Review Article

Design-Related Bias in Hospital Fall Risk Screening Tool Predictive Accuracy Evaluations: Systematic Review and Meta-Analysis

Terry P. Haines,[1] Keith Hill,[2] Willeke Walsh,[3] and Richard Osborne[4]

[1]The University of Queensland and Princess Alexandra Hospital, Australia.
[2]National Ageing Research Institute and The University of Melbourne, Australia.
[3]Western General Hospital, Melbourne, Australia.
[4]Arthritis Foundation of Victoria Centre for Rheumatic Disease and
Royal Melbourne Hospital, Australia.

Introduction. Fall risk screening tools are frequently used as a part of falls prevention programs in hospitals. Design-related bias in evaluations of tool predictive accuracy could lead to overoptimistic results, which would then contribute to program failure in practice.

Methods. A systematic review was undertaken. Two blind reviewers assessed the methodology of relevant publications into a four-point classification system adapted from multiple sources. The association between study design classification and reported results was examined using linear regression with clustering based on screening tool and robust variance estimates with point estimates of Youden Index (= sensitivity + specificity − 1) as the dependent variable. Meta-analysis was then performed pooling data from prospective studies.

Results. Thirty-five publications met inclusion criteria, containing 51 evaluations of fall risk screening tools. Twenty evaluations were classified as retrospective validation evaluations, 11 as prospective (temporal) validation evaluations, and 20 as prospective (external) validation evaluations. Retrospective evaluations had significantly higher Youden Indices (point estimate [95% confidence interval]: 0.22 [0.11, 0.33]). Pooled Youden Indices from prospective evaluations demonstrated the STRATIFY, Morse Falls Scale, and nursing staff clinical judgment to have comparable accuracy.

Discussion. Practitioners should exercise caution in comparing validity of fall risk assessment tools where the evaluation has been limited to retrospective classifications of methodology. Heterogeneity between studies indicates that the Morse Falls Scale and STRATIFY may still be useful in particular settings, but that widespread adoption of either is unlikely to generate benefits significantly greater than that of nursing staff clinical judgment.

FALLS among older adults are a serious public health concern. This is particularly so among hospitalized older adults for whom falls have been reported to be the most commonly occurring form of adverse incident (1,2). Many interventions have been trialed for the prevention of in-hospital falls. These interventions have ranged from single intervention strategies of bed alarms (3), patient sitters (4) and fall risk alert bracelets (5) to more complex multiple intervention strategies (6–15). A common feature of many of the interventions trialed has been their selective deployment of all or part of the intervention or intervention program (3,5–7,9–17). Targeted deployment of falls prevention interventions is a reasonable approach to ensure that scarce resources are not applied to persons who do not require them, provided that this can be done accurately.

Mechanisms for selective deployment of falls prevention interventions and intervention programs have varied widely. Commonly these mechanisms have been referred to as fall risk screening or risk assessment tools; however, they have varied in terms of their method of construction, content, and outcomes. The theory that a combination of two or more risk factors may predict falls more accurately than any one factor alone has driven much of the development of fall risk assessment tools. In keeping with the increasing number of publications in this area, a number of reviews of this field have recently been published (18–20). However, conclusions arising from these reviews have varied from support of nurse-completed screening tools being appropriate for the acute hospital setting (though questionable for the extended [subacute] care setting) and that a suitable tool should be able to be identified from existing literature, leaving little need for individual sites to develop their own tool (20), to the suggestion that none of the existing tools can be recommended for wholesale implementation and that it may be better to identify modifiable risk factors which could then be targeted for intervention in all patients (19), to highlighting the need for more research of existing tools with researchers independent of tool designers and in research locations separate from the setting of tool development (18). Thus, hospital clinicians and administrators appear to be

a little closer to being able to answer the questions of whether a fall risk screening tool should be used in their facility and, if so, which one and how.

Previous reviews have highlighted the need for prospective evaluation of fall risk assessment tools in multiple settings by researchers independent of the tool designers (18,19). These calls have been based on valid theoretical principles from the broader field of screening and prognostic tool development (21). Despite these concerns, little differentiation has been made in previous reviews of studies in terms of methodological quality. Indeed, few systematic reviews of observational studies involve quality assessment (22). The potential impact of study design on results presented in the hospital fall risk screening tool field has not previously been investigated. The sentinel work in this area systematically identified and reviewed 184 original articles, evaluating 218 medical diagnostic tests, employed a single descriptor of predictive accuracy for each test, and entered these along with design covariates into regression analyses to evaluate the effects of various design qualities on reported results (23). This work is recognized by the Standards for Reporting of Diagnostic Accuracy (STARD) initiative as the work that identified specific design features that can produce biased, overoptimistic estimates of diagnostic accuracy (24). This systematic review aims to follow this approach in the field of fall risk screening tools for the hospital setting to determine whether study design is associated with reported study outcomes and to present a pooled analysis of present literature based on the presence or absence of a methodology effect.

METHOD

Literature Review

Relevant publications for this review were initially identified through a search of MEDLINE (PubMed and Ovid), EMBASE, and CINAHL databases using the terms ''fall risk'' and ''hospital'' in abstract, title, keyword, and default fields. Other relevant publications were identified through searching reference lists at the end of each journal and review article identified, and from existing files of the authorship team. The review process was initiated in January 2006 and was completed in August 2006 with a final search of the above databases to identify any more recent publications since the commencement of the review process.

Publications included in this review had to meet the following criteria:

- Description of evaluation of a fall risk screening tool in the hospital (acute or subacute) setting
- Accuracy statistics (sensitivity and specificity) needed to be provided or able to be calculated from data presented. Ratio summary statistics (such as odds ratios) are not independent of both relative and absolute sizes of faller and nonfaller groups compared and are thus unsuitable for use in comparisons across studies (25).

