \phi\left(W_{\text{mix}}\left[h_i^{(\text{static})} \parallel h_{it}^{(\text{tv})}\right]\right), \eta_{it}^{(\text{base})} = c_{\text{base}}^{\top} z_{it} + \text{intercept}_{\text{base}} \quad \text{(Eq.S4)}$$

where $\phi(\cdot)$ denotes a nonlinear activation function and $\parallel$ indicates concatenation. This component captures population-structured mortality risk associated with age, race, and contemporaneous health through a constrained combination of static and time-varying information. By limiting depth to a single mixing layer followed by linear projection into the log-odds, the baseline specification implies age-based structure while allowing modest cross-domain interaction.

Incorporation of bio-physical and psychometric data

The architecture incorporates bio-physical and psychometric information through an additional mixing stage that allows representations from all branches to interact. Branch-level embeddings from static, time-varying, bio-physical, and psychometric domains are concatenated and transformed into a mixed representation,

$$z_{it}^{(\mathrm{mix})} = \phi\left(W_{\mathrm{mix}}\left[h_i^{(\mathrm{static})} \parallel h_{it}^{(\mathrm{tv})} \parallel h_{it}^{(\mathrm{bio})} \parallel h_{it}^{(\mathrm{psych})}\right]\right) \quad \text{(Eq.S5)}$$

This mixed representation is combined with the bio-physical and psychometric embeddings to form an auxiliary log-odds component,

$$\eta_{it}^{(\mathrm{add})} = c_{\mathrm{add}}^{\top}\left[h_{it}^{(\mathrm{bio})} \parallel h_{it}^{(\mathrm{psych})} \parallel z_{it}^{(\mathrm{mix})}\right] + \mathrm{intercept}_{\mathrm{add}},\ \eta_{it} = \eta_{it}^{(\mathrm{base})} + \eta_{it}^{(\mathrm{add})} \quad \text{(Eq.S6)}$$

This two-component structure allows population-structured hazard patterns (e.g., age and race gradients) to be represented through the baseline component while permitting additional data domains to contribute flexibly through nonlinear mixing. In this way, the age-embedded hazard scale is preserved through $\eta_{it}^{(\mathrm{base})}$, with additional information entering additively through $\eta_{it}^{(\mathrm{add})}$rather than entirely redefining the underlying hazard structure.

Two-Stage Estimation and Counterfactual Analyses

The estimation strategy follows directly from the architecture of the NN-hazard model. Because the model separates a baseline component derived from static covariates and time-varying self-reported health from auxiliary components derived from bio-physical and psychometric branches, optimization is aligned with the distinct roles these components play. The projection vectors in Eqs. 6–8 map each component into the log-odds scale, ensuring that all

predictions remain anchored to a discrete-time hazard formulation while allowing different sources of information to contribute separately.

We implement a Neyman–Pearson–inspired learning strategy (Huang et al. 2022) through a two-stage estimation procedure that mirrors this architectural decomposition. In the first stage, only the baseline branch is trained under symmetric classification loss. This stage establishes population-structured hazard patterns driven primarily by age, race, and contemporaneous health, analogous to conventional demographic hazard modeling but learned through nonlinear representations.

In the second stage, auxiliary branches capturing bio-physical and psychometric information are activated and combined through the mixing layer. The loss function is modified to place greater weight on missed death events, encouraging improved sensitivity to mortality risk. Because the auxiliary component enters additively into the baseline log-odds, it refines individual-level risk detection without redefining the underlying hazard scale or displacing the demographic structure learned in the first stage.

The resulting architecture yields two additive log-odds components, $\eta_{it}^{(\text{base})}$ and $\eta_{it}^{(\text{add})}$, whose sum defines the predicted hazard. The baseline component reflects mortality risk derived from the static covariates and time-varying self-reported health, while the auxiliary component captures adjustments associated with bio-physical and psychometric information and their interactions.

This separation facilitates counterfactual pattern recovery without imposing independence across data domains. By holding auxiliary representations fixed while varying demographic inputs, the model can recover age- and race-structured mortality hazard conditional

on constant levels of physiological and psychosocial information (set to standardized scores of 0, the population mean). Conversely, auxiliary contributions to prediction can be aggregated to describe any systemic divergences from the baseline hazard typical (by way of central tendencies) of comparable demographic profiles.

