Predicting the Risk of Mobility Difficulty in Older Women With Screening Nomograms

The Women’s Health and Aging Study II

Paulo H. M. Chaves, MD, PhD; Elizabeth S. Garrett, PhD; Linda P. Fried, MD, MPH

Background: A major obstacle to screening for early mobility disability (ie, mobility difficulty), a major public health concern, is the lack of a method that identifies those who are at high risk. The goal of this study was to develop easy-to-use clinical nomograms for estimation of the probability of incident mobility difficulty.

Methods: We conducted a population-based prospective study using data from 266 high physically and cognitively functioning older women, aged 70 to 80 years, who were free of mobility disability at the baseline evaluation of the Women’s Health and Aging Study II. The outcome measure was incident mobility disability within 18 months, defined as self-reported difficulty walking 0.8 km, climbing 10 steps, or transferring from or into a car or bus. Logistic regression and receiver operating characteristic curve analyses were used for evaluation of the optimal combination of self-reported and performance-based mobility measures. Bootstrap sampling and estimation was used for validation.

Results: Predictive nomograms were developed based on a final model that included 3 simple-to-obtain measures of preclinical disability: self-report of modification in mobility tasks without having difficulty with them, one-leg stance balance, and time to walk 1 m at a usual pace. Final model accuracy (as estimated by the area under the receiver operating characteristic curve) was 73% (SE=0.04). Validation analysis confirmed the high accuracy of these nomograms.

Conclusions: An original tool was developed for assessment of the risk of mobility difficulty in older women that can be used to assist physicians and researchers in deciding which women to target for preventive interventions.

Arch Intern Med. 2000;160:2525-2533

Mobility disability (ie, mobility difficulty) is a highly prevalent public health concern. Up to 50% of persons aged 65 years and older have disability in mobility-related tasks such as walking 0.4 km, climbing steps, transferring, or doing heavy housework. Mobility is but one of several types of physical disability. Nevertheless, it is a major risk factor for disability and dependency in other domains of physical functioning, causing decreased quality of life in older adults and substantial social and health care needs. Consequently, prevention or postponement of mobility disability is a high priority.

Screening of older adults for risk of mobility difficulty is an important step toward prevention. Identification of individuals at highest risk could provide relevant information for targeting those who are, theoretically, most likely to benefit from preventive interventions. However, how to identify the subset of older individuals at highest risk is yet to be established. In the clinical setting, for example, physicians have to rely on their subjective impressions when assessing patients’ risk of mobility difficulty. Effective methods for screening are needed.

During the past decade, extensive methodological and clinical research have laid the groundwork for the screening of older adults for risk of mobility difficulty. Important for the development of such a basis were (1) accumulation of evidence supporting the hypothesis of the existence of a preclinical stage of disability in the natural history of disablement, in which nondisabled older adults have “non-symptomatic” critical decline in physical function that constitutes a major risk factor for progression to mobility difficulty (such evidence suggests that preclinical mobility decrements in many cases precede overt mobility difficulty and are a marker of a high-risk group for whom interventions might potentially yield the greatest benefits in terms of decreasing the
PARTICIPANTS AND METHODS

STUDY POPULATION

The WHAS II is a population-based prospective study of 436 women aged 70 to 80 years with no or minimal physical disability living in the community. The WHAS II was designed to be a companion study to the WHAS I, a study of 1002 women representative of the one third most disabled women older than 65 years living in the community. The WHAS II sample was based on 3 replicate age-stratified random samples drawn from the Health Care Financing Administration Medicare files that listed all female beneficiaries who were 70 to 79 years old in 12 contiguous ZIP code areas in eastern Baltimore City and Baltimore County, Maryland, on March 1, 1994, October 1, 1994, and May 1, 1995. The methods of this study have been described previously. Briefly, a screening interview was used to assess scores on the Mini-Mental State Examination, a widely used measure of cognition, and to determine difficulty in 4 domains of physical functioning: mobility; upper extremity movement; performance of tasks indicative of complex functioning, such as instrumental activities of daily living (IADLs); and basic self-care tasks (Table 1). Eligibility for WHAS II consisted of difficulty in no more than one domain and a Mini-Mental State Examination score greater than 24. Of 1630 women screened, 880 were eligible for participation, and 436 agreed to participate in WHAS II. This apparent low initial response reflects the relatively low telephone recruitment rates (telephone recruitment was used because of fiscal constraints) and the difficulty in getting older adults to participate in an intensive prospective study. Compared with WHAS II participants, those who were eligible but decided not to participate had less education and lower income and were more likely to rate their health status as fair or poor. This study considers the 266 participants who were free of mobility disability at baseline.