Titles and abstracts were reviewed for all publications identified by one reviewer (TH), and those clearly not meeting the inclusion criteria were discarded. All remaining manuscripts were reviewed completely by one reviewer (TH) to determine if they met the inclusion criteria; those not meeting the criteria were excluded. Of the remaining publications, the study design used was rated according to the classification system described below by two raters (TH, WW) blinded to each others ratings. These ratings were compared by a third reviewer (KH), who provided a deciding vote where discrepancies in classifications arose. Additional data from each study (sample size, reported results, list of authors, hospital staff blinded to screening tool classifications) were then collected by one reviewer (TH) for use in the analysis.

DEFINITION OF TERMS

Fall Risk Screening and Fall Risk Factor Assessment

Previous literature has demarcated fall risk assessment tools into two types: those that predict a probability of falling (risk screening tools), and those that assess factors that contribute to the patient's risk of falling (risk factor assessment tools) (26). A majority of tools contain both a list of fall risk factors to be assessed and a scoring mechanism which attempts to measure a patient's probability of falling. In this review, any tool that contains a mechanism for predicting a patient's probability of falling, regardless of measurement scale being continuous or categorical, will be referred to as a fall risk screening tool. Typically, these tools list a number of fall risk factors, assign a numerical value to the presence or absence of the risk factor, and then sum the numerical values together to represent the individual's risk of falling during hospitalization. In development, inclusion and weighting of risk factors may have been determined by multiple logistic regression (27–34), another multiple regression approach (35,36), or designer opinion based on knowledge of risk factors from the literature or local data (37–45). A cutoff point is typically selected on the numerical fall risk scale to demarcate those persons at high or low risk of falling.

Several authors have noted that an important purpose of these tools lies in accurate deployment of falls prevention interventions, minimizing expenditure of resources on patients who do not require them (20,27,29,40,46,47). Hence a primary means for comparing the relative value of alternate tools lies in comparing their accuracy of prediction.

Examples of tools that do not provide a description of the probability of falling, but provide a list of risk factors for falls, are available in current literature (17,48) and are referred to in this review as being fall risk factor assessment tools. As these tools do not provide a measure of risk of falling, their accuracy in prediction cannot be calculated and cannot be compared through the same means as fall risk screening tools. Although these tools may play an important role in falls prevention, they could not be incorporated in this review.

Study Design Classifications

In their sentinel work, Lijmer and colleagues (23) identified eight design characteristics that they included in their analysis. Not all of these design characteristics are

relevant to the field of fall risk screening in the hospital setting. We therefore drew on the relevant items from this work and diagnostic accuracy study design considerations subsequently identified by other authors and developed one composite, four-level classification system to describe the study design characteristics of evaluations in this field.

1. Retrospective/autovalidation
2. Retrospective/internal validation
3. Prospective/temporal validation
4. Prospective/external validation

This classification system is closely related to STARD checklist items 6, 8, and 9 and is further described below.

Retrospective Versus Prospective

The terms ''prospective'' and ''retrospective'' can relate to a number of factors in an evaluation of a fall risk screening tool. To be classified as prospective in this review, the following must have all been collected and selected prior to the accuracy of the tool or factors within it being analyzed for that data set: (i) data required for screening tool completion. (ii) content of the screening tool, (iii) weighting of factors included in the screening tool, and (iv) screening tool cutoff point. Neither the content of a screening tool, weighting of factors within the screening tool, nor screening tool cutoff point should be influenced by the data from which the accuracy of the tool is being established. Theoretically, few tools will not fit the data from which they were custom made (49). Other design issues and results from such evaluations provide little generality beyond that data set (50); hence studies not meeting all these criteria for prospective evaluation will be classified as retrospective. It is noted that some studies with partially prospective designs have presented accuracy data on preexisting tools which were then modified by adding another risk factor, modifying the scoring, or changing the cutoff point, presumably to improve the reported accuracy of the tool. Such evaluations have been classified as retrospective unless all modifications were made prior to study.

Retrospective–Auto Versus Internal Validation

A subclassification of the retrospective study design classification is the internal validation approach. An internal validation collects the data and divides the sample into two groups before the tool is constructed, constructs the tool on one set, and then tests its validity on the other. In its simplest form, this is referred to as ''split-sample validation,'' though more complex cross-validation methods can be used (49). Although this approach is still retrospective using the design classification defined in the previous paragraph, the data with which the tool is being validated are independent from the data with which it was constructed; thus this approach is stronger than the auto validation approach into which all other retrospective designs are placed.

Prospective–Temporal Versus External Validation

Prospective validation studies have been further subclassified as temporal validation and external validation designs (51). Temporal validation uses a set of patients from the same location but at a time later than the time during which the tool was constructed to validate the tool. This is a prospective evaluation independent of the original data. A difficulty with this design is that the results may not be generalizable to sites other than the site of original development. External validation addresses the wider issue of generality by collecting new data from an appropriate patient population in a different center, so not only is the validation data set prospective and independent of the original data set, it is also independent of characteristics of the original testing location. This makes external validation designs theoretically more rigorous than temporal designs.

### Tool Construction Can Also Affect Classification

Where tools were developed through expert opinion by clinicians working in the research location, clinical knowledge gained from previous patients, quality improvement projects, and research at this location may have influenced tool design. These studies were therefore classified as temporal validation studies, and those developed by others not working at the research location were classified as external validation studies. Where the accuracy of nursing staff clinical judgment in classifying fallers was investigated in isolation (i.e., without being combined with other risk factors) and other conditions for prospective evaluation were met, these results were classified as coming from temporal validation studies.

