THE GERONTOLOGICAL SOCIETY OF AMERICA®

Journals of Gerontology: Medical Sciences
cite as: *J Gerontol A Biol Sci Med Sci*, 2022, Vol. 77, No. 5, 1072–1078
https://doi.org/10.1093/gerona/glab269
Advance Access publication September 16, 2021

OXFORD

Research Article

Predicting Future Mobility Limitation in Older Adults: A Machine Learning Analysis of Health ABC Study Data

Jaime L. Speiser, PhD, MS,[1,*] Kathryn E. Callahan, MD, MS,[2] Edward H. Ip, PhD,[1] Michael E. Miller, PhD,[1] Janet A. Tooze, PhD,[1] Stephen B. Kritchevsky, PhD,[2] and Denise K. Houston, PhD[2]

[1]Department of Biostatistics and Data Science, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA. [2]Department of Internal Medicine, Section on Gerontology and Geriatric Medicine, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA.

*Address correspondence to: Jaime L. Speiser, PhD, MS, Department of Biostatistics and Data Science, Wake Forest School of Medicine, Medical Center Blvd, Winston-Salem, NC 27157, USA. E-mail: jspeiser@wakehealth.edu

Received: June 4, 2021; Editorial Decision Date: September 12, 2021

Decision Editor: Lewis A. Lipsitz, MD, FGSA

Abstract

Background: Mobility limitation in older adults is common and associated with poor health outcomes and loss of independence. Identification of at-risk individuals remains challenging because of time-consuming clinical assessments and limitations of statistical models for dynamic outcomes over time. Therefore, we aimed to develop machine learning models for predicting future mobility limitation in older adults using repeated measures data.

Methods: We used annual assessments over 9 years of follow-up from the Health, Aging, and Body Composition study to model mobility limitation, defined as self-report of any difficulty walking a quarter mile or climbing 10 steps. We considered 46 predictors, including demographics, lifestyle, chronic conditions, and physical function. With a split sample approach, we developed mixed models (generalized linear and Binary Mixed Model forest) using (a) all 46 predictors, (b) a variable selection algorithm, and (c) the top 5 most important predictors. Age was included in all models. Performance was evaluated using area under the receiver operating curve in 2 internal validation data sets.

Results: Area under the receiver operating curve ranged from 0.80 to 0.84 for the models. The most important predictors of mobility limitation were ease of getting up from a chair, gait speed, self-reported health status, body mass index, and depression.

Conclusions: Machine learning models using repeated measures had good performance for identifying older adults at risk of developing mobility limitation. Future studies should evaluate the utility and efficiency of the prediction models as a tool in clinical settings for identifying at-risk older adults who may benefit from interventions aimed to prevent or delay mobility limitation.

Keywords: Mobility limitation, Prediction modeling, Random forest, Repeated measures analysis

Age-related mobility limitation carries a large burden for older adults, caregivers, hospitals, and communities (1,2). Limitations in mobility in older adults result in loss of independence, greater health care costs and health care utilization, poor health outcomes, and mortality (3). Several studies have evaluated risk factors for mobility limitation, including older age, race, low physical activity, obesity, muscle weakness or poor balance, and chronic diseases such as diabetes or arthritis (4–6). Mobility limitation is dynamic and older adults can transition to mobility-limited and back over time (7). Gill et al. (8) quantified the dynamic nature of mobility limitation in older persons over time, concluding that older age, female sex, and physical frailty were associated with a higher likelihood of transitioning into a mobility-limited state.

Despite the significance and prevalence of mobility limitation as the population ages, identification of at-risk individuals remains challenging. Some studies have proposed prediction models for determining the risk of mobility limitation in older adults, but most published models can only predict the first instance of mobility

limitation when compared with recurrent episodes (9–11). Given the dynamic nature of mobility limitation in older adults, published models predicting mobility limitation at a single time point may not accurately reflect this transitional process. There are limitations to traditional statistical methods (ie, regression), which require a priori knowledge about potential interactions among predictors and specification of nonlinear associations between predictors and outcome. Another challenge in predicting mobility limitation is that extant clinical assessments add time to already compressed clinical visits and may require expert knowledge. There is value in determining the fewest clinical factors needed to accurately identify those at risk for mobility limitation. To overcome these limitations, we employ a machine learning methodology that is robust to different types of data (eg, interactions among predictors and nonlinear associations between predictors and outcome) and offers built-in algorithms to determine optimal predictors (12).

The aim of this study is to develop prediction models to estimate the likelihood of future mobility limitation in older adults using repeated measures over time, where disability status may change from year to year. To accomplish this goal, we used data from a longitudinal study that includes annual assessments of mobility limitation, as well as annual and fixed risk factors, in a large cohort of community-dwelling older adults without mobility limitation at enrollment. The eventual goal of this research is to be able to incorporate a prediction model into electronic health record (EHR) systems as a decision support tool for identification of older adults who may be at risk for mobility limitation and who may benefit from additional mobility assessment and preventative interventions.

