**Supplementary Methods II. Methodological and Analytical**

*Study Population*

Clinical, protein, and genetic [biomarker](https://www.sciencedirect.com/topics/neuroscience/biomarkers) samples were from participants of the longitudinal study in aging *Invecchiare in Chianti* (Aging in Chianti, “InCHIANTI Study”), collected at baseline study entry from 1998-2000. This group is a representative sample (n=1,453) of the population of white European origin from two small towns in Tuscany, Italy. The primary aim of the InCHIANTI study was to evaluate physical function and mobility in older community-dwelling individuals. A detailed description of the study design, data collection, and sampling procedure are published elsewhere.18 This secondary study was approved by the ethics committee at *Centre de recherché Clinique du CHUS* and by the Institutional Review Board of the Azienda Sanitaria Area 10 Firenze; all participants consented to participate after having received a full description of the study project #548.

*Predictive Measures*

The International Consensus Group’s (I.A.A.A. /I.A.G.G.) list of potential biomarkers is not meant to be complete, accurate, or exhaustive.1 Our team used a previously published systematic review which identified shared biological markers for physical frailty and cognitive impairment, a total of 289 variables were available in the InCHIANTI database.11

*Genetic and Protein Markers*: After removal of any variables with > 12% missing data, there were 132 putative SNPs and 155 protein biomarkers. To build our model, we used protein markers with implications for clinical research and practice. Genetic risk score estimates (GRS) (i.e. the cumulative genetic risk burden estimated from SNPs of interest), were completed before including the individual single nucleotide polymorphisms (SNPs) in the final models (see GitHub supplementary material table III).

*Clinical Features*: Included age, sex, level of education, anticholinergic burden, and depressive symptoms. Anticholinergic medications were identified in the systematic review as a significant risk factor for cognitive impairment and physical frailty.11 Anticholinergic burden was calculated using the Anticholinergic Cognitive Burden Scale (ACB) and examined as a predictor.19 Current depressive symptomatology was measured using the Center for Epidemiologic Studies Depression Scale (CES-D) self-report scale (0-60), with cutoff points 16 or greater indicating depression.20 Reliability, validity, and factor structure have been similar across a diverse demographic and the scale has been used extensively in epidemiologic studies for depression and physical function.21

*Outcome Measure*

*Cognitive Frailty Measures*: Three instruments were used to measure neuropsychological function of cognitive frailty as determined by Delrieu et al.22 The Mini-Mental State Examination (MMSE) as a test of global mental status and the Trail Making Test, Part A and B (TMT). Attention was assessed using the Trail Making Test (TMT-A) and executive function was assessed using the Trail Making Test (TMT-B) scoring based on time in seconds to completion of a task with a score range of 0 to 300 seconds.23 Individuals who required additional time to complete TMT A and B, but were able to complete the task, were included and assigned a score of 300. Prior research has established precedent for including individuals who complete the TMT over the 300 second time limit to distinguish this group from individuals who cannot complete the task independent of the time limit.24,25 The rationale to include >300 completers in these analyses was made due to the possible presence of slowing without confusion in the target population of individuals with cognitive frailty. TMT, part A and B cut off scores are based on established norms for mild neurocognitive disorders.25 Normative data for time to complete the TMT tests in seconds was stratified by age and education category.25 The InCHIANTI criteria for frailty is defined by Fried et al. as exhaustion, slowness, low physical activity, weakness, and weight loss.3 Additional description of the InCHIANTI data collection and frailty classifications have been previously published.26,27

*Cognitive Frailty Phenotype:* Individuals with evidence of both physical frailty and cognitive impairment without a baseline clinical diagnosis of Alzheimer’s Disease or other dementia were defined as having the cognitive frailty phenotype.22 Phenotypic classification for this study included two models compared with healthy controls.

Healthy controls were defined as robust:

* No physical frailty (≦ 1 criteria) and absence of cognitive impairment (MMSE ≧ 24; Trail A ≦ 78, Trail B ≦ 106).

Model I considers participants with an MMSE ≦ 23 as having cognitive impairment and individuals with one or more of the physical frailty criteria are considered frail.3,25,28

* Physical frailty (≧ 1 criteria) and cognitive impairment (MMSE ≦ 23)

Model II considers participants who completed the MMSE with additional neuropsychiatric testing TMT, Part A and B.24,25 TMT cut off scores for cognitive impairment are based on cut off norms established by Ashendorf et al., 2008 and Strauss et al, 2006.

* Physical frailty (≧ 1 criteria) and cognitive impairment (Trail A ≧ 78, Trail B ≧106)

Numbers of participants with moderate and severe cognitive impairment were insufficient for inclusion as separate categories for pre-frail and frail phenotypes in statistical analyses.

