Decelerating Death: Estimating Changes in Life Expectancy from Dietary Risk Factors Using Accelerated Failure Time

Teresa Rebelo[1], Daniel Yoo[1], Chirag Patel[2], Olivier Jolliet[1]

Affiliations

[1] Quantitative Sustainability Assessment, Department of Environmental and Resource Engineering, Technical University of Denmark, 2800 Kgs Lyngby, Denmark

[2] Department of Biomedical Informatics, Medical School, Harvard University, 10 Shattuck St, Boston, MA 02115, USA

Corresponding author: Teresa Rebelo, Technical University of Denmark, 2800 Kgs Lyngby, Denmark. Email: tdevi@dtu.dk

Supplementary Information

Supplementary Figures

SI1. Study framework.

SI2. Pearson correlation matrix of study variables.

SI3. Bootstrap distributions (1,000 replicates) of model discrimination during training (calculated on out-of-bag (oob) data).

SI4. IPCW-based calibration performance during training across prespecified ages.

SI5. Top 30 predictors ranked by bootstrap selection frequency during training.

SI6: Rank-based evaluation of predicted age at death among deceased respondents.

SI7: SHAP summary (beeswarm) plot for the final penalized AFT model.

SI8: Raindrop plot of predicted age at death across exposure levels.

SI9: Discrimination by subgroup for the final model.

SI10: Sensitivity analysis: raindrop plots stratified by participant subgroups.

Supplementary Tables

SI1. Descriptive statistics of the study population and predictors included in the analysis (n=44,666 participants; 29 predictors).

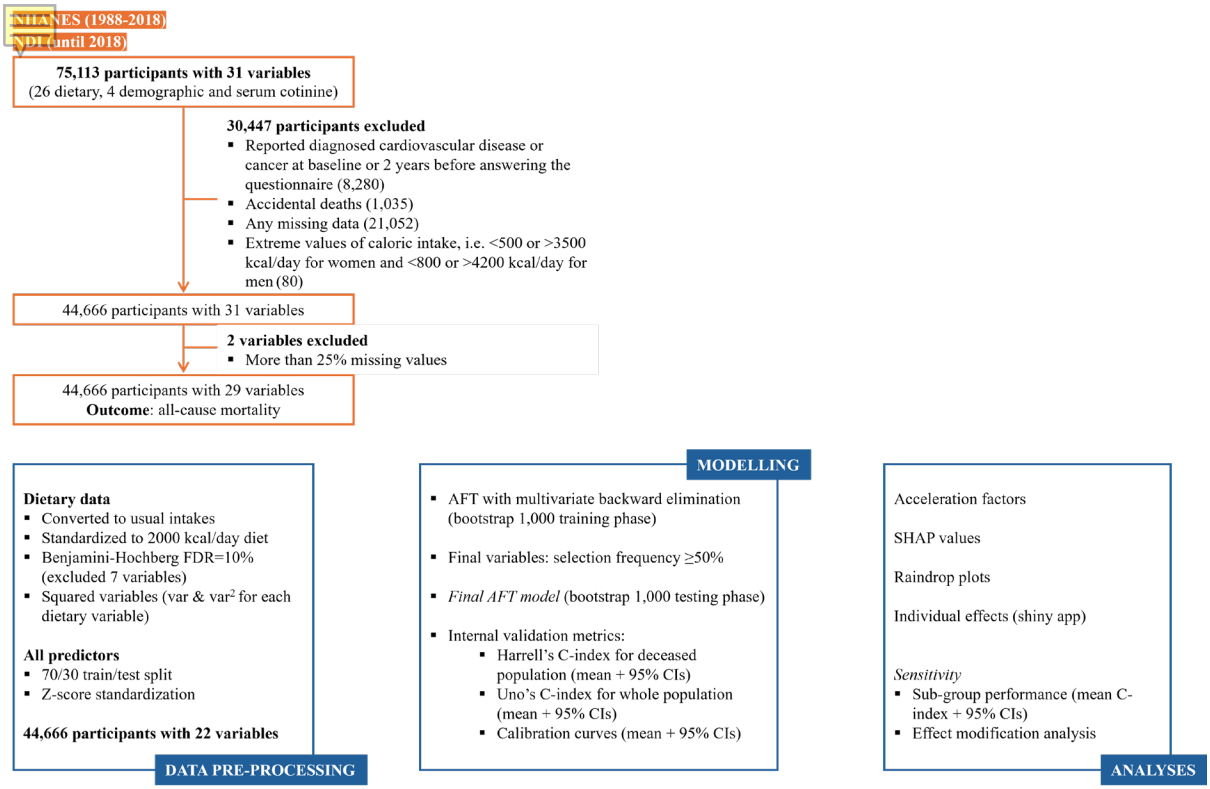
SI2. Comparison of participant characteristics between complete cases included in the analysis and participants excluded due to missing data (n=21,052).

SI3. Multicollinearity test using the Variance Inflation Factor (VIF).

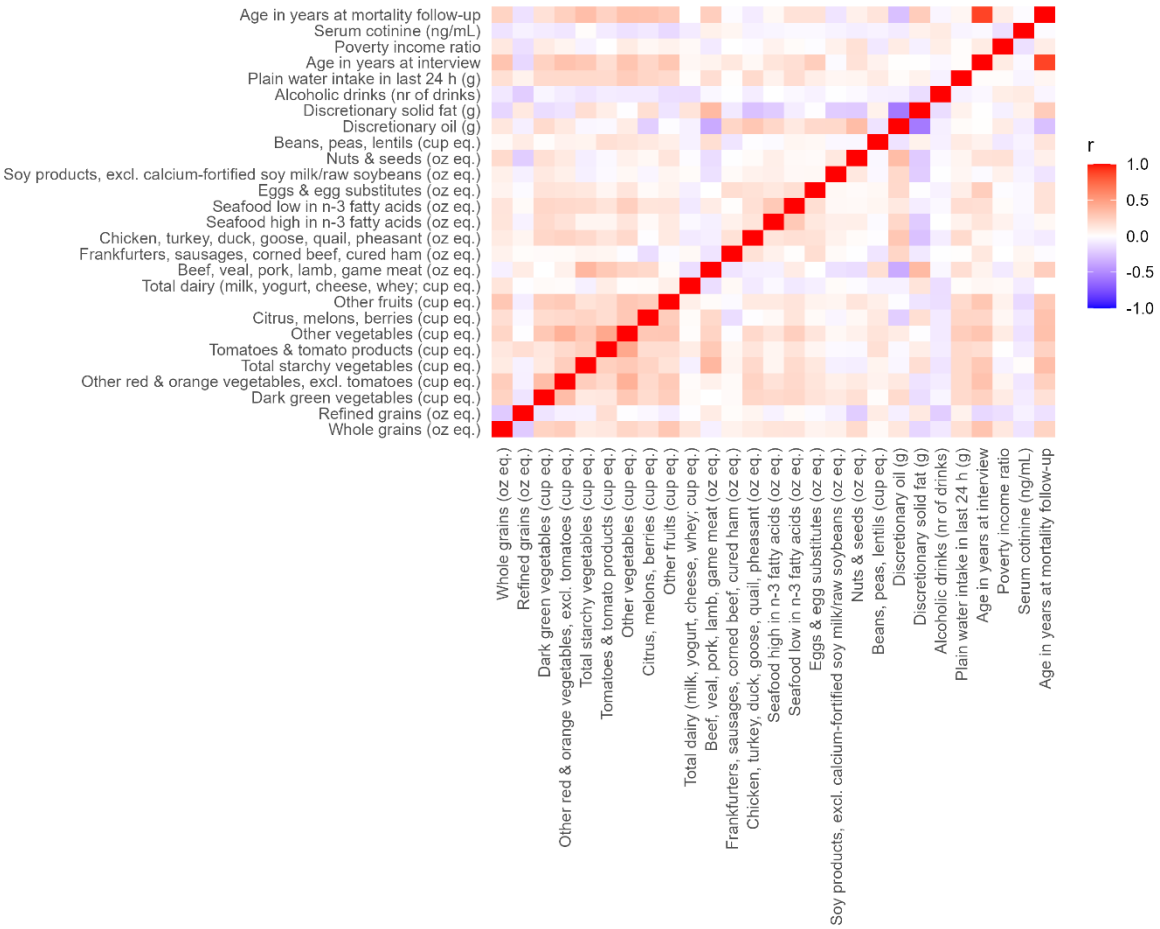
SI4. Multiple-testing correction for dietary exposures using the Benjamini-Hochberg false discovery rate (FDR) procedure.