Method

Study Population

Data from 3 075 community-dwelling older adults (aged 70–79 years at baseline) enrolled in the Health, Aging, and Body Composition (Health ABC) study from April 1997 to June 1998 were used in this longitudinal analysis. Participants were recruited from a random sample of White and Black Medicare-eligible residents in the Pittsburgh, PA and Memphis, TN metropolitan areas. Participants were eligible if they reported no difficulty walking a quarter mile or climbing 10 steps (ie, had no self-reported mobility limitation at baseline/Year 1), no difficulty performing basic activities of daily living, were free of life-threatening illness, planned to remain in the geographic area for at least 3 years, and were not enrolled in lifestyle intervention trials. In the Health ABC study, baseline was defined as Year 1 and annual follow-up visits began at Year 2. Data were collected yearly via questionnaires and in-person clinic visits. For this study, all participants with at least 2 years of mobility limitation data following baseline (Year 1) were included in order to develop models for mobility limitation over time (n = 2 825).

Mobility Limitation

Mobility limitation was the primary outcome of this study. Mobility limitation was defined as any self-reported difficulty (little/some difficulty or lot of difficulty/cannot do) walking a quarter mile and/or climbing 10 steps. We use yearly measures of mobility limitation (yes/no) in this study that aligned with annual participant visits, such that participants can change between these states for each given year. Although many analyses of Health ABC data use a definition of mobility limitation that includes difficulty walking or climbing steps for 2 consecutive visits (called persistent mobility limitation), in this study, we use yearly measures of mobility limitation so that we can model the associations between predictors and outcome in a given year. Follow-up includes mobility limitation ascertained through 9 years of follow-up with a mean (standard deviation [*SD*]) follow-up of 6.85 (1.75) years.

Predictor Variables

Candidate predictors considered for modeling mobility limitation were based on previous literature (4,6,8). Fixed predictors collected at baseline/Year 1 included demographics (age, race, sex, site, marital status, education, family income), self-reported lifestyle factors (number of people living in a household, walking/exercise kcal per week, pack-years exposure to cigarettes, drinking history, current drinking), previous comorbidities/conditions (coronary heart disease, coronary heart failure, cerebral/vascular disease, diabetes, knee osteoarthritis, knee pain, hypertension, depressive symptoms [yes/no based on Center for Epidemiological Studies—Depression scale {CES-D} cutoff of 16 and continuous CES-D score with higher levels indicating greater depressive symptoms (13)], pulmonary disease, cancer), and measures of health and function (self-reported health status [how is your health], gait speed over 6 m, pace of chair stands). Longitudinal predictors collected at multiple time points during the 9-year follow-up (Years 2, 3, 4, 5, 6, 8, and 10, when available) included age, physical/clinical measures (body mass index [BMI] from measured weight and height, self-reported ease of rising from a chair as a measure of leg strength [measured in 0–6 scale with 0 representing least easy and 6 representing easiest], blood pressure, fall within the past year), comorbidities/conditions at the time of the visit (knee pain, diabetes, depression, cancer, coronary heart disease, myocardial infarction, stroke, cardiovascular disease), cognition (Modified Mini-Mental State Examination [3MS] (14)), and self-reported health status. Additional longitudinal physical performance measures were considered (namely, standing balance time and pace of chair stands); however, these were omitted from modeling because these were not collected at regular yearly intervals over follow-up.

Statistical Methods

Analyses were conducted using R software (15). Participant characteristics were presented as mean (*SD*) or *N* (%) using the R package *tableone* (16). Sankey plots were developed to show the distribution of mobility limitation over time using the R package *ggalluvial* (17). All nonmissing observations of mobility limitation for participants over time were included in the analysis data set. Missing predictor values were imputed (filled in) using the R package *missForest* (18) under a Missing at Random assumption. This imputation method uses a random forest developed with observed values of the data set to predict missing values and is widely used for random forest applications. To develop the prediction models, we used an innovative machine learning approach called Binary Mixed Model (BiMM) forest (19,20). This is a mixed model random forest approach that allows for modeling of dynamic changes in mobility limitation in older adults over time. Justification for using the BiMM forest approach is included in Supplementary Methods. We used the one-iteration version of BiMM forest because of computational efficiency and accuracy for prediction compared to multiple iteration versions of BiMM forest (20). BiMM forest uses an unstructured correlation matrix and therefore does not make any simplifying assumptions about the correlation structure for the longitudinal outcomes. Relative variable importance of predictors included in the models was determined using the minimum depth criterion, which

reflects on average how close the variables are to the top position within the trees (21). Minimum depth was calculated using the R package *randomForestExplainer* (21). For comparison, we also developed a model using a standard generalized linear mixed model approach with the R package *lme4* (22). Random forest models were developed with the R package *randomForest* (23).

