Change in Physical Performance Over Time in Older Women: The Women’s Health and Aging Study

Graziano Onder,1,2 Brenda W.J.H. Penninx,1 Pablo Lapuerta,3 Linda P. Fried,4 Glenn V. Ostir,5 Jack M. Guralnik,5 and Marco Pahor1

1Sticht Center on Aging, Section on Gerontology and Geriatrics, Department of Internal Medicine, Wake Forest University, School of Medicine, Winston-Salem, North Carolina.
2Department of Gerontology and Geriatrics, Catholic University of the Sacred Heart, Rome, Italy.
3Pharmaceutical Research Institute, Bristol-Myers Squibb, Princeton, New Jersey.
4Departments of Medicine and Epidemiology, The Johns Hopkins Medical Institutions, Baltimore, Maryland.
5Epidemiology, Demography, and Biometry Program, National Institute on Aging, Bethesda, Maryland.

Background. Although lower and upper extremity performance measures are widely used and represent validated physical function measures in older adults, there is limited information regarding the magnitude of changes in these measures over time. This study (i) assesses prospective changes in physical performance measures, (ii) defines a summary score that demonstrates a significant amount of change over time, and (iii) examines rates of decline according to age and baseline performance levels.

Methods. Data from the Women’s Health and Aging Study (WHAS) were analyzed to assess change in the one third most disabled older women living in the community. Lower extremity function was assessed using walking speed, balance, and chair stands tests. The putting-on-blouse test, the lock and key test, the Purdue Pegboard test, and grip strength were used to gauge upper extremity function. Continuous and categorical summary performance scores were calculated using continuous and categorical data of lower and upper performance measures.

Results. After 3 years, lower extremity performance measures declined by 16%–27%, while upper extremity performance measures declined less (7%–24%). For lower extremity function, the continuous summary performance score showed a slightly greater 3-year decline from baseline (decline vs baseline mean: 23%; decline vs SD of the baseline mean: 9%); for all other tasks, with the exception of the balance test, a value corresponding to the 1st percentile of the task or who had a performance below the 1st percentile was assigned to participants who were unable to perform the task or who had a performance below the 1st percentile (walking speed: 9 cm/sec; grip strength: 5 kg). Similarly, for other tasks, with the exception of the balance test, a
value corresponding to the 99th percentile of baseline performance of participants completing the task was assigned to participants who were unable to perform the task or who had a performance above the 99th percentile (chair stands: 32.1 s; putting-on-blouse test: 233 s; lock and key test: 52.9 s; Purdue Pegboard test: 58.3 s).

Continuous Summary Performance Scores

After assigning arbitrary values as described above to worst performers and subjects unable to complete each task, individual measures were rescaled applying the following formulas (higher scores signify better performance):

(i) Walking speed: $1 - (9 / \text{speed in cm/s})$.
(ii) Chair stands test: $1 - (\text{time in s}/32.1)$.
(iii) Standing balance test: (time in s)/30).
(iv) Putting-on-blouse test: $1 - (\text{time in s}/233)$.
(v) Lock and key test: $1 - (\text{time in s}/52.9)$.
(vi) Purdue Pegboard test: $1 - (\text{time in s}/58.3)$.
(vii) Grip strength test: $1 - (5/\text{grip strength in kg})$.

Continuous summary performance scores for LE (baseline range 0–2.71) and UE (baseline range 0–3.49) were calculated by adding the rescaled scores for lower and upper tests.

Categorical Summary Performance Scores

To calculate a categorical score for the three LE measures, we used cut points derived from the Established Populations for Epidemiologic Studies of the Elderly (13) to construct separate 0 (unable to do test) to 4 (best performance) scales and one 0 to 12 summary score. Similarly, for the UE measures, 0 was assigned to those unable to do the test, and others received a score between 1 (worst performance) and 4 (best performance), based on quartiles of performance. The following cut-offs were used:

<table>
<thead>
<tr>
<th>Putting-on-blouse test (s)</th>
<th>Lock and key test (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. &gt;108.7</td>
<td>1. &gt;12.9</td>
</tr>
<tr>
<td>2. 78.8–108.7</td>
<td>2. 7.3–12.9</td>
</tr>
<tr>
<td>3. 50.6–78.7</td>
<td>3. 4.8–7.2</td>
</tr>
<tr>
<td>4. &lt;50.6</td>
<td>4. &lt;4.8</td>
</tr>
</tbody>
</table>
CHANGE IN PHYSICAL PERFORMANCE OVER TIME

