

Lower Extremity Function and Subsequent Disability: Consistency Across Studies, Predictive Models, and Value of Gait Speed Alone Compared With the Short Physical Performance Battery

Jack M. Guralnik,¹ Luigi Ferrucci,^{1,2} Carl F. Pieper,³ Suzanne G. Leveille,¹ Kyriakos S. Markides,⁴ Glenn V. Ostir,⁴ Stephanie Studenski,⁵ Lisa F. Berkman,⁶ and Robert B. Wallace⁷

¹Epidemiology, Demography and Biometry Program, National Institute on Aging, Bethesda, Maryland.

²Geriatric Department, National Institute for Research and Care of the Elderly (INRCA), Florence, Italy.

³Center for the Study of Aging and Human Development, Duke University Medical Center, Durham, North Carolina.

⁴Department of Preventive Medicine and Community Health, Center on Aging, Galveston, Texas.

⁵Center on Aging, University of Kansas Medical Center, Kansas City, Kansas.

⁶Harvard School of Public Health, Boston, Massachusetts.

⁷Department of Preventive Medicine and Environmental Health, University of Iowa, Iowa City.

Background. Although it has been demonstrated that physical performance measures predict incident disability in previously nondisabled older persons, the available data have not been fully developed to create usable methods for determining risk profiles in community-dwelling populations. Using several populations and different follow-up periods, this study replicates previous findings by using the Established Populations for the Epidemiologic Study of the Elderly (EPESE) performance battery and provides equations for the prediction of disability risk according to age, sex, and level of performance.

Methods. Tests of balance, time to walk 8 ft, and time to rise from a chair 5 times were administered to 4,588 initially nondisabled persons in the four sites of the EPESE and to 1,946 initially nondisabled persons in the Hispanic EPESE. Follow-up assessment for activity of daily living (ADL) and mobility-related disability occurred from 1 to 6 years later.

Results. In the EPESE, compared with those with the best performance (EPESE summary performance score of 10–12), the relative risks of mobility-related disability for those with scores of 4–6 ranged from 2.9 to 4.9 and the relative risk of disability for those with scores of 7–9 ranged from 1.5 to 2.1, with similar consistent results for ADL disability. The observed rates of incident disability according to performance level in the Hispanic EPESE agreed closely with rates predicted from models developed from the EPESE sites. Receiver operating characteristic curves showed that gait speed alone performed almost as well as the full battery in predicting incident disability.

Conclusions. Performance tests of lower extremity function accurately predict disability across diverse populations. Equations derived from models using both the summary score and the gait speed alone allow for the estimation of risk of disability in community-dwelling populations and provide valuable information for estimating sample size for clinical trials of disability prevention.

INCREASINGLY, standardized tests of physical performance have been applied in research and geriatric assessment settings. Performance tests have been found to be strongly associated with multiple measures of health status and are predictive of important outcomes such as hip fracture, nursing home admission, and death. It has been generally agreed that performance measures supplement, rather than replace, self-report of disability. Along the theoretical pathway from disease to disability proposed by Nagi (1) and the Institute of Medicine (2) and made operational by Verbrugge and Jette (3), most performance measures can be considered functional limitations. These functional limitations are proximal on the pathway to disability, which represents the interaction of performance ability with an individual's environment.

It has also been demonstrated that performance measures can capture a hierarchy of functioning in persons who are not disabled and even in those who are high functioning (4,5). Furthermore, in nondisabled persons performance measures have been shown to be predictive of the onset of incident disability (6–9). In research from Project Safety (6) a threshold effect was observed, whereby risk of future disability increased substantially below a certain critical level of performance. In contrast, in the Iowa cohort of the Established Populations for Epidemiologic Research in the Elderly (EPESE), a graded effect was seen, with increasing disability risk at each lower level of performance (9).

One important limitation of the findings from the latter study was that it was restricted to the Iowa EPESE site, a ru-

ral, all-white cohort. Data have recently become available that allow for the replication of the study in other EPESE sites and for the comparison of findings both between sites and for different lengths of follow-up. This study evaluates findings across these sets of studies, assesses whether a simple measure of gait speed has the same predictive value for disability as the full performance battery used in the initial EPESE studies (9,10), presents predictive models of risk of disability and death according to performance level, age, sex, and number of years of follow-up, and examines how well these models predict disability in an external, Hispanic population. It is then demonstrated how estimates of disability outcomes can be used to determine sample size requirements for clinical trials, information that is critical in planning for interventions to prevent disability.

METHODS

Study Populations

Data are from the four sites of the EPESE and from the Hispanic EPESE. The East Boston, Iowa, New Haven (Connecticut), and North Carolina sites of the EPESE comprise a collaborative study of aging initiated by the Epidemiology, Demography, and Biometry Program of the National Institute on Aging. After recruitment of representative populations, baseline interviews were conducted in 1981 and 1982 in East Boston, Iowa, and New Haven and in 1985 and 1986 in North Carolina. Details of the baseline assessments have been described in detail (11,12). Annual contacts were made for 6 years and at the sixth follow-up measures of physical performance were added to the protocol for the first time. Subsequent to follow-up 6, which is considered the baseline for these analyses, interviews were performed at 1 year in Iowa and New Haven, at 4 years in Iowa and North Carolina, and at 6 years in New Haven. Mortality subsequent to follow-up 6 was ascertained at all sites through interviews, obituaries, and linkage with the National Death Index. No interviews were performed in East Boston after follow-up 6 and East Boston data contribute to only estimates of baseline distribution of physical performance and estimates of mortality according to baseline performance.

At follow-up 6, the number of persons not known to have died totaled 9,974 in all sites and 9,130 (91.5%) were interviewed (East Boston 2,345/2,563; Iowa 2,547/2,711; New Haven 1,671/1,819; North Carolina 2,567/2,881). Of these, 1,778 (19.5%) were excluded because they were living in institutions, living at home but needing a proxy respondent because of cognitive or physical impairment, or living outside the area and requiring a telephone interview. Of the remaining 7,352 persons interviewed in person in their homes, 141 (1.9%) refused the performance tests. To select an initially nondisabled cohort, we excluded persons reporting disability in activities of daily living (ADLs) or mobility-related disability (defined below). To further reduce disabled persons being misclassified as nondisabled at baseline, we also excluded those 4% of participants reporting no disability but having scores of 0 to 3 on the summary performance scale (defined below). A large proportion of these persons were unable to walk 8 ft or maintain a simple side-by-side stand, which is incompatible with a report of ability

to walk 0.5 mile and climb stairs. This left 4,588 persons who were initially classified as nondisabled (East Boston 1,298; Iowa 1,363; New Haven 700; North Carolina 1,227).

At follow-up, interviews to assess disability status were performed using proxies and in nursing homes if necessary. Among those not known to have died, follow-up disability information was missing at 1 year in 3.2% and 3.8% of those in Iowa and New Haven, respectively, at 4 years in 4.2% and 6.7% of those in Iowa and North Carolina, respectively, and at 6 years in 5.0% of those in New Haven. After death and loss to follow-up, subsequent information on disability status was obtained at 1 year on 1,342 persons in Iowa and 655 persons in New Haven; at 4 years on 1,121 persons in Iowa and 962 persons in North Carolina; and at 6 years on 455 persons in New Haven.