Models developed

Four models were developed for mobility limitation, 1 with all predictors and 3 using variable selection. Variable selection was employed to reduce the number of predictors needed within models while maintaining similar predictive performance to a model with all predictors. This allows for analysis of which predictors are useful for predicting outcomes versus which predictors are superfluous. BiMM forest was used in the following: Model 1 used all predictors available, Model 2 used an automated variable selection procedure (backward elimination, implemented in the BiMM forest method using the approach proposed by Díaz-Uriarte and Alvarez de Andrés (24)) to optimize predictors included, and Model 3 used the top 5 variables according to the random forest variable importance criteria of minimum depth from Model 1. We chose to use the top 5 variables because this emulates the case where a clinician only has time to collect 5 measures, which may be reasonable in real-world practice. Regardless of variable selection for Models 2 and 3, we included age in the models to allow for age-varying covariates through interactions among age and other predictors that naturally occur in the decision tree framework. Essentially, each decision tree represents an interaction among all predictors included in the tree. Thus, including age as a predictor in the models allows for changes in the predictors that vary by age. In addition to the 3 BiMM forest models, Model 4 used a generalized linear mixed model approach with the same predictors as Model 3, included as main effects without interaction terms. Finally, to assess the impact of using longitudinal, repeated measures data versus cross-sectional data, we developed standard random forest models for each time point separately.

Training and testing data set split

A split sampling approach was used to develop and internally validate the models for mobility limitation (Figure 1). Models were developed using training data, which consisted of all observations up until the last year for an individual, and models were evaluated using testing data, which consisted of the outcome for the last year available and the predictors from the previous year for an individual. For example, for a participant with data for Years 2, 3, 4, 5, and 6, the Years 2, 3, 4, and 5 data would be included in the training data set. The outcome for Year 6 and the predictors from Year 5 would be included in the testing data set. Therefore, the performance characteristics of the models can be interpreted as the models' ability to predict next year's outcome using present data. As a sensitivity analysis, we randomly split the data into training data (all observations for 2/3 of participants) and testing data (all observations for 1/3 of participants). This analysis can be interpreted as the models' ability to predict outcomes for independent participants not included in the training data set.

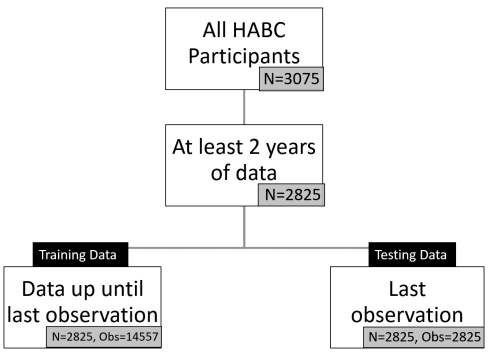

Figure 1. Diagram showing split of Health ABC data into training and testing data sets. This diagram displays how the Health ABC data were split into training and testing data. The training data consisted of all observations for participants' longitudinal measures from year to year, except for the last observation. The testing data consisted of the outcome for the last year observed and the predictor data from the previous observation. Health ABC = the Health, Aging, and Body Composition.

Evaluation metrics

Performance of the models for the testing data set was evaluated using area under the receiver operating curve (AUC), accuracy (percent of correct predictions), sensitivity (percent of correct predictions for the mobility limitation outcome group), specificity (percent of correct predictions for the mobility limitation-free outcome group), positive predicted value (PPV, the percent of correct predictions considering all predictions of mobility limitation) and negative predicted value (NPV, the percent of correct predictions considering all predictions of no mobility limitation), and their associated 95% binomial confidence intervals. AUC and its corresponding confidence interval were calculated using a cross-validation approach in the R package *cvAUC* (25). Receiver operating curve (ROC) plots and precision–recall plots were used to compare model performance and were developed using the R package *ROCR* (26). Precision–recall plots are useful for analyzing predictions when the outcome variable is imbalanced, meaning that the percent of outcomes in each group is not close to 50%.

Results

Summary Data

Baseline participant characteristics and prevalent chronic conditions are presented in Table 1. Of the 2 825 participants with at least 2 years of data, 52.6% were women and 40.5% identified as Black. The mean age at baseline was 73.6 years. The proportion of participants reporting mobility limitation over 9 years of follow-up ranged from 31% to 49% (Figure 2). Mobility limitation dynamically changed over time, with participants transitioning in and out of this state during the course of the study. In general, the proportion of participants reporting mobility limitation increased over time.

Models for mobility limitation were developed using the training data set and evaluated using the testing data set (Figure 1). The training data set included 2 825 participants with a total of 14 557 observations collected from annual visits from year to year, and 33.7% of observations for participants represented follow-up years where mobility limitation was reported. The testing data set included mobility limitation from the last year of observed data and predictors from the second to the last year of observed data for the 2 825 participants, and 53.7% of participants had mobility limitation.

