Within the US, 30 up to nearly 50% of dogs may be categorized as geriatric based on various survey analyses. The process of aging involves a reduction in organ and systemic functional reserve due to degeneration and cellular senescence, resulting in an increased risk of multiple age-related comorbidities. Many diseases of aging compromise mobility, including metabolic, degenerative, musculoskeletal, neurological, cardiovascular, respiratory, and neoplastic processes. Sarcopenia and obesity represent a set of commonly identified comorbidities of aging closely linked to inactivity, reduced functional capacity, and sedentary behavior. A disease diagnosis alone, however, is often a poor indicator of daily function since many diseases of aging are progressive in nature, resulting in a spectrum of impairment. The assessment of mobility is considered an essential component of the geriatric exam, with

**Timed up and go demonstrates strong interrater agreement and criterion validity as a functional test in geriatric dogs**

Kelly M. McMullin, DVM; Patrick C. Carney, DVM, DACVIM, PhD; Ivanka G. Juran, BS; Kurt R. Venator, DVM, PhD; Allison V. Miller, DVM; Margret I. Lenfest, VMD; Brittany Jean Carr, DVM, DACVSMR; Christopher W. Frye, DVM, DACVSMR*

*Corresponding author: Dr. Frye (chris.frye@cornell.edu)

**OBJECTIVE**

To measure interobserver agreement for 4 functional tasks and their summed geriatric functional score (GFS) and correlate tasks and GFS with client-specific outcome measurements (CSOMs): Canine Brief Pain Inventory (CBPI) pain severity, CBPI pain interference, and Liverpool Osteoarthritis in Dogs.

**ANIMALS**

89 geriatric dogs were recruited between April and September 2023 from staff, friends, and clients of the Cornell University College of Veterinary Medicine with a median age of 11.0 years and weight of 26.4 kg.

**METHODS**

Dogs underwent 4 sequential functional tests: timed up and go (TUG), cavallettis, figure 8s, and down to stands. Two observers independently scored each dog. The GFS was calculated based on the summed scores of the individual tests. Additional information collected included signalment, weight, measurements reflecting the comorbidities of aging (body condition score and muscle condition score), and CSOMs.

**RESULTS**

Strong interrater agreement was found for all functional tests. The TUG in seconds (sTUG) and figure 8s demonstrated significant (P < .05) moderate to strong correlations to all CSOMs. The GFS showed similar significant correlations with all CSOMs except CBPI pain severity; however, when correlating individual tests to CSOMs, only figure 8s and TUG were significantly contributing to GFS results. Receiver operating characteristic curve analysis defined highly functional dogs as those completing the sTUG in under 3.83 seconds. The sTUG represented the best test for geriatric function given it was objective, reliable, correlated well to CSOMs, and could help identify highly functioning dogs.

**CLINICAL RELEVANCE**

The sTUG appears to be the first practical and reliable functional test of canine geriatric mobility.

**Keywords:** TUG, dog, geriatric, mobility, function

Received March 08, 2024
Accepted April 30, 2024
doi.org/10.2460/ajvr.24.03.0062

© 2024 THE AUTHORS. Published by the American Veterinary Medical Association as an Open Access article under Creative Commons CCBY-NC license.
clinical functional testing providing different benefits compared to survey scoring or physical examination.5,6 Therefore, standardized clinical functional tests have been developed for geriatric humans to measure physical attributes including strength, balance, and coordination to provide prognosis or monitor response to treatment. Poor scores in validated tests, including the timed up and go (TUG), have been associated with increased frailty, sarcopenia, disability, falls, fractures, hospital stays, and mortality.7,8 In contrast to human medicine, a paucity of clinical functional tests have been evaluated in canine medicine, including the 6-minute walk test and ratings of the perceived exertion scale.9,10

Neither test has been applied to assess geriatric mobility, and both tests are cumbersome to employ. Recently, owner videos guided by specific instruction were used to grade functionality in dogs suffering from osteoarthritis (OA); however, the tests were not standardized (obstacle, flooring, and dog size) nor validated before the trial.5 Despite indirect measurements of mobility having been correlated to quality of life (QOL) and time to death in aging dogs,5 there remains a need for a validated, practical, in-clinic instrument that directly assesses geriatric function.

We have previously proposed a testable framework for clinical canine geriatric functional scoring (GFS) consisting of a summed score of 4 sequential standardized tasks: TUG, cavalletti, figure 8s, and down to stands (Supplementary Table S1).11,12 The tasks are multidimensional: aimed at testing balance, proprioception, strength, and endurance. Furthermore, the GFS is practical in that it does not require special facilities, additional dog training, large amounts of time and personnel, or complex handling. Both the TUG and figure 8s were derived from validated human geriatric tests.7 All GFS scoring was ordinal in nature; however, TUG and down-to-stand scoring was categorized from measured objective interval data. TUG time in seconds (sTUG) should be assessed independently from GFS because it provides an objective performance measure, is not impacted by the other subsequent tests in our model, and is well-established in human medicine. Geriatric functional scoring has never been investigated in dogs and lacks a gold standard against which to compare our clinical testing. Therefore, 2 well-established client-specific outcome measures (CSOMs) with a heavy focus on canine mobility were employed to assess criteria validity: the Canine Brief Pain Inventory (CBPI) and Liverpool Osteoarthritis in Dogs (LOAD).13,14

The primary aims of this study are to assess interrater agreement between each of the 4 individual tests as well as overall GFS and sTUG and to assess criterion validity by correlating sTUG, GFS, and its subtests against CBPI and LOAD questionnaires.