Supplementary Information

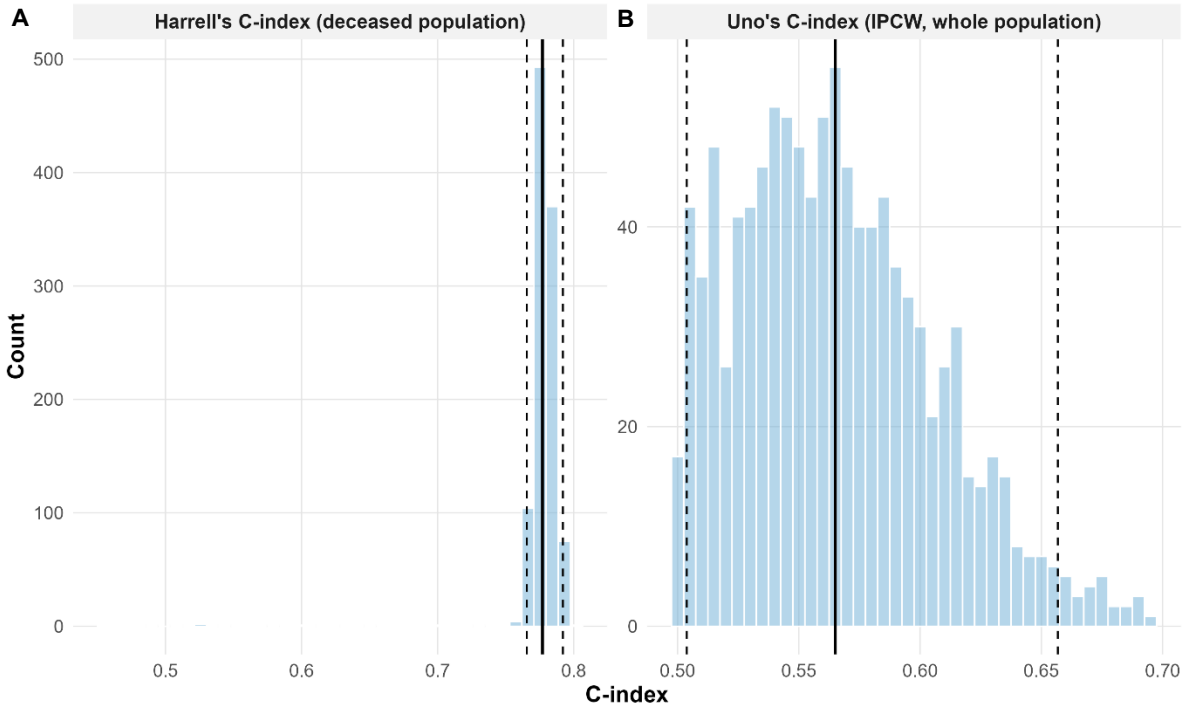
SI1. Penalized AFT model shows robust performance across resampling.



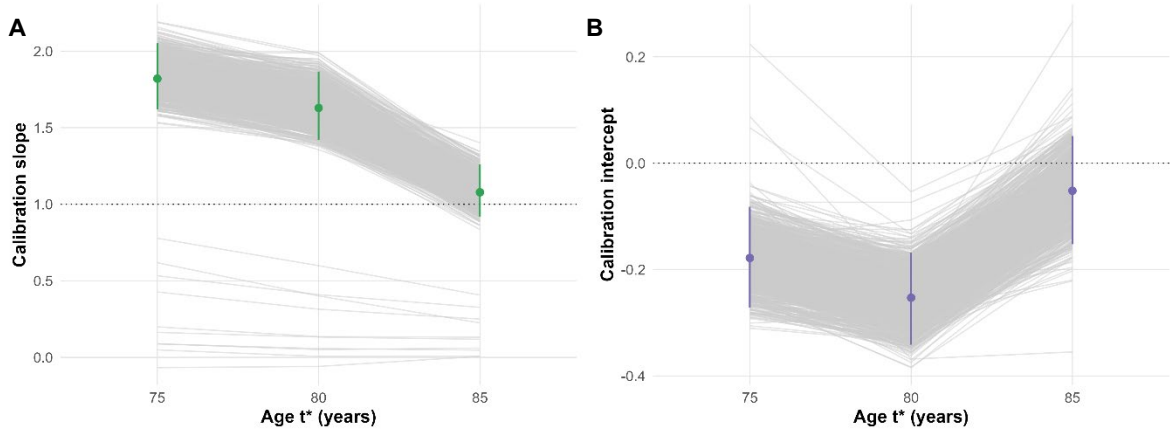
Supplementary Figure SI1: Study framework, population, variables and main steps of data processing, modeling and analyses.



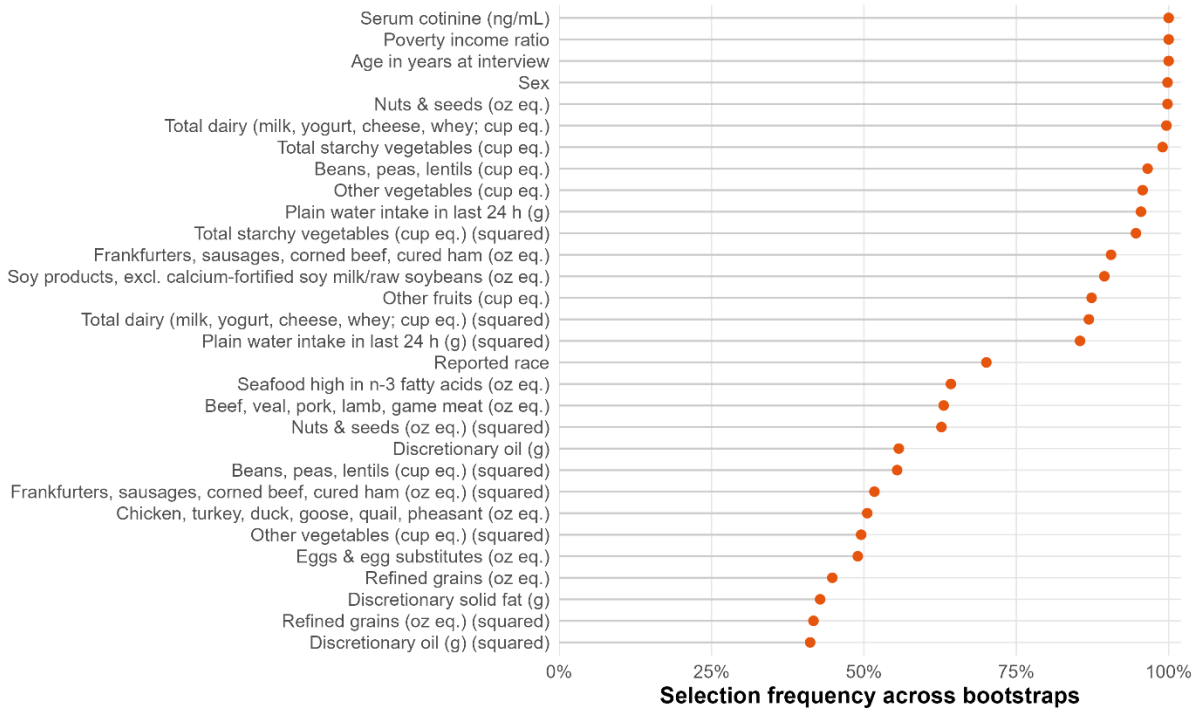
Supplementary Figure SI2: Pearson correlation matrix of study variables. No strong correlations (|r| > 0.60) were observed, with the exception of age at questionnaire and age at death.