Variables Included in Models

Four models were developed using repeated measures of mobility limitation as the outcome variable. Model 1 contained all 46 predictor variables, whereas Models 2, 3, and 4 used variable selection

Table 1. Baseline Characteristics of Participants: The Health ABC Study (n = 2 825)

Characteristics	Number of Missing	*N* (%) or Mean (*SD*)
Black race	0	1 142 (40.4)
Female sex	0	1 484 (52.5)
Age	0	73.59 (2.87)
Lives alone	10	839 (29.8)
Less than high school education	7	690 (24.5)
Pack-years cigarettes exposure	39	18.3 (27.7)
Current drinker	12	1 405 (49.9)
Exercise kcal/week	0	1 065 (1 931)
BMI, kg/m^2	0	27.4 (4.8)
Health status, % excellent or very good	3	1 281 (45.4)
3MS score	12	90.2 (8.4)
Ease standing from a chair, % very easy	213	1 347 (23.5)
CES-D score	23	4.6 (5.3)
Chair stands pace, stand/s	38	0.37 (0.12)
Gait speed, m/s	0	1.18 (0.23)
Prevalent chronic conditions		
CHD	51	458 (16.5)
CVD	27	196 (7.0)
Diabetes	0	417 (14.8)
Knee osteoarthritis	41	160 (5.7)
Hypertension	23	1 236 (44.1)
Depression (CES-D ≥16)	14	59 (2.1)
Cancer	12	533 (18.9)

Notes: Health ABC = Health, Aging, and Body Composition; BMI = body mass index; 3MS = Modified Mini-Mental State Examination; CES-D = Center for Epidemiologic Studies—Depression scale; CHD = coronary heart disease; CVD = cardiovascular disease; SD = standard deviation. Variables considered for modeling but not shown in the table include site, marital status, family income, drinking history, coronary heart failure, pulmonary disease, blood pressure, and fall within the past year.

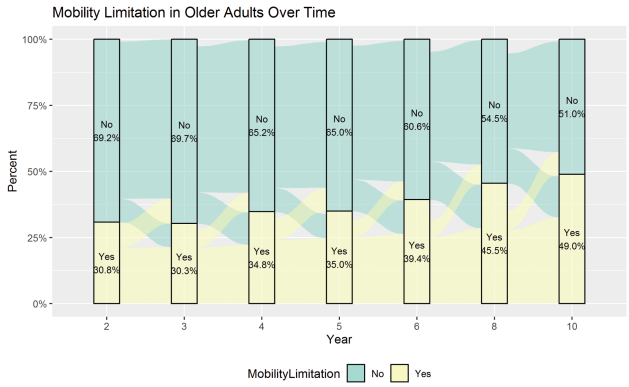

Figure 2. A plot of mobility limitation over 9 years of follow-up: the Health ABC study. This plot shows the percent of participants in the Health ABC data set with mobility limitation (yellow) and without mobility limitation (green) over follow-up Years 1–9. Each bar represents 1 year of the study. The lines between the bars show transitions of participants from year to year. The lines that cross in between the bars indicate transitions between the states. For example, the line going from the green bar on the top connecting to the yellow bar on the bottom of the next year indicates the participants who transitioned from no mobility limitation to having mobility limitation. The thickness of these connecting lines between the bars indicates the percent of participants going from one state to the other one in the subsequent year. Health ABC = the Health, Aging, and Body Composition.

to identify predictors from these. Model 2 included 37 predictors, including age, identified using an automated backward elimination approach. The 5 most important predictors in both Models 1 and 2 included ease of getting up from a chair, self-reported health status, baseline 6 m gait speed, BMI, and depressive symptoms measured by CES-D score, in that order (Supplementary Figure 1). Models 3 and 4 contained only these 5 variables, as well as age. The order of most important variables as determined by minimum depth was similar for the 3 BiMM forest models (Supplementary Figure 1).

Performance of Models

Models for mobility limitation over time were evaluated in terms of AUC, accuracy, sensitivity, specificity, PPV, and NPV (Table 2). Overall, model performance exhibited AUCs of 0.84, 0.84, 0.81, and 0.83 for Models 1, 2, 3, and 4, respectively. Accuracy for the 3 BiMM forest models was close to 75% and was 73% for Model 4. Sensitivity for predicting mobility limitation was lower (ranged from 60% to 68%); however, this was coupled with higher specificity and moderate PPV and NPV values for the models. ROC plots visualized the balance of false positive rates and true positive rates for the models (Figure 3). Models 1 and 2 had slightly better ROC curves compared to Models 3 and 4. Precision–recall curves are useful for evaluating the sensitivity and specificity of the models simultaneously for data sets that have an imbalanced outcome, which is the case for mobility limitation as depicted in Figure 2. Similar to the ROC plots, Models 1 and 2 had slightly better precision–recall curves compared to Models 3 and 4 (Figure 3). Models developed using cross-sectional data from each year separately had AUCs ranging from 0.78 to 0.80, with model sensitivities ranging from 0.47 to 0.64 (Supplementary Table 1).

