Balance Impairment Not Predictive of Falls in Geriatric Rehabilitation Wards

Terry Haines,1,2 Suzanne S. Kuys,2,3 Greg Morrison,2 Jane Clarke,4 and Paul Bew5

1Division of Physiotherapy, School of Health and Rehabilitation Sciences, University of Queensland, Brisbane, Australia.  
2Physiotherapy Department, Princess Alexandra Hospital, Brisbane, Australia.  
3School of Physiotherapy and Exercise Science, Griffith University, Gold Coast, Australia.  
4St. Andrews War Memorial Hospital, Brisbane, Australia.  
5The Prince Charles Hospital, Brisbane, Australia.

Background. Falls are common among hospital inpatients, particularly in rehabilitation wards. Standing balance impairment is widely held to be a contributing factor to falls, is a component of several falls risk screening tools, and has motivated the development of balance retraining programs for the reduction of in-hospital falls. Little rigorous investigation of the link between standing balance impairment and in-hospital falls has been undertaken.

Methods. We identified optimal cut-off points of four commonly used balance measures (functional reach, Timed Up and Go, step test, and timed static stance) in a prospective multicenter cohort study. Admission data (n = 1373) were clustered and matched by center then randomly allocated to development and validation data sets.

Results. Optimal cut-off points for each test were identified from the development data set. The predictive accuracy of all four balance tests was poor when the optimal cut-off was applied to the validation data set (Youden Index scores ranged between 0.02 and 0.15).

Conclusions. These findings do not support an association between admission standing balance and falls in a geriatric rehabilitation setting. This result has implications for content of falls risk screening tools and interventions to prevent falls in a geriatric rehabilitation population.

Key Words: Falls—Predicting falls—Predictive value of tests—Rehabilitation.
these measures that best predict fallers and fall ERs in geriatric rehabilitation wards.

METHODS

Design
A multicenter prospective cohort study was conducted with the cohort randomly divided into development and validation data sets for investigation of predictive accuracy.

Participants and Setting
A total of 1373 persons from 17 inpatient geriatric and rehabilitation wards in Australia were recruited to this study. For inclusion, persons had to be admitted for rehabilitation and be referred for physiotherapy. Persons were excluded if they had paraplegia, tetraplegia, or lower limb amputation. Hospital and university ethics committees provided approval for this study.

Measures
The primary outcome variable was number of patient in-hospital falls as recorded in patient histories and through hospital incident reporting systems. A fall was defined as "any event where a patient unexpectedly comes to rest on the ground, the floor, or another lower level" (30). The classification variables examined were four routinely used measures of standing balance: functional reach (31–33), Timed Up and Go (25), step test (34), and timed static stance with feet together and eyes closed, a component of the clinical test of sensory integration for balance (35,36). These measures were selected by an expert working party of clinicians from participating centers prior to project commencement based on their clinical utility; collectively they sampled a range of domains (e.g., a reaching task, a stepping task, a walking task, a sensory challenge) that make up the standing balance construct. Patient demographic data and length of stay in inpatient geriatric rehabilitation were also collected.

Procedure
Detailed instructions for each balance test based on original sources were provided to each center in a standardized study protocol manual. Participants performed only one trial of the functional reach, as this has demonstrated high reliability ($r > 0.9$) with the mean of three trials (33). A "walk through" was provided (where possible) to familiarize participants with the Timed Up and Go before completion of the actual test (25). Participants performed three trials of the static stance test up to a maximum of 30 seconds each, and the duration of the three trials was added to generate the final score. For this test, eyes were closed, feet placed together, and arms folded across the chest; if the patient successfully lasted for 30 seconds in one trial, all remaining subsequent trials were awarded 30 seconds as per the original procedure. The step test was performed with each lower limb, a 7.5 cm step, and 15 seconds for each test; the results were averaged to produce one score for this test. The testing was normally performed the day following admission to the geriatric rehabilitation ward, although there may have been as much as 72 hours following admission in the case of patients admitted over the weekend. Data were collected at participating centers for between 1 and 12 months from May 2005.