DATA COLLECTION

Study participants underwent a comprehensive baseline evaluation at the Johns Hopkins Functional Status Laboratory, Baltimore, Md, between August 1994 and February 1996, and then returned to the clinic for a similar follow-up assessment visit 18 months later. Trained interviewers conducted standardized face-to-face interviews and collected data on demographic characteristics, health behaviors, cognitive and functional status, and medical history. Two independent experts using standardized algorithms subsequently validated the latter.

Participants also underwent a thorough physical examination and had blood samples drawn. The procedures used were approved by the institutional review board of The Johns Hopkins Medical Institutions. Informed consent was obtained from all participants.

PHYSICAL FUNCTION ASSESSMENT

Details for obtaining the 3 measures for risk assessment using nomograms are provided in Table 2. Self-reports of difficulty were obtained for 3 mobility-related and 24 other activities of daily life via questions such as, “For health or physical reasons, do you have difficulty climbing up 10 steps?” Self-reports of task modification in mobility tasks, a novel marker of preclinical disability, were obtained by asking individuals who reported no difficulty in mobility tasks whether they had modified the way they were performing them, by changing the method or reducing the frequency of task performance, because of underlying health problems. Recently, Fried et al demonstrated that this is a strong, independent predictor of developing incident mobility difficulty within 18 months that correlates well with performance-based measures of mobility functioning. Besides, it has good reliability (same-day test-retest weighted k = 0.74).

Objective and standardized performance-based measures were adapted from WHAS I protocols and included (1) time to walk 1 m at usual and rapid paces; (2) time to rise from a chair 3 times as rapidly as possible with arms crossed in front of chest; (3) one-leg stance balance based on the length of time participants were able to maintain balance while standing on one leg up to a maximum of 30 seconds (using 3 categories: <10 seconds, 10 to <30 seconds, and 30 seconds); (4) the static balance measure, created by Guralnik et al based on the participant’s ability to maintain their feet in the side-by-side, semitandem, and tandem positions for 10 seconds; (5) maximal strength of the hip flexor muscles, assessed using a handheld dynamometer (2 trials for each leg were carried out, and the average strength of the strongest leg was used for analysis); (6) maximal grip strength of the dominant and nondominant hands using a handheld Jamar dynamometer (2 trials for each hand were carried out, and the average was used as the final measure for each hand); and (7) the Mini-Mental State Examination score. The intraclass correlation coefficient for the time to walk 1 m at a usual pace for measures obtained 1 week apart in WHAS I, which used the same protocol, was 0.90 (95% confidence interval, 0.89-0.91) and confirms the high reliability of this measure (data not published). Intraclass correlation coefficients for the one-leg stance balance, static balance measure, and Jamar dynamometer were 0.67, 0.74, and 0.74, respectively. Each of these measures was grouped into tertiles, and diagnostic accuracy was estimated in terms of sensitivity and specificity.
correlation coefficients for repeated chair stands time and grip and hip strength were also recently reported (0.76-0.89, 0.92-0.95, and 0.93, respectively).

PRIMARY OUTCOME

Incident mobility difficulty within 18 months was the outcome of interest. For the purposes of these analyses, all participants at baseline were nondisabled, ie, they reported no difficulty in each of 3 tasks: walking 0.8 km, climbing 10 steps, and transferring from and into a car or bus. Those reporting difficulty in at least 1 of the 3 tasks at the 18-month follow-up visit were considered to have incident mobility difficulty. The rationale for using these 3 tasks in our disability definition was based on (1) our clinical judgment that the broad concept of mobility functioning includes different components that have distinct biomechanical and energy demands that are best captured by 3 tasks instead of 1 and (2) previous work showing that these tasks are likely to be affected early in the mobility disablement process and constitute different and nonhierarchical entry points into this process (P.H.M.C., E.S.G., and L.P.F., unpublished data, 1994-1996).