In the sensitivity analysis that had independent participants in the training and testing data sets, the 4 models had AUCs ranging from 0.80 to 0.82 (Supplementary Table 2). ROC and precision–recall plots showed that Models 1 and 2 had slightly better performance compared to Models 3 and 4 (Supplementary Figure 2).

Table 2. Performance Statistics (95% confidence intervals) for the Mobility Limitation Prediction Models: The Health ABC Study

Statistic (95% confidence interval)	Model 1 (46 predictors)	Model 2 (37 predictors)	Model 3 (6 predictors)	Model 4 (6 predictors)
AUC	0.84 (0.82–0.85)	0.84 (0.82–0.85)	0.81 (0.80–0.83)	0.83 (0.81–0.84)
Accuracy	0.76 (0.74–0.77)	0.76 (0.74–0.77)	0.75 (0.74–0.77)	0.73 (0.71–0.75)
Sensitivity	0.68 (0.66–0.71)	0.68 (0.66–0.71)	0.67 (0.65–0.70)	0.60 (0.58–0.63)
Specificity	0.84 (0.82–0.86)	0.84 (0.82–0.86)	0.85 (0.83–0.87)	0.88 (0.86–0.90)
PPV	0.70 (0.67–0.72)	0.70 (0.67–0.72)	0.69 (0.67–0.71)	0.66 (0.63–0.68)
NPV	0.83 (0.81–0.85)	0.83 (0.81–0.85)	0.84 (0.81–0.86)	0.85 (0.83–0.87)

Note: Health ABC = Health, Aging, and Body Composition; AUC = area under the receiver operating curve; PPV = positive predictive value; NPV = negative predictive value.

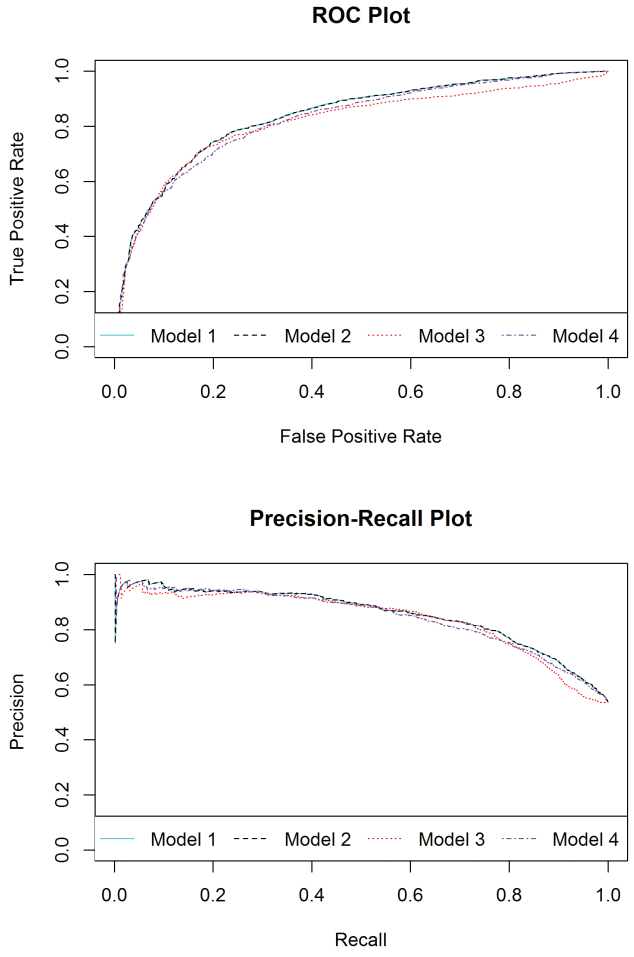

Figure 3. ROC and precision–recall curve plots of mobility limitation: the Health, Aging, and Body Composition study. These plots display ROC curves and precision–recall curves for the 3 models for the testing data. ROC curves extending into the upper left quadrant of the plot indicate superior performance. Precision–recall curves extending into the upper right quadrant of the plot indicate superior performance. ROC = receiver operating curve.

Discussion

In this study, we developed prediction models for mobility limitation in older adults using an innovative machine learning approach called BiMM forest and repeated measures of mobility limitation collected over 9 years of follow-up from the Health ABC study. This allows for accurate prediction of future mobility limitation in older adults by using past annual visit data, a unique aspect of this study because most analyses of Health ABC data use the first occurrence of persistent mobility limitation. We developed 3 BiMM forest models: one with all predictors considered (Model 1), one with automated variable selection based on backward elimination (Model 2), and one with variable selection by an ad hoc method of analyzing random forest variable importance using the minimum depth criteria (Model 3). For comparison, we also developed Model 4 using a generalized linear mixed model with the same predictors as Model 3. Models had good overall performance for predicting next year's mobility limitation status using the previous year's predictor data. Specificity was higher than sensitivity for the 3 models, meaning that models had better accuracy predicting participants without mobility limitation compared to those with mobility limitation; however, the values of sensitivity and specificity can be altered using different threshold cutoff values if it was of interest to maximize one measure over the other. Use of a mixed modeling approach and repeated measures data resulted in better predictive performance (AUC and accuracy) compared to models developed using standard random forest and cross-sectional data from 1 year at a time. With 2 types of test data sets, one for the last observation of participants in the training data and one for independent participants, all models had AUCs higher than 0.80.