The secondary aims of this study are as follows to differentiate highly functioning versus lower functioning dog groups based on sTUG and GFS and to assess the relationships between signalment, muscle condition score (MCS), and body condition score (BCS) with sTUG, GFS, and its subtests.

We hypothesize that interrater agreement will be strong for all individual tests as well as total GFS; that lower GFS score and higher sTUG will strongly correlate to higher pain status, lower function, and reduced QOL as measured by CBPI and LOAD; that GFS or sTUG can be used to distinguish highly functional from less functional dog populations; and that lower muscle condition, increased age, and higher body condition will correlate to lower GFS and higher sTUG values.

Methods

Recruitment, inclusion, and exclusion

Dogs were recruited from Cornell University College of Veterinary Medicine between April 21 and September 8, 2023, by soliciting staff, students, faculty, friends, and clients of the canine sports medicine service. Inclusion criteria consisted of dogs amenable to handling and classified as geriatric based on a previously published model of age to body weight (≥ 12 years and ≤ 9.1 kg; ≥ 10 years and between 9.2 and 22.7 kg; ≥ 8 years and between 22.8 and 54.5 kg; ≥ 6 years and > 54.5 kg).15,16 Exclusion criteria consisted of dogs who underwent surgery within the preceding 3 months or had any changes in pain medications, including therapeutic joint injections, within the preceding week. Study approval was obtained by the Cornell University Veterinary Clinical Studies Committee for IACUC exemption, and informed client consent was collected before enrollment. All enrollees were provided a $100 incentive to participate.

Study population

Initial information gathered included dog signalment (age, breed, sex, and neuter status), the owner’s name and contact information, and the presence of any veterinarian-diagnosed musculoskeletal or neurological disease contributing to decreased mobility. Before testing, body weight and body length (defined as the distance from the tip of the nose to the base of the tail) were measured for each dog. The BCS was assigned by a veterinarian on a scale of 1 to 9 as previously described.17 The MCS was determined by assessing generalized muscle loss over the epaxial musculature by veterinarian palpation and designated as none (4), mild (3), moderate (2), or severe (1).18,19

Testing procedures

All tests were conducted on level nonslip flooring away from outside distractions including owners. Any harnesses, collars, or accessories were removed, and a standard slip lead was placed around the dog’s neck for use throughout the test. Gentle tension to direct the dog was allowed, but no additional forces or pulling was permitted. Dogs moved from one test to the next without allowing for rest. At the beginning of each test, a timer was immediately started by...
an evaluator after finishing a countdown to “go.” A
dog was considered incapable of completing an indi-
vidual test if they fell during the test or if they failed
to complete it within 60 seconds, in which case they
were immediately moved to the next test. To mini-
mize behavioral and training bias, verbal commands
and hand signals were limited to encouraging dogs
to move forward only. No rewards (treats, toys) were
provided as motivation given inherently varying lev-
eels of food and play drive. As a safety precaution,
an appropriately sized basket muzzle (Jorvet nylon
plastic basket muzzle; Jorgensen Laboratories) was
applied in advance of testing for dogs that appeared
averse to handling, provided it did not result in any
observable behavioral interference with the func-
tional testing. If an observer believed a dog’s behav-
ior undermined testing results for any reason, the
test would be marked “behavioral interference” and
no numerical score assigned. Two observers inde-
dependently scored each test in real time for all dogs
using a tablet with access to a web-based survey
platform based on the previously proposed frame-
work (Supplementary Table S1). Each of the 4 indi-
vidual tests conducted in succession was assigned
a point score ranging from 0 to 4, with 4 being the
most functional score achievable and 0 being inca-
 capable of completing the task. The GFS was then
calculated based on the summed scores of each indi-
vidual test and could range from 0 to 16. Observers
were blinded to the other’s scores.

Test area setup
The tests were set up to accommodate the most
efficient chronological and spatial succession to mini-
mize time and effort between the tests. First, 10 body
lengths were measured against a wall with start and
finish points marked by tape. Second, 2 caval-
letti rail obstacles were set at the dog’s approximate
tarsal height and spaced apart by 1 body length. Third, a set of 2 traffic cones were spaced apart by
1 body length.

Timed up and go
A handler positioned the dog in sternal recum-
bency with its front paws at the starting line, applying
a hand atop the dog’s withers to maintain position. A
second person (caller) stood 5 meters beyond the fin-
ish line and encouraged the dog to move at its quick-
est pace across the finish line in a straight line
using the adjacent wall as a guide on one side
and the handler on the other. Both observers were
positioned looking down the finish line, each with a
stopwatch. Immediately upon finishing the call out,
“ready-set-go,” the stopwatches were started and
the handler guided the dog into rising and travel-
ing across the finish line (Supplementary Video V1).
Time was recorded in seconds to 2 decimals.

