**Design and analysis (study design, inclusion/exclusion, outcome and other variables of interest with specific reference to the time of their collection, summary of data analysis, and anticipated methodologic limitations or challenges).**

***1. Methods/Study Design***

A cross-sectional predictive model comparing predictors with robust, pre-frail and frail participants developed from the baseline InCHIANTI data will be used to replicate a population predictive model in ARIC. The ARIC model is run separately with a comparison of the AUC. The results discuss the attributable AUC differences in part based on outcome derivations. An Extreme Gradient Boosting (xgboost) in R12 for the validation model in ARIC. While boosting was initially developed for machine learning, ‘xgboost’ in R is based in boosted trees. Xgboost is an open-source tool and a variant of the gradient boosting machine and uses a tree-based model. Xgboost is used in this study for a supervised learning problem where the variables identified are used to predict pre-frail and frail individuals. Statistical distances for model adequacy were performed in the ARIC models based on biomarker distance from the phenotype measurement V5-timepoint (S.Table 2). Figure 1 represents the workflow process that is used to replicate the frailty prediction model in the ARIC dataset.

Study approach and workflow diagram

Test

Tune

Train

1.1. Model generation

The predictive clinical and laboratory biomarkers were identified in InCHIANTI Model 1 is analyzed using an Extreme Gradient Boosting (xgboost) in R1 for the replication model in ARIC. While boosting was initially developed for machine learning, ‘xgboost’ in R is based in boosted trees. Xgboost is an open-source tool and a variant of the gradient boosting machine and uses a tree-based model. Xgboost is used in this study for a supervised learning problem where the variables identified from a previous study are used to predict pre-frail and frail individuals.2

1.2. Evaluation of the model

With the use of any predictive model in machine learning there is a chance for inflated risk of capitalizing on chance features (overfitting) in the data. Overfitting of the model is mitigated in two ways: 1) having a distinct training and validation process for the model and 2) using xgb in R which has built-in parameter settings for selection to reduce poor predictive performance. *Internal validation:* A randomly assigned training subset is be used to validate the model.

1.3. Calibration of the model

Parameter estimates for each predictive factor and associated descriptive statistics will be evaluated to provide biological insight into the underpinnings of the classification algorithm. We evaluated the calibration by partitioning the data into 5, 10, 20, 30, 40, 50, 75, 100 and 200 groups and then run the calibration test. Next, we repeated tests for all possible values between 5-200 groups and evaluated the distribution of the test statistic. The best prediction thresholds were determined using AUC. Table 3. Represents the univariate frailty prediction model fit InCHIANTI and ARIC for used for testing and calibration. Multivariate model used for the final prediction models (details described in the manuscript) have been saved for public use in transfer learning and are available on GitHub for this report.

1.4. Frailty Phenotype

The frailty phenotype defined in three categories—robust (0), pre-frail (1-2), and frail (3-5).3,4 The outcome measure of frailty is used from V5 in the ARIC dataset. For this study the V5 derived variable frailty52 is used. The frailty52 variable uses a missing at random (MAR) assumption to recover information amongst those participants with some but not all of the 5 frailty characteristics. If the participant has all 5 indicators observed, they are considered frail, 3 characteristics observed they are classified as frail, 1-2 characteristics they are considered pre-frail. If the participant has only 4 observed characteristics than 3-4 is considered frail and 1-2 is considered pre-fail. Frailty52 yields a frailty sample size of *N*=433. The same protocol for deriving the frailty variables was used in InCHANTI; there were 298 that were <65 and not included. InCHANTI yields a frailty sample size of *N*=85. ARIC contains a frailty51 variable at V5 derived which does not use the missing at random (MAR) assumption. The composite frailty is classified by the frailty characteristics and same number of components but participants with missing information on all component characteristics were classified as missing the frailty phenotype yielding a frailty sample size of *N*=368.4

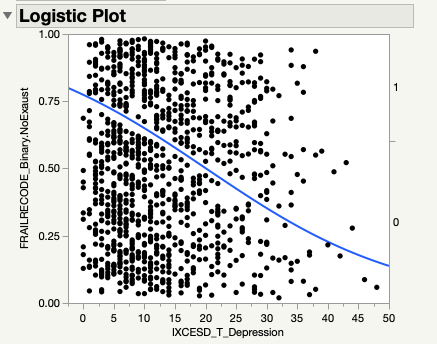
1.5. Predictors

1.5.1. Anticholinergic Burden Calculation

The Anticholinergic Cognitive Burden (ACB) scale is the most clinically validated scale for evaluating adverse health outcomes including cognitive and physical function5,6. The anticholinergic properties of each medication will be quantified using the ACB scale based on each drug’s serum anticholinergic activity7. ACB scores for each participants’ medications will be assigned points (0, 1, 2, 3) according to the published 2012 update and summed for a total anticholinergic burden score. Higher scores indicate higher anticholinergic properties. An example of medications with ACB scores include: Amitriptyline = 3, Amantadine = 2, and Atenolol = 1.

