

Supplementary Materials

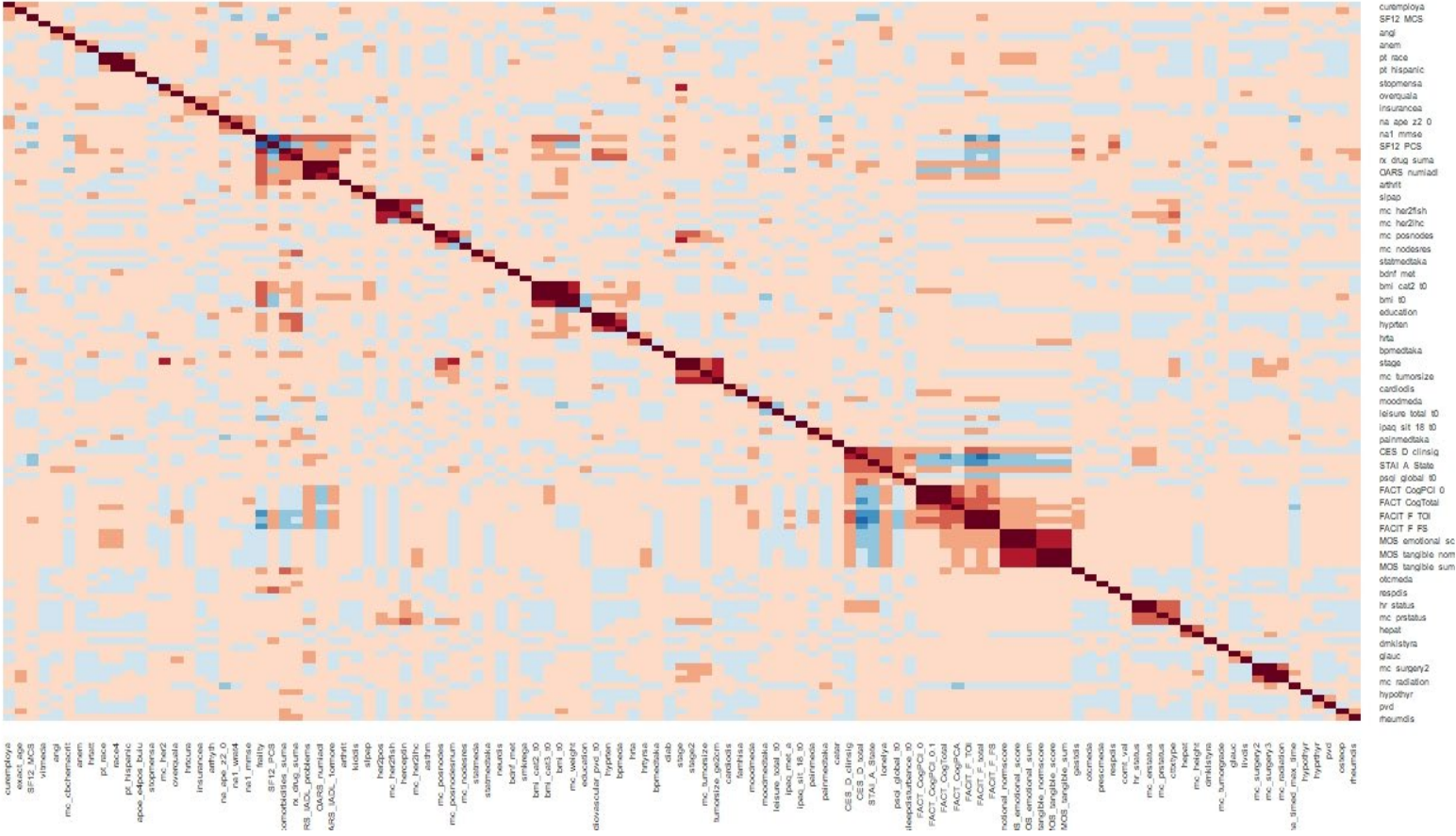
Supplementary Methods

Cross Validation: We used a k-fold cross-validation for LASSO hyperparameter tuning and other model optimization, a common procedure in machine learning to reduce overfitting. After assessing each cross-validation, we chose four-fold cross-validation which led to the most parsimonious set of predictors, which were also consistent with those selected in other machine learning approaches. As such, our analytic sample of 228 survivors was split into 4 equal sized “folds” of 57. The average error and standard deviation over the folds for a sequence of penalty parameter values are computed and we select the penalty parameter value that yields the minimum cross-validated error for the final LASSO model.