By counterfactually constraining bio-physical and psychometric inputs to the population mean and examining the resulting change in predicted hazard surfaces, the model yields contrasts that are analytically familiar yet not computationally analogous to net effects in regression settings. Representations are learned jointly through nonlinear mixing across branches so that the contribution of any single input cannot be derived in isolation unlike regression-based hazard models. All prediction results summarize 100 independent trainings with different random seeds under a single fixed person-level 80/20 train–test split, so variation reflects estimation and initialization stochasticity rather than differences in data partitioning. The 20% portion is referred to throughout as the held-out test set.

Supplemental description 1: Representations of Longitudinal Data

Variables in all branches except those that are time-constant require explicit consideration of how within-person variation and missingness are represented in the form of inputs. In the NN-hazard specification, longitudinal information is encoded as structured blocks rather than as variable-specific trajectories. The main time-varying branch records information across 15 intervals, while the biophysical and psychometric branches record information across three panel waves with about four rather than two-year intervals. Because person–intervals are dropped following death, missingness is defined only over intervals in which individuals are at risk, ensuring that missingness indicators do not mechanically encode the occurrence of death itself.

To summarize information across waves for each participant, the model employs an attention-based aggregation mechanism (Kino et al., 2021). Within each data branch, wave-specific inputs are first transformed using shared parameters, yielding a sequence of latent representations that retain within-person variation across observation periods (intervals and Waves). Attention weights are then learned over these representations to determine how much each wave contributes to the individual-level summary used for mortality prediction. Unlike latent growth or trajectory models, this weighting does not estimate parametric change over time (e.g., intercepts or slopes), nor does it assume that interval/wave-specific effects lie along a common developmental dimension. Instead, attention operates on learned representations of interval/wave-specific information, allowing the model to flexibly summarize longitudinal histories without imposing a trajectory form or requiring temporal smoothness.

Chronological age (not log-transformed) is incorporated explicitly by encoding the temporal distance between each wave-specific measurement and the current hazard interval, enabling recency to be treated as a learned feature rather than a prespecified rule. Importantly, attention in this context functions as a data-driven aggregation operator rather than an interpretive device: it determines how longitudinal information is combined for prediction, not which variables or waves are substantively "important" in a causal sense.

Supplemental description 2: Learning Depth and Hyperparameters

We distinguish the above description of the model architecture—defining the analytic task and how data branches are combined into NN-based hazard—from analytic choices referred to as hyperparameters. Hyperparameters specify learning depth and can be altered for other purposes and future research. We do not claim the selected hyperparameters as optimal or robust solely for the purpose of prediction, but rather a reasonable choice based on convention and

computational/implementational ease. Others may repurpose the hyperparameters for their own research goals.

We report all analytic choices that govern representational capacity and training behavior (hyperparameters) to support replication rather than to claim optimality. These include (i) preprocessing rules (log transformation of age, z-scoring of non-age continuous variables, treatment of binary indicators, and the encoding of missingness via masks), (ii) representation capacity (within-branch depth, hidden dimensionality, dropout rates, attention-based aggregation over repeated measures, and mixing-layer dimensionality), and (iii) training settings (class-weighting scheme, optimizer, learning rate, weight decay, batch size, number of epochs, and random seeds). Full specifications are provided in the accompanying README and code repository referenced in the Data Availability Statement in text.

Citations

Huang, J., Galal, G., Etemadi, M., & Vaidyanathan, M. (2022). Evaluation and mitigation of racial bias in clinical machine learning models: Scoping review. *JMIR Medical Informatics*, *10*(5), e36388.

Kino, S., Hsu, Y.-T., Shiba, K., Chien, Y.-S., Mita, C., Kawachi, I., & Daoud, A. (2021). A scoping review on the use of machine learning in research on social determinants of health: Trends and research prospects. *SSM-Population Health*, *15*, 100836.

Krieger, N., Testa, C., Chen, J. T., Johnson, N., Watkins, S. H., Suderman, M., Simpkin, A. J., Tilling, K., Waterman, P. D., & Coull, B. A. (2024). Epigenetic aging and racialized, economic, and environmental injustice: NIMHD Social Epigenomics Program. *JAMA Network Open*, *7*(7), e2421832–e2421832.