STATISTICAL ANALYSIS

Step 1: Data Exploration

Descriptive statistics were used to characterize the study population and measures of physical function. Bivariate associations between incident mobility difficulty and potential predictors were examined using scatter plots, box plots, cross-tabulations, and logistic regression analyses, with and without adjustment for age. Threshold and nonlinear relationships were evaluated using quadratic and spline terms during modeling (F and \( \chi^2 \) statistics at \( \alpha = .05 \) were used for evaluation of statistical significance).

Step 2: Multiple Logistic Regression

Modeling was carried out for selection of the most parsimonious combination of predictors of incident mobility disability 18 months later. The goal in using such an approach was to maximize data collection efficiency in the clinical setting. Only variables that achieved a significance level of \( P < .05 \) in step 1 were considered. A stepwise model was fit. Variables were entered into the model in an order that would reproduce the information collection process most likely to occur in the clinical setting. Likelihood ratio tests were used to assess whether the addition of variables would provide supplemental information above and beyond that already obtained by a simpler model (\( P < .05 \) is necessary for statistical significance).

Step 3: Receiver Operating Characteristic Curve Analyses

Receiver operating characteristic curves and their respective areas under the curve (AUCs) were used for estimation of the overall accuracy of different models in predicting disability. Differences in AUCs were then compared according to the method developed by Hanley and McNeil.

Step 4: Nomograms

Taking into account results from the receiver operating characteristic curve analyses and clinical feasibility, one model was identified as optimal for developing predictive nomograms. We then constructed fitted plots of the predicted probability of incident mobility difficulty, and their associated 95% confidence intervals, given the measures included in this optimal model. By connecting the points in these plots, we obtained the probability and confidence bands that define our nomograms.

Step 5: Validation Analysis

The prediction rule used for developing these nomograms was internally validated by taking 1000 bootstrap samples from the original cohort and determining the median and 95% confidence intervals of 3 accuracy measures: sensitivity, specificity, positive and negative predictive values, and correct classification (ie, the percentage of individuals whose predicted and observed probabilities are the same). Briefly, fitted probabilities were computed according to the nomogram prediction rule and compared with the observed probabilities of incident disability. The latter assumed values of 1 or 0 depending on whether the individual developed or did not develop disability, respectively. For comparison, the fitted probabilities were dichotomized: they were defined as 0 if the fitted were lower than the probability cutoff point or 1 if higher. For example, if the computed probability was 0.40, her predicted probability was considered 0 for cutoff levels less than 0.40 and 1 for cutoff levels greater than 0.40. Bootstrap analyses were then conducted considering different cutoff points. Stata 6.0 and S-Plus 3.3 were used for analyses.
ators included static balance summary score, hip and grip strength, and Mini-Mental State Examination score. Age-adjusted and unadjusted analyses yielded almost identical results.

Table 5 shows the results of forward stepwise logistic regression analyses to identify the most predictive, yet parsimonious, combination of measures that predicted mobility difficulty within 18 months. Because self-report of modification of task performance can be easily obtained through clinical interview and this measure strongly predicted incident mobility difficulty in step 1, it was included in all models. Walking and chair stand times were entered next. Time to walk 1 m at a usual pace and time to complete 5 chair stands added significant and nonredundant information to models containing only self-report of task modification (models 2 and 4; \( P = .006 \) and \(.003 \), respectively); time to walk 1 m at a rapid pace did not (model 3; \( P = .09 \)). Time to complete 5 chair stands did not significantly improve fit over a model with task modification and time to walk 1 m at a usual pace (model 5; \( P = .11 \)). On the other hand, addition of one-leg stance balance categories to models 2 and 4 resulted in significant improvement in fit (models 6 and 7; \( P = .002 \) for both). No interaction terms were statistically significant (data not shown). Based on these results, models 6 and 7 seemed to be the best final model candidates.

Figure 1 displays receiver operating characteristic curves for models 6 and 7 and for the simpler models 1 and 2 (the latter 2 were included for illustration of the gain in accuracy obtained with incorporation of additional measures). The AUCs were 73%, 72%, 70%, and 62%, respectively (SE = 0.04 for all). The difference between AUCs of models 2 and 1 was 8% but achieved only marginal statistical significance (\( P = .09 \)). The difference between AUCs of models 6 and 7 represented a statistically significant gain in predictive accuracy over the simplest model 1 (11%; \( P = .02 \) and 10%; \( P = .03 \), respectively). The difference between the AUCs of models 6 and 7 was not significant (1%; \( P = .40 \)).