STATISTICAL ANALYSIS

A standard approach for describing accuracy of hospital fall risk screening tools has been through sensitivity (the proportion of fallers correctly classified as high fall risk) and specificity (the proportion of nonfallers correctly classified as low fall risk). Calculation of these statistics and confidence intervals of these for published tools have been described in a previous review (19). A difficulty with comparing tools on the basis of these statistics is that there are two statistics instead of one single statistic which singularly describes the accuracy of the tool. Singular statistics that could be used for this purpose include the area under the receiver operating characteristic curve, the diagnostic odds ratio (23), the proportion of cases appropriately identified, an overall accuracy statistic, or the Youden Index (25). To pool data from separate evaluations, it is important to select a measure that is independent of the relative sizes of faller and nonfaller groups in each study. The Youden Index, diagnostic odds ratio, and area under the receiver operating characteristic curve all meet this requirement; however, insufficient data are presented in many studies for area under the receiver operating characteristic curve to be calculated, and the diagnostic odds ratio cannot be calculated if either sensitivity = 1 or 0 or specificity = 0. Thus, in this study, the Youden Index with 95% confidence intervals was used to represent tool accuracy (25). The Youden Index is equal to sensitivity plus specificity minus one, has a range from 1 to −1, with 1 representing perfect accuracy and 0 the accuracy that would asymptotically be achieved through chance.

To assess the influence of study design on reported results, the mean (standard deviation) Youden Index of tools evaluated under each study design classification were
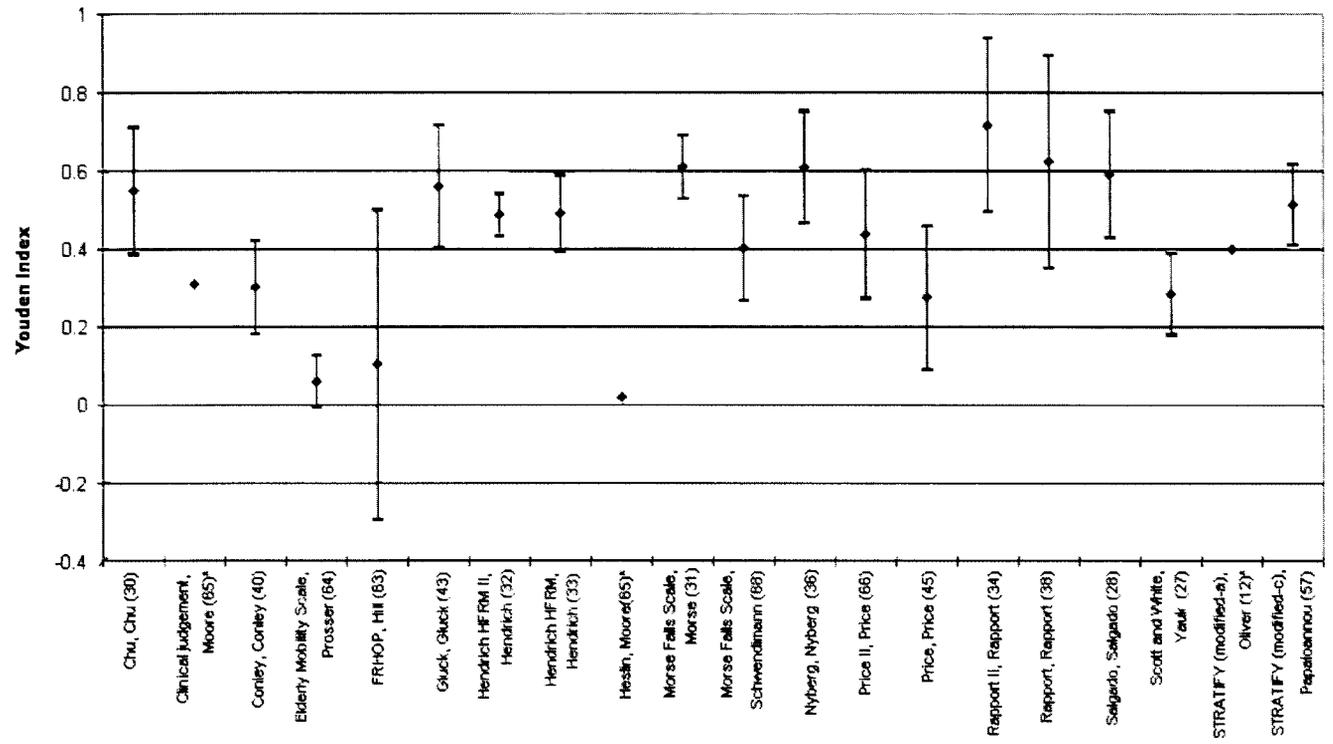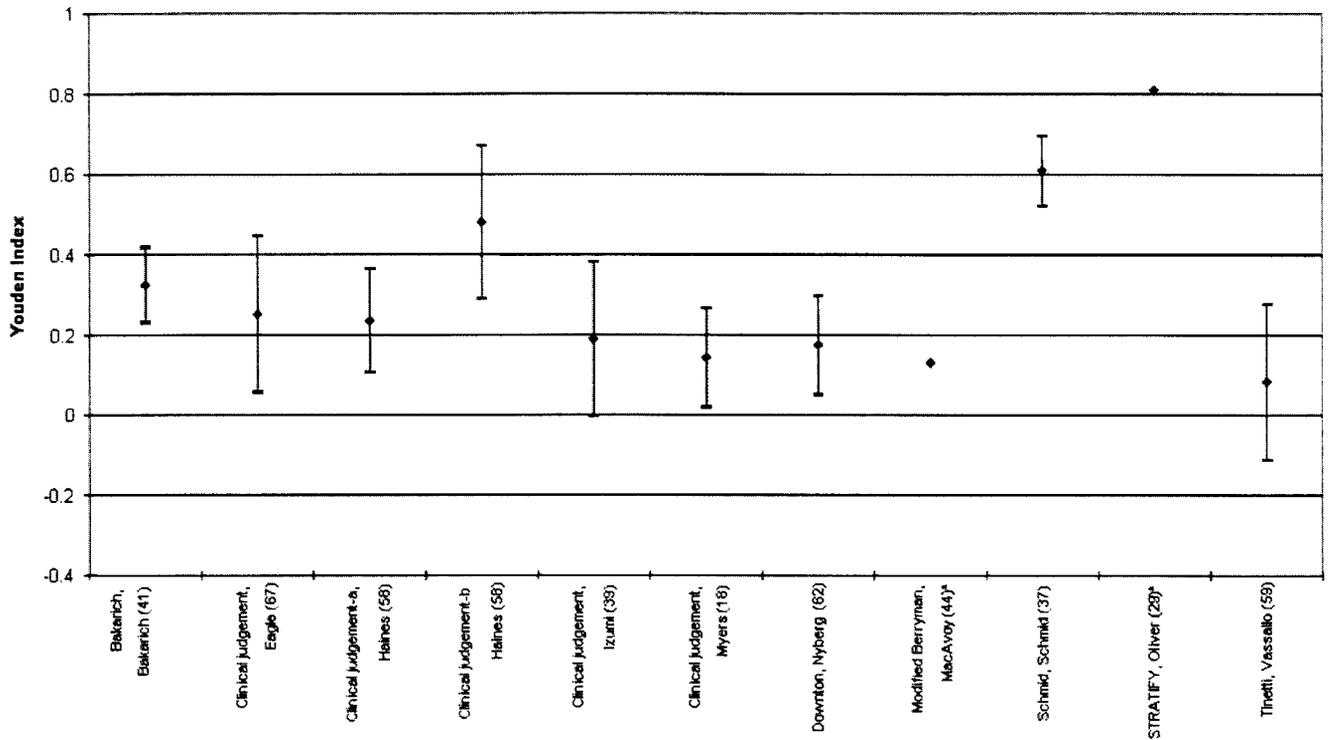
Figure 1. **A,** Youden Index (95% confidence interval [CI]) for evaluations classified as retrospective (auto). *95% CIs unable to be calculated due to insufficient data, sensitivity, and specificity point estimates used to calculate Youden Index. *x*-axis: Tool, author (reference). **B,** Youden Index (95% CI) for evaluations classified as prospective (temporal). **C,** Youden Index (95% CI) for evaluations classified as prospective (external).
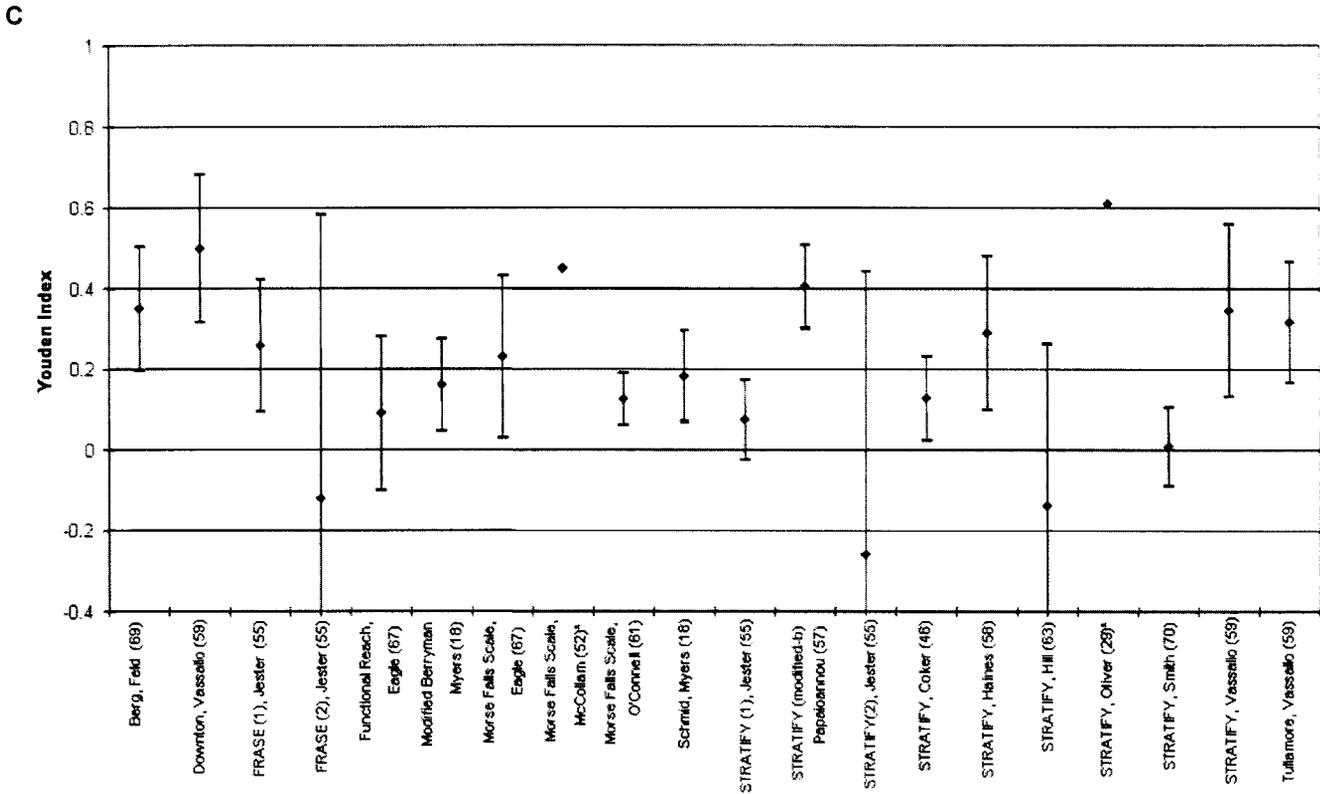
C