Attention, processing speed, and executive functioning (APE): Attention, processing speed, and executive functioning (APE) measure consists of following 6 tests;

1. Digits Forward - Number of sequences correctly recalled;
2. Digits Backward - Number of sequences correctly recalled;
3. Trail Making A - Time in seconds;
4. Trail Making B - Time in seconds;
5. Controlled Oral Word Association Test (COWAT) - Number of words given in allowed time;
6. Digit Symbol Test raw score - Number of symbols produced in allowed time.

Supplementary Figure 1. Heatmap showing the correlation between each variable included in the analysis. For comparison of two nominal variables, bias corrected Cramer’s V statistic was computed. For the comparison of two numeric variables, association was measured as the absolute value of the Pearson’s correlation coefficient. Red indicates positive correlation and blue indicates negative correlation. Measures from the same test (e.g. FACIT) are highly correlated, as are multiple categorizations (different cut-offs) of the same outcome measure. Additionally, we observe a correlation between various comorbidities, particularly cardiovascular comorbidities, and certain medications.



Supplementary Table 1

Variable Summary	Min	Max	Percent Missing	Type of variable
Key Predictors (breast cancer related variables)				
Treatment Type (Chemotherapy with or without hormonal therapy vs. hormonal therapy alone)	1	2	0%	Binary
Hormonal Replacement Therapy Use (Yes/No)	0	1	0%	Binary
Currently Using Hormonal Replacement Therapy (Yes/No)	0	1	48%	Binary
Years of Past Menopausal Hormonal Therapy Use	1	5	49%	Continuous
Estrogen Receptor Status (Positive/Negative)	1	2	0.4%	Binary
Progesterone Receptor Status (Positive/Negative)	1	2	0.9%	Binary
Hormone Receptor Status	0	1	0.4%	Binary
Positive Nodes (Yes/No)	0	1	0.4%	Binary
Nodes Resected (Yes/No)	0	1	88%	Binary
Number of Positive Nodes	0	15	0.4%	Continuous
Tumor Grade (Poor/Moderate/Well)	1	3	16%	Categorical
Tumor Size	0.1	12	2%	Continuous
Tumor Size Greater Than or Equal to 2cm (Yes/No)	0	1	2%	Binary
Radiation (Yes/No)	0	1	0%	Binary
Stage (DCIS/I/II/III/IV)	0	3	0%	Categorical
Categorical Stage Factor (0-1/2-3)	1	2	0%	Binary
Surgery (BCS/Mastectomy)	1	2	0.4%	Binary
Surgery and Radiotherapy (BCS Only/BCS with Radiation/Mastectomy)	1	3	0.4%	Categorical
Herceptin	0	1	0%	Binary
APOE E4 Positivity	0	1	4.4%	Binary
HER2 Positivity	0	1	4.4%	Binary
Non-Cancer Clinical Variables (medicine, comorbidity, frailty)				
Age at Menopause	23	63	3.1%	Continuous
Smoke Regularly (Currently/Used to/Never)	0	2	1%	Categorical
Drank Alcoholic Beverages in Last Year (Yes/No)	0	1	6.1%	Binary
Family History of Dementia or Cognitive Decline (Yes/No)	0	1	0%	Binary
Regularly Take Statin Medication (Yes/No)	0	1	0%	Binary
Took Statin Medication Yesterday or Today (Yes/No)	0	1	65%	Binary
Regularly Take Medication for Mood or Nerves (Yes/No)	0	1	0%	Binary
Took Medication for Mood or Nerves Yesterday or Today (Yes/No)	0	1	81%	Binary
Regularly Take Blood Pressure Medication (Yes/No)	0	1	0.9%	Binary
Took Blood Pressure Medication Yesterday or Today (Yes/No)	0	1	65%	Binary
Regularly Take Pain Medication (Yes/No)	0	1	0.4%	Binary
Took Pain Medication Yesterday or Today (Yes/No)	0	1	81%	Binary
Regularly Take Other Prescription Medication (Yes/No)	0	1	1%	Binary
Regularly Take Over the Counter Medication (Yes/No)	0	1	0.4%	Binary
Regularly Take Multivitamins (Yes/No)	0	1	3%	Binary
Sum of Number of Prescription Drugs	0	11	0%	Continuous
Height	0.6	167	15%	Continuous
Weight	4	360	15%	Continuous
Hematocrit	27.4	363.7	21%	Continuous