Cavalletti
The dog was positioned facing the 2 caval-
letti rails, the stopwatch was started, and the dog
was led over both rails at a walk, turned around
after completion, and led back over the same rails at a walk.

Figure 8s
The dog was positioned in alignment with the
2 cones, the stopwatch was started, and the dog
was walked between the 2 cones in a figure-8 pat-
tern until 4 figure 8s were completed. The handler
remained outside the cones.

Down to stands
A handler positioned on the floor behind the dog
initially guided the dog into a seated position using
one hand to gently push down on the caudal dorsum
with the other hand positioned just caudal to the
stifles. Once in a seated position, the forelimbs were
gently grasped at the level of the antebrachium, and
the dog was assisted down into sternal recumbency.
A second handler was positioned in front of the dog
holding its leash. Once the stopwatch was started,
the dog was encouraged into a standing position
(defined by normal forelimb and hindlimb standing
angles as observed by the handler on the floor). This
process was immediately repeated as many times
as the dog was capable within a 60-second period.
Each evaluator was equipped with a manual pitch
counter (Uxcell mechanical 4-digit number click
pitch counter) to silently keep track of the number of
stands from a down position. Only complete transi-
tions from down to stand were counted.

Owner surveys
While the dogs were completing functional
testing, their owners were seated in a separate
location with a tablet and instructed to fill out the
following web-based surveys regarding their dogs’
status over the last 7 days: LOAD (questions M1 to
M13) and CBPI (questions 1 to 11). The LOAD
is a summed 13-question instrument with a 5-point
Likert-like scale (0 to 4) with higher scores rela-
ting to worse mobility. The CBPI contains 3 sections:
4 questions that are averaged regarding pain sever-
ity (psCBPI), 6 questions that are averaged regard-
ing pain interference (piCBPI), and 1 question
regarding QOL (qlCBPI). The psCBPI and piCBPI
questions have an 11-point scale (0 to 10) that is
averaged per section. The qlCBPI has a 5-point
Likert-type scale (0 to 4). Higher scoring reflects
worse pain status, more pain interfering with daily
function, and improved QOL.

Statistics
A statistical software program SAS 9.4
(SAS Institute Inc) was used to analyze all data.
Continuous variables were summarized as mean
and SD or median, IQR range, and range based on
normality as determined by the Shapiro-Wilk test.
Categorical variables were summarized as per-
centages. Performance between 2 evaluators was
assessed using the Shrout-Fleiss intraclass correla-
tion coefficient for continuous measures and the
Cohen weighted $k$ for categorical measures.
Bivariate associations between primary measures of interest and BCS, MCS, total LOAD mobility score, and average CBPI pain score were assessed using simple linear regression and the Pearson correlation coefficient. Homoskedasticity was visually assessed via plots of predicted values by residuals, and normality of residuals was assessed by visual inspection of normal quantile-quantile (QQ) plots; if the normality assumption was violated, the primary measure was natural log transformed, and the model was refitted. If residuals of the natural log-transformed model were not normally distributed, then the Spearman rank correlation coefficient and associated 95% CI were calculated. For associations between primary measures of interest and breed, ANOVA was used; if residuals were not normally distributed and log transformation of the primary measure did not normalize residuals, the Kruskal-Wallis test was used. Multiple linear regression was assessed for collinearity by calculating the variance inflation factor for each independent variable, with a rule-of-thumb cutoff of less than 10 suggesting negligible collinearity. Backward stepwise selection was used for model fitting, with a P value less than .2 as the retention criterion.

Bivariate associations between selected individual items of the LOAD and CBPI plausibly related to mobility and function and the primary measures of interest were assessed via Spearman rank correlation coefficients.

Partial Spearman correlation was used to assess the association between GFS subtests and selected CSOMs while controlling for all other GFS subtests. Subtests with significant associations, as well as GFS, were then assessed for significant associations with selected CSOMs while controlling for those patient characteristics found to be significantly associated with GFS in multiple linear regression.

A high-functioning dog population was defined by having a sum score of 0 or 1 on selected general mobility questions (LOAD M1, LOAD M2, and CBPI 5). Receiver operating characteristic curve analysis was performed on sTUG to distinguish between high-functioning and mobility-impaired dogs, with optimal cutoff determined by the Youden index. P values less than .05 were considered statistically significant.

Results

Population

There were 89 subjects enrolled in the study with a median age of 11.0 years (IQR, 4.0; range, 8 to 17), a median weight of 26.4 kg (IQR, 17.3; range, 2.6 to 46.4), and a median ideal weight of 26.0 kg (IQR, 16.3; range, 3.3 to 46.4). Forty (44.9%) were male neutered, 47 (52.8%) were female spayed, and 2 (2.3%) were intact males. Twenty-three breeds were identified. Mixed-breed dogs were the most common (37.1%), followed by Labrador Retrievers (13.5%) and Golden Retrievers (9.0%). Each of the remaining breeds made up less than 5% of the total population. Of the dogs, 37.1% exhibited mild muscle loss (MCS 3), 36% showed no loss (MCS 4), 21.4% showed moderate loss (MCS 2), and 5.6% had marked loss (MCS 1). Of the dogs, 42.7% were in ideal body condition (BCS 5/9), 24.7% had a BCS of 4/9, 22.5% had a BCS of 6/9, and 3 each (3.4%) had a score of 3/9, 7/9, or 8/9. No dogs scored a 1, 2, or 9/9. Forty-nine dogs (55.1%) had previously been diagnosed by a veterinarian with orthopedic and neurologic diseases affecting their mobility.