The most important predictors of mobility limitation were consistent across the 3 BiMM forest models and included a surrogate for muscle strength (self-reported ease of getting up from a chair), baseline gait speed, self-reported health status, BMI, and depressive symptoms (measured by CES-D). Of these predictors, one was considered a fixed variable (baseline gait speed) whereas the others were collected on an annual basis. This highlights the importance of having multiple years of data for making accurate predictions for future mobility limitation. A total of 37 predictors were selected for Model 2 from the 46 considered. Model 2 predictors included physical factors (eg, BMI, ease getting up from a chair), medical conditions (eg, depression, knee pain, diabetes), demographics (eg, age, income), cognition (eg, 3MS score), and lifestyle factors (eg, exercise, drinking, smoking). Inclusion of many predictors from a variety of aspects of health demonstrates the complexity of predicting mobility limitation in older adults over time. In an effort to simplify the prediction model, we used an ad hoc variable selection approach in Models 3 and 4, which included only 5 predictors identified as most important, as well as age. While this greatly reduces the burden of data collection, these models had slightly worse performance compared to Models 1 and 2. However, the simplicity of Models 3 and 4 make it appealing because collecting data for all variables in Models 1 and 2 may be clinically laborious and having fewer predictors may result in fewer issues with missing data. It was interesting that age was not identified as the top most important variable in Models 1 and 2, and it was the least important variable included in Model

3. This may indicate that other predictors contain more predictive value compared to age alone, especially considering that the age range of participants for the data included is somewhat narrow (10-year spread from baseline/Year 1 to Year 10).

Using these models for predicting mobility limitation in older adults has enormous potential utility for helping clinicians determine at-risk older adults who may benefit from additional assessment and/or intervention (eg, physical therapy). Models 3 and 4 may be particularly useful for obtaining these predictions because fewer measures are needed compared to Models 1 or 2. The BiMM forest model (Model 3) had slightly worse AUC compared to the generalized linear mixed model (Model 4) with the same predictors, although these differences were not statistically significant. Validation with external data is needed to determine which of these models is preferable.

Many of the variables included in the models are typically present in EHR data from annual Medicare wellness visits, so there is potential for leveraging readily available data in order to predict future mobility limitation in older adults. Although some variables may not be available, most EHR data contain similar information that could be matched to the Health ABC predictors included in models so that predictions of outcome can be made. For instance, CES-D may not be available in EHR data, but there are other questions or analogous measures of depression that could be used for modeling instead. On the other hand, some variables, such as gait speed over 6 m, will likely not appear in EHR data in any form. The outcome variable used in this study, mobility limitation, may not appear in the EHR exactly as measured (ie, with ability to climb 10 steps and walk a quarter mile), so this may present a challenge for validating these models with EHR data. However, other similar measures may be available, such as self-reported difficulty walking a block. Future work is needed to evaluate the potential for using prediction models for mobility limitation with EHR data.

A few studies have investigated repeated measures of mobility over the aging process. Here, we provide summaries of previous models and their performance characteristics, although we note that AUCs for models developed with different data sets are not directly comparable. Gill et al. (8) used time to event Cox regression for repeated events to analyze a cohort of 754 community-living older people who were nondisabled over the course of 5 years and concluded that mobility disability in older adults is a dynamic process in which frequent transitions between independence and disability occur over time. Similar to some of the variables selected in our models, Gill et al. found that older age and physical frailty were associated with a greater likelihood of disability. Female sex was also determined as a risk factor for disability, whereas in our study, sex was considered but not identified as one of the most important variables in any of the BiMM forest models. Two recent studies used latent class growth modeling to predict trajectories of functional decline in older adults. Jonkman et al. (27,28) used age, living alone, economic satisfaction, tandem balance stands, gait speed, physical activity, BMI, and cardiovascular disease as predictors of functional decline and achieved AUCs of 0.63 and 0.74 for females and males, respectively. A study by Hoekstra et al. (29) used a similar latent class analysis framework and concluded that trajectories of functional decline are heterogeneous across different physical performance measures (gait speed, chair stands, and handgrip strength). A benefit of the models we proposed in our study is that they include subject-specific effects, meaning that they can handle heterogeneity across varying characteristics.

Several studies provided prediction models for incident or future mobility disability or limitation using standard regression methodology, although these use cross-sectional prediction data collected at one time point rather than annual prediction data across multiple years as in this study. Models proposed by Nüesch et al. (10) had AUCs of 0.73 and 0.74 and models proposed by Taş et al. (11) had AUCs of 0.67 and 0.69. A model developed with activities of daily living variables proposed by den Ouden et al. (9) had an AUC of 0.83. Similar to predictor variables used in our study, Nüesch et al. (10) included self-rated health and BMI. Taş et al. (11) found that blood markers did not increase predictive performance of models that already included age, history of arthritis, and physical activity. Aside from depression and knee pain, the models in our study did not identify comorbidity measures as most important for predicting mobility limitation in older adults.