Supplementary Figure SI3: Bootstrap distributions (1,000 replicates) of model discrimination during training (calculated on out-of-bag (oob) data). **A** Harrell's C-index evaluated in the deceased subset (mean=0.777; 95% CI: 0.766-0.792), reflecting discrimination conditional on observed events (death). **B** Uno's C-index computed in the full cohort using inverse probability of censoring weighting (IPCW) (mean=0.565; 95% CI: 0.504-0.657), accounting for the high degree of right-censoring in the population. Both metrics exceed chance level (C=0.5), indicating meaningful discriminative performance.

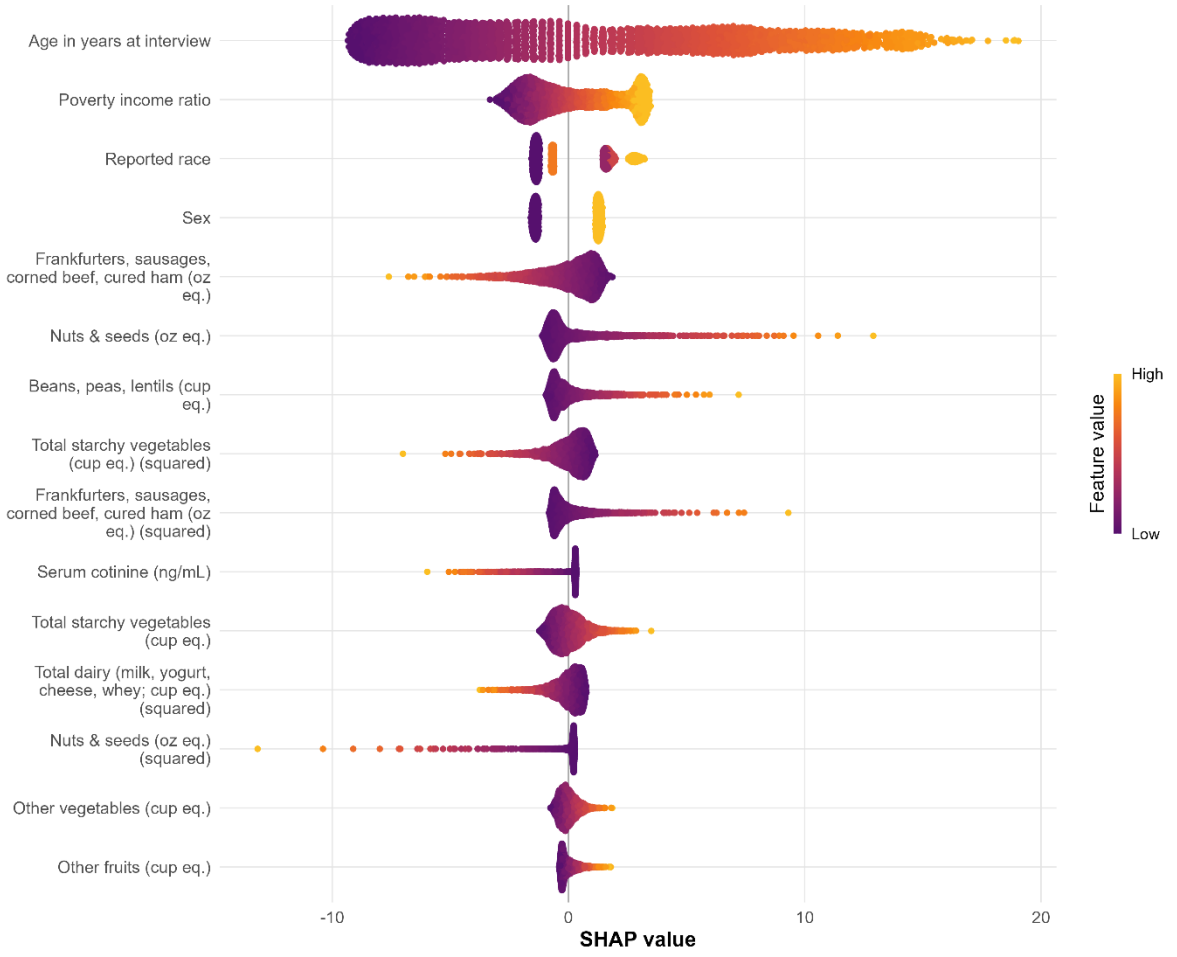


Supplementary Figure SI4: IPCW-based calibration performance during training across prespecified ages. **A** Bootstrap distributions (1,000 replicates) of calibration slopes evaluated at ages at interview 75, 80 and 85 years, with points indicating the mean and error bars the 95% confidence intervals. **B** Corresponding calibration intercepts. The dotted horizontal lines denote the ideal values (slope=1; intercept=0). Calibration slopes approach unity with increasing age, while intercepts remain close to zero, indicating overall good calibration of predicted survival times at 85 years under inverse probability of censoring weighting (IPCW).

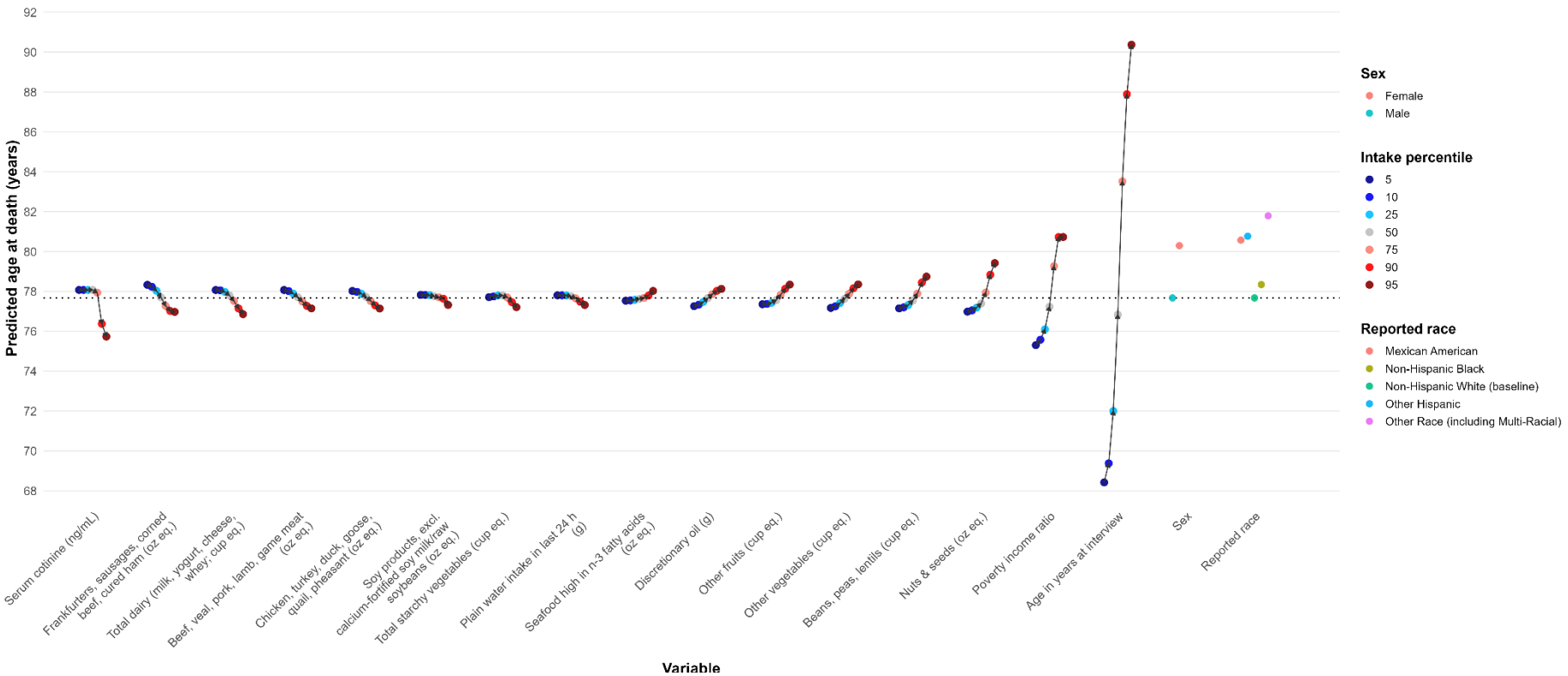


Supplementary Figure SI5: Top 30 predictors ranked by bootstrap selection frequency during training. Points indicate the proportion of bootstrap replicates in which each variable was selected in the penalized AFT model, reflecting the stability and relative importance of predictors across resamples. Both linear and squared terms are shown, highlighting non-linear contributions of dietary factors. Variables with higher selection frequencies represent more consistently retained predictors across bootstrap iterations.