Results of CBPI and LOAD

Mean total LOAD score was 17.00 ± 8.48, median psCBPI was 2.00 (IQR, 3.75; range, 0.00 to 7.75), median piCBPI was 2.17 (IQR, 5.00; range, 0.00 to 10.00), and median qCBPI score was 1.00 (IQR, 2.00; range, 0.00 to 4.00).

Median scores and interrater agreement for sTUG and GFS

Median scores as well as interobserver agreement for sTUG, GFS, and each subtest are shown (Table 1). Based on excellent interobserver agreement, the primary measures of interest (sTUG and GFS or its individual components) were averaged between the 2 evaluators for the remainder of the analysis.

Associations of sTUG and GFS with demographics and comorbidities of aging

The natural logarithm of sTUG was significantly associated with age (P < .01) in simple linear regression (Table 2). For each 1 unit increase in age, sTUG on average increased by 7.7% (95% CI, 2.9 to 12.7). The sTUG was not significantly associated with BCS but the natural logarithm of sTUG was significantly (P < .001) positively correlated with MCS; for each 1 unit increase in MCS, the sTUG decreased on average by 20.4% (95% CI, 11.0 to 28.7). There were no significant correlations between sTUG and breed, sex, weight, or BCS.

There was a significant (P < .0001) moderate to strong negative correlation between age and GFS. The MCS and GFS were also significantly (P < .0001) correlated; for every increase in the MCS by 1 unit, the score increased by 2.32 (95% CI, 1.37 to 3.27). The BCS was weakly but significantly (P = .029) correlated to GFS. There were no significant correlations noted in GFS by breed, sex, or weight.

A multiple linear regression model predicting GFS using age, sex, breed, BCS, MCS, and weight as independent variables was assessed. No strong evidence of collinearity was noted (variance inflation factors of 1.24 to 1.67). Three individuals were identified as marked outliers/leverage points both graphically and based on the Cook D and were found to be dogs with neurologic disease that rendered them nonambulatory and incapable of completing any of the tasks; removal of these 3 from the analysis resulted in approximately normally distributed residuals based on normal QQ plot inspection. Independent variables retained at the P < .2 threshold were age (for each additional year of age, score decreased by 0.91 [95% CI, –1.28 to –0.54], P < .0001), MCS (for each additional MCS
category, score increased by 1.73 [95% CI, 0.94 to 2.52], *P* < .0001), and weight (for each additional kg, score decreased by 0.12 [95% CI, –0.19 to –0.06], *P* = .016).

**Table 1**—Median scores and interobserver analysis results for timed up and go (TUG) measured in seconds (sTUG), geriatric functional score (GFS) calculated by the sum of all individual GFS tests scores, and all GFS subtests based on a scoring rubric ranging from 0 to 4 (0 is incapable, and 4 is highest capability) for 89 client and student-owned geriatric dogs between April 21 and September 8, 2023.

<table>
<thead>
<tr>
<th>Score/examiner</th>
<th>Bl</th>
<th>Median</th>
<th>IQR</th>
<th>Range</th>
<th>Interrater</th>
</tr>
</thead>
<tbody>
<tr>
<td>sTUG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>5.03</td>
<td>3.02</td>
<td>1.90 to &gt; 60.00</td>
<td>0.994ab</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>5.00</td>
<td>3.40</td>
<td>1.54 to &gt; 60.00</td>
<td></td>
</tr>
<tr>
<td>GFS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>19</td>
<td>11.50</td>
<td>6.00</td>
<td>0.00 to 16.00</td>
<td>0.977ab</td>
</tr>
<tr>
<td>B</td>
<td>19</td>
<td>12.00</td>
<td>6.00</td>
<td>0.00 to 16.00</td>
<td></td>
</tr>
<tr>
<td>GFS subtest</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TUG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>3.00</td>
<td>1.00</td>
<td>0.00 to 4.00</td>
<td>0.980ab</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>4.00</td>
<td>1.00</td>
<td>0.00 to 4.00</td>
<td></td>
</tr>
<tr>
<td>Cavalletti</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>2</td>
<td>3.00</td>
<td>2.00</td>
<td>0.00 to 4.00</td>
<td>0.910ab</td>
</tr>
<tr>
<td>B</td>
<td>3</td>
<td>4.00</td>
<td>2.00</td>
<td>0.00 to 4.00</td>
<td></td>
</tr>
<tr>
<td>Figure 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>5</td>
<td>3.00</td>
<td>2.00</td>
<td>0.00 to 4.00</td>
<td>0.840ab</td>
</tr>
<tr>
<td>B</td>
<td>2</td>
<td>3.00</td>
<td>2.00</td>
<td>0.00 to 4.00</td>
<td></td>
</tr>
<tr>
<td>Down to stand</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>14</td>
<td>2.00</td>
<td>0.00</td>
<td>0.00 to 4.00</td>
<td>0.880ab</td>
</tr>
<tr>
<td>B</td>
<td>15</td>
<td>2.00</td>
<td>0.00</td>
<td>0.00 to 4.00</td>
<td></td>
</tr>
</tbody>
</table>

Values are reported as median, IQR, and range.
Bl = Number of subjects classified as “behavioral interference.”
abShrout-Fleiss intraclass correlation coefficient. abcCohen weighted κ statistic.