BiMM forest prediction models in this study had good predictive performance for the validation data set, with AUCs ranging from 0.81 to 0.84. A major strength of the study was the use of Health ABC data, a large prospective cohort with validated measures of predictors. We used self-reported mobility limitation as the outcome for our study, which is a validated measure and clinically significant (30). Positive aspects of modeling in this study included consideration of a variety of predictors simultaneously and use of annually collected data which allowed for changes in mobility limitation over time. The repeated measures analysis in our study also allowed for capturing health events related to mobility that change over time (eg, if a participant had a fall that resulted in mobility limitation in 1 year, but subsequently the participant was without mobility limitation). Analysis of incidence of first mobility limitation or mobility limitation at a single time point cannot capture these dynamic changes, whereas our models can handle changes from year to year. Models account for all previous years of data for participants through the random intercepts in the BiMM forest framework, so specific transition paths in and out of mobility limited states are not captured (eg, a participant has no mobility limitation in Years 2 and 3, becomes mobility-limited in Years 4 and 5 and then back to no mobility limitation for Year 6). A future study could analyze specific transition paths across multiple visits using a multistate modeling approach. A novel aspect of the study was that innovative machine learning methodology was used to develop prediction models. Three models were developed, with and without variable selection, to ascertain important predictors of mobility limitation.

There are some limitations of the study. Participants were recruited to be well functioning and free of mobility limitation at baseline; thus, these results may not be generalizable to the general older population. Although the models had good performance (AUCs of 0.81 and 0.84), it would be desirable to improve performance further. Future studies could investigate if inclusion of additional predictors, such as accelerometer data, blood-based biomarkers like inflammation or kidney function, and medications, can increase predictive performance. Findings from our study should also be validated with external data sets to determine the generalizability of prediction models. Overall, despite these limitations, our study was conducted in a rigorous manner using a large data set, performing internal validation, and employing machine learning methodology that allowed for comparison of prediction models for mobility limitation.

We developed models for mobility limitation in older adults over time using fixed and repeated measurements of predictors with an innovative machine learning approach. Only 5 measures and age are needed for making predictions to achieve an AUC of 0.81 (Model 3). Future studies should evaluate the utility and efficiency of the prediction models as a tool for identifying at-risk older adults who may benefit from interventions aimed to prevent or delay mobility limitation.

Supplementary Material

Supplementary data are available at *The Journals of Gerontology, Series A: Biological Sciences and Medical Sciences* online.

Funding

J.L.S. is the recipient of a KL2 award from the National Association for Clinical and Translational Sciences to the Wake Forest School of Medicine Clinical and Translational Science Institute (KL2TR001421, UL1TR001420). This research was supported by the National Institute on Aging (NIA) contracts N01-AG-6-2101, N01-AG-6-2103, N01-AG-6-2106; NIA grant R01-AG028050; and National Institute of Nursing Research grant R01-NR012459. This research was funded in part by the Intramural Research Program of the NIH, NIA.

Conflict of Interest

None declared.

Author Contributions

J.L.S. conceived the idea for the study, conducted the data analysis, and drafted the manuscript. All other authors contributed to the study design, advised about data analysis, and edited the manuscript.