Supplementary Figure SI6: Rank-based evaluation of predicted age at death among deceased respondents. Scatter plot comparing the rank of predicted age at death (x-axis) with the rank of observed age at death (y-axis) in the deceased sub-population. The red dashed line denotes the line of identity (perfect rank agreement). Point colors indicate observed rank. The coefficient of determination (R^2) summarizes agreement between predicted and observed ranks.

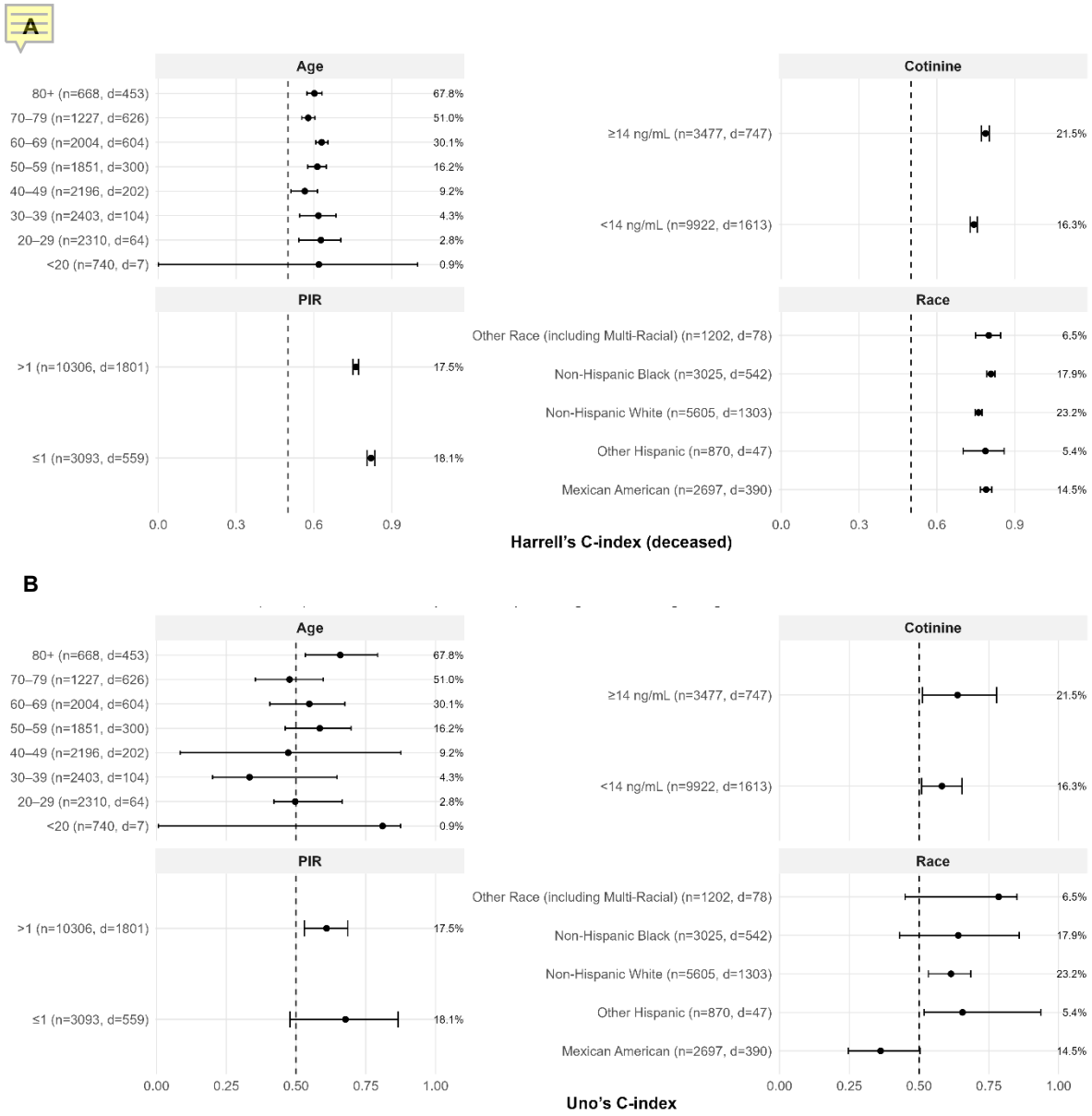


Supplementary Figure SI7: SHAP summary (beeswarm) plot for the final penalized AFT model. Features are ranked by mean absolute SHAP value, indicating their overall contribution to the model's predictions. Each point represents one participant; the horizontal position shows the SHAP value (direction and magnitude of the feature's contribution) and color denotes the feature value (low to high). Positive versus negative SHAP values indicate shifts toward the model's predicted outcome on the time scale (i.e., longer versus shorter predicted survival, respectively), with squared terms capturing non-linear contributions. For exposures modeled with both linear and squared terms, the two SHAP components should be interpreted jointly, as considering either term in isolation may give an incomplete or potentially misleading interpretation of the exposure-survival relationship.