*Statistical Analysis (*Figure 2 shows a summary of our workflow)

We used a cross-sectional (study baseline) data to develop boosted ML models for identification of features associated with cognitive frailty. Then we further qualified the clinical and biological predictors identified in the model by examining their significance by healthy control and cognitive frailty phenotype. Covariates were selected to control for potential confounding effects, including sex, age, education, baseline diagnosis of dementia (n=82), vascular dementia (n=41), depression (n=412), and Parkinson’s disease (n=16).

Model Development: Mechanisms that contribute to the development of cognitive frailty were determined by evaluation of genetic variability, protein and clinical markers as predictors of the development and persistence of cognitive frailty. *Model I* tested prediction of genetic, protein and clinical markers on cognitive frailty with the use of criteria from the MMSE while *Model II* tested prediction of genetic, protein, and clinical markers on cognitive frailty with use of additional neuropsychological testing TMT, Part A and B.24,25

Using a Boosted Tree Approach for Data Pruning: Boosted trees, a machine learning technique for supervised learning, are ensembles of regression trees, similar to decision trees and are used for prediction or classification. The advantage of using a tree boosting approach model for the evaluation of multiple variables simultaneously is that it provides a high predictive value with low bias.29 Additionally, parameters are set to prevent over fitting for the models. Extreme Gradient Boosting (xgboost) in R, statistical software, is an effective method for building a reproducible predictive model for the detection of a complex heterogeneous phenotype such as cognitive frailty with large numbers of predictors. *Xgboost* is based in boosted trees and provides more efficient and accurate predictive modeling with large datasets and a rapid / robust framework for variable selection. Statistical modeling is used to design, test, and validate an accurate method for classifying patients into phenotypic outcomes. The statistical analysis was completed in three steps: 1) analysis of all available variables for feature selection and data reduction, 2) model discovery followed by model validation, and 3) determination of significance in the model features between phenotype and healthy control.

Step 1) The data were randomly divided, two thirds were assigned to the training cohort, and one third was assigned to the validation cohort. One of the features that is central to *xgboost* is its ability to combine multiple trees or “weak predictors” to reach maximum prediction performance while reducing bias. This approach uses large amounts of data from different aspects of clinical, genetic, and biomarker research, strengthening the models’ generalizability and classification power. *Xgboost* iteratively re-weighs the variables, taking a weighted majority; the parameters identified after pruning comprised the final predictive model.30 None of the candidate features in the models are used in the diagnosis of cognitive frailty. This standard technique prevents circularity, overestimation, and over fitting for both the models generated. Parameters for the model include: max depth = “10”, nthread = “12”, nrounds = 5-200, objective = “binary:logistic”, evaluation metric = “auc”, silent =”1”, gamma = default =“0” to control the number of trees, and eta default= “0.3” to prevent over fitting. We used the default setting for all other parameters which can be found in the *xgboost* 0.6 documentation.14 The *xgboost* algorithm iteratively determines the maximum function of a model based on a tree building algorithm (quadratic problem) which creates a node then assigns a prediction point to each leaf; the assigned number is termed “gain” (figure 1). Once the model has reached maximum depth, pruning occurs by taking out the nodes with a negative gain and keeping those with a positive gain. Results from the population predictive model are ranked by gain which is a metric based on each feature’s contribution in the model. When comparing top features to other features in the model, the higher the gain the more important the feature is for prediction of the outcome. *Cover* is a measure of the relative quantity of observations found by one feature and *frequency* is the percentage representing the relative number of time a feature is used in the trees of the model.14 Gain is the most relevant metric to interpreting the rank and importance of each feature.

Step 2. To evaluate the models, we used the evaluation metric area under the curve (AUC). AUC were calculated from each model and used to determine discrimination of participants with cognitive frailty (case) from healthy individuals (control) in the training cohort. An AUC of 0.5 was considered chance, > 0.8 informative, and > 0.9 clinically relevant.

Step 3. Univariate analysis, *t-*tests for continuous and chi-squared tests for binomial traits, were used to determine the significance of the predictor with a Bonferroni correction (to account for multiple comparisons). Of note, no further adjustment for multiple comparison was carried out due to the exploratory, hypothesis-generating nature of this study to protect against type II error (false negative) rather than type I. To evaluate additive effects of SNPs, a positive regression coefficient means that each copy of the allele of interest increases the risk for the cognitive frailty phenotype.31 Our study used the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health (Bethesda, MD, USA).

Boosted ML models use “ensembling”, a technique that can combine many independent predictors to model interacting systems and determine the top combined predictive factors associated with cognitive frailty.17 At least in theory, the determination of genetic and biological markers that define a clinical group should facilitate a better understanding of the interrelated pathology for cognitive impairment and physical frailty and promote new ideas for understanding complex interacting biological systems. The ML boosted model employs an unbiased logic that does not make assumptions about the relationship among predictors or apply weights in the scoring system. Boosting is an ensembling technique that employs a sequential algorithm which repetitively learns and improves as the model reaches a final prediction. Parameter estimates for each predictive factor and associated descriptive statistics were evaluated to provide biological insight into the underpinnings of the classification algorithm.