Analysis
Participants unable to complete the Timed Up and Go were classified as "unable" and ascribed a time of 999 seconds for this test in analyses of predictive accuracy. A score of zero was ascribed for other balance tests in which the participant was unable to attain the minimum score in the test.

Using a split sampling technique, centers were matched based on the number of participants at each center, then randomly allocated to the development or validation data set using a computer-generated random number generator. This was necessary as optimal cut-off points for the balance tests being examined had not been established for this population, and previous research has identified that elements of study design (such as identifying optimal cut-off points and calculating predictive accuracy from the same data set) increase the likelihood of producing "over optimistic" estimates of predictive accuracy (15,37). Comparisons between data sets in patient demographics were made using independent groups $t$ tests (continuous data), Chi-square tests for dichotomous data, and negative binomial regression for falls rates (Table 1).

The development set was used to identify optimal cut-off points for each of the classification (balance test) variables. Predictive accuracy of the balance tests in classifying patients as fallers or nonfallers (the incidence rate perspective) was assessed using sensitivity, specificity, and Youden Index (38). Sensitivity equals the proportion of all fallers that were correctly classified as high risk. Specificity equals the proportion of all nonfallers correctly classified as low risk. Youden Index equals sensitivity plus specificity minus one. The optimal cut-off point for the purposes of this study was defined as that which maximized the Youden Index. Sensitivity and specificity data were presented on receiver operating characteristic (ROC) curves.

Limitations of assessing the predictive accuracy of classification instruments using in-hospital falls as an outcome have previously been identified (39). In this study, conventional sensitivity, specificity, and Youden Index calculations are insensitive to multiple falls by individual patients and variation in length of hospital stay. Therefore, we also selected cut-off points based on the ER (falls per time) perspective by maximizing the Youden Index$^E_R$ where: Youden Index$^E_R = \text{sensitivity}^E_R + \text{specificity}^E_R - 1$, sensitivity$^E_R = \frac{\text{the proportion of all falls where the faller was classified as being at high risk at the time of the fall and }}{\text{specificity}^E_R = \frac{\text{the proportion of all patient-time that patients were classified as being low risk (39). Sensitivity}^E_R \text{ and specificity}^E_R \text{ data were also presented on ROC curves.}}}$

Optimal cut-off points identified from the development data set were applied to the validation data set. This data set was independent from the development data set both in terms of patients and research locations. Sensitivity, specificity, Youden Index, sensitivity$^E_R$, specificity$^E_R$, and Youden Index$^E_R$ statistics were calculated using their
RESULTS

Demographic details of the development and validation data sets are presented (Table 1). Development and validation data sets differed in several areas, including mix of patient primary diagnosis, patient cognitive status (validation poorer), admission functional status (validation poorer), premorbid ambulatory assistance (validation required more), length of stay (validation longer), standing balance (Timed Up and Go; validation poorer), and proportion of patients who became fallers (development set more fallers).

ROC curves displaying the range of sensitivity, specificity, sensitivityER, and specificityER outcomes for possible cut-off points for each balance measure are presented (Figure 1). The ROC curves for each measure were similar, although those displaying sensitivityER and specificityER outcomes appeared closer to the reference line (indicating results attainable through random allocation to the high risk group).

The optimal cut-off points (which maximized Youden Index) for each test identified from the development data set are displayed (Table 2). Also displayed are the accuracy statistics for these cut-off points when applied to the development and validation data sets. All cut-offs selected produced Youden Indices and Youden IndicesER significantly more accurate than chance when applied to the development data set. However, the Youden Index for the Timed Up and Go was not significantly different from that of chance when the same cut-offs were applied to the validation data set, unlike those for the step test, functional reach, and static standing tests (although point estimates did decrease relative to the development data set). None of the Youden IndicesER were significantly different from chance in the validation data set, with the point estimates from all four tests providing results less accurate than chance.