Figure 1. (*Continued*).

compared. By treating study design as a categorical independent variable, the Youden Index as a continuous outcome variable, and each evaluation as a subject, the effect of methodology on accuracy reported was investigated using linear regression analyses with clustering of data based on screening tool evaluated, and each study design (retrospective-auto, retrospective-internal, prospective-temporal, prospective-external) was entered into the model as a dummy variable. Other variables entered into the model included sample size, whether hospital staff were blinded to screening tool outcomes ("staff blinding"), and whether a member of the authorship team was involved in tool development ("author independence").

Based on the findings of the above analysis, a meta-analysis was undertaken pooling patient level data from prospective studies for screening tools where more than one evaluation has been undertaken. Pooled Youden Indices with 95% confidence intervals were calculated. To be able to be pooled, publications had to present directly the frequency of true positives, false positives, true negatives, and false negatives from their evaluations or sufficient data from which these could be calculated. Five evaluations arising from four studies were excluded (29,44,52,53) because insufficient data were presented, and attempts to obtain further data from authors were unsuccessful.

RESULTS

From the search strategy, 137 publications were identified for possible inclusion. Of these, 73 were excluded at the abstract review stage and 28 were excluded following review of the complete article, leaving a total of 36 publications to be included in the review (a list of publications identified with reason for exclusion, where applicable, is available from the corresponding author). Agreement on classification of study design was reached between the two blinded reviewers on 45 of the 52 evaluations, with the remaining 7 requiring a decision from the third assessor.

The classification of each of the 52 evaluations from studies included in this review, along with Youden Index (95% confidence interval) results are presented in Figure 1A–C. These 52 evaluations arose from 36 separate publications. A total of 20 evaluations were classified as retrospective validation evaluations, 11 as prospective (temporal) validation evaluations, and 21 as prospective (external) validation evaluations. Linear regression output examining the effect of prospective (temporal) design and retrospective (auto) design on Youden Index point estimates from each evaluation relative to prospective (external) design is presented in Table 1. This analysis indicates that the Youden Index scores calculated are significantly lower in prospective (external) designs than in retrospective (auto) designs by a magnitude of 0.22 on the Youden Index scale range (−1 to +1). This model explained 18% of variance (R^2) in Youden Index scores calculated, and the model was significant ($p < .01$). A mild trend for higher Youden Index scores was apparent for prospective (temporal) designs relative to prospective (external) designs (magnitude = 0.11), though this was not significant.

Table 1. Linear Regression of Youden Index by Trial Design Classification (Clustered by Tool With Robust Variance Estimates)

Design	β Coefficient	Robust 95% CI	p Value
Retrospective (auto)	0.22	0.11 to 0.32	<.001
Prospective (temporal)	0.11	−0.06 to 0.28	.18

Notes: Total number of observations = 52, $F(2,27) = 8.97$, $R^2 = 0.18$, number of clusters (tool) = 28. β coefficients relative to prospective (external) design.

CI = confidence interval.

Incorporating the variables ''sample size'' and ''staff blinding'' into the model did not considerably affect β coefficient point estimates or the significance of variable associations of the study design variables. Sample size had a significant univariate association with Youden Indices calculated. Beta coefficients did change considerably when the ''author independence'' variable was entered into the model, though this is likely to be due to the high level of multicolinearity between this variable and the design variable (logistic regression odds ratio of ''author independence'' with prospective [external] design = 15.0, $p = .001$, and with retrospective [auto] design = 0.08, $p < .001$).

Pooled Youden Index scores for screening tools with two or more prospective evaluations are presented in Figure 2, and indicate that the Schmid and Downton fall risk screening tools have the highest predictive accuracy, though these tools have only been subjected to two prospective evalua-

tions each, whereas the STRATIFY (nine evaluations) and nursing staff clinical judgment (five evaluations) have been subjected to many more.

DISCUSSION

It is apparent from this investigation that study design does have a significant effect on predictive accuracy results reported. This finding confirms findings from similar investigations in other fields, simulated studies (54), and statements made by the STARD initiative that studies with specific (particularly retrospective) design characteristics can provide biased and overly optimistic results (24). Consequently, conclusions arising from reviews of this area may be distorted unless due consideration is given to this effect.

In light of this evidence, this study completed the first meta-analysis of predictive accuracy data in this field. Only data from prospective studies were pooled to eliminate the potentially overoptimistic data provided from studies with retrospective design characteristics. From this pooled analysis, it appears that the Schmid and Downton fall risk scales potentially offer the greatest accuracy in predicting which patients will become in-hospital fallers. However, it is noted that these tools have been subjected to substantially fewer prospective evaluations than the STRATIFY and nursing staff clinical judgment, that the pooled number of participants is substantially lower than the STRATIFY, Morse Falls Scale, and clinical judgment, and that there is considerable variation in results within the individual pro-
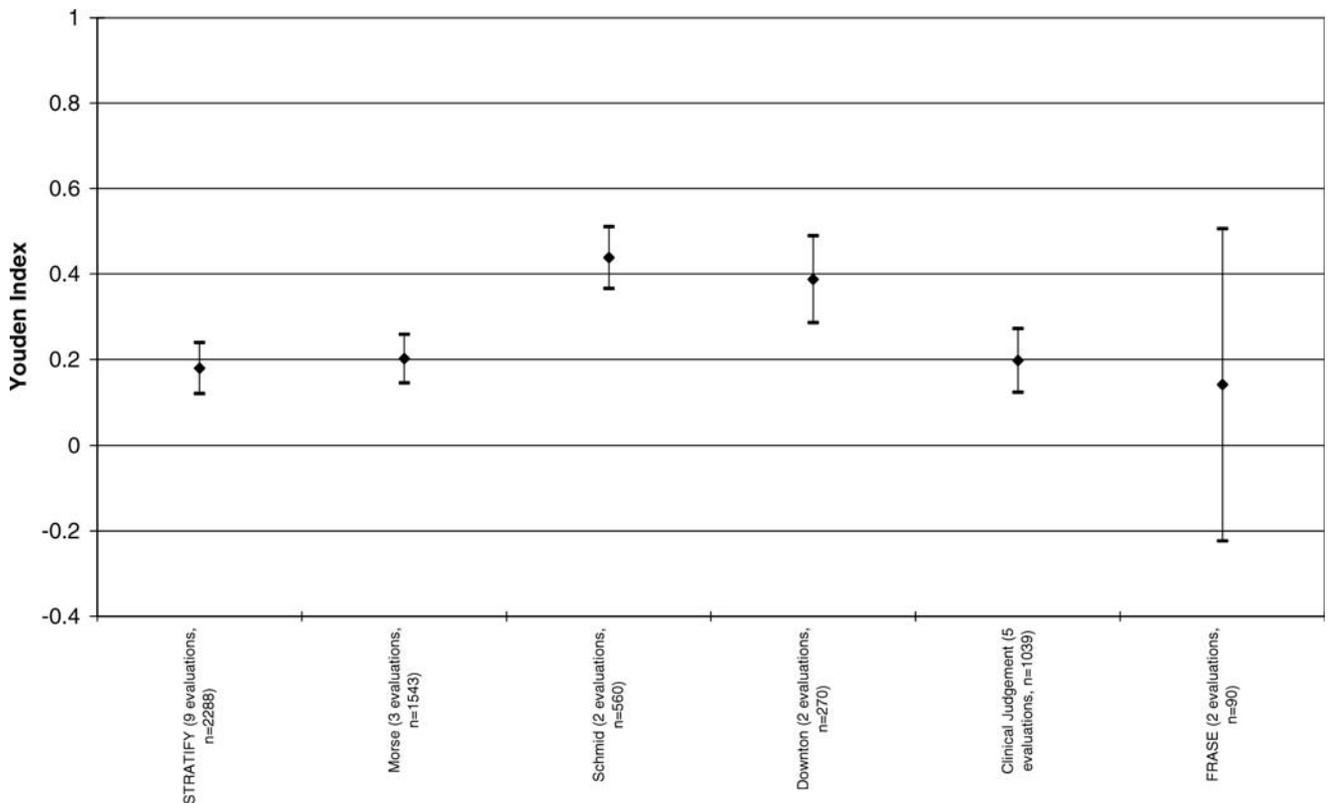