M291

Table 1. One- and Three-Year Decline in Lower and Upper Extremity Function*

<table>
<thead>
<tr>
<th>Extremity Function</th>
<th>Baseline Scores</th>
<th>1-Year Change</th>
<th>3-Year Change</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>Mean (SD)</td>
<td>% Change vs Baseline Mean</td>
</tr>
<tr>
<td>Lower Balance (s)</td>
<td>1002</td>
<td>18.1 (10.3)</td>
<td>7.4 (13.3)</td>
</tr>
<tr>
<td>Chair stands (s)</td>
<td>998</td>
<td>20.1 (8.5)</td>
<td>11.2 (25.4)</td>
</tr>
<tr>
<td>Walking speed (cm/s)</td>
<td>987</td>
<td>61.2 (30.8)</td>
<td>5.2 (10.4)</td>
</tr>
<tr>
<td>Lower extremities categorical</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>summary performance score</td>
<td>982</td>
<td>5.9 (3.3)</td>
<td>6.6 (12.4)</td>
</tr>
<tr>
<td>Lower extremities continuous</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>summary performance</td>
<td>982</td>
<td>1.71 (0.71)</td>
<td>8.8 (21.1)</td>
</tr>
<tr>
<td>Upper Put on and button blouse (s)</td>
<td>975</td>
<td>115.6 (72.4)</td>
<td>3.9 (5.9)</td>
</tr>
<tr>
<td>Lock and key test (s)</td>
<td>994</td>
<td>13.9 (14.2)</td>
<td>-8.4 (-7.9)</td>
</tr>
<tr>
<td>Purdue Pegboard (s)</td>
<td>997</td>
<td>32.9 (10.7)</td>
<td>22.6 (14.5)</td>
</tr>
<tr>
<td>Grip strength (kg)</td>
<td>930</td>
<td>19.7 (5.9)</td>
<td>1.0 (4.0)</td>
</tr>
<tr>
<td>Upper extremities categorical</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>summary performance score</td>
<td>891</td>
<td>9.4 (3.8)</td>
<td>0.2 (0.5)</td>
</tr>
<tr>
<td>Upper extremities continuous</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>summary performance</td>
<td>891</td>
<td>2.57 (0.65)</td>
<td>0.8 (2.6)</td>
</tr>
</tbody>
</table>

*aNegative values signify improvement.

**Percent change vs standard deviation (SD) of the mean was calculated with the following formula: 100 × mean change/SD of mean at baseline. SD of all baseline participants was used for these analyses.

Table 2. Average Annual Decline in Lower Extremity Function Among 927 Participants With Baseline Data for All Three Lower Extremity Tests and Their Summary Scores*

<table>
<thead>
<tr>
<th>Baseline Lower Extremity Performance</th>
<th>All N = 927</th>
<th>Good Performers n = 168</th>
<th>Intermediate Performers n = 444</th>
<th>Poor Performers n = 315</th>
<th>p Intermediate vs Good</th>
<th>p Intermediate vs Poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual Decline In</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Young-old (65–79 years) (n = 540)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Balance test (s)</td>
<td>1.6</td>
<td>1.7</td>
<td>1.8</td>
<td>1.1</td>
<td>.575 (.049)</td>
<td></td>
</tr>
<tr>
<td>Chair stands (s)</td>
<td>1.5</td>
<td>1.6</td>
<td>2.0</td>
<td>0.3</td>
<td>.132 (.001)</td>
<td></td>
</tr>
<tr>
<td>Usual walking speed (cm/s)</td>
<td>3.4</td>
<td>2.7</td>
<td>2.9</td>
<td>5.3</td>
<td>.849 (.055)</td>
<td></td>
</tr>
<tr>
<td>Lower extremities categorical</td>
<td>0.4</td>
<td>0.4</td>
<td>0.5</td>
<td>0.3</td>
<td>.600 (.411)</td>
<td></td>
</tr>
<tr>
<td>Lower extremities continuous</td>
<td>0.12</td>
<td>0.11</td>
<td>0.13</td>
<td>0.11</td>
<td>.351 (.322)</td>
<td></td>
</tr>
<tr>
<td>Old-old (80 years) (n = 387)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Balance test (s)</td>
<td>2.1**</td>
<td>2.9</td>
<td>2.8</td>
<td>1.4</td>
<td>.892 (&lt;.001)</td>
<td></td>
</tr>
<tr>
<td>Chair stands (s)</td>
<td>2.2†</td>
<td>2.2</td>
<td>3.4</td>
<td>1.0</td>
<td>.154 (&lt;.001)</td>
<td></td>
</tr>
<tr>
<td>Usual walking speed (cm/s)</td>
<td>6.2‡</td>
<td>6.3</td>
<td>6.5</td>
<td>6.2</td>
<td>.665 (.617)</td>
<td></td>
</tr>
<tr>
<td>Lower extremities categorical</td>
<td>0.6†</td>
<td>0.9</td>
<td>0.9</td>
<td>0.3</td>
<td>.500 (&lt;.001)</td>
<td></td>
</tr>
<tr>
<td>Lower extremities continuous</td>
<td>0.20†</td>
<td>0.19</td>
<td>0.25</td>
<td>0.16</td>
<td>.294 (.001)</td>
<td></td>
</tr>
</tbody>
</table>