BDNF	0	1	33%	Binary
COMT	0	1	34%	Binary
Frailty	0	0.4	0%	Continuous
Angina In Last 6 Months	0	1	0%	Binary
Arrhythmia In Last 6 Months	0	1	2%	Binary
Congestive Heart Failure In Last 6 Months	0	0	0%	Binary
Hypertension In Last 6 Months	0	1	0.4%	Binary
Heart Attack In Last 6 Months	0	1	0%	Binary
Other Cardiovascular Disease In Last 6 Months	0	1	0.4%	Binary
Asthma In Last 6 Months	0	1	0%	Binary
Other Respiratory Disease In Last 6 Months	0	1	0.4%	Binary
Arthritis In Last 6 Months	0	1	1%	Binary
Other Rheumatological Disease In Last 6 Months	0	1	1%	Binary
Hip Fracture In Last 6 Months	0	0	0.4%	Binary
Osteoporosis In Last 6 Months	0	1	1%	Binary
Cataracts In Last 6 Months	0	1	1%	Binary
Glaucoma In Last 6 Months	0	1	0%	Binary
Other Gastrointestinal Disease In Last 6 Months	0	1	0%	Binary
Hepatitis In Last 6 Months	0	1	0.40%	Binary
Other Liver Disease In Last 6 Months	0	1	0%	Binary
Other Kidney In Last 6 Months	0	1	0.4%	Binary
Hyperthyroidism In Last 6 Months	0	1	0%	Binary
Hypothyroidism In Last 6 Months	0	1	0%	Binary
Peripheral Vascular Disease In Last 6 Months	0	1	0.4%	Binary
Anemia In Last 6 Months	0	1	1%	Binary
Other Neurological Disease In Last 6 Months	0	1	1%	Binary
Diabetes In Last 6 Months	0	1	0.4%	Binary
Sleep Apnea In Last 6 Months	0	1	66%	Binary
Sum of Comorbidities In Last 6 Months	0	8	0%	Continuous
Cardiovascular Comorbidities In Last 6 Months	0	1	0%	Binary
BMI Baseline	16.8	61.8	0%	Continuous
Obesity Baseline (BMI <30 Not Obese/BMI >30 Obese)	1	2	0%	Binary
Obesity Baseline (BMI <30 Not Obese/BMI 30-<35 Low Risk Obese/BMI >35 High Risk Obese)	1	3	0%	Categorical
Socio-Demographics				
Age	60	91.1	0.4%	Continuous
Patient Race (White/Black or African American/Other)	1	3	0%	Categorical
Hispanic (Yes/No)	0	1	0%	Binary
Education (In Years)	10	18	0%	Continuous
Current Employment Status (Employed/Not Employed)	1	2	0%	Binary
Insurance (Private/Medicaid/Medicare/Uninsured)	1	4	1%	Categorical
Survey & Evaluation Based (cognition, depression, anxiety, emotional, and etc.)				
MMSE	24	30	0%	Continuous