Figure 2. Pooled Youden Index (95% confidence interval [CI]) for screening tools from prospective studies (temporal or external) with > 1 prospective evaluation. *x*-axis: Tool (number of evaluations, sum number of participants).

spective evaluations that have looked at the Schmid and Downton fall risk scales. Further prospective evaluations of the Schmid and Downton fall risk scales are required in a range of settings before stronger recommendations can be made regarding the potential use of these tools as "across the board" fall risk screening tools in hospitals. Other factors that should also be considered in such a recommendation are the time cost required for tool completion, completion rates specific to that tool when used in the clinical context, and if the tool identifies modifiable risk factors for which interventions are readily available in that clinical context.

Our pooled analysis revealed only moderate predictive accuracy for the STRATIFY, Morse Falls Scale, Falls Risk Assessment Scale for the Elderly, and nursing staff clinical judgment, though the STRATIFY had two prospective evaluations with favorable results excluded based on insufficient data (which would have increased the pooled accuracy estimate if included), while the Morse Falls Scale had one evaluation with more favorable and one with less favorable results than the pooled analysis excluded (which is unlikely to have affected the pooled accuracy estimate if included). There was also substantial heterogeneity in the ward and patient types in which the prospective evaluations were undertaken, and also large variations in the reported results from individual evaluations of the STRATIFY and Morse Falls Scale. Thus it is suitable to conclude that these two tools will provide moderate accuracy when used across the range of settings from which their evaluations have arisen, but they may provide greater or lesser accuracy in specific patient groups. For example, when used among acute hip fracture patients on an orthopedic ward, the STRATIFY performed poorly [Youden Index $= -0.26$ (55,56)], yet among older patients admitted to general medical units [Youden Index $= 0.51$ (57)] and among those admitted to Elderly Care Units [Youden Index $= 0.81$ (29)], it performed very well. Conversely, nursing staff clinical judgment was consistently demonstrated to have moderate levels of predictive accuracy across the range of settings in which it was evaluated. A moderate level of accuracy may be sufficient to guide the selective deployment of some interventions, as previous research has found that in-hospital falls can be reduced through provision of a multifactorial intervention program selectively deployed largely through clinical judgment of hospital staff (7).

Other factors identified as having significant univariate associations with Youden Index scores calculated were sample size and whether the authors of the evaluation did not include an author involved in the development of the tool. The association between sample size and Youden Index scores calculated may be attributable to authors prematurely ceasing studies with poorer outcomes at smaller sample sizes. The association between the "author independence" variable and Youden Index scores may have been attributable to its multicolinearity with the study design variable, but conceivably also due to a publication bias whereby authors are less willing to proceed to publication with results of their trials if results are poor. Analysis of these factors was not the primary intent of this investigation, and these findings were somewhat serendipitous; however,

the potential bias they might introduce is of concern. To account for a sample size bias, one could model the results of smaller studies to be of equivalent sample size to the largest study, pool these results, and then examine if the pooled results from this adjusted model varied markedly from the results arising from the unadjusted pooled analysis.

There are some factors that are likely to further affect the validity of the pooled analysis findings. First, some tools were adapted for specific health care settings (57), and the research conditions within which the evaluation took place varied. For example, missing data (where screening tools were not completed) were excluded in some evaluations (46), but were able to be incorporated in others (58), and interventions were directly provided on the basis of the screening tool in some studies (55), but not in others where clinical staff were blinded to the screening tool classifications for their patients (59). These variations increase heterogeneity of the data that contributed to the pooled data and must be recognized when considering the validity of the pooled results.

Other study limitations require acknowledgment. First, only results relating to whether patients became fallers or nonfallers could be used in the meta-analysis, rather than also analyzing the ability of tools to predict fall event rates. Although arguments for the importance of this latter outcome in the hospital setting are emerging [particularly due to multiple falls being incurred by individual patients, screening tools being applied on multiple occasions, and patients having variable lengths of stay in hospital (60)], there is presently insufficient documentation of these data to justify meta-analysis. Recent advances in methodologies to evaluate the ability of screening tools to predict falls rates (58,60) may promote greater reporting of fall event rate outcomes and predictive accuracy, such that a meta-analysis of these data could be completed in the future.