1.5.2. Depression Score

The CES-D self-report scale (0-60) is used to measure depressive symptoms. 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.8 The CES-D score will be used in the predictive model. Because the exhaustion criterion from the physical frailty definition is derived from the depression scale CES-D, a sensitivity analysis was performed by excluding the exhaustion criterion from the frailty definition. The frailty scores were recalculated on the other 4 frailty components and the analyses were repeated. The final outcomes remained statistically significant p-value < .001 (S.Figure 2.).



S.Figure 1. Frail No Exhaustion Regression Plot

1.5.3. Table 1. represents the validated biological and clinical markers used to predict frailty in the InCHIANTI data and the clinical and biological markers available for the prediction of frailty in the ARIC dataset. In ARIC, 3 biological markers were used as substitutes (serum TNF used instead of TNF-a receptor I&II, total testosterone used instead of Free testosterone, and serum creatinine used instead of Creatinine clearance, 24-hr urine/24-hour urinary creatinine). InCHAINTI model has a total of 19 frail clinical and biological markers and 13 total frail markers are available in ARIC. InCHAINTI model has a total of 22 pre-frail clinical and biological markers and 16 total prefrail markers are available in ARIC.

**1.6 Potential Limitations**

1.6.1 Several of the biomarker measurements come from different time points than the outcome measure of frailty at V5. The model is built with data as close to the outcome diagnosis (V5 Frailty) as possible. Additionally, we examine model parameters and AUC adding variables from visits 1 through 5. As variables are added, parameters (model fit and AUC) were examined for best fit. Initial rebuild in InCHIANTI with the predictors from the ARIC model maintained an AUC of .897. Although we cannot completely control for the varying temporal differences between some predictors and the frailty outcome, we find the findings to be informative and useful for future biomarker-based modeling studies.

1.6.2. Frailty measure differs slightly between databases. Often there are variations in how frailty is measured between longitudinal studies which may affect the AUC and the ability of the InCHIANTI model to “fit” or accurately predicted frailty in the ARIC data. However, frailty in ARIC was based on the same frailty phenotype that is used in InCHIANTI and has been validated.

1.6.3. Demographics and ethic/racial differences between databases. The InCHIANTI databases is a geographically homogeneous population with a large White European population. We consider it a strength rather than a limitation that ARIC is a more diverse populations and will allow external validation to be conducted in different cohorts. Models explored ethic and racial differences in the model fit as the model was being built.

References

1. Chen, Tianqi, He, Tong, Benesty, Michael, Khotilovich, Vadim, Tang Y. ’xgboost’-Extreme Gradient Boosting. Published online 2017. doi:10.1145/2939672.2939785>

2. Sargent L, Nalls M, Starkweather A, et al. Shared biological pathways for frailty and cognitive impairment: A systematic review. *Ageing Res Rev*. 2018;47:149-158. doi:10.1016/J.ARR.2018.08.001

3. Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: evidence for a phenotype. *J Gerontol A Biol Sci Med Sci*. 2001;56(3):M146-56.

4. Kucharska-Newton AM, Palta P, Burgard S, et al. Operationalizing Frailty in the Atherosclerosis Risk in Communities Study Cohort. *Journals Gerontol Ser A*. 2017;72(3):382-388. doi:10.1093/GERONA/GLW144

5. Salahudeen MS, Duffull SB, Nishtala PS. Anticholinergic burden quantified by anticholinergic risk scales and adverse outcomes in older people: a systematic review. Published online 2015. doi:10.1186/s12877-015-0029-9

6. Campbell NL, Boustani M, Lane K, et al. Use of anticholinergics and the risk of cognitive impairment in an African-American population. *Neurology*. 2010;75:152-159.