WRAT4 Word Reading	37	70	0.4%	Continuous
MMSE	24	30	0%	Continuous
WRAT4 Word Reading	37	70	0.4%	Continuous
Reported Sleep Disturbance (Yes/No)	0	1	0%	Binary
Felt Lonely in Past Week (Rarely/Some or Little/Moderate/Most of the time)	1	4	1%	Categorical
IPAQ Average Daily Sedentary Time	1	18	8%	Continuous
IPAQ Average METS per Week	0	0.9	11%	Continuous
Timed Up & Go Max Seconds	7.3	24.7	7%	Continuous
Total Number of Leisure Activities (0-11)	2	11	0%	Continuous
SF12 Mental Component Score	38.8	68.7	4%	Continuous
SF12 Physical Component Score	20.6	59.9	4%	Continuous
FACT Cog PCI sub-scale	20	72	0%	Continuous
FACT Cog PCA sub-scale	5	28	0%	Continuous
FACT Cog Total	46	148	0.4%	Continuous
MOS Emotional Support Score	1.9	5	3%	Continuous
MOS Emotional Support Normalized Score	21.9	100	3%	Continuous
MOS Emotional Support Sum	15	40	3%	Continuous
MOS Tangible Support Score	1.5	5	3%	Continuous
MOS Tangible Support Normalized Score	12.5	100	3%	Continuous
MOS Tangible Support Sum	6	20	3%	Continuous
CES D Total of Items	0	43	2%	Continuous
CES D Clinically Significant Depression (≥ 16 Depressed/ < 16 Not Depressed)	0	1	2%	Binary
STAI A-Scale State	22	61	1%	Continuous
OARS Instrumental ADL 1 or More Impairments (Yes/No)	0	1	0%	Binary
OARS Instrumental ADL Impairments (None/Mild/Moderate/Severe)	0	3	0%	Categorical
OARS Instrumental ADL Total Impairments	0	7	0%	Continuous
FACIT-F Fatigue Subscale	7	52	0.4%	Continuous
FACIT-F Trial Outcome Index	12.3	100	0.4%	Continuous
FACIT-F Total Score	42.3	136	1%	Continuous
APE Z-Score (NAB Digits Forward, NAB Digits Backward, Trail Making A & B, Digit Symbol Test, & COWAT)	-3.1	1.7	0%	Continuous

Supplementary Table 2: The top 10 influential variables for other machine learning explored in this analysis are recorded for classifying transient and persistent CRCD. For stochastic gradient boosting machines (SGB) the variable importance is determined by calculating the change in squared error, over all trees, when a variable is chosen to split on in the tree building process. For random forest (RF) models, the variable importance is determined by calculating the mean decrease in model accuracy on out of bag samples when a given variable is excluded from the model.

Table 2A: Influential Variables for Other Machine Learning Models to Predict Transient 12-Month Cognitive Decline	
Stochastic Gradient Boosting (SGB)	Random Forest (RF)
BMI Baseline	APOE E4 Pos
APE Baseline	BMI Baseline
IPAQ Sitting Hours	FACIT F TOI
APOE E4 Pos	BMI Obesity (3 Category)
SF12 PCS	APE Baseline
Age	IPAQ Sitting Hours
NA Timed – Max	Regularly Takes Pain Meds
FACIT F Total	STAI A State
IPAQ Met. Avg	FACIT F FS
SF12 MCS	Hispanic (Y/N)

Table 2B: Influential Variables for Other Machine Learning Models to Predict Persistent 24-Month Cognitive Decline	
SGB	RF
BMI Baseline	BMI Baseline
SF12 MCS	CES D Total
IPAQ Sitting Hours	Sum of Presc. Drugs
Frailty	BMI Obesity
IPAQ Met Avg	FACT Cog PCI Baseline
FACIT F FS	FACIT F FS
APE Baseline	Cataracts
Regularly Take Pain Meds	IPAQ Sitting Hours
NA Timed - Max	APE Baseline
FACIT F Total	Frailty

Supplementary Table 3. LASSO selected Predictors of the Risk of 12-Month Transient Cognitive Decline among Older Women with Breast Cancer¹

Model with Baseline APE Domain Score ²				
Variable	Regression Coefficient	Standard Error	Z-Value	p-value
Intercept	-1.23	0.50	-2.46	0.01
Treatment Type – Chemo ³	-0.01	0.37	-0.02	0.98
Cardiovascular Disease Comorbidity	0.50	0.36	1.37	0.17
BMI >30 ⁴	-0.73	0.34	-2.13	0.03
ApoE E4 Positivity (vs. E4 negative)	1.28	0.38	3.40	<0. 01
Number of Prescription Drugs Regularly Taken, per one drug	0.12	0.09	1.29	0.20
HER2 Positivity (vs. negative)	-1.67	0.86	-1.95	0.05
Baseline APE Domain Z-Score	0.85	0.27	3.09	<0.01
Model without Baseline APE Domain Score				
Variable	Regression Coefficient	Standard Error	Z-Value	p-value
Intercept	-1.15	0.49	-2.36	0.02
Treatment Type – Chemo	-0.02	0.36	-0.05	0.96
Cardiovascular Disease Comorbidity	0.44	0.35	1.27	0.21
BMI >30	-0.68	0.33	-2.05	0.04
ApoE E4 Positivity (vs. E4 negative)	1.26	0.37	3.43	<0. 01
Number of Prescription Drugs Regularly Taken, per one drug	0.10	0.09	1.14	0.25
HER2 Positivity (vs. negative)	-1.40	0.82	-1.70	0.09
Baseline APE Domain Score				