Chronic Conditions at Baseline

We assessed chronic conditions at baseline and each annual interview through follow-up 6 by asking participants if a physician had told them they had a heart attack, stroke, cancer, hip fracture, or diabetes. Cancer was considered present only if the participants also reported spending one or more nights in the hospital for the cancer. Chronic conditions were considered present if mentioned at baseline or at any follow-up interview.

Measures of Physical Performance

The physical performance battery used in the EPESE studies, previously described in detail (9,10), evaluated lower extremity function by using tests of gait speed, standing balance, and time to rise from a chair five times. Identical assessments were done in all sites by specially trained interviewers, with a training videotape used to optimize standardization of the procedures across sites. Good to excellent test-retest reliability of these tests has been demonstrated (4,13–15). Based on data from over 5,000 persons from the original three EPESE sites (East Boston, Iowa, and New Haven), a five-level categorical score was created for each test, with 0 representing inability to complete the test and 4 representing the highest level of performance (10).

For tests of standing balance, participants attempted to maintain the side-by-side, semitandem, and tandem positions for 10 seconds. Participants were scored 1 if they could hold a side-by-side stand for 10 seconds but were unable to hold a semitandem stand for 10 seconds, 2 if they held a semitandem stand for 10 seconds but were unable to hold a full tandem stand for more than 2 seconds, 3 if they held the full tandem stand for 3 to 9 seconds, and 4 if they held the full tandem stand for 10 seconds.

A usual pace, 8-ft walk was timed from a standing start, and participants were scored according to quartiles of performance. Time on the faster of two walks was used to define scores: score of 1: ≥ 5.7 seconds (≤ 0.43 m/s); score of 2: 4.1–5.6 seconds (0.44–0.60 m/s); score of 3: 3.2 to 4.0 seconds (0.61–0.77 m/s); score of 4: ≤ 3.1 seconds (≥ 0.78 m/s).

Participants were asked to fold their arms across their chest and to stand up once from a chair. If successful they were asked to stand up and sit down five times as quickly as possible. Quartiles of performance for the repeat chair stands were used to define scores as follows: score of 1:

> 16.7 seconds; score of 2: 16.6–13.7 seconds; score of 3: 13.6–11.2 seconds; score of 4: \leq 11.1 seconds.

A summary performance score was created by summation of the scores for tests of standing balance, gait speed, and rising from a chair 5 times. This scale has been demonstrated to have predictive validity in analyses showing a gradient of risk for mortality, nursing home admission, and incident disability (7,9,10). The internal consistency of the summary scale as assessed by Cronbach's alpha was 0.76 (10). The subjects in this report were nondisabled at baseline and had summary performance scores ranging from 4 to 12.

In addition to the summary performance score, gait speed alone was assessed for its relation to incident disability. An 8-ft (2.44 m) walk was used in EPESE because of concern about limited space in participants' homes. However, for the assessment of gait speed over a short distance we believe the 4-m walk is now the distance of choice because it has been demonstrated to be feasible in the home as well as in the clinical setting and its longer distance may improve measurement accuracy. In the Women's Health and Aging Study it was demonstrated that in small houses and apartments in Baltimore a 4-m walk was possible for 90% of subjects, with 9% requiring a 3-m walk and 0.8% not having space for a 3-m walk (16). Because of the advantage of using the 4-m walk for both home and clinic gait speed tests, we converted velocity in the 8-ft walk into estimated velocity for a 4-m walk in all analyses presented here. This conversion was accomplished with data from the University of Kansas Prediction of Elder Performance Study. In that study, 700 assessments were done in which time was measured for both an 8-ft and a 4-m walk. Velocity for both tests was converted to meters per second. The correlation between velocities was 0.97. Piecewise linear regression was used to evaluate whether the relationship between 8-ft walk velocity and 4-m walk velocity was linear. A model with a single break point and change in slope had a significantly better fit than a straight linear model. The break point that maximized the R^2 , determined iteratively, was at an 8-ft gait speed of 1 m/s ($R^2 = 0.96$). The observed data showed that, on average, the ratio of 4-m walk velocity to 8-ft walk velocity (both expressed in meters per second) was 1.065 for 8-ft walk velocity \leq 1.0 m/s and 1.016 for 8-ft walk velocity $>$ 1.0 m/s. The following equations were used to convert 8-ft gait speed to 4-m gait speed:

For 8-ft gait speed \leq 1.0 m/s
 $4\text{-m gait speed} = 0.01 + (8\text{-ft gait speed})(1.052).$

For 8-ft gait speed $>$ 1.0 m/s
 $4\text{-m gait speed} = 0.481 + (8\text{-ft gait speed})(0.581).$

When these equations are used, the quartiles used for the categorization of gait speed when a participant is performing a 4-m walk would become 1: $<$ 0.46 m/s; 2: 0.47–0.64 m/s; 3: 0.65–0.82 m/s; 4: \geq 0.83 m/s.

Disability Status at Follow-up

Disability status at follow-up, based on self-report or proxy report, used a three-level hierarchical scale (9,17) that contains a subset of questions from a scale originally developed by Berkman and colleagues (18). This scale, used to

exclude those with disability at baseline, was used to classify subjects at follow-up as having no disability, having mobility-related disability (inability to walk 0.5 mile or climb stairs without help), or having disability in ADLs (having mobility-related disability plus the inability to perform one or more of the following activities without help from another person: moving from a bed to a chair, using the toilet, bathing, and walking across a small room). Because the baseline performance tests measured lower extremity function, disability items related to lower extremity function were used. At each follow-up in all sites less than 2% of persons did not fit the hierarchical pattern, reporting disability in ADLs but not in mobility-related disability. These persons were classified as having disability in ADLs.

Hispanic EPESE

Predictive models developed from the original EPESE sites were evaluated with data from the Hispanic EPESE, a population-based study of 3,050 community-dwelling Mexican Americans living in five Southwestern states (19). The baseline assessment, including the same self-report items on disability and the same performance battery, was completed in 1993–1994. A follow-up interview was performed 2 years later, in 1995–1996, at which time mortality and disability data were ascertained (7).

Statistical Analysis

Rates of ADL and mobility disability were calculated according to summary performance score for each site and according to number of follow-up years. Multiple logistic models, adjusting for age, sex, and number of chronic conditions, were used to predict the relative risk of disability for those with summary performance scores of 4–6 and 7–9 compared with the reference group, those with scores of 10–12. Interactions between performance scores and other variables in the models were examined but there was no consistency seen across sites or follow-up times for the few interactions that were significant. Meta-analytic techniques were used to test for homogeneity for the two sites with 1-year follow-up data and the two sites with 4-year follow-up data (20).

To compare the relative predictive ability of gait speed with that of the summary performance score, receiver-operator characteristic (ROC) curves were constructed and tested with AccuROC software (Accumetric Corp., Montreal). Gait speed was categorized as deciles for these analyses. Areas under the curves were calculated, and comparison of these areas was done with the method of Hanley and McNeil (21,22).