7. Cai X, Campbell N, Khan B, Callahan C, Boustani M. Long-term anticholinergic use and the aging brain. *Alzheimer’s Dement*. 2013;9(4):377-385. doi:10.1016/j.jalz.2012.02.005

8. Radloff LS. A Self-Report Depression Scale for Research in the General Population. *Appl Psychol Meas*. 1977;1(3):385-401. doi:10.1177/014662167700100306

S.Table 1. Biological and Clinical Predictors: InCHIANTI and ARIC (supplemental table)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **InCHIANTI** n=1,026  Baseline Data | **ARIC** | | | **Frailty**  **Feature** |
| **Clinical** | Variable | Variable | n | Time Point |  |
| Age | X | X | 6,508 | V5 | Frail/Prefrail |
| Anticholinergic Burden | X | X | 6,508 | V5 | Frail/Prefrail |
| Depression/CES-D self-report scale | X | X | 6,508 | V5 | Frail/Prefrail |
| **Inflammatory/Immunity** |  |  |  |  |  |
| 24-hour urinary cortisol (µg/24 hours) | X | NA | NA | NA | Prefrail |
| Erythrocyte sedimentation rate (ESR) (mm/hour) | X | NA | NA | NA | Frail/ Prefrail |
| Homocysteine via FPIA analysis (Âµmol/L) | X | X | 328 | V1 | Frail/ Prefrail |
| Interleukin-1 via ELISA (pg/mL) | X | X | 442 | V1 | Prefrail |
| Interleukin-6 via ELISA ultrasensitive (pg/mL) | X | X | 572 | V1 | Frail/ Prefrail |
| Monocyte chemoattractant protein-1 via Bio-Plex (pg/mL) | X | X | 604 | V1 | Prefrail |
| Soluble TNF-a receptor I via quantitative sandwich EIA (pg/mL) | X | NA\*  (serum TNF) | 170 | V4 | Frail/ Prefrail |
| Soluble TNF-a receptor II via quantitative sandwich EIA (pg/mL) | X | NA\*  (serum TNF) | 170 | V4 | Frail/ Prefrail |
| **Hematology/Liver** |  |  |  |  |  |
| Folate via RIA (ng/mL) | X | X | 6,166 | V3 | Frail/ Prefrail |
| Mean corpuscular volume (MCV) (fL) | X | X | 6,281 | V5 | Prefrail |
| Retinol via high performance liquid chromatography (µmol/L) | X | NA | NA | NA | Prefrail |
| GPT (also known as ALT) (U/L) | X | X | 5,997 | V4 | Frail |
| **Endocrine/Hormones** |  |  |  |  |  |
| 25(OH)-D (25-hydroxyvitamin D) via RIA (nmol/L) | X | X | 6,026 | V3 | Frail/ Prefrail |
| Free testosterone (ng/dL), Vermeulen | X | NA\*  (total testosterone ng/dL) | 5,555 | V4 | Prefrail |
| Blood glucose (mg/dL) | X | X | 6,108 | V5 | Frail |
| Free thyroxine, fT4 (ng/dL) | X | X | 6,409 | V5 | Frail |
| Parathyroid hormone, two-site immunoradiometric assay (pg/mL) | X | X | 6,054 | V2 | Frail |
| **Metabolomics(plasma lipids)** |  |  |  |  |  |
| Fatty acid C24:0 weight (mg/L) | X | NA | NA | NA | Prefrail |
| Lipids: HDL cholesterol (mg/dL) | X | X | 6,404 | V5 | Frail |
| **Renal/Electrolyte** |  |  |  |  |  |
| Creatine phosphokinase (U/L) | X | NA | NA | NA | Prefrail |
| Creatinine clearance, 24-hr urine (mL/minute) | X | NA\*  (urine creatinine) | 6,317 | V5 | Prefrail |
| Urine proteins (mg/dL) | X | NA | NA | NA | Frail/ Prefrail |
| 24-hour urinary creatinine (mg/24 hours) | X | NA\*  (urine creatinine) | 6,317 | V5 | Frail |
| Blood urea nitrogen (mg/dL) | X | X | 6,461 | V1 | Frail |
| **Nutrient Biomarker** |  |  |  |  |  |
| Vitamin B6 via high performance liquid chromatography (ng/mL) | X | X | 6,166 | V3 | Frail/ Prefrail |
| Vitamin E gamma tocopherol, high performance liquid chromatography (Âµmol/L) | X | X | 6,166 | V3 | Prefrail |
| Lycopene via high performance liquid chromatography (Âµmol/L) | X | NA | NA | NA | Frail |

\*substitute biomarker was used because exact biomarker match was not available, red noted >15% missing removed from model

S.Table 2. Stepwise logistic regression to measure temporal biomarker differences ARIC