### Table 1. Domains of Physical Disability in Which Difficulty Performing Certain Tasks Was Used for Screening of Participants Into Women’s Health and Aging Study II

<table>
<thead>
<tr>
<th>Domain of Physical Function</th>
<th>Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mobility</td>
<td>Walking 0.4 km, walking up 10 steps without resting, getting into and out of a bed or chair, doing heavy housework</td>
</tr>
<tr>
<td>Upper extremity movement</td>
<td>Raising arms up overhead, grasping or handling, lifting something as heavy as 4.5 kg</td>
</tr>
<tr>
<td>Complex</td>
<td>Using the telephone, doing light housework, preparing meals, shopping for personal items</td>
</tr>
<tr>
<td>Self-care functioning</td>
<td>Bathing or showering, dressing, eating, using the toilet</td>
</tr>
</tbody>
</table>

### Table 2. Obtaining the 3 Measures for Risk Assessment Using Nomograms

#### Self-report of mobility task modification
1. Presence or absence of mobility task modification among those who report no mobility difficulty can be obtained based on the answer to the following questions “Does your health or physical condition, do you have any difficulty in [task]?” where tasks are (1) walk 0.8 km, (2) climb 10 steps, and (3) transfer from/into car/bus. Possible answers are “yes” or “no.”

2. Mobility modifications are assessed in steps.
   - Step 1: “In the previous set of questions, I was interested in whether you have difficulty with a task. Now, I will be interested in whether you have changed the way or changed how often you do a task. Due to a health or physical condition, have you changed how frequently you (task)?”, where tasks were the same as above. Possible answers are “no,” “yes, do it less frequently,” or “yes, don’t do it anymore.”
   - Step 2: “Due to a health or physical condition, have you changed (modified) the method that you used to (task)?”, where changing the method meant doing tasks slower, using body in a different position or using a different method of moving, using different types of clothing, or having omitted some parts of the task, or having changed the home. Possible answers are “no” or “yes.”

3. Self-report of mobility task modification should be coded as follows. Subjects who answer “no” to the 3 questions in all 3 tasks are considered as “no report of task modification.” If they answer “no” to the first question (task difficulty), but “yes” either to the frequency or method changes questions, they are considered as having mobility task modification.

#### One-leg stance balance
Help participants into the appropriate position: standing on one leg, while keeping the other leg off the floor, and holding arms out to the side. When the subject is in position and ready, instruct to let go of the clinician’s arm, and begin timing. The timing should be stopped when the subject moves her foot, grasps the clinician for support, or when 30 s elapses. One trial for each leg should be performed. The following one-leg balance categories should then be created based on the participant’s best performance (ie, the longer time between the 2 trials): (1) category 1: able to hold balance for \( > 10 \) s; (2) category 2: able to hold balance for \( < 10 \) s, but \( > 3 \) s; and (3) category 3: able to hold balance for the full 30 s.

#### Time to walk 1 m at usual pace
Place 2 other tapes on the floor to serve as the starting line. Tell the subject to start walking after the prompt “Ready, go!” The clinician should walk beside the subject. Timing should be started as the participant begins walking and stopped when the participant’s toe crosses the 1-m mark.

### ASSESSING THE RISK OF INCIDENT MOBILITY DISABILITY WITH NOMOGRAMS

Taking into account the AUC of each model and clinical feasibility issues (eg, asking an older adult to walk 1 m and timing it is easier to implement than the 5 chair stands test; besides, the latter also depends on chair specifica-

(Reprinted) Arch Intern Med/Vol 160, Sep 11, 2000

©2000 American Medical Association. All rights reserved.
1 m at her usual pace in 2 seconds, her estimated probability of developing significantly mobility disability within 18 months is 56% (95% confidence interval, 40%-72%). Overall, risk is greater in individuals reporting task modification in mobility tasks decreases as individuals are able to maintain their one-leg balance for longer times, and increases with longer walking times.