DISCUSSION

The causal association between impaired standing balance and falls underpins a number of falls prevention strategies both in hospitals and the community. This study is the first to simultaneously examine the predictive accuracy of multiple balance assessments in the inpatient rehabilitation setting. It
is also the first to examine the accuracy of such tests in predicting both fallers (incidence rate perspective) and falls rates (ER perspective). The findings arising from this research have provided strong evidence that balance tests administered on patient admission have little association with falls in geriatric rehabilitation and are of little value in identifying individuals likely to fall in this setting.

The accuracy of balance tests arising from the development data set indicated that the balance tests may have been of moderate utility in identifying individuals at high risk of experiencing in-hospital falls. However, the accuracy of these tests arising from the validation data set was poorer, especially when considering the point estimates of the Youden Indices\textsuperscript{ER}. Previous meta-analyses have identified that the reported accuracy of a screening tool can be over-optimistic if the optimal cut-off is not identified a priori (15,37). In the present investigation, the optimal cut-off points were identified from the development data set, and thus were not identified a priori in that evaluation, whereas they were for the evaluation using the validation data set. Hence it is the results from the validation data set which are more likely to be generalizable to what could be expected in clinical practice.

The accuracy of the four balance tests was comparable to

![Figure 1. Receiver operating characteristic curves of standing balance tests for predicting fallers (a, sensitivity and 1 – specificity) and falls event rates (b, sensitivity\textsuperscript{ER} and 1 – specificity\textsuperscript{ER}) from the development data set.](https://academic.oup.com/biomedgerontology/article-abstract/63/5/523/590643/fig1)

**Table 2. Predictive Accuracy Statistics (Sensitivity, Specificity, Youden Index, Sensitivity\textsuperscript{ER}, Specificity\textsuperscript{ER}, Youden Index\textsuperscript{ER}) With 95% CIs for Balance Measures in Development and Validation Data Sets**

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Step Test (Average)*</th>
<th>Timed Up and Go\textsuperscript{i}</th>
<th>Functional Reach\textsuperscript{j}</th>
<th>Static Standing\textsuperscript{g}</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Development data set</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.82 (0.75–0.89)</td>
<td>0.76 (0.68–0.83)</td>
<td>0.60 (0.52–0.69)</td>
<td>0.62 (0.53–0.70)</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.40 (0.35–0.44)</td>
<td>0.44 (0.40–0.49)</td>
<td>0.60 (0.56–0.65)</td>
<td>0.57 (0.53–0.62)</td>
</tr>
<tr>
<td>Youden Index</td>
<td>0.22 (0.14–0.30)</td>
<td>0.20 (0.11–0.29)</td>
<td>0.21 (0.11–0.31)</td>
<td>0.19 (0.10–0.29)</td>
</tr>
<tr>
<td>Sensitivity\textsuperscript{ER}</td>
<td>0.93 (0.87–0.97)</td>
<td>0.87 (0.81–0.93)</td>
<td>0.81 (0.72–0.88)</td>
<td>0.83 (0.75–0.89)</td>
</tr>
<tr>
<td>Specificity\textsuperscript{ER}</td>
<td>0.16 (0.13–0.19)</td>
<td>0.20 (0.17–0.24)</td>
<td>0.29 (0.24–0.33)</td>
<td>0.30 (0.26–0.35)</td>
</tr>
<tr>
<td>Youden Index\textsuperscript{ER}</td>
<td>0.09 (0.03–0.13)</td>
<td>0.08 (0.02–0.14)</td>
<td>0.09 (0.01–0.16)</td>
<td>0.13 (0.06–0.19)</td>
</tr>
<tr>
<td><strong>Validation data set</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.87 (0.79–0.93)</td>
<td>0.80 (0.71–0.89)</td>
<td>0.70 (0.61–0.79)</td>
<td>0.67 (0.57–0.77)</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.24 (0.20–0.28)</td>
<td>0.22 (0.19–0.26)</td>
<td>0.43 (0.38–0.47)</td>
<td>0.48 (0.44–0.53)</td>
</tr>
<tr>
<td>Youden Index</td>
<td>0.11 (0.02–0.18)</td>
<td>0.02 (–0.07 to 0.11)</td>
<td>0.13 (0.03–0.24)</td>
<td>0.15 (0.04–0.25)</td>
</tr>
<tr>
<td>Sensitivity\textsuperscript{ER}</td>
<td>0.81 (0.59–0.95)</td>
<td>0.74 (0.53–0.89)</td>
<td>0.63 (0.45–0.81)</td>
<td>0.57 (0.40–0.75)</td>
</tr>
<tr>
<td>Specificity\textsuperscript{ER}</td>
<td>0.16 (0.13–0.20)</td>
<td>0.16 (0.13–0.20)</td>
<td>0.29 (0.25–0.34)</td>
<td>0.38 (0.33–0.43)</td>
</tr>
<tr>
<td>Youden Index\textsuperscript{ER}</td>
<td>–0.03 (–0.23 to 0.10)</td>
<td>–0.09 (–0.28 to 0.05)</td>
<td>–0.08 (–0.25 to 0.09)</td>
<td>–0.05 (–0.20 to 0.12)</td>
</tr>
</tbody>
</table>