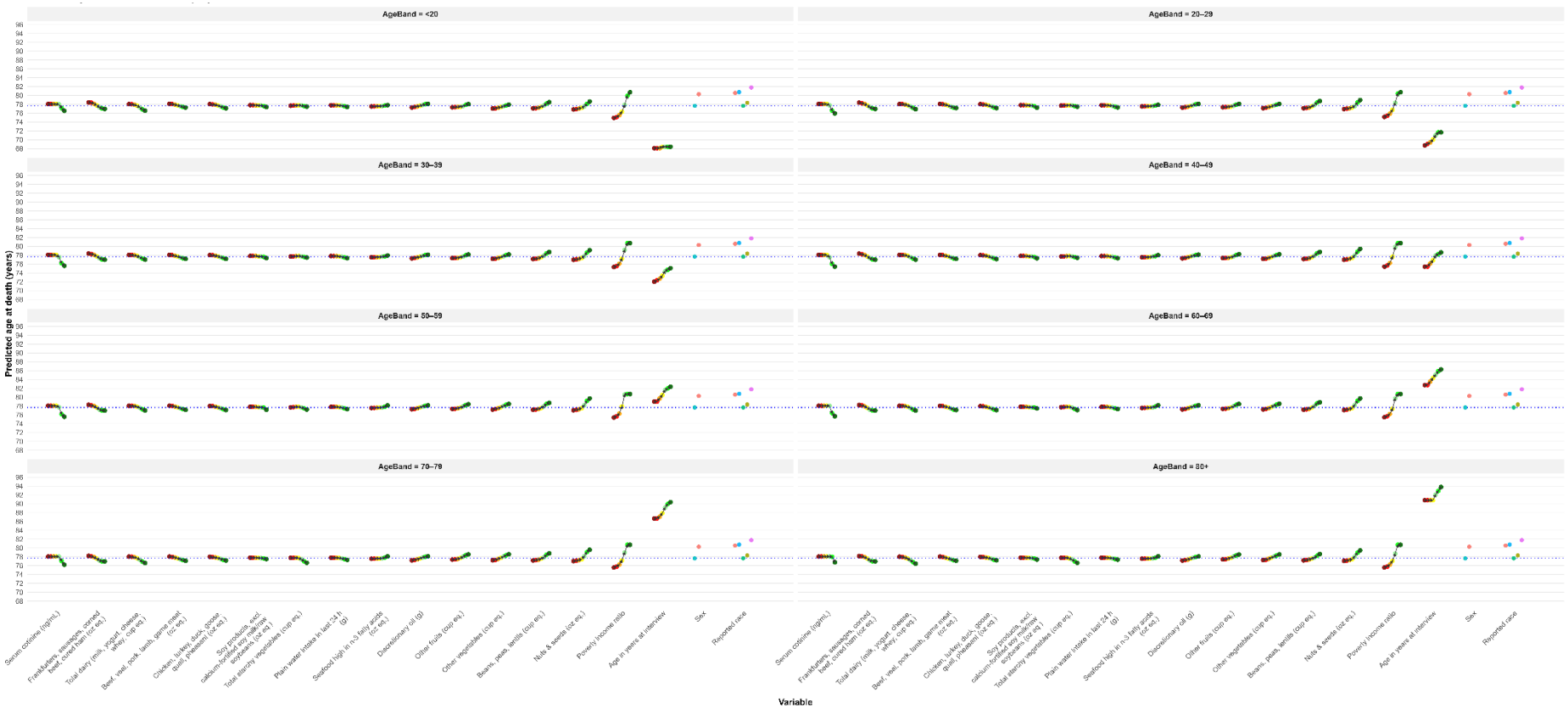
Supplementary Figure SI8: Raindrop plot of predicted age at death across exposure levels. Predicted age at death from the final penalized AFT model is shown for each retained variable, evaluated at selected intake percentiles for continuous exposures (5th, 10th, 25th, 50th, 75th, 90th and 95th) and across category levels for categorical variables (sex and reported race). For each continuous variable, points are connected to visualize how predicted age at death changes across the exposure distribution, thereby summarizing potentially non-linear effects. The horizontal blue dashed line indicates the cohort mean predicted life expectancy (77.7 years). Variables are ordered by the difference between the 95th and 5th percentile predictions, from the most detrimental to the most beneficial. When varying a given variable across percentiles (or category levels),

all other continuous predictors are held at their median (50th percentile) and categorical predictors at their reference level. Race results are de facto corrected by PIR, gender and food intakes.

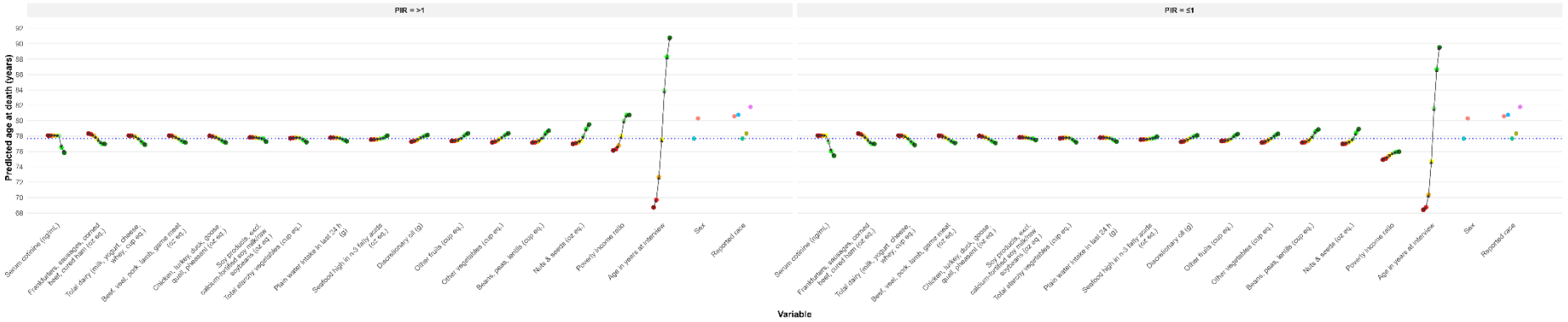


Supplementary Figure SI9: Discrimination by subgroup for the final model. **A** Harrell's C-index estimated among deceased participants only. **B** Uno's C-index (IPCW) estimated in the full cohort within prespecified subgroups to account for right-censoring. In both panels, points indicate subgroup-specific concordance estimates and horizontal bars denote 95% bootstrap confidence intervals. Subgroups are defined by age at questionnaire, serum cotinine (<14 vs ≥14 ng/mL), poverty income ratio (≤1 vs >1) and reported race. Sample size (n) and number of deaths (d) are shown in parentheses, and the death percentage within each subgroup is reported on the right side. The vertical dashed line marks the overall reference value for comparison across strata (chance level, C = 0.5).