The classification system used for this analysis grouped together a range of study characteristics and merged them into a four-point classification system. As a result, this study cannot identify the potential impact of each of these characteristics individually. However, in contrast to the sentinel work by Lijmer and colleagues (23), in which 218 evaluations were reviewed, to have done so using a multiple regression analysis in the present study where only 52 evaluations were reviewed would have made overfitting of such a model likely. Counting each evaluation as an individual subject for inclusion in the regression analysis could be considered a crude method for evaluating the effect of study design on reported results, which did not take into account variances in sample size of differences between tools. However, data were clustered according to screening tool evaluated, robust variance estimates were used, and regression coefficients were largely unchanged when the sample size variable was entered into the model (adjusting for studies of different sizes).

This study provides several implications for selection of a fall risk screening tool for use in clinical practice and future research. First, although retrospective evaluations still hold value in generating initial results and identifying tools and cutoff points that may be useful in the clinical setting, their results should not be weighted as heavily as those

arising from prospective studies when selecting a screening tool for use in the clinical setting. Second, results from initial studies should not be viewed as being definitive, even if a prospective design was used, until it has been repeated in similar populations elsewhere (if selecting a tool for a similar patient group) and in wider hospital-based populations (if selecting a tool for generic hospital-wide use). Third, the meta-analysis has indicated that, of the tools with multiple prospective evaluations and large pooled participant numbers, the STRATIFY, Morse Falls Scale, and nursing staff clinical judgment provide comparable levels of accuracy. For future research, the Schmid and Downton fall risk screening tools are worthy of further investigation. Nursing staff clinical judgment, STRATIFY, or Morse Falls Scale could all be used as comparison instruments, though the nursing staff clinical judgment had less variation in results reported in individual prospective studies and, logically, a screening tool is only of use if it improves the accuracy of prediction above that which could be predicted by the staff member without its use. Finally, further research of this nature should be undertaken in fields of fall risk assessment in residential aged care facilities and among community-dwelling elders to guide health care practitioners in their selection of fall risk screening tools.

CORRESPONDENCE

Address correspondence to Terry Haines, PhD, Physiotherapy Department, GARU, Princess Alexandra Hospital, Ipswich Rd., Woolloongabba, Queensland, 4112, Australia. E-mail: terrence_haines@health.qld.gov.au

REFERENCES

1. Gaebler S. Predicting which patient will fall again . . . and again. *J Adv Nurs*. 1993;18:1895–1902.
2. Rawsky E. Review of the literature on falls among the elderly. *Image J Nurs Sch*. 1998;30:47–52.
3. Tideiksaar R, Feiner C, Maby J. Falls prevention: the efficacy of a bed alarm system in an acute care setting. *Mt Sinai J Med*. 1993;60:522–527.
4. Boswell D, Ramsey J, Smith M, Wagers B. The cost-effectiveness of a patient-sitter program in an acute care hospital: a test of the impact of sitters on the incidence of falls and patient satisfaction. *Qual Manag Health Care*. 2001;10:10–16.
5. Mayo N, Gloutney L, Levy R. A randomised trial of identification bracelets to prevent falls among patients in a rehabilitation hospital. *Arch Phys Med Rehabil*. 1994;75:1302–1308.
6. Dempsey J. Falls prevention revisited: a call for a new approach. *J Clin Nurs*. 2004;13:479–485.
7. Haines T, Bennell K, Osborne R, Hill K. Effectiveness of targeted falls prevention programme in subacute hospital setting: randomised controlled trial. *Brit Med J*. 2004;328:676–679.
8. Healey F, Monro A, Cockram A, Adams V, Heseltine D. Using targeted risk factor reduction to prevent falls in older in-patients: a randomised controlled trial. *Age Ageing*. 2004;33:390–395.
9. Brady R, Chester F, Pierce L, Salter J, Schreck S, Radziewicz R. Geriatric falls: prevention strategies for the staff. *J Gerontol Nurs*. 1993;19:26–32.
10. Foster K, Kohlenberg E. Patient falls in a tertiary rehabilitation setting. *Rehabil Nurs Res*. 1996;5:23–29.
11. Grenier-Sennelier C, Lombard I, Jeny-Loeper C, Maillet-Gouret M, Minvielle E. Designing adverse event prevention programs using quality management methods: the case of falls in hospital. *Int J Qual Halth Care*. 2002;14:419–426.
12. Oliver D, Martin F, Seed P. Preventing patient falls. *Age Ageing*. 2002;31:75–82.
13. Sweeting H. Patient fall prevention—a structured approach. *J Nurs Manag*. 1994;2:187–192.
14. Tuffnell C. Falling can seriously damage one's health—even in hospital. *World Health Forum*. 1990;11:282–285.
15. Vassallo M, Vignaraja R, Sharma J, et al. The effect of changing practice on fall prevention in a rehabilitative hospital: the Hospital Injury Prevention Study. *J Am Geriatr Soc*. 2004;52:335–339.
16. Brandis S. A collaborative occupational therapy and nursing approach to falls prevention in hospital inpatients. *J Qual Clin Pract*. 1999;19:215–220.
17. Uden G, Ehnfors M, Sjostrom K. Use of initial risk assessment and recording as the main nursing intervention in identifying risk of falls. *J Adv Nurs*. 1999;29:145–152.
18. Myers H. Hospital fall risk assessment tools: a critique of the literature. *Int J Nurs Pract*. 2003;9:223–235.
19. Oliver D, Daly F, Martin F, McMurdo M. Risk factors and risk assessment tools for falls in hospital in-patients: a systematic review. *Age Ageing*. 2004;33:122–130.
20. Perell K, Nelson A, Goldman R, Luther S, Prieto-Lewis N, Rubenstein L. Fall risk assessment measures: an analytic review. *J Gerontol*. 2001;56A:M761–M766.
21. Simon R, Altman D. Statistical aspects of prognostic factor studies in oncology. *Brit J Cancer*. 1994;69:979–985.
22. Mallen C, Peat G, Croft P. Quality assessment of observational studies is not commonplace in systematic reviews. *J Clin Epidemiol*. 2006;59:765–769.
23. Lijmer J, Mol B, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA*. 1999;282:1061–1066.
24. Bossuyt P, Reitsma J, Bruns D, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD Initiative. *Ann Intern Med*. 2003;138:40–44.
25. Youden W. Index for rating diagnostic tests. *Cancer*. 1950;3:32–35.
26. Morse J. Nursing research on patient falls in health care institutions. *Ann Rev Nurs Res*. 1993;11:299–316.
27. Yauk S, Hopkins B, Phillips C, Terrell S, Bennion J, Riggs M. Predicting in-hospital falls development of the Scott and White falls risk screener. *J Nurs Care Qual*. 2004;20:128–133.
28. Salgado R, Lord S, Packer J, Ehrlich F. Factors associated with falling in elderly hospital patients. *Gerontology*. 1994;40:325–331.
29. Oliver D, Britton M, Seed P, Martin F, Hopper A. Development and evaluation of evidence based risk assessment tool (STRATIFY) to predict which elderly inpatients will fall: case-control and cohort studies. *Brit Med J*. 1997;315:1049–1053.
30. Chu L, Pei C, Chiu A, et al. Risk factors for falls in hospitalized older medical patients. *J Gerontol*. 1999;54A:M38–M43.
31. Morse J, Morse R, Tylko S. Development of a scale to identify the fall-prone patient. *Can J Aging*. 1989;8:366–377.
32. Hendrich A, Bender P, Nyhuis A. Validation of the Hendrich II fall risk model: a large concurrent case/control study of hospitalised patients. *Appl Nurs Res*. 2003;16:9–21.
33. Hendrich A, Nyhuis A, Kippenbrock T, Soja M. Hospital falls: development of a predictive model for clinical practice. *Appl Nurs Res*. 1995;8:129–139.
34. Rapport L, Hanks R, Millis S, Deshpande S. Executive functioning and predictors of falls in the rehabilitation setting. *Arch Phys Med Rehabil*. 1998;79:629–633.
35. Halfon P, Eggli Y, Melle G, Vagnair A. Risk of falls for hospitalized patients: a predictive model based on routinely available data. *J Clin Epidemiol*. 2001;54:1258–1266.
36. Nyberg L, Gustafson Y. Fall prediction index for patients in stroke rehabilitation. *Stroke*. 1997;28:716–721.
37. Schmid N. Reducing patient falls: a research-based comprehensive fall prevention program. *Mil Med*. 1990;155:202–207.
38. Rapport L, Webster J, Flemming K, et al. Predictors of falls among right-hemisphere stroke patients in the rehabilitation setting. *Arch Phys Med Rehabil*. 1993;74:621–626.
39. Izumi K, Makimoto K, Kato M, Hiramatsu T. Prospective study of fall risk assessment among institutionalized elderly in Japan. *Nurs Health Sci*. 2002;4:141–147.
40. Conley D, Schultz A, Selvin R. The challenge of predicting patients at risk for falling: development of the Conley Scale. *Medsurg Nurs*. 1999;8:348–354.