**VALIDATION ANALYSIS**

Results from bootstrap analyses conducted for validation purposes are presented in Table 6 and should be interpreted according to the following example. For the .30 cutoff point, the nomograms prediction rule have an estimated sensitivity of 44% (ie, the percentage of all the individuals who will develop mobility difficulty within 18 months have baseline fitted probabilities > .30), a specificity of 80% (ie, the percentage of those who will remain free of mobility difficulty have baseline fitted probabilities < .30), a positive predictive value of 40% (ie, the percentage of those whose baseline fitted probabilities were > .30 who will develop incident difficulty), and a negative predictive value of 83% (ie, the percentage of those whose baseline fitted probabilities were < .30 who will remain free of difficulty); overall correct classification will be achieved 72% of the time. Validity measures are also shown for other cutoff points. As shown in Table 6, correct classification of individuals regarding future difficulty status can be expected 72% to 79% of the time for cutoff points of .30 or higher. In our study, 25% of older women who were free of mobility difficulty at baseline had fitted risks greater than .30.

**COMMENT**

We presented nomograms for straightforward estimation of the probability of developing mobility difficulty within 18 months for nondisabled, community-dwelling women aged 70 to 80 years. To estimate such risk, 3 measures need to be obtained and “plugged” into the nomograms: (1) self-report of whether the individual has changed the method or frequency of performing mobility tasks, while having no difficulty with them;
preclinical disability and identify functional decline be-
sessment point of view, these measures are markers of
time-consuming (Table 2). From the functional as-
stained in clinical settings and are safe, inexpensive, and
1 m at a usual pace. These 3 measures can be easily ob-
(data not shown).

Table 5. Multivariate Logistic Regression Models Used to Identify the Most Parsimonious Combination of Predictors of Developing Mobility Difficulty Within 18 Months

<table>
<thead>
<tr>
<th>Measure of Physical Function</th>
<th>Participants, No.* (N = 266)</th>
<th>Unadjusted</th>
<th>Age Adjusted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-report of task modification (yes vs no)</td>
<td>266</td>
<td>3.16 (1.71-5.74)</td>
<td>.001</td>
</tr>
<tr>
<td>Time to walk 1 m at a usual pace (1 s)</td>
<td>256</td>
<td>4.47 (1.91-10.48)</td>
<td>.001</td>
</tr>
<tr>
<td>Time to walk 1 m at a rapid pace (1 s)</td>
<td>263</td>
<td>4.41 (1.31-14.8)</td>
<td>.02</td>
</tr>
<tr>
<td>Time to complete 5 chair stands (5 s)</td>
<td>258</td>
<td>2.26 (1.37-3.72)</td>
<td>.001</td>
</tr>
<tr>
<td>One-leg stance balance†</td>
<td>258</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category 1 (&lt;10 s)</td>
<td>1.00</td>
<td></td>
<td>1.00</td>
</tr>
<tr>
<td>Category 2 (10 to &lt;30 s)</td>
<td>0.51 (0.26-0.98)</td>
<td>.04</td>
<td>0.51 (0.26-0.99)</td>
</tr>
<tr>
<td>Category 3 (30 s)</td>
<td>0.09 (0.02-0.39)</td>
<td>.001</td>
<td>0.09 (0.02-0.39)</td>
</tr>
<tr>
<td>Static balance‡</td>
<td>265</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Score 1</td>
<td>1.00</td>
<td></td>
<td>1.00</td>
</tr>
<tr>
<td>Score 2</td>
<td>1.27 (0.24-6.82)</td>
<td>.78</td>
<td>1.32 (0.24-7.11)</td>
</tr>
<tr>
<td>Score 3</td>
<td>0.51 (0.11-2.38)</td>
<td>.39</td>
<td>0.52 (0.11-2.43)</td>
</tr>
<tr>
<td>Score 4</td>
<td>0.57 (0.14-2.39)</td>
<td>.45</td>
<td>0.60 (0.14-2.53)</td>
</tr>
<tr>
<td>Hip strength§ (1 kg)</td>
<td>259</td>
<td>1.00 (0.96-1.05)</td>
<td>.95</td>
</tr>
<tr>
<td>Grip strength (1 kg)</td>
<td>261</td>
<td>0.98 (0.92-1.05)</td>
<td>.55</td>
</tr>
</tbody>
</table>

* Number of individuals with measures available for each test. Data are not available for 0 to 10 people because of missing data or exclusion of outliers.
† One trial on each leg was performed. Analyses are shown for the best performance, ie, the longer time.
‡ For explanation, see the last footnote to Table 3.
§ Two trials were performed on each leg; average strength for the stronger leg is used. Analyses were also carried out using the average strength for the weaker leg and the average between them, but results were virtually identical (data not shown).