In addition to evaluating relative risk, these data were used to estimate absolute risk of disability. Risk of disability was modeled in multiple logistic equations in which age, sex, and either summary performance score as an ordinal variable or gait speed as a continuous variable were entered. Coefficients from these models were used to construct equations that predict the absolute risk of disability according to these variables. Estimated risks predicted from the models are compared graphically with actual observed risk. For specific levels of performance, we calculated prediction of disability risk by entering into the predictive equation

that level of performance and the average age and sex distribution for that level of performance. In the same way, actual rate of disability after 2 years in the Hispanic EPESE cohort was compared with predicted risk for each baseline performance score after 1 and 4 years by use of the EPESE predictive equations.

Additional data are furnished to make it possible to estimate transitions along all steps in the pathway from encountering a community-dwelling population, to estimating the proportion initially disability free, to predicting the rate of survival and incident disability in survivors. The distribution of summary performance scores and categories of gait speed according to age and sex were estimated by being treated as ordinal scaled dependent variables in polychotomous logistic regression models (SAS procedure CATMOD) (23). The estimates calculated from this model for men and women of specific ages are provided in additional tables, which also provide information on the proportion expected to be nondisabled at each performance level. Rather than being adjusted for chronic disease status, which would make results for different age and sex subgroups more similar to each other, the analyses are not adjusted in order to present each age and sex group's performance distribution as you would find it in the community. Multiple logistic models were used to estimate risk of mortality according to age, sex, and performance, and equations are presented that predict absolute risk of mortality according to these variables. Preliminary to these analyses we performed site-specific analyses for mortality and found that the coefficients for age, sex, and performance were similar across sites.

Examples illustrate how these findings can be used to estimate mortality and disability outcomes for 80-year-old men and for a mixed cohort selected for a clinical trial.

RESULTS

Lower Extremity Function as a Predictor of Disability

Figure 1 illustrates the proportion of initially nondisabled persons developing ADL and mobility disability according

to summary performance score, site, and years of follow-up. For each group there is a decline in risk of disability with increasing performance. There is a higher risk for mobility disability than for ADL disability, and a higher proportion of persons become disabled with longer follow-up periods. There is consistency across sites for the 1-year follow-up results (Iowa and New Haven) and 4-year follow-up results (Iowa and North Carolina).

Table 1 shows the adjusted risk of initially nondisabled participants developing mobility disability at follow up. Compared with those with a baseline performance score of 9–12, persons with a score of 4–6 had a relative risk ranging from 2.9 to 4.9, whereas those with a score of 7–9 had a relative risk ranging from 1.5 to 2.1. Similar consistency in findings is demonstrated for risk of ADL disability in Table 2. Meta-analysis demonstrated the homogeneity of findings for ADL and mobility outcomes at 1 year and at 4 years.

Gait Speed Alone Versus the Summary Performance Score

Of the three components of the summary performance scale, the steepest gradient of risk for disability has been observed across categories of gait speed (7,9). Because gait speed is easy to measure and may be done quickly in the clinical setting, it is useful to evaluate whether measuring gait speed alone may capture the predictive power of a more comprehensive battery. To examine this we created ROC curves for the prediction of ADL and mobility disability, examining the gait speed component of the summary performance score and the summary performance score itself. Figure 2 illustrates that the area under all ROC curves is 65% or greater, with all curves significantly different from a diagonal line that indicates zero predictive ability of the test. In each case the summary score performs better than the gait speed. However, for predicting ADL disability at 4 years the two curves are not substantively or significantly different from each other. For the other three comparisons the curves do differ significantly from each other but the difference in area under the curves is not substantial, rang-

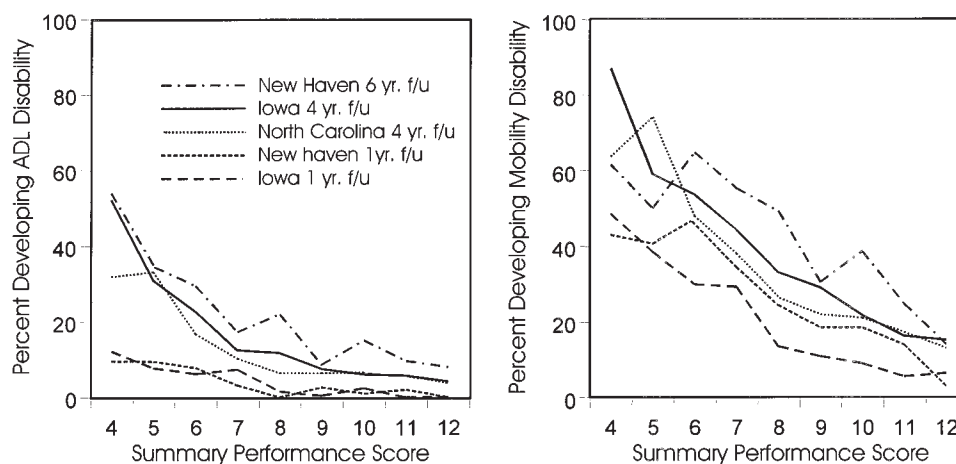


Figure 1. Percentage of participants with no disability at baseline who developed ADL disability and mobility disability according to summary performance score, site, and number of years of follow-up (f/u).

Table 1. Adjusted Relative Risk for Developing Mobility-Related Disability According to Baseline Summary Performance Score

Site (Follow-up Time)	<i>n</i> with Disability/Total <i>n</i>			Relative Risk (95% CI)	
	Summary Performance Scores				
	4–6	7–9	10–12	4–6 vs 10–12	7–9 vs 10–12
Iowa (1 y)	65/182	92/570	42/590	4.9 (3.1–7.8)	2.1 (1.4–3.2)
New Haven (1 y)	47/104	69/279	36/272	3.7 (2.1–6.4)	1.8 (1.1–2.8)
Iowa (4 y)	88/146	171/506	95/523	4.5 (3.0–6.9)	1.9 (1.4–2.5)
North Carolina (4 y)	114/194	120/440	60/328	4.9 (3.2–7.4)	1.5 (1.0–2.1)
New Haven (6 y)	39/65	82/192	54/198	2.9 (1.6–5.5)	1.7 (1.1–2.7)

Notes: Relative risks adjusted for age, sex, number of chronic conditions. Higher summary performance scores denote better performance on tests. Subjects with scores of 10–12 serve as the reference group. CI = confidence interval.

ing from 3% to 5%. These findings indicate that gait speed alone is nearly as good a predictor of disability outcomes as the full performance battery. Predictive models for disability and mortality will thus be presented for both measures.

Predicted and Observed Disability Outcomes

Table 3 combines data from the two sites with 1-year follow-up (Iowa and New Haven) and the two sites with the 4-year follow-up (Iowa and North Carolina) to present equations that estimate probability of disability. These equations result from multiple logistic models that adjust for age and sex. When age, sex, and summary performance score or gait speed are entered into the equations, an estimate of disability risk can be obtained. To examine how well these models fit the data and predict observed disability risk, the probability of disability associated with specific summary performance scores and gait velocities was calculated with the equations in Table 3. Figure 3 illustrates the predicted risk of disability and the actual risk observed in the study cohorts. The fit is excellent, with minor deviations between predicted and observed occurring mainly at the lower end of performance, where the data are more sparse than at the upper end.