**Notes:** Sensitivity, specificity, and Youden Index statistics were calculated using optimal cut-off point from fallers incidence perspective. Sensitivity\textsuperscript{ER}, specificity\textsuperscript{ER}, and Youden Index\textsuperscript{ER} were calculated using optimal cut-off point from falls event rate perspective.

\*Optimal cut-off points for the step test were <2 steps (high risk) for incidence rate data and <5 steps (high risk) for event rate data.

\textsuperscript{i}Optimal cut-off points for the timed up and go were >30 seconds or unable (high risk) for incidence rate data and >25 seconds (high risk) or unable for event rate data.

\textsuperscript{j}Optimal cut-off points for the functional reach were <4 cm (high risk) for incidence rate data and <14 cm (high risk) for event rate data.

\textsuperscript{g}Optimal cut-off points for the static stance test were <8 seconds (high risk) for incidence rate data and <35 seconds (high risk) for event rate data.

CI = confidence interval.
results arising from previous investigations of the Berg Balance Scale (28), Get Up and Go test (a version of the Timed Up and Go where patients were classified as being safe, unsafe, or unable) (40), Timed Up and Go test (26,41), and the functional reach tests (29) in the hospital setting. Three of these investigations were not conducted on comparable participant groups, with one conducted among patients with acquired brain injury (28), and two among patients on acute geriatric care units (26,41). All of these investigations were conducted at individual facilities and did not identify an “optimal” cut-off point for each test prior to application. Hence it can be anticipated that, even when modestly favorable results were identified, these were likely to have been overly optimistic. Accordingly, it was necessary to proceed in the present study with a large, multicenter, prospective evaluation using both a development and validation data set. Given these recent results, it is now evident that admission balance tests have little utility for predicting in-hospital falls in isolation. Rather than screening for overall falls risk, previous authors have suggested identifying reversible risk factors in every patient and addressing these (17). The tests investigated in this research may still be useful in the assessment and development of treatment programs for balance and gait disorders more generally.