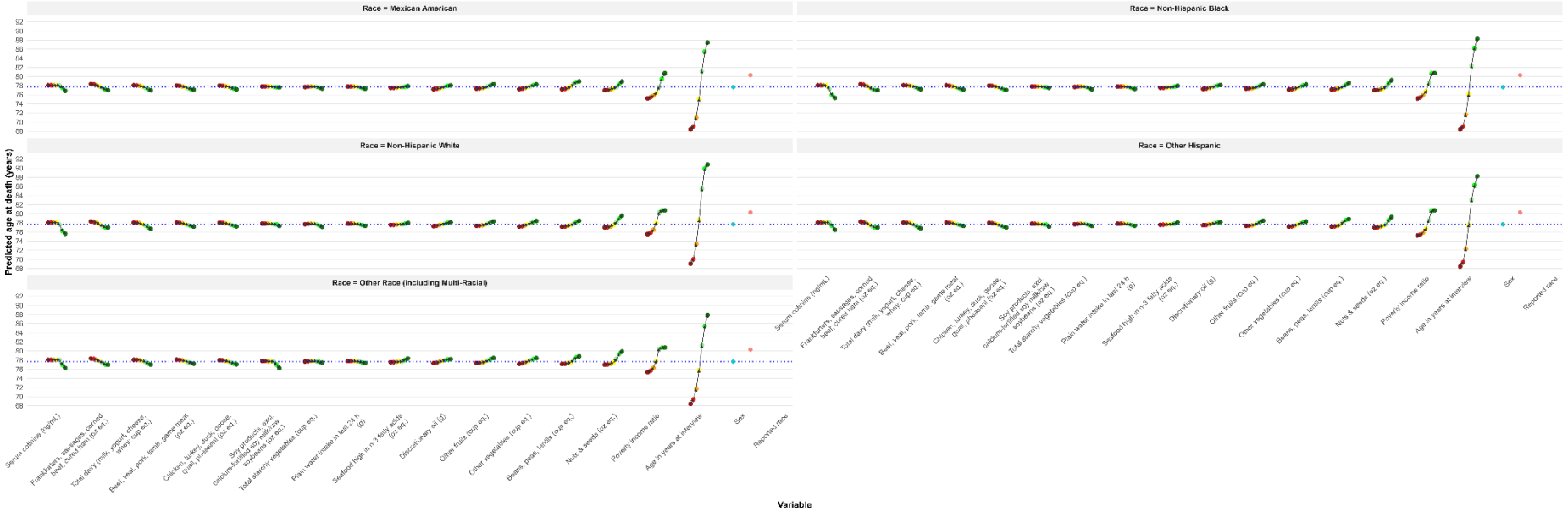
A

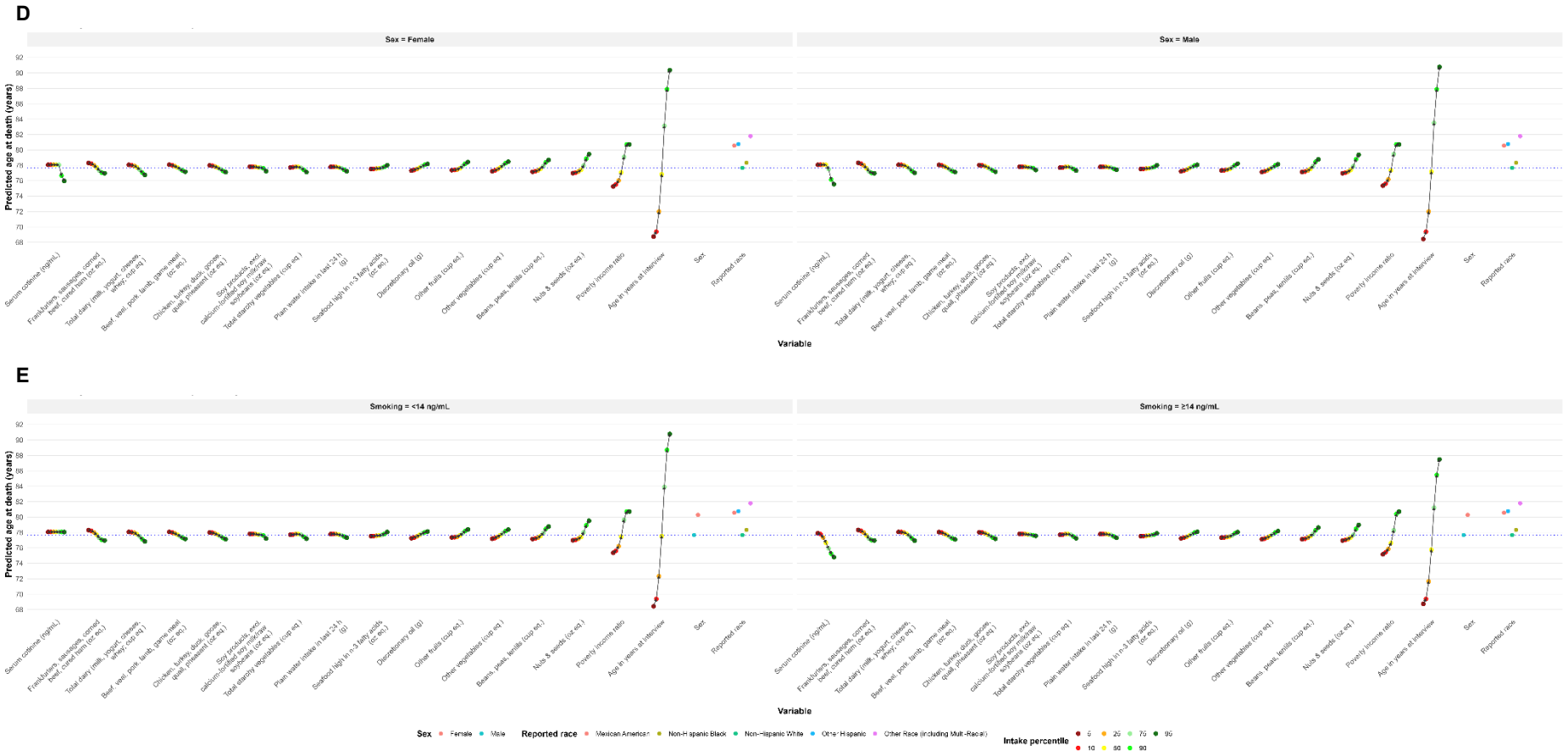

AgeBand = <20
AgeBand = 20–29
AgeBand = 30–39
AgeBand = 40–49
AgeBand = 50–59
AgeBand = 60–69
AgeBand = 70–79
AgeBand = 80+
Predicted age at death (years)
Serum cotinine (ng/mL)
Frankfurters, sausages, corned beef, cured ham (oz eq.)
Total dairy (milk, yogurt, cheese, whey, cup eq.)
Beef, veal, pork, lamb, game meat (oz eq.)
Chicken, turkey, duck, goose, quail, pheasant (oz eq.)
Soy products, excl. calcium-fortified soy milk (oz eq.)
Total starchy vegetables (cup eq.)
Plain water intake in last 24 h (g)
Seafood high in n-3 fatty acids (oz eq.)
Discretionary oil (g)
Other fruits (cup eq.)
Other vegetables (cup eq.)
Beans, peas, lentils (cup eq.)
Nuts & seeds (oz eq.)
Poverty income ratio
Age in years at interview
Sex
Reported race
Variable

B


PIR = >1
PIR = ≤1
Predicted age at death (years)
Variable
Serum cotinine (ng/mL)
Frankfurters, sausages, corned beef, cured ham (oz eq.)
Total dairy (milk, yogurt, cheese, whey; cup eq.)
Beef, veal, pork, lamb, game meat (oz eq.)
Chicken, turkey, duck, goose, quail, pheasant (oz eq.)
Soy products, excl. calcium-fortified soy milk/raw soybeans (oz eq.)
Total starchy vegetables (cup eq.)
Plain water intake in last 24 h (g)
Seafood high in n-3 fatty acids (oz eq.)
Discretionary oil (g)
Other fruits (cup eq.)
Other vegetables (cup eq.)
Beans, peas, lentils (cup eq.)
Nuts & seeds (oz eq.)
Poverty income ratio
Age in years at interview
Sex
Reported race

C


Race = Mexican American
Race = Non-Hispanic Black
Race = Non-Hispanic White
Race = Other Hispanic
Race = Other Race (including Multi-Racial)
Predicted age at death (years)
Variable
Serum cotinine (ng/mL)
Frankfurters, sausages, corned beef, cured ham (oz eq.)
Total dairy (milk, yogurt, cheese, whey; cup eq.)
Beef, veal, pork, lamb, game meat (oz eq.)
Chicken, turkey, duck, goose, quail, pheasant (oz eq.)
Soy products, excl. calcium-fortified soy milk/raw soybeans (oz eq.)
Total starchy vegetables (cup eq.)
Plain water intake in last 24 h (g)
Seafood high in n-3 fatty acids (oz eq.)
Discretionary oil (g)
Other fruits (cup eq.)
Other vegetables (cup eq.)
Beans, peas, lentils (cup eq.)
Nuts & seeds (oz eq.)
Poverty income ratio
Age in years at interview
Sex
Reported race



Supplementary Figure SI10: Sensitivity analysis: raindrop plots stratified by participant subgroups. Multi-panel raindrop plots display predicted age at death from the final penalized AFT model across selected intake percentiles for continuous exposures (5th, 10th, 25th, 50th, 75th, 90th, 95th) and across category levels for categorical variables, with all other continuous covariates held at their median (50th percentile) (and categorical covariates at their reference levels) within each stratum. Points are connected within each exposure to show changes in predicted

age at death across the exposure distribution; the horizontal dashed line denotes the overall mean predicted life expectancy of the cohort (77.7 years). Panels show stratification by **A** age band, **B** poverty income ratio (PIR; ≤1 vs >1), **C** reported race, **D** sex and **E** smoking status (serum cotinine <14 vs ≥14 ng/mL).

Supplementary Table SI1: Descriptive statistics of the study population and predictors included in the analysis (n=44,666 participants; 29 predictors).