The predictive accuracy of the equations was further assessed by examination of data from an external source, the Hispanic EPESE. This cohort had a 2-year follow-up after the baseline assessment of performance. Figure 4 shows the observed Hispanic EPESE ADL and mobility disability rates at 2 years and compares them with predicted rates for 1 and 4 years generated from the equations derived from

EPESE data. The observed rates generally fall within the 1- and the 4-year predictions. Overall, in the Hispanic EPESE cohort with no ADL or mobility disability at baseline with a summary performance score of 4–12 ($n = 1319$), 52 cases of ADL disability and 211 cases of mobility disability were observed at 2 years of follow-up. The predictive models projected that at 1 and 4 years, ADL disability would be present in 29 and 69 persons and mobility disability would be present in 181 and 272 persons, respectively.

Characterizing Performance and Predicting Functional Outcomes in a Community-Dwelling Population

With confirmation that performance measures can accurately predict disability outcomes in groups, it is valuable to further exploit the rich EPESE database to characterize performance levels and outcomes that can be expected to be found in a representative population. These descriptive data and predictive models may then be applied to other populations. Available EPESE data provide information on the distributions of the summary performance score and gait speed in representative older populations, and, according to level of performance, the proportion of the population that is expected to be nondisabled when first encountered, the proportion expected to die over time, and, finally, the proportion that is predicted to become disabled. With data from all four EPESE sites, Appendix A shows the distribution of the summary performance score according to age and sex. It also shows the proportion of persons in each performance category that is initially nondisabled in ADLs and mobility and thus at risk of future disability. Appendix B shows sim-

Table 2. Adjusted Relative Risk of Developing ADL Disability According to Baseline Summary Performance Score

Site (Follow-up Time)	<i>n</i> with Disability/Total <i>n</i>			Relative Risk (95% CI)	
	Summary Performance Scores				
	4–6	7–9	10–12	4–6 vs 10–12	7–9 vs 10–12
Iowa (1 y)	14/182	14/570	6/590	5.1 (1.8–14.3)	2.0 (0.7–5.3)
New Haven (1 y)	9/104	5/279	3/272	7.4 (1.8–30.5)	1.5 (0.3–6.3)
Iowa (4 y)	43/146	51/506	29/523	4.5 (2.6–8.0)	1.5 (0.9–2.5)
North Carolina (4 y)	48/194	33/440	19/328	4.0 (2.2–7.2)	1.2 (0.6–2.1)
New Haven (6 y)	23/65	29/192	23/198	3.4 (1.7–7.1)	1.2 (0.7–2.2)

Notes: Relative risks adjusted for age, sex, number of chronic conditions. Higher summary performance scores denote better performance on tests. Subjects with scores of 10–12 serve as the reference group. CI = confidence interval.

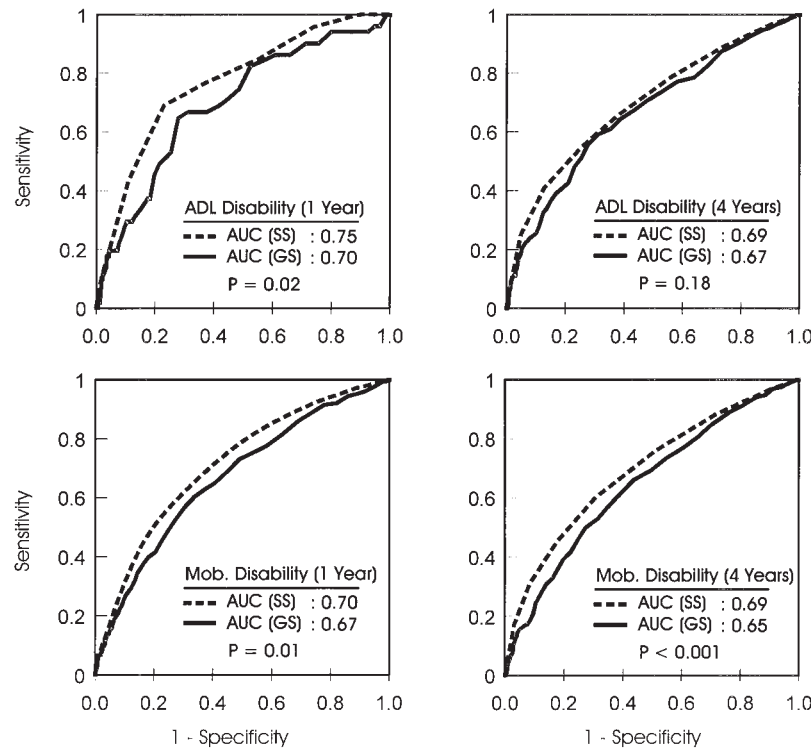


Figure 2. Receiver-operator characteristic curves showing sensitivity and 1-specificity for prediction of ADL and mobility (mob.) disability according to varying cut points for summary performance score (SS) and gait speed (GS). AUC indicates area under the curve; *p* values are for the test of null hypothesis that the areas under the curves are the same.

ilar data for gait speed. By using these data one can calculate the proportion of persons of a specific age that have a gait speed within a certain range who might be candidates for a specific intervention or clinical trial. It should be noted that these gait speeds are lower than those measured in many previous studies of gait speed and aging (24). Most of those studies, performed in gait laboratories, made measurements after subjects reached their full speed, as opposed to the home assessment done in the EPESE, in which subjects were timed from a standing start. Appendix C presents equations resulting from multiple logistic analyses that predict probability of death according to age, sex, and performance.

Table 4 demonstrates how the data provided here can be used to characterize a population and predict outcomes over time. With data from Appendix A, the distribution of the summary performance score in one-thousand 80-year-old men is shown in column 2. Further, with data from Appendix A on the proportion of nondisabled persons, column 3 shows the actual number expected to be nondisabled. The probability of death at 1 year is estimated when age, sex, and summary performance score are inserted into the first equation listed in Appendix C. The number of men surviving at each level of performance score can then be calculated. The probability of mobility disability is then estimated when age, sex, and summary performance score are inserted

Table 3. Equations Estimating Probability of Developing Disability in Mobility and Activities of Daily Living Over 1 Year and 4 Years in Survivors According to Age, Sex, and Performance

Measure of Performance	Years of Follow-up	Outcome (Type of Disability)	Probability of Disability*
Summary score	1	Mobility	$1/1 + e^{-[-5.20 + 0.08(\text{age}) + 0.08(\text{sex}) - 0.31(\text{SS})]}$
Summary score	1	ADL	$1/1 + e^{-[-8.03 + 0.10(\text{age}) - 0.18(\text{sex}) - 0.38(\text{SS})]}$
Summary score	4	Mobility	$1/1 + e^{-[-5.61 + 0.09(\text{age}) + 0.21(\text{sex}) - 0.29(\text{SS})]}$
Summary score	4	ADL	$1/1 + e^{-[-8.39 + 0.11(\text{age}) - 0.31(\text{sex}) - 0.29(\text{SS})]}$
Gait speed	1	Mobility	$1/1 + e^{-[-7.00 + 0.09(\text{age}) + 0.10(\text{sex}) - 2.64(\text{GS})]}$
Gait speed	1	ADL	$1/1 + e^{-[-10.76 + 0.12(\text{age}) - 0.15(\text{sex}) - 2.96(\text{GS})]}$
Gait speed	4	Mobility	$1/1 + e^{-[-7.28 + 0.10(\text{age}) + 0.28(\text{sex}) - 2.36(\text{GS})]}$
Gait speed	4	ADL	$1/1 + e^{-[-9.59 + 0.12(\text{age}) - 0.23(\text{sex}) - 2.81(\text{GS})]}$

*Sex = 0 for men, 1 for women; SS = summary performance score. GS = gait speed (in meters per seconds). ADL = activity of daily living. All coefficients are statistically significant ($p < .05$) except for the sex coefficient, which is significant for only the model for gait speed predicting mobility loss at 4 years.