**Table 2**—Results of bivariate associations between the mean functional test results (timed up and go [TUG] measured in seconds [sTUG] and geriatric functional score [GFS], averaged between examiners) and the demographics and comorbidities of aging for the dogs described (Table 1).

<table>
<thead>
<tr>
<th>Variable</th>
<th>r²</th>
<th>Point estimate</th>
<th>95% CLL</th>
<th>95% CLU</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>sTUG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Body condition score (BCS) (Ln)a</td>
<td>0.005</td>
<td>0.032</td>
<td>-0.073</td>
<td>0.138</td>
<td>.544</td>
</tr>
<tr>
<td>Muscle condition score (MCS) (Ln)a</td>
<td>0.169</td>
<td>-0.228</td>
<td>-0.339</td>
<td>-0.117</td>
<td>&lt; .001a</td>
</tr>
<tr>
<td>Age (Ln)a</td>
<td>0.114</td>
<td>0.074</td>
<td>0.029</td>
<td>0.120</td>
<td>.002a</td>
</tr>
<tr>
<td>Breed b,c</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.706</td>
</tr>
<tr>
<td>Sex d,e</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.411</td>
</tr>
<tr>
<td>Weight f</td>
<td>-</td>
<td>0.193</td>
<td>-0.017</td>
<td>0.385</td>
<td>.070</td>
</tr>
<tr>
<td>GFS</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>BCS’</td>
<td>-</td>
<td>0.261</td>
<td>0.025</td>
<td>0.465</td>
<td>.029a</td>
</tr>
<tr>
<td>MCS’</td>
<td>-</td>
<td>0.539</td>
<td>0.344</td>
<td>0.685</td>
<td>&lt; .001a</td>
</tr>
<tr>
<td>Age’</td>
<td>-</td>
<td>-0.459</td>
<td>-0.624</td>
<td>-0.248</td>
<td>&lt; .001a</td>
</tr>
<tr>
<td>Breed’</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.667</td>
</tr>
<tr>
<td>Sex e</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.176</td>
</tr>
<tr>
<td>Weight’</td>
<td>-</td>
<td>0.063</td>
<td>-0.293</td>
<td>0.175</td>
<td>.605</td>
</tr>
</tbody>
</table>

CLL = Lower quality of life limit. CLU = Upper quality of life limit. Ln = Associated variable correlated to the natural logarithm of TUG time.
- = Not applicable.
aLinear regression. bLimited to breeds with n > 4. cKruskal-Wallis test. dLimited to categories with n > 2. eWilcoxon rank sum.
fSpearman rank correlation. gCorrelation demonstrates significance.

category, score increased by 1.73 [95% CI, 0.94 to 2.52], *P* < .0001), and weight (for each additional kg, score decreased by 0.12 [95% CI, –0.19 to –0.06], *P* = .016).

**Associations of sTUG, figure 8s, and GFS with CSOMs**

The sTUG, figure 8s, and GFS were all correlated with CSOMs when taking into account age, MCS, and weight (Table 3). The sTUG showed significant (*P* < .05) moderate to strong correlations with all CSOMs (LOAD, psCBPI, piCBPI, and qICBPI). The GFS showed significant (*P* < .05) moderate to strong negative correlations with LOAD, piCBPI, and qICBPI; however, no significant correlation was noted for psCBPI. Finally, figure 8s demonstrated moderate to strong correlations with all CSOMs, similar to sTUG.
Table 3—Spearman partial correlation between the mean functional test results (timed up and go [TUG]) measured in seconds [sTUG] and geriatric functional score [GFS] and figure 8s, averaged between examiners) and the client-specific outcome measures (Canine Brief Pain Inventory pain severity [psCBPI], CBPI pain interference [piCBPI], quality of life CBPI [qlCBPI], and Liverpool Osteoarthritis in Dogs [LOAD]), controlling for age, muscle condition score, and body weight for the dogs described (Table 1).