Variable	Variable description	Type	Level	Mean	Median [25th, 75th]	SD	Counts (%)
GENDERRC	Sex (Male)	Demographics	1	-	-	-	21440 (48%)
	Sex (Female)	Demographics	2	-	-	-	23226 (52%)
RIDRETH1	Reported race and Hispanic origin information (non-Hispanic White)	Demographics	3	-	-	-	18940 (42.4%)
	Reported race and Hispanic origin information (Mexican American)	Demographics	1	-	-	-	8760 (19.6%)
	Reported race and Hispanic origin information (other Hispanic)	Demographics	2	-	-	-	2852 (6.4%)
	Reported race and Hispanic origin information (non-Hispanic Black)	Demographics	4	-	-	-	10280 (23%)
	Reported race and Hispanic origin information (other Race including multi-Racial)	Demographics	5	-	-	-	3834 (8.6%)
RIDAGEYR	Age in years at interview	Demographics	-	46.42	45 [30, 62]	18.61	-
INDFMPIR	Poverty income ratio	Demographics	-	2.43	2.04 [1.07, 3.76]	1.59	-
MORTSTAT	Mortality status until Dec. 31, 2018 (Alive)	Mortality	0	-	-	-	36797 (82.4%)
	Mortality status until Dec. 31, 2018 (Deceased)	Mortality	1	-	-	-	7869 (17.6%)
End	Age in years at mortality follow-up	Mortality	-	58.82	58.92 [45.17, 73]	18.33	-
LBXCOT	Serum cotinine (ng/mL)	Chemicals	-	60.66	0.11 [0.02, 21.28]	131.67	-
DRX.320Z	Plain water intake in last 24 h (g)	Dietary	-	1104.68	987.12 [572.79, 1472.25]	715.55	-

DRXTKCAL	Total energy intake (kcal)	Dietary	-	2053.13	1993.22 [1680.73, 2369.44]	518.67	-
DRXT_A_DRINKS	Alcoholic drinks (number of drinks)	Dietary	-	0.5	0.13 [0.06, 0.35]	0.92	-
DRXT_D_TOTAL	Total dairy (milk, yogurt, cheese, whey; cup eq.)	Dietary	-	1.43	1.36 [0.96, 1.79]	0.66	-
DRXT_F_CITMLB	Citrus, melons, berries (cup eq.)	Dietary	-	0.27	0.18 [0.1, 0.33]	0.25	-
DRXT_F_OTHER	Other fruits (cup eq.)	Dietary	-	0.51	0.38 [0.24, 0.69]	0.34	-
DRXT_G_REFINED	Refined grains (oz eq.)	Dietary	-	5.71	5.55 [4.53, 6.7]	1.7	-
DRXT_G_WHOLE	Whole grains (oz eq.)	Dietary	-	0.8	0.61 [0.42, 1.06]	0.51	-
DRXT_OILS	Discretionary oil (g)	Dietary	-	20.66	19.84 [15.24, 25.55]	7.76	-
DRXT_PF_CUREDMEAT	Frankfurters, sausages, corned beef, cured ham (oz eq.)	Dietary	-	0.83	0.7 [0.54, 1.03]	0.39	-
DRXT_PF_EGGS	Eggs and egg substitutes (oz eq.)	Dietary	-	0.57	0.48 [0.36, 0.71]	0.28	-
DRXT_PF_MEAT	Beef, veal, pork, lamb, game meat (oz eq.)	Dietary	-	1.7	1.58 [1.23, 2.06]	0.63	-
DRXT_PF_NUTSDS	Nuts and seeds (oz eq.)	Dietary	-	0.53	0.33 [0.26, 0.58]	0.5	-
DRXT_PF_POULT	Chicken, turkey, duck, goose, quail, pheasant (oz eq.)	Dietary	-	1.5	1.41 [1.14, 1.73]	0.44	-
DRXT_PF_SEAFD_HI	Seafood high in n-3 fatty acids (oz eq.)	Dietary	-	0.13	0.1 [0.09, 0.13]	0.11	-
DRXT_PF_SEAFD_LOW	Seafood low in n-3 fatty acids (oz eq.)	Dietary	-	0.44	0.35 [0.32, 0.47]	0.2	-
DRXT_PF_SOY	Soy products, excl. calcium-fortified soy milk/raw soybeans (oz eq.)	Dietary	-	0.05	0.03 [0.01, 0.04]	0.12	-
DRXT_SOLID_FATS	Discretionary solid fat (g)	Dietary	-	39.08	37.3 [28.99, 47.35]	14.05	-
DRXT_V_DRKGR	Dark green vegetables (cup eq.)	Dietary	-	0.12	0.08 [0.07, 0.12]	0.09	-
DRXT_V_LEGUMES	Beans, peas, lentils (cup eq.)	Dietary	-	0.13	0.1 [0.08, 0.15]	0.1	-

DRXT_V_OTHER	Other vegetables (cup eq.)	Dietary	-	0.52	0.51 [0.39, 0.64]	0.18	-
DRXT_V_REDOR_OTHER	Other red and orange vegetables, excl. tomatoes (cup eq.)	Dietary	-	0.09	0.07 [0.06, 0.1]	0.05	-
DRXT_V_REDOR_TOMATO	Tomatoes and tomato products (cup eq.)	Dietary	-	0.28	0.27 [0.22, 0.34]	0.09	-
DRXT_V_STARCHY_TOTAL	Total starchy vegetables (cup eq.)	Dietary	-	0.45	0.43 [0.35, 0.52]	0.13	-

Supplementary Table SI2: Comparison of participant characteristics between complete cases included in the analysis and participants excluded due to missing data (n=21,052). Categorical variables are summarized as counts (percentages) and continuous variables as mean (standard deviation). P values correspond to group comparisons to assess potential selection bias associated with complete-case analysis.

Variable description	Type	Level	Complete cases*	Excluded cases*	p value**
Sex (Male)	Categorical	1	21504 (48.1%)	31710 (49.2%)	<0.001
Sex (Female)	Categorical	2	23242 (51.9%)	32802 (50.8%)	<0.001
Reported race and Hispanic origin information (non-Hispanic White)	Categorical	3	18970 (42.4%)	20931 (32.4%)	<0.001
Reported race and Hispanic origin information (Mexican American)	Categorical	1	8770 (19.6%)	17807 (27.6%)	<0.001
Reported race and Hispanic origin information (other Hispanic)	Categorical	2	2856 (6.4%)	3805 (5.9%)	<0.001
Reported race and Hispanic origin information (non-Hispanic Black)	Categorical	4	10311 (23%)	17354 (26.9%)	<0.001
Reported race and Hispanic origin information (other Race including multi-Racial)	Categorical	5	3839 (8.6%)	4615 (7.2%)	<0.001
Age in years at interview	Continuous	-	46.4 (18.6)	20.8 (21.4)	<0.001
Poverty income ratio	Continuous	-	2.4 (1.6)	2.1 (1.5)	<0.001
Mortality status until Dec. 31, 2018 (Alive)	Categorical	0	36869 (82.4%)	16314 (77.5%)	<0.001
Mortality status until Dec. 31, 2018 (Deceased)	Categorical	1	7877 (17.6%)	4738 (22.5%)	<0.001
Age in years at mortality follow-up	Continuous	-	58.8 (18.3)	60.8 (18.5)	<0.001
Serum cotinine (ng/mL)	Continuous	-	60.8 (131.8)	22.2 (77.3)	<0.001
Plain water intake in last 24 h (g)	Continuous	-	1105.1 (716.1)	776.9 (610.4)	<0.001
Total energy intake (kcal)	Continuous	-	2056.5 (527.9)	1919.8 (512.7)	<0.001
Alcoholic drinks (number of drinks)	Continuous	-	0.5 (0.9)	0.2 (0.7)	<0.001
Total dairy (milk, yogurt, cheese, whey; cup eq.)	Continuous	-	1.4 (0.7)	1.9 (0.8)	<0.001
Citrus, melons, berries (cup eq.)	Continuous	-	0.3 (0.3)	0.4 (0.3)	<0.001
Other fruits (cup eq.)	Continuous	-	0.5 (0.3)	0.6 (0.3)	<0.001