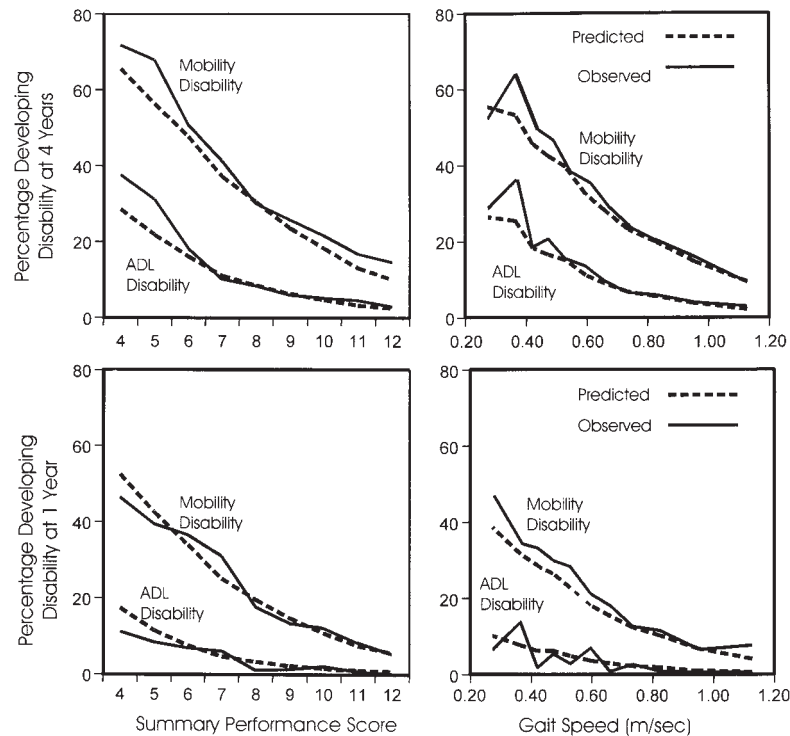


Figure 3. Predicted and observed rates of developing ADL and mobility disability at 1 and 4 years according to summary performance score and gait speed. Predicted rates are based on equations in Table 3 and utilize average age and sex distribution for the specific level of performance.

into the first equation in Table 3. Applying this probability (Table 4, column 6) to the number surviving for each level of performance gives an estimate of the number predicted to be disabled by level of performance (column 7). Summing across all levels of performance, for one-thousand 80-year-

old men, 705 are expected to be nondisabled at baseline, 675 to survive for 1 year, and 137 to develop mobility disability. Calculations for 4-year outcomes are done in a similar way by use of the appropriate equations. Appendix D illustrates how rates of death and disability incidence can be

Table 4. Distribution of Performance and Predicted Outcomes in 1,000 Community-Dwelling 80-Year-Old Men

Summary Performance Score	Baseline		1-Year Projections [†]				4-Year Projections [†]			
	<i>n</i>	<i>n</i> nondisabled	Percentage surviving*	<i>n</i>	Percentage mobility disabled [‡]	<i>n</i>	Percentage surviving*	<i>n</i>	Percentage mobility disabled [‡]	<i>n</i>
0	45	0	—				—			
1	31	0	—				—			
2	30	0	—				—			
3	41	9	93.3	9	56.7	5	50.2	5	67.3	3
4	50	17	93.8	16	49.0	8	54.2	9	60.6	6
5	60	30	94.3	28	42.3	12	58.2	17	53.5	9
6	90	58	94.8	55	34.1	19	62.0	36	46.3	17
7	109	84	95.2	80	27.5	22	65.7	55	39.2	22
8	130	112	95.6	107	21.8	23	69.2	78	32.5	25
9	142	131	96.0	126	16.9	21	72.5	95	26.5	25
10	126	120	96.3	116	13.0	15	75.6	91	21.2	19
11	90	88	96.6	85	9.9	8	78.4	69	16.8	12
12	55	54	96.9	53	7.4	4	81.0	44	13.1	6
Total	1,000	705	95.8	675	20.4	137	70.9	500	28.8	144

*Of those nondisabled at baseline.

[†]All estimated *n*'s are rounded to nearest whole number.

[‡]Of those surviving.

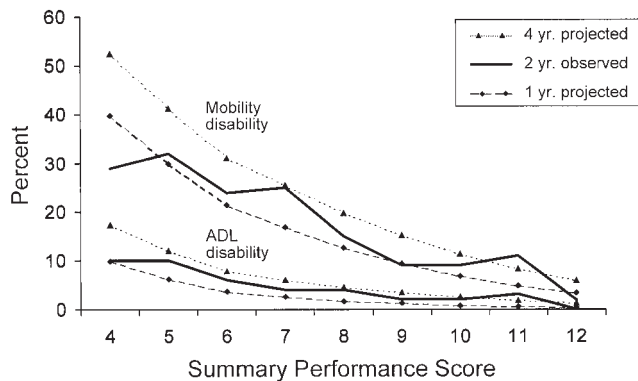


Figure 4. Predicted rates of developing ADL and mobility disability at 1 and 4 years and observed rates of developing ADL and mobility disability in the Hispanic EPESE at 2 years, according to summary performance score. Predicted rates are based on equations in Table 3 and utilize Hispanic EPESE average age and sex distribution for the specific level of performance.

estimated for a population encountered in the community. These rates are critical for making power calculations for clinical trials of disability prevention.

DISCUSSION

This article reproduces findings from previous studies that physical performance measures of lower extremity function predict the onset of disability in those initially reporting no disability in ADLs, walking a half mile and climbing stairs. The performance battery and gait speed alone were significant predictors of incident disability as long as 6 years after they were done. The absolute risk of disability increased with longer follow-up time (Figure 1) but, irrespective of time to follow-up, there was excellent consistency among the relative risks comparing the low- and medium-performance individuals with those with high performance (Tables 1 and 2). It appears that the predictive power of the performance measures wanes somewhat at 6 years, as the relative risk for the low- versus the high-performance groups was somewhat lower than for the shorter follow-ups. However, it is striking that after 6 years the increased risk associated with low performance on this simple, brief performance battery done by an interviewer in the participant's home is still of substantial magnitude and is statistically significant. This is especially impressive considering the high rate of intervening and catastrophic conditions occurring in persons in their 70s and 80s followed over a long period of time.