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Frail Stepwise Logistic Regression** | | | | |  | **Pre-Frail Stepwise Logistic Regression** | | | |
| **V1** | **V2** | **V3** | **V4** | **V5** |  | **V1** | **V3** | **V4** | **V5** |
| AGE | AGE | AGE | AGE | AGE |  | AGE | AGE | AGE | AGE |
| BUN | BUN | BUN | BUN | BUN |  | **AUC=.60** | Folate | Folate | Folate |
| **AUC=.67** | Parathyroid | Parathyroid | Parathyroid | Parathyroid |  |  | Vit D | Vit D | Vit D |
|  | **AUC= .67** | Vit D | Vit D | Vit D |  |  | Vitamin B6 | Vitamin B6 | Vitamin B6 |
|  |  | Folate | Folate | Folate |  |  | Vitamin E IU | Vitamin E IU | Vitamin E IU |
|  |  | B6 | B6 | B6 |  |  | **AUC= .60** | Testosterone | Testosterone |
|  |  | **AUC=.68** | ALT | ALT |  |  |  | **AUC = .60** | Urine Creatinine |
|  |  |  | **AUC= .69** | HDL |  |  |  |  | Depression |
|  |  |  |  | Urine Creatinine |  |  |  |  | ACB Score |
|  |  |  |  | Fasting glucose |  |  |  |  | WBC |
|  |  |  |  | T4 |  |  |  |  | **AUC = .63** |
|  |  |  |  | ACB Score |  |  |  |  |  |
|  |  |  |  | CESD |  |  |  |  |  |
|  |  |  |  | **AUC=.72** |  |  |  |  |  |

S.Table 3. Univariate Frailty Prediction Models Fit InCHIANTI and ARIC

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **InChianti** | | |  | **ARIC** | | |
| **Measure** | **Robust/Pre.Frail** | **Pre.Frail/Frail** | **Frail/Robust** |  | **Robust/Pre.Frail** | **Pre.Frail/Frail** | **Frail/Robust** |
| **AUC 95% CI** | **0.77 (0.71-0.83)** | **0.74 (0.69-0.85)** | **0.95 (0.84-0.98)** |  | **0.65 (0.63-0.68)** | **0.62 (0.56-0.67)** | **0.80 (0.76-0.84)** |
| **Model Accuracy** | **0.7266** | **0.8253** | **0.8956** |  | **0.6249** | **0.6238** | **0.8726** |
| **Sensitivity** | **0.7647** | **0.5435** | **0.8696** |  | **0.5502** | **0.5455** | **0.5036** |
| **Specificity** | **0.7045** | **0.9333** | **0.8987** |  | **0.6955** | **0.6351** | **0.9443** |
| **Positive Predictive Value (Precision)** | **0.6000** | **0.7576** | **0.5556** |  | **0.6309** | **0.1769** | **0.5798** |
| **Negative Predictive Value** | **0.8378** | **0.8421** | **0.9793** |  | **0.6204** | **0.9067** | **0.9258** |
| **F1 Score** | **0.6724** | **0.6329** | **0.678** |  | **0.5878** | **0.2672** | **0.5391** |
| **Balance Accuracy** | **0.7346** | **0.7384** | **0.8841** |  | **0.6228** | **0.5903** | **0.7239** |

S.Table 4. ARIC Univariate Frailty Prediction Models Fit by Race

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Black** | | | |  | **White** | | |
| **Measure** | **Robust/Pre.Frail** | **Pre.Frail/Frail** | **Frail/Robust** | |  | **Robust/Pre.Frail** | **Pre.Frail/Frail** | **Frail/Robust** |
| **AUC 95% CI** | **0.60 (0.53-0.64)** | **0.57 (0.52-0.63)** | | **0.71 (0.62-0.79)** |  | **0.64 (0.61-0.67)** | **0.66 (0.64-0.69)** | **0.79 (0.74-0.84)** |
| **Model Accuracy** | **0.5672** | **0.5687** | **0.8042** | |  | **0.6332** | **0.6122** | **0.8213** |
| **Sensitivity** | **0.5732** | **0.5283** | **0.5135** | |  | **0.6407** | **0.4894** | **0.5124** |
| **Specificity** | **0.7951** | **0.6108** | **0.8571** | |  | **0.8147** | **0.7250** | **0.8764** |
| **Positive Predictive Value (Precision)** | **0.5802** | **0.5864** | **0.7958** | |  | **0.6628** | **0.6207** | **0.5001** |
| **Negative Predictive Value** | **0.7904** | **0.5536** | **0.9062** | |  | **0.7996** | **0.6070** | **0.9172** |
| **F1 Score** | **0.5767** | **0.5558** | **0.6788** | |  | **0.6516** | **0.5473** | **0.5339** |
| **Balance Accuracy** | **0.6842** | **0.5696** | **0.6853** | |  | **0.7277** | **0.6072** | **0.6445** |