References

1. Guralnik JM, Fried LP, Salive ME. Disability as a public health outcome in the aging population. *Annu Rev Public Health*. 1996;17:25–46. doi:10.1146/annurev.pu.17.050196.000325
2. Satariano WA, Guralnik JM, Jackson RJ, Marottoli RA, Phelan EA, Prohaska TR. Mobility and aging: new directions for public health action. *Am J Public Health*. 2012;102(8):1508–1515. doi:10.2105/AJPH.2011.300631
3. Beswick AD, Rees K, Dieppe P, et al. Complex interventions to improve physical function and maintain independent living in elderly people: a systematic review and meta-analysis. *Lancet*. 2008;371(9614):725–735. doi:10.1016/S0140-6736(08)60342-6
4. Brown CJ, Flood KL. Mobility limitation in the older patient: a clinical review. *JAMA*. 2013;310(11):1168–1177. doi:10.1001/jama.2013.276566
5. Ip EH, Zhang Q, Rejeski WJ, Harris TB, Kritchevsky S. Partially ordered mixed hidden Markov model for the disablement process of older adults. *J Am Stat Assoc*. 2013;108(502):370–380. doi:10.1080/01621459.2013.770307
6. Yeom HA, Fleury J, Keller C. Risk factors for mobility limitation in community-dwelling older adults: a social ecological perspective. *Geriatr Nurs*. 2008;29(2):133–140. doi:10.1016/j.gerinurse.2007.07.002
7. Rejeski WJ, Ip EH, Marsh AP, Zhang Q, Miller ME. Obesity influences transitional states of disability in older adults with knee pain. *Arch Phys Med Rehabil*. 2008;89(11):2102–2107. doi:10.1016/j.apmr.2008.05.013
8. Gill TM, Allore HG, Hardy SE, Guo Z. The dynamic nature of mobility disability in older persons. *J Am Geriatr Soc*. 2006;54(2):248–254. doi:10.1111/j.1532-5415.2005.00586.x
9. den Ouden ME, Schuurmans MJ, Mueller-Schotte S, van der Schouw YT. Identification of high-risk individuals for the development of disability in activities of daily living. A ten-year follow-up study. *Exp Gerontol*. 2013;48(4):437–443. doi:10.1016/j.exger.2013.02.002
10. Nüesch E, Pablo P, Dale CE, et al. Incident disability in older adults: prediction models based on two British prospective cohort studies. *Age Ageing*. 2015;44(2):275–282. doi:10.1093/ageing/afu159
11. Taş U, Steyerberg EW, Bierma-Zeinstra SM, Hofman A, Koes BW, Verhagen AP. Age, gender and disability predict future disability in older people: the Rotterdam Study. *BMC Geriatr*. 2011;11:22. doi:10.1186/1471-2318-11-22
12. Speiser JL, Callahan KE, Houston DK, et al. Machine learning in aging: an example of developing prediction models for serious fall injury in older adults. *J Gerontol A Biol Sci Med Sci*. 2021;76(4):647–654. doi:10.1093/gerona/glaa138
13. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. *Appl Psychol Meas*. 1977;1(3):385–401. doi:10.1177/014662167700100306
14. Teng EL, Chui HC. The Modified Mini-Mental State (3MS) examination. *J Clin Psychiatry*. 1987;48(8):314–318.
15. R Development Core Team. *R: A Language and Environment for Statistical Computing*. T.R.F.f.S. Computing; 2008.
16. Yoshida K, Bohn J. *tableone: Create" Table 1" to Describe Baseline Characteristics*. R package version 0.7; 2015.
17. Brunson JC. Ggalluvial: layered grammar for alluvial plots. *J Open Source Softw*. 2020;5(49):2017. doi:10.21105/joss.02017
18. Stekhoven DJ, Bühlmann P. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*. 2012;28(1):112–118. doi:10.1093/bioinformatics/btr597
19. Speiser JL. A random forest method with feature selection for developing medical prediction models with clustered and longitudinal data. *J Biomed Inform*. 2021;117:103763. doi:10.1016/j.jbi.2021.103763
20. Speiser JL, Wolf BJ, Chung D, Karvellas CJ, Koch DG, Durkalski VL. BiMM forest: a random forest method for modeling clustered and longitudinal binary outcomes. *Chemometr Intell Lab Syst*. 2019;185:122–134. doi:10.1016/j.chemolab.2019.01.002
21. Paluszynska A, Biecek P. randomForestExplainer: explaining and visualizing random forests in terms of variable importance. R package version 0.9. https://CRAN. R-project. org/package=randomForestExplainer. 2017.
22. Bates D, Maechler M, Bolker B, et al. Package 'lme4'. 2015.
23. Liaw A, Weiner M. Classification and regression by randomForest. *R News*. 2002;2:18–22.
24. Díaz-Uriarte R, Alvarez de Andrés S. Gene selection and classification of microarray data using random forest. *BMC Bioinformatics*. 2006;7:3. doi:10.1186/1471-2105-7-3
25. LeDell E, Petersen M, van der Laan M, LeDell ME. Package 'cvAUC'. 2014.
26. Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCR: visualizing classifier performance in R. *Bioinformatics*. 2005;21(20):3940–3941. doi:10.1093/bioinformatics/bti623
27. Jonkman NH, Colpo M, Klenk J, et al. Development of a clinical prediction model for the onset of functional decline in people aged 65–75 years: pooled analysis of four European cohort studies. *BMC Geriatr*. 2019;19(1):179. doi:10.1186/s12877-019-1192-1
28. Jonkman NH, Del Panta V, Hoekstra T, et al. Predicting trajectories of functional decline in 60- to 70-year-old people. *Gerontology*. 2018;64(3):212–221. doi:10.1159/000485135
29. Hoekstra T, Rojer AGM, van Schoor NM, Maier AB, Pijnappels M. Distinct trajectories of individual physical performance measures across 9 years in 60- to 70-year-old adults. *J Gerontol A Biol Sci Med Sci*. 2020;75(10):1951–1959. doi:10.1093/gerona/glaa045
30. Fried LP, Young Y, Rubin G, Bandeen-Roche K, WHAS II, Collaborative Research Group. Self-reported preclinical disability identifies older women with early declines in performance and early disease. *J Clin Epidemiol*. 2001;54(9):889–901. doi:10.1016/s0895-4356(01)00357-2