<table>
<thead>
<tr>
<th>Item/CSOM</th>
<th>Partial ρ</th>
<th>95% CLL</th>
<th>95% CLU</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>sTUG (n = 84)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>psCBPI</td>
<td>0.284</td>
<td>0.064</td>
<td>0.475</td>
<td>.011a</td>
</tr>
<tr>
<td>piCBPI</td>
<td>0.301</td>
<td>0.082</td>
<td>0.489</td>
<td>.007a</td>
</tr>
<tr>
<td>qlCBPI</td>
<td>0.337</td>
<td>0.122</td>
<td>0.519</td>
<td>.002a</td>
</tr>
<tr>
<td>LOAD</td>
<td>0.353</td>
<td>0.139</td>
<td>0.532</td>
<td>.001a</td>
</tr>
<tr>
<td>Figure 8 (n = 84)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>psCBPI</td>
<td>-0.247</td>
<td>-0.440</td>
<td>-0.029</td>
<td>.026a</td>
</tr>
<tr>
<td>piCBPI</td>
<td>-0.283</td>
<td>-0.471</td>
<td>-0.067</td>
<td>.010a</td>
</tr>
<tr>
<td>qlCBPI</td>
<td>-0.443</td>
<td>-0.602</td>
<td>-0.247</td>
<td>&lt;.0001a</td>
</tr>
<tr>
<td>LOAD</td>
<td>-0.423</td>
<td>-0.585</td>
<td>-0.223</td>
<td>&lt;.0001a</td>
</tr>
<tr>
<td>GFS (n = 70)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>psCBPI</td>
<td>-0.190</td>
<td>-0.410</td>
<td>0.054</td>
<td>.124</td>
</tr>
<tr>
<td>piCBPI</td>
<td>-0.304</td>
<td>-0.506</td>
<td>-0.067</td>
<td>.012a</td>
</tr>
<tr>
<td>qlCBPI</td>
<td>-0.475</td>
<td>-0.640</td>
<td>-0.262</td>
<td>&lt;.001a</td>
</tr>
<tr>
<td>LOAD</td>
<td>-0.486</td>
<td>-0.648</td>
<td>-0.275</td>
<td>&lt;.001a</td>
</tr>
</tbody>
</table>

Data are presented as partial ρ with 95% lower quality of life limit (CLL) and 95% upper quality of life limit (CLU). CSOM = Client-specific outcome measurement. *Significant correlation at the P < .05 level.

Individual GFS subtest correlations with CSOMs

To account for the influence of each individual component of the summed GFS, each individual functional test was correlated against piCBPI, psCBPI, qlCBPI, and LOAD while taking the covariates identified in multiple regression and all other GFS components into account (Table 4). Figure 8s retained significant (P < .05) correlations to all CSOMs. TUG scoring retained similar significant correlations (P < .05) with LOAD and qlCBPI. Neither cavalletti nor down-to-standing test demonstrated correlations to any CSOMs.

Receiver operating characteristic curve analysis to define high- versus low-functioning dog populations

There were 21 dogs (23.6%) who qualified as highly functional. When comparing the high-functioning group to the low-functioning group, age (median for high function = 10.0 years, median for low function 12.0 years) differed significantly (Wilcoxon rank sum P = .034) but weight (median for high function = 25.1 kg, median for low function = 27.6 kg) did not (Wilcoxon rank sum, P = .10). There were no significant differences in sex (Fisher exact, P = .342), MCS (Fisher exact, P = .094), or BCS (Wilcoxon rank sum, P = .653) between groups. Based on this, the average total geriatric function score was compared between the high and low function groups while controlling for age via analysis of covariance, which found both age (P = .005) and

Table 4—Results for Spearman partial correlation analysis to identify associations between individual geriatric functional score (GFS) subtests (results as described [Table 1], averaged between examiners) and client-specific outcome measures (Canine Brief Pain Inventory pain severity [psCBPI], CBPI pain interference [piCBPI], quality of life CBPI [qlCBPI], and Liverpool Osteoarthritis in Dogs [LOAD]) for 70 of the dogs described (Table 1), with other subtests as covariates.

<table>
<thead>
<tr>
<th>Item/average subtest score</th>
<th>Partial ρ</th>
<th>95% CLL</th>
<th>95% CLU</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBPI</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TUG</td>
<td>-0.176</td>
<td>-0.398</td>
<td>0.068</td>
<td>.155</td>
</tr>
<tr>
<td>Cavalletti</td>
<td>0.117</td>
<td>-0.128</td>
<td>0.346</td>
<td>.349</td>
</tr>
<tr>
<td>Figure 8</td>
<td>-0.297</td>
<td>-0.495</td>
<td>-0.053</td>
<td>.016a</td>
</tr>
<tr>
<td>Down to stand</td>
<td>0.192</td>
<td>-0.052</td>
<td>0.412</td>
<td>.120</td>
</tr>
<tr>
<td>piCBPI</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TUG</td>
<td>-0.219</td>
<td>-0.435</td>
<td>0.024</td>
<td>.075</td>
</tr>
<tr>
<td>Cavalletti</td>
<td>0.042</td>
<td>-0.201</td>
<td>0.279</td>
<td>.737</td>
</tr>
<tr>
<td>Figure 8</td>
<td>-0.241</td>
<td>-0.453</td>
<td>0.001</td>
<td>.049a</td>
</tr>
<tr>
<td>Down to stand</td>
<td>0.113</td>
<td>-0.132</td>
<td>0.343</td>
<td>.363</td>
</tr>
<tr>
<td>qlCBPI</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TUG</td>
<td>-0.315</td>
<td>-0.514</td>
<td>-0.079</td>
<td>.009a</td>
</tr>
<tr>
<td>Cavalletti</td>
<td>0.058</td>
<td>-0.185</td>
<td>0.294</td>
<td>.642</td>
</tr>
<tr>
<td>Figure 8</td>
<td>-0.245</td>
<td>-0.457</td>
<td>-0.003</td>
<td>.045a</td>
</tr>
<tr>
<td>Down to stand</td>
<td>-0.076</td>
<td>-0.310</td>
<td>0.168</td>
<td>.542</td>
</tr>
<tr>
<td>LOAD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TUG</td>
<td>-0.352</td>
<td>-0.544</td>
<td>-0.119</td>
<td>.003a</td>
</tr>
<tr>
<td>Cavalletti</td>
<td>0.069</td>
<td>-0.175</td>
<td>0.304</td>
<td>.581</td>
</tr>
<tr>
<td>Figure 8</td>
<td>-0.297</td>
<td>-0.500</td>
<td>-0.059</td>
<td>.014a</td>
</tr>
<tr>
<td>Down to stand</td>
<td>0.068</td>
<td>-0.175</td>
<td>0.303</td>
<td>.584</td>
</tr>
</tbody>
</table>