Refined grains (oz eq.)	Continuous	-	5.7 (1.7)	5.7 (1.9)	0.1
Whole grains (oz eq.)	Continuous	-	0.8 (0.5)	0.6 (0.4)	<0.001
Discretionary oil (g)	Continuous	-	20.7 (7.8)	17 (6.7)	<0.001
Frankfurters, sausages, corned beef, cured ham (oz eq.)	Continuous	-	0.8 (0.4)	0.7 (0.3)	<0.001
Eggs and egg substitutes (oz eq.)	Continuous	-	0.6 (0.3)	0.5 (0.2)	<0.001
Beef, veal, pork, lamb, game meat (oz eq.)	Continuous	-	1.7 (0.6)	1.4 (0.6)	<0.001
Nuts and seeds (oz eq.)	Continuous	-	0.5 (0.5)	0.4 (0.3)	<0.001
Chicken, turkey, duck, goose, quail, pheasant (oz eq.)	Continuous	-	1.5 (0.4)	1.2 (0.4)	<0.001
Seafood high in n-3 fatty acids (oz eq.)	Continuous	-	0.1 (0.1)	0.1 (0.1)	<0.001
Seafood low in n-3 fatty acids (oz eq.)	Continuous	-	0.4 (0.2)	0.3 (0.1)	<0.001
Soy products, excl. calcium-fortified soy milk/raw soybeans (oz eq.)	Continuous	-	0 (0.1)	0 (0.1)	<0.001
Discretionary solid fat (g)	Continuous	-	39.2 (14.2)	40.7 (14.4)	<0.001
Dark green vegetables (cup eq.)	Continuous	-	0.1 (0.1)	0.1 (0.1)	<0.001
Beans, peas, lentils (cup eq.)	Continuous	-	0.1 (0.1)	0.1 (0.1)	<0.001
Other vegetables (cup eq.)	Continuous	-	0.5 (0.2)	0.4 (0.2)	<0.001
Other red and orange vegetables, excl. tomatoes (cup eq.)	Continuous	-	0.1 (0)	0.1 (0)	<0.001
Tomatoes and tomato products (cup eq.)	Continuous	-	0.3 (0.1)	0.3 (0.1)	<0.001
Total starchy vegetables (cup eq.)	Continuous	-	0.4 (0.1)	0.4 (0.1)	<0.001

* counts (percentage) for categorical variables and mean (standard deviation) for continuous variables

** Welch's two-sample t-test for continuous variables and Pearson's chi-square test for categorical variables; all tests are unadjusted

Supplementary Table SI3: Multicollinearity test using the Variance Inflation Factor (VIF). The generalized variance inflation factor (GVIF) is reported, together with its associated degrees of freedom (df). To allow comparison across variables with different df, the normalized GVIF is presented. A VIF value of 1 indicates no correlation with other predictors, implying the absence of multicollinearity in the model.

Variable description	Type	GVIF	Df*	Normalized GVIF**
Whole grains (oz eq.)	Dietary	1.373	1	1.172
Refined grains (oz eq.)	Dietary	1.335	1	1.156
Dark green vegetables (cup eq.)	Dietary	1.354	1	1.164
Other red and orange vegetables, excl. tomatoes (cup eq.)	Dietary	1.411	1	1.188
Total starchy vegetables (cup eq.)	Dietary	1.562	1	1.250
Tomatoes and tomato products (cup eq.)	Dietary	1.468	1	1.212
Other vegetables (cup eq.)	Dietary	1.706	1	1.306
Citrus, melons, berries (cup eq.)	Dietary	1.431	1	1.196
Other fruits (cup eq.)	Dietary	1.359	1	1.166
Total dairy (milk, yogurt, cheese, whey; cup eq.)	Dietary	1.277	1	1.130
Beef, veal, pork, lamb, game meat (oz eq.)	Dietary	1.580	1	1.257
Frankfurters, sausages, corned beef, cured ham (oz eq.)	Dietary	1.356	1	1.164
Chicken, turkey, duck, goose, quail, pheasant (oz eq.)	Dietary	1.423	1	1.193
Seafood high in n-3 fatty acids (oz eq.)	Dietary	1.191	1	1.091
Seafood low in n-3 fatty acids (oz eq.)	Dietary	1.269	1	1.126
Eggs and egg substitutes (oz eq.)	Dietary	1.267	1	1.126
Soy products excl. calcium-fortified soy milk/raw soybeans (oz eq.)	Dietary	1.085	1	1.042
Nuts and seeds (oz eq.)	Dietary	1.197	1	1.094
Beans, peas, lentils (cup eq.)	Dietary	1.169	1	1.081
Discretionary oil (g)	Dietary	2.530	1	1.591
Discretionary solid fat (g)	Dietary	1.944	1	1.394
Alcoholic drinks (number of drinks)	Dietary	1.198	1	1.095

Plain water intake in last 24 h (g)	Dietary	1.362	1	1.167
Age in years at interview	Demographics	1.865	1	1.365
Reported race and Hispanic origin information	Demographics	1.827	4	1.078
Poverty income ratio	Demographics	1.232	1	1.110
Serum cotinine (ng/mL)	Chemicals	1.121	1	1.059
Sex	Demographics	1.549	2	1.245

*Number of coefficients

**Normalized GVIF = GVIF^(1/(2*Df)); comparable across different numbers of parameters

Supplementary Table SI4: Multiple-testing correction for dietary exposures using the Benjamini-Hochberg false discovery rate (FDR) procedure. Raw p values are derived from comparing a lifestyle-only Weibull AFT model with models additionally including each dietary exposure. Adjusted p values reflect FDR correction across all tested exposures. The Selected (FDR) column indicates whether each variable met the prespecified FDR threshold of 10% and was retained for further modeling (TRUE) or excluded (FALSE).

Variable description	Raw p value*	BH-FDR p value**	Selected (FDR)
Nuts and seeds (oz eq.)	<0.001	<0.001	TRUE
Total starchy vegetables (cup eq.)	<0.001	<0.001	TRUE
Total dairy (milk, yogurt, cheese, whey; cup eq.)	<0.001	<0.001	TRUE
Discretionary solid fat (g)	<0.001	<0.001	TRUE
Other fruits (cup eq.)	<0.001	0.003	TRUE
Beef, veal, pork, lamb, game meat (oz eq.)	<0.001	0.003	TRUE
Beans, peas, lentils (cup eq.)	0.001	0.003	TRUE
Plain water intake in last 24 h (g)	0.002	0.005	TRUE
Frankfurters, sausages, corned beef, cured ham (oz eq.)	0.003	0.007	TRUE
Discretionary oil (g)	0.003	0.007	TRUE
Seafood high in n-3 fatty acids (oz eq.)	0.025	0.053	TRUE
Refined grains (oz eq.)	0.032	0.057	TRUE
Chicken, turkey, duck, goose, quail, pheasant (oz eq.)	0.032	0.057	TRUE
Soy products, excl. calcium-fortified soy milk/raw soybeans (oz eq.)	0.036	0.059	TRUE
Other vegetables (cup eq.)	0.047	0.072	TRUE
Eggs and egg substitutes (oz eq.)	0.056	0.08	TRUE
Whole grains (oz eq.)	0.255	0.345	FALSE
Citrus, melons, berries (cup eq.)	0.282	0.36	FALSE
Tomatoes and tomato products (cup eq.)	0.314	0.38	FALSE
Dark green vegetables (cup eq.)	0.538	0.619	FALSE
Other red and orange vegetables, excl. tomatoes (cup eq.)	0.587	0.643	FALSE
Alcoholic drinks (number of drinks)	0.782	0.818	FALSE
Seafood low in n-3 fatty acids (oz eq.)	0.845	0.845	FALSE

*Likelihood-ratio tests comparing a lifestyle-only Weibull AFT model versus the model additionally including each exposure

**Benjamini-Hochberg adjusted across all exposures to control the false discovery rate

Supplementary Information SI1: Penalized AFT model shows robust performance across resampling

Across repeated resampling iterations within the training data, the penalized AFT framework showed stable discrimination. Concordance assessed on the out-of-bag (OOB) data among deceased participants using Harrell's C-index was 0.777 (95% CI: 0.766-0.792), whereas whole population-wide discrimination assessed using Uno's C-index with IPCW was 0.565 (95% CI: 0.504-0.657), a more modest value consistent with the high proportion of right-censoring (~80%) and sparse observed events in the full cohort (Supplementary Figure SI3); discrimination was considered together with calibration. Calibration was age-dependent. IPCW-based calibration slopes exceeded 1 at ages 75 and 80, while calibration intercepts were negative at these ages, indicating some miscalibration. Calibration approached the ideal (slope ~1, intercept ~0) by age 85 (Supplementary Figure SI4).