(2) one-leg stance balance categories; and (3) time to walk 1 m at a usual pace. These 3 measures can be easily ob-
tained in clinical settings and are safe, inexpensive, and
not time-consuming (Table 2). From the functional as-
stment point of view, these measures are markers of
preclinical disability and identify functional decline be-
before it might be symptomatically apparent. The develop-
ment of these nomograms provides an original screen-
ing tool for risk assessment of early and clinically relevant
mobility disability in older women.

Previously, Guralnik et al14 used mobility-related per-
formance-based measures to predict self-reported mo-

Model Predictors Odds Radio (SE) Likelihood Ratio Test*†

1 Self-report of task modification (yes vs no) 3.16 (0.97) Not applicable
2 Self-report of task modification (yes vs no) 2.83 (0.92) Models 2 and 1: 7.49 (1 df); P = .006
Walking time for 1 m at a usual pace (1 s) 3.36 (1.50)
3 Self-report of task modification (yes vs no) 2.99 (0.95) Models 3 and 1: 3.84 (1 df); P = .08
Walking time for 1 m at a rapid pace (1 s) 3.02 (1.90)
4 Self-report of task modification (yes vs no) 3.13 (1.00) Models 4 and 1: 8.67 (1 df); P = .003
Time to complete 5 chair stands (5 s) 2.17 (0.58)
5 Self-report of task modification (yes vs no) 3.06 (1.00) Models 5 and 2: 2.56 (1 df); P = .11
Walking time for 1 m at a rapid pace (1 s) 2.41 (1.20)
Time to complete 5 chair stands (5 s) 1.60 (0.47)
6 Self-report of task modification (yes vs no) 2.37 (0.81) Models 6 and 2: 12.4 (2 df); P = .002
Walking time for 1 m at a usual pace (1 s) 2.23 (1.10)
One-leg balance category 2 (10 to <30 s)† 0.60 (0.21)
One-leg balance category 3 (30 s)† 0.13 (0.09)
7 Self-report of task modification (yes vs no) 2.55 (0.86) Models 7 and 4: 12.9 (2 df); P = .002
Time to complete 5 chair stands (5 s) 1.89 (0.55)
One-leg balance category 2 (10 to <30 s)† 0.59 (0.21)
One-leg balance category 3 (30 s)† 0.12 (0.09)

* Testing statistical significance of fit improvement resulting from the addition of a new variable(s).
† Reference category is category 1 (<10 s).
of mobility difficulty, an earlier stage of the disableness process. Regarding other screening tools, Moore and Stu\textsuperscript{44} recently proposed the timed “Up & Go” test (the time for the individual to rise from an armchair, walk 3 m, turn, walk back, and sit down again). Although simple and appealing, it has yet to have its prospective screening accuracy established.

In addition to the major results of this study, other findings should be noticed. The static balance component of the tool developed by Guralnik et al\textsuperscript{14} was not a statistically significant predictor of mobility difficulty in our study, but the one-leg stance balance measure was. This suggests that the latter might be more sensitive and appropriate for the screening of higher-functioning populations. Also, time to complete 5 chair stands did not provide statistically significant information beyond that provided by 1-m walking time regarding short-term prediction of mobility difficulty. Finally, hip strength was not a significant predictor of incident mobility difficulty. This was likely because of a ceiling effect (ie, these high-functioning women’s strength was beyond a threshold above which increases in strength do not add further protective effect).\textsuperscript{35}

The decision to consider only measures of mobility functioning for inclusion in the prediction rule was based on 2 points: first, the conception that they represent not only the effects of chronic disease but also disease status–related deconditioning (ie, lower level of fitness) on functioning decline and should be among the most proximal measures to the outcome measure of this study, mobility difficulty, and second, an a priori expectation that use of only a few selected measures of mobility would allow reasonably accurate risk prediction, therefore obviating the need for additional measures from other domains. Indeed, increases in the time needed for obtaining such additional measures in the clinical setting and in the complexity for communicating risk would certainly represent a threat to the clinical usefulness of our nomograms. Nonetheless, we conducted a post hoc analysis to evaluate the impact of 6 established risk factors for mobility difficulty not considered as candidates for our prediction rule on model fit and predictive accuracy: (1) definite diagnosis of angina or myocardial infarction, (2) definite diagnosis of knee osteoarthritis (symptomatic or asymptomatic), (3) presence of depression (\(\geq 11\) points on the Geriatric Depression Scale),\textsuperscript{45} (4) body mass index (\(<18.5\) and \(>18.5 \leq 25\) and \(>25\), and \(>30\) and \(>30\) kg/m\(^2\)), (5) history of falls within the past 12 months, and (6) self-reported fear of falling within the past 12 months. When added to the prediction rule one at a time, none of these measures improved the predictive ability of our predictive rule in a statistically significant way (data
not shown). These findings can be explained in part by the relatively low variability in disease severity in high-functioning populations.