Data are presented as partial ρ with 95% lower quality of life limit (CLL) and 95% upper quality of life limit (CLU). *Significant correlation at the P < .05 level.

function group (P = .002) to be significant predictors of sTUG. Receiver operating characteristic curve analysis helped discern less functional dogs from highly functional dogs with a sTUG cutoff of 3.83 seconds (area under the curve, 0.90; sensitivity, 64.3%; specificity, 95.1%).

Discussion

The initial aim of the study was to determine interrater agreement. The GFS demonstrated strong interrater agreement within each subtest and for the total GFS. The sTUG was examined separately as an objective measure, showing the strongest interobserver agreement among all tests. Repeatability testing (intranater agreement and test-retest) was not conducted; however, clinical signs or a patient’s functional capacity often show variability over time depending on a multitude of factors including disease progression, medications, physical fitness status, and other comorbidities. Furthermore, compromised patients often physically fatigue and consequentially score worse upon immediate repeat testing.20 Within
our sequential test framework, isolating a single test’s repeatability score would have proven impossible. On the other hand, the exceptionally high interrater agreement among all testing would lead to expectations of strong intrarater agreements. The objective measurement of time (seconds) would be subject to negligible, if any, variability amongst raters, making sTUG the most reliable test.

A second major goal was to establish criterion validity by correlating GFS and sTUG to CBPI (piCBPI, psCBPI, and qICBPI) and LOAD scoring. Dogs with higher sTUG were found to have more severe pain, worse functional scoring, and reduced QOL (Table 3). The GFS showed similarly trending significant strong correlations to all CSOM categories except for psCBPI. However, when assessing each individual GFS test against the CSOMs, only the TUG and figure 8s remained significantly correlated (Table 4). Therefore, TUG and figure 8 scoring, rather than the entirety of all 4 tasks, appears to be driving significant GFS correlations to CSOMs, prompting further examination of the relationship of figure 8s alone against CSOMs.

Despite being correlated, like sTUG, to all CSOMs (Table 3), figure 8s were the third test performed in sequence, and their results may therefore be influenced by the prior tasks. Figure 8s also had a 5.6% (5/89 dogs) behavioral interference compared to 0% for the TUG, making TUG easier to perform in practice (Table 1). Finally, given the subjectivity in scoring figure 8s, there may be more unwanted variability by veterinarians who are not specialists in mobility disorders. Despite contributing greatly to the GFS, figure 8s need to independently be researched further before making conclusions regarding its value as a test for geriatric function. sTUG logically becomes the preferred geriatric mobility test as it is comprised of objective data and underlies all significant correlations regarding criterion validation. For the remainder of this article, sTUG will be the focus.

No standard exists to establish criterion validity for canine geriatric functional testing, but a strong argument can be made that CBPI and LOAD CSOMs are the most appropriate tools currently available for comparison. These CSOMs have been previously validated for assessing dogs with OA and their response to treatment.15,14,21 Discordance exists between studies regarding the prevalence of canine OA (20% to 90% of the dog population). Given that about 60% of younger dogs exhibited OA in a recent prospective study,22 it is reasonable to assume the OA prevalence in geriatric dogs would be at least 60% and likely much higher.23,24 A high OA prevalence in older dogs lends more credence to our selection of CBPI and LOAD CSOMs as a general reflection of geriatric dog mobility.

Although originally designed to assess OA-affected dogs, LOAD and CBPI questions regarding comfort, QOL, and mobility are not necessarily specific to OA, potentially imparting broader applications. For example, CBPI and LOAD have been previously implemented to assess response to treatment for canine spinal disease.25 Although pain and immobility greatly overlap, dogs may exhibit little pain but have high levels of immobility (degenerative myelopathy, congestive heart failure, etc) or exhibit higher levels of discomfort but still be perceived as mobile. Therefore, selecting CSOMs that capture both mobility and discomfort, such as CBPI and LOAD, are ideal for our study. Interestingly, when comparing sTUG against CSOMs, partial ρ correlations were strongest for LOAD (Table 3). This finding may reflect LOAD’s ability to better capture a portion of our population with nonpainful mobility issues because LOAD is less pain specific than CBPI.