Note: Good performers = categorical lower extremity score 10–12; Intermediate performers = categorical lower extremity score 4–9 and able to perform all lower extremity tests; Poor performers = categorical lower extremity score 0–3 or unable to perform one or more lower extremity tests.

*aAll the analyses are adjusted for baseline performance score.

**p vs decline in young-old participants = .02.

†p vs decline in young-old participants = .002.

‡p vs decline in young-old participants = .001.

Data Analyses

We examined the average decline/year in performance measures after stratification by age and baseline performance subgroups, using mixed model analysis of covariance (SAS Version 6.12, SAS Institute, Cary, NC). We used random intercept and random slope in a growth curve model. Analyses were adjusted for baseline value of the outcome variable.
Young-old (65–79 years) (in the group older than 80, presented a higher decline
performers. The only exception was walking speed, which, from both age groups with an intermediate level of base-
line performance were more likely to decline than poor
performers. One possible explanation for this observation, given that the poor perfor-

did not account for this observation, given that the poor perfor-

Intermediate performers may have pre-
dedimental decline in function. Alternatively, a floor effect may
represent a useful outcome measure for clinical studies
of physical function. For this reason, and in consideration of their ability to pre-
dict incident disability (6,13,18,19), LE measures seem
preferable outcomes for studies that examine prospective
changes in physical function. More specifically, the contin-
uous summary score of LE performance, which showed a
larger decline from baseline SD of the mean than other tests,
may represent a useful outcome measure for clinical studies
of physical function.

Participants with intermediate baseline levels of performance were more likely to decline in LE measures and
scores than poor performers. One possible explanation for this finding is that intermediate performers may have pre-
clinical disabilities that will eventually trigger more precipi-
tous declines in function. Alternatively, a floor effect may
account for this observation, given that the poor perfor-

Table 3. Average Annual Decline in Upper Extremity Function Among 842 Participants With Baseline Data for All Four Upper Extremity
Tests and Their Summary Scores*

<table>
<thead>
<tr>
<th>Annual Decline In</th>
<th>All N = 842</th>
<th>Good performers n = 209</th>
<th>Intermediate performers n = 407</th>
<th>Poor performers n = 226</th>
<th>Intermediate vs good</th>
<th>Intermediate vs poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Young-old (65–79 years) (n = 495)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Put on and button blouse (s)</td>
<td>6.9</td>
<td>9.1</td>
<td>12.0</td>
<td>-11.2</td>
<td>.242</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Lock and key test (s)</td>
<td>0.1</td>
<td>0.1</td>
<td>-0.1</td>
<td>0.5</td>
<td>.691</td>
<td>.530</td>
</tr>
<tr>
<td>Purdue Pegboard (s)</td>
<td>1.6</td>
<td>1.3</td>
<td>1.8</td>
<td>1.3</td>
<td>.173</td>
<td>.350</td>
</tr>
<tr>
<td>Grip strength (kg)</td>
<td>0.5</td>
<td>0.4</td>
<td>0.5</td>
<td>0.5</td>
<td>.816</td>
<td>.625</td>
</tr>
<tr>
<td>Upper extremities categorical summary performance score</td>
<td>0.1</td>
<td>0.3</td>
<td>0.1</td>
<td>-0.2</td>
<td>.024</td>
<td>.135</td>
</tr>
<tr>
<td>Upper extremities continuous summary performance</td>
<td>0.03</td>
<td>0.04</td>
<td>0.05</td>
<td>-0.07</td>
<td>.131</td>
<td>.040</td>
</tr>
<tr>
<td>Old-old (≥ 80 years) (n = 347)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Put on and button blouse (s)</td>
<td>13.0**</td>
<td>12.0</td>
<td>18.2</td>
<td>-4.6</td>
<td>.364</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Lock and key test (s)</td>
<td>1.5†</td>
<td>2.1</td>
<td>2.0</td>
<td>0.9</td>
<td>.960</td>
<td>.641</td>
</tr>
<tr>
<td>Purdue Pegboard (s)</td>
<td>2.5†</td>
<td>2.9</td>
<td>2.8</td>
<td>2.4</td>
<td>.820</td>
<td>.403</td>
</tr>
<tr>
<td>Grip strength (kg)</td>
<td>0.6</td>
<td>0.6</td>
<td>0.7</td>
<td>0.6</td>
<td>.791</td>
<td>.474</td>
</tr>
<tr>
<td>Upper extremities categorical summary performance score</td>
<td>0.5†</td>
<td>1.3</td>
<td>0.6</td>
<td>0.0</td>
<td>.029</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Upper extremities continuous summary performance performance score</td>
<td>0.09†</td>
<td>0.11</td>
<td>0.13</td>
<td>0.01</td>
<td>.102</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