The strengths of these nomograms include (1) prediction of incident mobility difficulty that is as accurate as current National Cholesterol Education Program guidelines and the single ratio of total plasma cholesterol level to high-density lipoprotein cholesterol level for prediction of coronary heart disease mortality (AUC=0.73, 0.74, and 0.72, respectively)48 and as accurate as Papanicolaou smears for prediction of cervical cancer (AUC=70%);49 (2) risk characterization as a probability, which is easier to communicate to patients than an odds ratio; and (3) provision of a range for the estimated probability.

Four important limitations of this study must be pointed out. First, initial nonresponse of individuals chosen for inclusion in this population-based cohort is likely to have limited the external validity of our results. Whether they can be generalized to other age-sex groups, which have overall lower risk of mobility disability than the population targeted here, or to less educated populations remains to be determined. Second, nonresponse is likely to have biased our results toward underestimation of the true risk of mobility difficulty onset, especially among those who reported mobility task modification or were classified in the worst categories of one-leg stance balance, considering that those who were eligible but decided not to participate had lower levels of functioning than did study participants. Third, validation of the predictive rule used for developing these nomograms, which showed a relatively high degree of predictive validity for the proposed screening tool, was internal rather than external. The latter is an inherently stronger approach, but given the nonexistence of another data set containing all the measures included in our prediction rule, and the relatively small sample size of this study, which could have led to significant loss of power and generation of unstable and imprecise estimates had a sample-splitting technique been used, validation of the prediction rule in an independent population was not performed at this time. Finally, the confidence intervals presented in our nomograms do not incorporate day-to-day variability of the measures used in the prediction rule, such as 1-m walking time. Nonetheless, there is evidence in the literature that the size of these short-term fluctuations are relatively small, and individuals’ locations in the population distribution of performance tend to be preserved for at least 6 months,48 thus not constituting a threat for risk stratification purposes, which was the broad goal of this study.

In conclusion, this article offers a new approach to screening high-functioning older women for subsequent mobility difficulty. There are 2 contexts in which these nomograms have potential for application. First, they could be used for clinical screening and decision making. Evidence of a low disability risk can be used to reassure patients of their high functioning status, whereas further clinical evaluation might be sought for patients at high risk. Providing patients with knowledge about their short-term difficulty risk might encourage them to engage in the primary prevention activities currently recommended.49-52 Second, by permitting identification of a preclinically disabled, high-risk group of older adults, these nomograms could contribute to the delineating effective preventive interventions for forestalling the cascade of functional decline that leads to the onset of early mobility disability.

Accepted for publication March 17, 2000.

This research was supported by grant RO1 AG11703-01 from the National Institute on Aging and National Institutes of Health—National Center for Research Resources OPD-GCRC grant RR00722. Dr Chaves was supported by the PhD Program Award from the Brazilian Federal Agency for Post-Graduate Education (CAPES), Brasilia, Brazil.

We are indebted to Karen Bandeen-Roche, PhD, for statistical review of the manuscript; Jack M. Guralnik, MD, and Luigi Ferrucci, MD, for critical input regarding study design; and Carol Han for excellent assistance in manuscript preparation.

Presented in part at the 50th Annual Scientific Meeting of The Gerontological Society of America, Cincinnati, Ohio, November 17, 1997.

Corresponding author: Paulo H. M. Chaves, MD, Center on Aging and Health, The Johns Hopkins University School of Medicine, 2024 E Monument St, Suite 2-600, Baltimore, MD 21205-2223 (e-mail: pchaves@jhsp.edu).

Reprints: Linda P. Fried, MD, MPH, Center on Aging and Health, The Johns Hopkins University, 2024 E Monument St, Suite 2-600, Baltimore, MD 21205-2223.