The perception of a strong association between in-hospital falls and impaired standing balance must now be challenged and raises the question as to why there should be a difference in this relationship between the hospital and community settings. We offer three likely explanations. First, to generalize, in the community there are two key factors in a majority of falls; the demands of the activity being performed (including environmental) and the capacity of the individual (of which standing balance is a key component) to meet the demands of that activity. In hospitals there is a third factor, the role of nursing, medical, and therapy staff, who can compensate for the poor capacity of a patient (attributable to impaired standing balance) to perform a functional task by providing advice and physical assistance. The presence of this third factor reduces the impact that impaired standing balance may have on the risk of falls while performing functional tasks. Second, a period of hospitalization is often characterized by rapid change in standing balance, particularly during inpatient rehabilitation where patients commonly receive daily sessions of physical therapy. It is likely that admission standing balance test results of fallers and nonfallers alike are not reflective of results that would have been gleaned had the test been conducted at the time of the fall. Third, it is possible that hospital staff provided falls prevention strategies to a greater extent to people who they believe to be at high risk of falling, a factor related to the first explanation offered, which may confound investigations of falls risk screening accuracy (42,43). It is possible and even likely that hospital staff limit the activity of a person identified as a high falls risk. By providing fewer opportunities for people to be transferring or mobilizing unsupervised, hospital staff will be reducing the falls risk and also reducing the activity levels of hospital inpatients. It should also be acknowledged that the rehabilitation hospital environment differs from the community setting in that patients are recovering from significant illness and are being encouraged to mobilize in unfamiliar surroundings.

It should be noted that the majority of patients in this study had very poor scores on balance tests. This indicates that this population has a high risk of falls, as confirmed by the high falls rate observed in the present study. However, the ability of these tests to discriminate between fallers and nonfallers within this high risk demographic may be reduced to the consistently low scores.

There were two primary limitations of this study. First, the balance tests selected could be criticized as not adequately measuring standing balance; this is a reason why no significant relationship with in-hospital falls was identified. However, these tests were selected by an expert clinician working party partly because they sampled a range of domains that make up the standing balance construct. Additionally, previous research has identified that some of the measures correlated highly with other tests not included (e.g., four quadrant step test) (44), thus the results were likely to have been similar had they been included. Second, this study sampled only inpatients on rehabilitation wards. Although falls occur very frequently on these wards relative to other wards (3–5,45), there are other ward types that also have high rates of falls, for which identification of factors that predict in-hospital falls would be important as these factors may vary considerably from those in the rehabilitation setting.

Several issues for future research have been raised by this work. There is need to establish whether impaired standing balance as measured closer to the time of an in-hospital fall is associated with falling. For example, measuring standing balance at multiple time points during a hospital admission and using the most recent measures of balance as the classification variable would largely overcome the problem inherent in the present study that admission balance measures may not have been reflective of the balance ability of the patient at the time of the fall. If such an investigation similarly demonstrates little association between falls and standing balance, research should be directed toward developing a falls prevention paradigm specific to the hospital setting on which falls prevention activities may be based rather than include strategies targeting balance impairment. It is also possible that other measures of balance may be predictive of falls as some balance tests have been shown to be better predictors than others in other populations (46).

Conclusion

In geriatric patients admitted for inpatient rehabilitation, standing balance as measured at admission has limited use to identify those people at high risk of falling.

Acknowledgments

This work was supported by The University of Queensland New Staff Research Fund.

We acknowledge the support of the staff and patients of all participant centers (Mater Private Hospital, Greenslopes Private Hospital, Peter James Centre, Angliss Health Service, Townsville Hospital, Cairns Base, QEII Hospital, Princess Alexandra Hospital, Gold Coast Hospital, Prince Charles Hospital, Mt. Olivett Private Hospital, Ipswich Hospital, Redlands Hospital,
Redcliffe Hospital, Toowoomba Hospital, St. Andrew’s War Memorial Hospital, Royal Brisbane and Women’s Hospital) without which this study would not have been feasible. We also thank Steven McPhail for his assistance during this project.

CORRESPONDENCE
Address correspondence to Suzanne Kuyès, B Phty, School of Physiotherapy and Exercise Science, Griffith University, Gold Coast Campus, PMB 50, Gold Coast Mail Centre, QLD 9726, Australia. E-mail: s.kuyes@griffith.edu.au

REFERENCES

Received March 7, 2007
Accepted July 16, 2007
Decision Editor: Luigi Ferrucci, MD, PhD