Our findings may be directly compared with the results in a recent report by Ostir and colleagues on the 2-year follow-up of the Hispanic EPESE, for which the performance battery was performed in the same way (7). The relative risk of mobility-related disability was 4.8 in their low-performance group and 2.4 in their middle-performance group, compared with the high-performance group. Relative risks for ADL disability were 6.2 and 2.0, respectively. These relative risks are very similar to the estimates shown in Tables 1 and 2 for three EPESE sites with different lengths of follow-up.

It is essential to examine whether the models developed to test the relationship between performance and incident

disability not only indicate an elevated *relative risk* but also accurately predict the actual rate of disability. We first found that predictions made from the models worked very well in predicting the observed outcomes in the cohorts used to develop the models (Figure 3), indicating that the models fit the data from which they were derived quite well. The true test of a model's predictive ability, however, is to test it by using an external data set. This was done with the Hispanic EPESE. Although we could not make direct predictions for the 2-year Hispanic EPESE follow-up period, it was possible to compare these observations with 1- and 4-year predictions. Considering the differences in the ethnicity of the cohorts and the regions in which the studies were done, it was remarkable how well the models predicted the range of the observed rates of disability in the Hispanic EPESE at 2 years (Figure 4). This evidence supports the validity and the robustness of these simple measures of performance and supports their use in developing disability risk estimates for diverse populations. Analyses within the North Carolina cohort also revealed no clear evidence for differences in the relationship of performance with disability between whites and African Americans (data not shown).

This research also demonstrated that assessing gait speed alone is nearly as good as performing the full battery of performance tests in the prediction of incident disability. This is a valuable finding to support the routine measurement of gait speed in older persons in the clinical setting, in which it may be quite useful to have an objective measure of lower extremity functioning, but in which there may not be time to perform a full performance battery. Gait speed assessment may also be an efficient tool as the first step in screening a large number of older persons to identify and recruit into clinical trials persons with a specific level of functioning. The equations presented in this report allow for estimation of disability and mortality risk in persons who are initially nondisabled, as determined from the few self-report questions used here, and for whom a gait speed has been measured.

The question of whether to abandon the full performance battery and use gait speed alone in all clinical and research settings is a difficult one. Certainly the assessment of rising from a chair and balance give information that may be of value in understanding the pathway from disease to disability (3). Specific diseases and impairments may affect specific aspects of lower extremity function, which may then determine the characteristics of an individual's disability. Furthermore, measuring a specific construct with multiple measures increases reliability (25), so some measurement accuracy may be gained by using the full battery. For this reason, the full battery is likely to be a better instrument in which to assess change over time. In the Iowa EPESE cohort, the summary performance battery had a mean decline of 1.6 points over 4 years, with 67% of the sample declining, 17% improving, and 16% remaining unchanged (26). Validity of the change scores was supported by the findings that, with adjustment for baseline performance, the summary performance score declined more in those ≥ 80 years old, having less education, and having coronary heart disease and chronic lung disease. Furthermore, declines in summary performance score were sensitive to the presence of depres-

sion at baseline, even after adjustment for baseline performance, demographic characteristics, behavioral risk factors, and chronic disease status.

Ultimately, knowledge gained from observational epidemiologic studies must be translated into strategies to prevent or reduce disease and disability. The real value of the findings on performance measures presented here is that they provide a strategy to target disability prevention interventions to high-risk groups that can be identified with a simple screening battery or even just a test of walking speed. The reference tables here provide the information needed to estimate the distribution of performance levels that can be anticipated to be present in the community and the probability that persons of a specific age, sex, and performance level will develop disability. This information allows for the rational development of estimates of disability risk that will put the calculation of sample sizes needed for these studies on a much more empirical footing. It should be noted that these estimates are for persons living in the community who are in their usual state of health and are not appropriate for estimating risk for persons who have acute illnesses or are just being discharged from the hospital.

The largest benefit for the effort expended in disability prevention programs may come from intervening not on those at the best level of functioning but on persons with lower levels of functioning. Thus, screening a nondisabled population and recruiting persons with summary performance scores ranging from 4 to 9 would yield a population with a high risk of disability. The higher rate of events in this population would allow for a reduced sample size compared with intervening on very high functioning individuals. Furthermore, nondisabled persons who are objectively demonstrated to have reduced performance may be on a progressively downward trajectory, so their functional abilities and quality of life could be substantially improved with a timely and well-focused intervention. This approach can support future efforts to reduce disability at both an individual and a public health level.

Acknowledgments

This research was supported by National Institute on Aging contracts NO1-AG-0-2105, NO1-AG-0-2106, NO1-AG-0-2107, and NO1-AG-1-2102, National Institute on Aging grant AG-10639, and a grant from Merck Research Laboratories for the University of Kansas Prediction of Elder Performance Study.

Address correspondence to Jack Guralnik, M.D., Ph.D., Epidemiology, Demography, and Biometry Program, National Institute on Aging, Gateway Building, Suite 3C-309, 7201 Wisconsin Avenue, Bethesda, MD 20892-9205. E-mail: jg48s@nih.gov

References

1. Nagi SZ. An epidemiology of disability among adults in the United States. *Milbank Mem Fund Q Health Soc.* 1976;54:439-467.
2. Institute of Medicine. Committee on a National Agenda for Prevention of Disabilities. *Disability in America: Toward a National Agenda for Prevention.* Washington, DC: Institute of Medicine, National Academy Press; 1991.
3. Verbrugge LM, Jette AM. The disablement process. *Soc Sci Med.* 1994;38:1-14.
4. Seeman TE, Charpentier PA, Berkman LF, et al. Predicting changes in physical performance in a high-functioning elderly cohort: MacArthur studies of successful aging. *J Gerontol Med Sci.* 1994;49:M97-M108.
5. Guralnik JM, Seeman TE, Tinetti ME, Nevitt MC, Berkman LF. Vali-