¹Logistic regression model with least absolute shrinkage and selection operator (LASSO) was used to identify predictors indicative of risk of cognitive decline. 4-fold cross-validation was used during the model optimization of tuning/penalty parameters to obtain the model selected variables and corresponding regression coefficient estimates for the displayed LASSO model outputs. Regression coefficients, standard errors, z-values, and p-values are summarized. The model output in the top panel corresponds to an Area Under the Curve (AUC) of 0.750, while the model without a measure of baseline cognitive reserve (APE domain score) corresponds to an AUC of 0.710.

²The APE domain is a summary Z-score of performance on six neuropsychologic tests of attention, processing speed and executive function.

³Systemic treatment categories are for chemotherapy with or without endocrine therapy vs. the referent group of endocrine therapy alone.

⁴BMI = weight (kg) / height (m²) at baseline

Supplementary Table 4. LASSO selected Predictors of the Risk of 24-Month Persistent Cognitive Decline among Older Women with Breast Cancer¹

Model with Baseline APE Domain Score ²				
Variable	Regression Coefficient	Standard Error	Z-Value	p-value
Intercept	6.33	2.97	2.13	0.03
Treatment Type – Chemo ³	-0.28	0.49	-0.57	0.57
Cardiovascular Disease Comorbidity	0.30	0.54	0.56	0.58
BMI >30 ⁴	-0.79	0.45	-1.74	0.08
ApoE E4 Positivity (vs. E4 negative)	0.90	0.50	1.80	0.07
Number of Prescription Drugs Regularly Taken, per one drug	0.29	0.12	2.32	0.02
HER2 Positivity (vs. negative)	-0.42	0.96	-0.44	0.66
IPAQ Daily Sitting Hours	-0.15	0.07	-2.22	0.03
SF12 Mental Component Score	-0.14	0.05	-2.79	0.01
Baseline APE Domain Score	0.57	0.36	1.58	0.11
Baseline Mini Mental Status Exam ⁵				
Years Of Education				
Model without Baseline APE Domain Score				
Variable	Regression Coefficient	Standard Error	Z-Value	p-value
Intercept	22.66	6.18	3.67	<0.01
Treatment Type – Chemo	-0.54	0.52	-1.04	0.30
Cardiovascular Disease Comorbidity	0.48	0.56	0.86	0.39
BMI >30	-0.99	0.47	-2.09	0.04
ApoE E4 Positivity (vs. E4 negative)	0.96	0.52	1.83	0.07
Number of Prescription Drugs Regularly Taken, per one drug	0.31	0.13	2.33	0.02
HER2 Positivity (vs. negative)	-0.48	0.93	-0.51	0.61
IPAQ Daily Sitting Hours	-0.15	0.07	-2.09	0.04
SF12 Mental Component Score	-0.17	0.05	-3.38	<0.01
Baseline APE Domain Score				
Baseline Mini Mental Status Exam ⁵	-0.59	0.18	-3.32	<0.01
Years Of Education	0.17	0.12	1.45	0.15

¹Logistic regression model with least absolute shrinkage and selection operator (LASSO) was used to identify predictors indicative of risk of cognitive decline. 4-fold cross-validation was used during the model optimization of tuning/penalty parameters to obtain the model selected variables and corresponding regression coefficient estimates for the displayed LASSO model outputs. Regression coefficients, standard errors, z-values, and p-values are summarized. The model output in the top panel corresponds to an Area Under the Curve (AUC) of 0.788, while the model on the bottom panel with MMSE and Years of Education as a substitute measure of baseline cognitive reserve corresponds to an AUC of 0.832.

²The APE domain is a summary Z-score of performance on six neuropsychologic tests of attention, processing speed and executive function.

³Systemic treatment categories are for chemotherapy with or without endocrine therapy vs. the referent group of endocrine therapy alone.

⁴BMI = weight (kg) / height (m²) at baseline

⁵The Mini Mental Status Exam scores are truncated at >24 based on study eligibility.