Quality of life is challenging to assess in companion animals due to differences in opinion amongst caregivers and veterinarians as well as a lack of consensus regarding the best means of measurement.26 Despite differences in our study design, prior research has similarly demonstrated correlations between reduced mobility and worsening QOL.3 adding justification for implementing mobility scoring to support end-of-life decisions.

Objective measures are often preferred for test validation. In contrast to CSOMs, common objective outcome measures for canine lameness and mobility have significant limitations for capturing various aspects of patient mobility, including task-dependent movements (posture transitions, turns, obstacle navigation, and play), balance, spatial awareness, and endurance.27 For example, kinetic gait analysis may fail to differentiate dogs affected by OA from healthy populations28 and was designed to test load-bearing limb lameness as opposed to general mobility.14 On the other hand, weak correlations have been noted when comparing ground reaction forces to LOAD and CBPI instruments.14 Accelerometer or pedometer data may reflect gross activity; however, they are not task specific and may be restricted by factors unrelated to a dog’s willingness to move12,29 such as weather, owner work schedule, and an owner’s capacity or desire to exercise a dog.

A secondary objective was to identify highly functional from less functional dog populations using GFS and sTUG results. Ideally, a population of dogs free of mobility-related disease would act as a control group. However, functional tests are not designed to be diagnostic tests, and diagnosis alone is a poor reflection of function. For example, a highly functional dog may still have mild or subclinical disease affecting mobility.5,30 We, therefore, defined a highly functioning dog population as having a total sum score of 0 or 1 on 3 selected CSOM questions most reflective of general mobility. Receiver operating characteristic curve analysis of these criteria indicates that dogs completing the TUG in under 3.83 seconds can be considered highly functional.

The comorbidities of aging (obesity and sarcopenia) are common to geriatric populations, often occurring together or with other age-related diseases, and have been correlated to poor prognosis, increased age, and reduced mobility in people.4,31,32 The BCS and MCS were used to represent these comorbidities respectively in our study. The sTUG was significantly associated with MCS in simple linear
prior epidemiological studies. On the other hand, measurements of body condition when compared to have been most notable in our lower than average of typical geriatric populations. This difference may and university hospital may not be representative Recruitment of dogs from the veterinary community and patient mobility (sTUG) may have complex rela- it appears that increased age, reduced muscle scoring, and patient mobility (sTUG) may have complex relationships that warrant further investigation.

This study possesses some limitations. Recruitment of dogs from the veterinary community and university hospital may not be representative of typical geriatric populations. This difference may have been most notable in our lower than average measurements of body condition when compared to prior epidemiological studies. On the other hand, our study population was evenly divided between dogs with and without a diagnostic history of neurological or orthopedic disease, providing grounds for exploring a diverse spectrum of patient function. Our study also lacked a true control population of healthy dogs identified by physical examination and various diagnostic tests. We accounted for a lack of a control by defining a population of highly functioning dogs based on select stringent CSOM scoring. Although we attempted to minimize the influence of behavior (including motivation) through our methodology, it is impossible to discern what role, if any, behavior played in some dogs and their scoring. Finally, sTUG provides an objective measure that should be minimally impacted by the observer; however, we recommend pursuing repeatability testing in the future.

In summary, all individual functional tests, total GFS, and sTUG showed extremely strong inter-rater agreement, supporting our first hypothesis. Increases in sTUG correlated strongly with decreases in comfort, function, and QOL as measured by all CBPI and LOAD scoring, supporting our second hypothesis. Similar functional relationships were noted with GFS; however, individual analysis of each GFS subtest demonstrated that only TUG and figure 8s were significantly contributing, rendering GFS unnecessary. Despite figure 8s correlating with all CSOMs similarly to sTUG, their results may have been influenced by the prior sequential tasks, there were behavioral interferences, and their scoring is subjective. For these reasons, sTUG is proposed as the best functional test and can be implemented easily in most clinical practice settings. Furthermore, sTUG was able to define highly functioning from lower functioning dogs, as predicted, with a cutoff of 3.83 seconds. Lower muscle condition and increasing age correlated to decreased function as measured by sTUG; however, BCS did not. These data partially support our fourth hypothesis regarding the complex associations of mobility and comorbidities of aging and lend greater support for the use of sTUG to assess geriatric mobility.

Before this study, canine functional testing was limited, albeit advocated for and applied by rehabilitation practitioners through various unstandardized methods. Timed up and go appears to be a practical and reliable test of physical mobility in a relatively diverse group of canine geriatrics with the capability of distinguishing highly functional from less functional populations. Future investigations should target repeatability testing and explore clinical applications such as prognosis and outcome monitoring.

Acknowledgments
None reported.

Disclosures
The authors have nothing to disclose. No AI-assisted technologies were used in the generation of this manuscript.

Funding
Funding was provided by Nestle Purina PetCarein support of Cornell’s “Canine Healthy Aging and Mobility Program” (CHAMP).

References
8. Braun T, Thiel C, Peter RS, et al. Association of clini- cal outcome assessments of mobility capacity and


**Supplementary Materials**

Supplementary materials are posted online at the journal website: avmajournals.avma.org.