- dation and use of performance measures of functioning in a non-disabled older population: MacArthur Studies of Successful Aging. *Ag-ing Clin Exp Res.* 1994;6:410-419.
6. Gill TM, Williams CS, Tinetti ME. Assessing risk for the onset of functional dependence among older adults: the role of physical performance. *J Am Geriatr Soc.* 1995;43:603-609.
7. Ostir GV, Markides KS, Black SA, Goodwin JS. Lower body functioning as a predictor of subsequent disability among older Mexican Americans. *J Gerontol Med Sci.* 1998;53A:M491-M495.
8. Furuta T, Nagasaki H, Nishizawa S, et al. Longitudinal change in the physical performance of older adults in the community. *J Jpn Phys Ther Assoc.* 1999;1:1-5.
9. Guralnik JM, Ferrucci L, Simonsick EM, Salive ME, Wallace RB. Lower-extremity function in persons over the age of 70 years as a predictor of subsequent disability. *N Engl J Med.* 1995;332:556-561.
10. Guralnik JM, Simonsick EM, Ferrucci L, et al. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. *J Gerontol Med Sci.* 1994;49:M85-M94.
11. Cornoni-Huntley J, Brock DB, Ostfeld A, Taylor JO, Wallace RB, eds. *Established Populations for the Epidemiologic Study of the Elderly, Resource Data Book.* Bethesda, MD: National Institutes of Health; 1986.
12. Cornoni-Huntley J, Blazer DB, Lafferty ME, Everett DF, Brock DB, Farmer ME, eds. *Established Populations for the Epidemiologic Study of the Elderly. Vol. 2.* Bethesda, MD: National Institutes of Health; 1990.
13. Nevitt MC, Cummings SR, Kidd S, Black D. Risk factors for recurrent nonsyncopal falls. A prospective study. *JAMA.* 1989;261:2663-2668.
14. Ferrucci L, Guralnik JM, Salive ME, et al. Effect of age and severity of disability on short-term variation in walking speed: The Women's Health and Aging Study. *J Clin Epidemiol.* 1996;49:1089-1096.
15. Jette AM, Jette DU, Ng J, Plotkin DJ, Bach MA. Are performance-based measures sufficiently reliable for use in multicenter trials? *J Gerontol Med Sci.* 1999;54A:M3-M6.
16. Simonsick EM, Maffeo CE, Rogers SK, et al. Methodology and feasibility of a home-based examination in disabled older women: The Women's Health and Aging Study. *J Gerontol Med Sci.* 1997;52A: M264-M274.
17. Corti MC, Guralnik JM, Salive ME, Sorkin JD. Serum albumin level and physical disability as predictors of mortality in older persons. *JAMA.* 1994;272:1036-1042.
18. Berkman LF, Berkman CS, Kasl S, et al. Depressive symptoms in relation to physical health and functioning in the elderly. *Am J Epidemiol.* 1986;124:372-388.
19. Markides KS, Rudkin L, Angel R, Espino DV. Health status of Hispanic elderly. In Martin LJ, Soldo B, eds: *Racial and Ethnic Differences in the Health of Older Americans.* Washington, DC: National Academy Press; 1997.
20. Hedges L, Olkin I. *Statistical Methods for Meta-Analysis.* San Diego, CA: Academic; 1985.
21. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology.* 1982;143:29-36.
22. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology.* 1983;148:839-843.
23. SAS Institute I. *SAS/STAT User's Guide, Version 6, 4th ed. Vol. 1.* Cary, NC: SAS Institute; 1989.
24. Alexander NB. Gait disorders in older adults. *J Am Geriatr Soc.* 1996; 44:434-451.
25. Stewart AL, Hays RD, Ware JE. The MOS short-form general health survey. Reliability and validity in a patient population. *Med Care.* 1988;26:724-735.
26. Penninx BW, Guralnik JM, Ferrucci L, Simonsick EM, Deeg DJ, Wallace RB. Depressive symptoms and physical decline in community-dwelling older persons. *JAMA.* 1998;279:1720-1726.
27. Fleiss JL. *Statistical Methods for Rates and Proportions.* New York: Wiley; 1981.

Received March 22, 1999

Accepted August 18, 1999

Decision Editor: William B. Ershler, MD

Appendix A. Percent Distribution of Summary Performance Score in Men and Women of Specific Ages

Summary Performance Score	Men (Age)					Women (Age)				
	70	75	80	85	90	70	75	80	85	90
0	1.8 (0.0)	2.9 (0.0)	4.5 (0.0)	7.1 (0.0)	11.0 (0.0)	2.4 (0.0)	3.7 (0.0)	5.9 (0.0)	9.2 (0.0)	14.0 (0.0)
1	0.8 (0.0)	1.6 (0.0)	3.1 (0.0)	5.5 (0.0)	9.2 (0.0)	1.4 (0.0)	2.6 (0.0)	4.7 (0.0)	8.1 (0.0)	12.7 (0.0)
2	0.8 (0.0)	1.6 (0.0)	3.0 (0.0)	5.4 (0.0)	8.6 (0.0)	1.7 (0.0)	3.3 (0.0)	5.8 (0.0)	9.4 (0.0)	13.4 (0.0)
3	1.2 (28.7)	2.3 (25.4)	4.1 (22.4)	6.6 (19.6)	9.3 (17.1)	2.3 (21.7)	4.1 (19.0)	6.7 (16.5)	9.6 (14.4)	11.7 (12.4)
4	1.3 (42.8)	2.7 (38.7)	5.0 (34.9)	8.0 (31.2)	10.7 (27.7)	2.7 (33.9)	4.9 (30.3)	8.0 (26.9)	10.8 (23.7)	11.9 (20.8)
5	2.2 (58.1)	3.8 (54.0)	6.0 (49.8)	8.0 (45.7)	8.7 (41.6)	4.1 (48.8)	6.5 (44.6)	8.8 (40.5)	9.7 (36.6)	8.5 (32.8)
6	3.6 (72.0)	6.1 (68.5)	9.0 (64.8)	10.8 (60.9)	10.4 (56.9)	6.4 (63.8)	9.4 (59.9)	11.3 (55.8)	10.9 (51.7)	8.5 (47.5)
7	7.2 (82.7)	9.7 (80.2)	10.9 (77.4)	9.9 (74.3)	7.1 (71.0)	10.8 (76.6)	12.3 (73.5)	11.2 (70.1)	8.2 (66.5)	4.9 (62.7)
8	9.3 (89.9)	12.1 (88.2)	13.0 (86.4)	11.6 (84.3)	8.6 (82.0)	11.5 (85.9)	12.3 (83.7)	10.8 (81.3)	8.0 (78.6)	5.2 (75.7)
9	14.3 (94.3)	15.8 (93.3)	14.2 (92.2)	10.8 (90.9)	7.1 (89.4)	15.7 (91.9)	14.1 (90.5)	10.6 (89.0)	7.0 (87.2)	4.2 (85.3)
10	16.8 (96.8)	16.1 (96.3)	12.6 (95.6)	8.6 (94.9)	5.3 (94.0)	15.1 (95.4)	11.9 (94.7)	8.1 (93.7)	5.0 (92.7)	2.9 (91.5)
11	23.7 (98.3)	15.6 (98.0)	9.0 (97.6)	4.6 (97.2)	2.4 (96.7)	16.6 (97.5)	9.6 (97.0)	5.1 (96.5)	2.6 (95.9)	1.3 (95.2)
12	16.9 (99.1)	9.8 (98.9)	5.5 (98.7)	3.0 (98.5)	1.7 (98.2)	9.3 (98.6)	5.2 (98.4)	2.9 (98.1)	1.6 (97.8)	0.8 (97.4)

Note: Numbers in parentheses indicate percentage of those free of mobility and ADL disability for each age, sex, and performance subgroup. ADL = activity of daily living.