*All the analyses are adjusted for baseline performance score. Negative values signify improvement.

**p vs decline in young-old participants = .04.
†p vs decline in young-old participants = .008.
‡p vs decline in young-old participants < .001.

Results

The mean age of the 1002 participants was 78.9 ± 8.1
years, 28.3% were black, and at baseline, 31.5% reported a
lot of difficulty or were unable to perform one or more ac-

divities of daily living. The decline in the LE measures (ex-
pressed as percent change from baseline mean) ranged from
5.2% to 11.2% after 1 year and from 16.3% to 26.6% after 3
years (Table 1). The lock and key test was the only UE task
presenting a 1-year improvement from baseline. The 1- and
3-year declines from baseline mean of the other UE mea-

ures ranged from 1.0% to 7.4% and from 6.8% to 23.9%.

For both LE and UE, the magnitude of the decline expressed
as percent change versus the baseline mean in categorical and
continuous scores was similar. However, for LE perfor-
mance, the average 1- and 3-year declines of the categorical
summary score, expressed as percent change versus the
baseline standard deviation of the mean, were substantially
lower (12.4% and 41.0%) than those of the continuous sum-
mary score (21.1% and 59.2%).

Participants older than 80 years experienced greater de-
cline in all performance measures and summary scores than
women younger than 80 years (Tables 2 and 3). These
results were virtually unchanged after adjustment for
MMSE score. Regarding LE performance, participants from both age groups with an intermediate level of base-
line performance were more likely to decline than poor
performers. The only exception was walking speed, which,
in the group older than 80, presented a higher decline
among poor performers than among both good and inter-
mediate performers.

Discussion

Compared with healthier populations, we described larger
changes in LE measures (3,14,15), probably because the
WHAS participants are all disabled and, therefore, have a
higher risk of declining in function (16). The decline in UE
tests was not linear, in particular for the lock and key test
and the put-on-blouse test, probably because these two tests
have a lower test-retest reliability than other measures (17).
For this reason, and in consideration of their ability to pre-
dict incident disability (6,13,18,19), LE measures seem
preferable outcomes for studies that examine prospective
changes in physical function. More specifically, the contin-
uous summary score of LE performance, which showed a
larger decline from baseline SD of the mean than other tests,
may represent a useful outcome measure for clinical studies
of physical function.

Participants with intermediate baseline levels of performance were more likely to decline in LE measures and
scores than poor performers. One possible explanation for this finding is that intermediate performers may have pre-
clinical disabilities that will eventually trigger more precipi-
tous declines in function. Alternatively, a floor effect may
account for this observation, given that the poor perfor-


among poor performers than among both good and inter-
mediate performers.
age can identify those at greatest risk for physical performance decline.

Acknowledgments

The Women’s Health and Aging Study was supported by the National Institute on Aging (NIA; Contract No. N01AG12112). The work of Dr. Onder is supported by the Wake Forest Claude D. Pepper Older Americans Independence Center (NIA Grant 5P60 AG 10484-07).

Address correspondence to Graziano Onder, MD, Section on Gerontology and Geriatrics—Sticht Center on Aging, Wake Forest University—Baptist Medical Center, Medical Center Boulevard, Winston-Salem, NC 27157. E-mail: graziano_onder@rm.unicatt.it

References


Received October 15, 2001
Accepted January 14, 2002