Appendix B. Percent Distribution of Gait Speed in 4-m Walk in Men and Women of Specific Ages

Gait Speed (m/s)	Men (Age)					Women (Age)				
	70	75	80	85	90	70	75	80	85	90
<0.2	1.5 (32.6)	2.6 (24.7)	4.5 (18.2)	7.5 (13.1)	12.3 (9.3)	2.8 (23.9)	4.7 (17.6)	7.9 (12.6)	12.9 (8.9)	20.5 (6.2)
0.2–0.29	1.3 (45.2)	2.3 (35.9)	3.8 (27.5)	6.0 (20.4)	9.1 (14.8)	2.6 (34.9)	4.3 (26.7)	6.8 (19.8)	10.2 (14.3)	13.9 (10.2)
0.3–0.39	2.5 (58.4)	4.1 (48.8)	6.3 (39.2)	9.1 (30.5)	12.1 (22.9)	4.7 (47.7)	7.1 (38.3)	10.2 (29.6)	13.3 (22.2)	15.2 (16.2)
0.4–0.49	5.5 (70.5)	7.9 (61.9)	10.7 (52.4)	13.3 (42.7)	14.7 (33.6)	9.0 (60.9)	12.0 (51.4)	14.4 (41.7)	15.5 (32.7)	14.5 (24.8)
0.5–0.59	9.1 (80.3)	12.3 (73.5)	15.3 (65.2)	17.1 (56.0)	16.9 (46.3)	12.0 (72.6)	14.5 (64.3)	15.8 (55.0)	15.1 (45.3)	12.9 (36.0)
0.6–0.69	14.3 (87.4)	16.1 (82.5)	16.4 (76.2)	14.8 (68.4)	12.0 (59.5)	16.7 (81.9)	16.6 (75.4)	14.7 (67.5)	11.7 (58.5)	8.4 (48.9)
0.7–0.79	17.8 (92.2)	16.7 (88.9)	14.2 (84.5)	11.0 (78.7)	7.8 (71.5)	19.0 (88.5)	15.9 (84.0)	12.1 (78.0)	8.6 (70.6)	5.7 (62.0)
0.8–0.89	20.4 (95.3)	17.5 (93.2)	14.2 (90.3)	10.8 (86.3)	7.9 (81.0)	17.0 (92.9)	13.3 (89.9)	9.9 (85.8)	7.1 (80.4)	5.0 (73.5)
0.9–0.99	12.8 (97.2)	9.2 (95.9)	6.2 (94.1)	4.0 (91.5)	2.5 (87.9)	8.3 (95.7)	5.6 (93.8)	3.6 (91.2)	2.3 (87.5)	1.4 (82.6)
≥1.0	14.7 (98.3)	11.2 (97.6)	8.5 (96.4)	6.4 (94.8)	4.7 (92.5)	8.1 (97.4)	6.1 (96.3)	4.5 (94.6)	3.4 (92.3)	2.5 (89.0)

Note: Numbers in parentheses indicate percentage of those free of mobility and ADL disability for each age, sex, and gait speed subgroup. Gait speed measured from a standing start. ADL = activity of daily living.

Appendix C. Equations Estimating Probability of Death Over 1 and 4 Years According to Age, Sex, and Performance in Those Nondisabled at Baseline

Measure of Performance	Years of Follow-up	Probability of Death*
Summary score	1	$1/1 + e^{-[-6.36 + 0.05(\text{age}) - 0.96(\text{sex}) - 0.09(\text{SS})]}$
Summary score	4	$1/1 + e^{-[-4.33 + 0.06(\text{age}) - 1.12(\text{sex}) - 0.16(\text{SS})]}$
Gait speed	1	$1/1 + e^{-[-6.40 + 0.05(\text{age}) - 0.96(\text{sex}) - 1.16(\text{GS})]}$
Gait speed	4	$1/1 + e^{-[-4.98 + 0.06(\text{age}) - 1.09(\text{sex}) - 1.60(\text{GS})]}$

*Sex = 0 for men, 1 for women; SS = summary performance score, GS = gait speed (m/s). All coefficients are significant at $p < .05$ in all models.

Appendix D

Example of Use of EPESE Data to Estimate Required Sample Sizes in Clinical Trials

The InChianti Study is an observational study of 1,340 persons in the Chianti region of Italy. A subset of nondis-

abled participants will also be invited to participate in a demonstration project that tests the impact of physical activity on disability incidence. This limited project is funded for interventions on 100 persons. The targets of this intervention will be persons with a summary performance score ranging from 4 to 9, chosen because they are at high risk of incident disability. Data and predictive equations presented in this report are used to estimate the distribution of performance in this cohort, expected survival, and expected disability rates. When these results are used, the power of the study to show a significant intervention effect can be estimated.

Table 1A shows the expected distribution of participants according to the sample drawn from the local census (column 2). The distribution of summary performance score at each age is estimated from Appendix A, and column 3 shows the total number of persons expected to have a summary score of 4 to 9 by age group. For each specific age, sex, and summary score level, the proportion of nondisabled persons is estimated from data in Appendix A, the proportion of surviving persons is estimated from equations in Ap-

Table 1A. 1-Year Projected Survivorship and Incident Mobility Disability in a Community-Dwelling Cohort of 1,340 Men and Women for those Initially Nondisabled and with a Summary Performance Score of 4 to 9. Projections for the InChianti Study, Italy

Sex and Age Group	<i>n</i>	<i>n</i> with Summary Score = 4–9	<i>n</i> Nondisabled	<i>n</i> Surviving (%)	<i>n</i> Disabled (%)
Men (Age)					
65–69	170	64	55	54 (97.8)	5 (8.5)
70–74	162	61	52	51 (97.2)	6 (12.1)
75–79	109	55	44	43 (96.4)	8 (17.7)
80–84	58	34	25	24 (95.4)	6 (25.5)
85–89	41	24	16	15 (94.0)	5 (35.2)
90+	30	16	9	9 (92.2)	4 (46.3)
Total	570	217	174	168 (96.6)	28 (16.7)
Women (Age)					
65–69	221	113	87	87 (99.1)	8 (9.6)
70–74	168	86	66	66 (98.9)	9 (13.6)
75–79	159	95	67	66 (98.6)	13 (20.1)
80–84	107	65	41	40 (98.1)	11 (28.7)
85–89	80	44	24	23 (97.6)	9 (39.0)
90+	35	15	7	7 (96.8)	3 (50.0)
Total	770	345	246	242 (98.7)	44 (18.0)

pendix C, and the proportion of persons becoming disabled in 1 year is estimated from equations in Table 3. This is done in the same way that the estimates were made for a specific age and sex in Table 4. Numbers shown in the table aggregate estimates across summary scores of 4 to 9 for each age/sex group.

Overall, of 168 men in the cohort who are predicted to have a summary performance score of 4–9 and survive 1 year, 28 (16.7%) are expected to have mobility disability at one year. Of 242 surviving women with an initial performance score of 4–9, 44 (18.0%) are expected to have mobility disability at 1 year. The intervention is hypothesized to reduce disability by half from the 17% expected. For an alpha of .05 and a power of 50%, this magnitude of effect could be demonstrated in a study with over 120 persons in both the intervention and the control groups (27). Thus this demonstration project, which is limited to 100 persons, is substantially underpowered to show a significant effect of exercise, even with a proposed halving of disability rates. To reach an acceptable power as a formal clinical trial this demonstration project would require an increased sample size.