41. Bakarich A, McMillan V, Prosser R. The effect of a nursing intervention on the incidence of older patient falls. *Aust J Adv Nurs*. 1997;15:26–31.

42. Barbieri E. Patient falls are not accidents. *J Gerontol Nurs*. 1983;9:165–173.

43. Gluck T, Wientjes H, Rai G. An evaluation of risk factors for in-patient falls in acute and rehabilitation elderly care wards. *Gerontology*. 1996;42:104–107.

44. MacAvoy S, Skinner T, Hines M. Clinical methods: fall risk assessment tool. *Appl Nurs Res*. 1996;9:213–218.

45. Price C, Suddes M, Maguire L, Harrison S, O'Shea D. Single assessment of risk predicted which elderly patients would fall. *Brit Med J*. 1998;316:1750.

46. Coker E, Oliver D. Evaluation of the STRATIFY falls prediction tool on a geriatric unit. *Outcomes Manag*. 2003;7:8–14.

47. Myers H, Nikoletti S. Fall risk assessment: a prospective investigation of nurses' clinical judgement and risk assessment tools in predicting patient falls. *Int J Nurs Pract*. 2003;9:158–165.

48. Spellbrig A. Assessing elderly patients at high risk for falls: a reliability study. *J Nurs Care Qual*. 1992;6:30–35.

49. Steyerberg E, Harrell F, Borsboom G, Eijkemans M, Vergouwe Y, Habbema J. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol*. 2001;54: 774–781.

50. Harrell F, Lee K, Mark D. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15:361–387.

51. Altman D, Royston P. What do we mean by validating a prognostic model? *Stat Med*. 2000;19:453–473.

52. McCollam M. Evaluation and implementation of a research-based falls assessment innovation. *Nurs Clin North Am*. 1995;30:507–514.

53. Chow S, Lai C, Wong T, et al. Evaluation of the Morse Fall Scale: applicability in Chinese hospital populations. Int J Nurs Stud. In press.

54. Ewald B. Post hoc choice of cut points introduced bias to diagnostic research. *J Clin Epidemiol*. 2006;59:798–801.

55. Jester R, Wade S, Henderson K. A pilot investigation of the efficacy of falls risk assessment tools and prevention strategies in an elderly hip fracture population. *J Orthop Nurs*. 2005;9:27–34.

56. Haines T. All not lost on falls risk assessment for hip fracture patients. *J Orthop Nurs*. 2006;10:56.

57. Papaioannou A, Parkinson W, Cook R, Ferko N, Coker E, Adachi J. Prediction of falls using a risk assessment tool in the acute care setting. *BMC Med*. 2004;2:1.

58. Haines T, Bennell K, Osborne R, Hill K. A new instrument for targeting falls prevention interventions was accurate and clinically applicable in a hospital setting. *J Clin Epidemiol*. 2006;59:168–175.

59. Vassallo M, Stockdale R, Sharma J, Briggs R, Allen S. A comparative study of the use of four fall risk assessment tools on acute medical wards. *J Am Geriatr Soc*. 2005;53:1034–1038.

60. Haines T, Hill K, Bennell K, Osborne R. Recurrent events counted in evaluations of predictive accuracy. *J Clin Epidemiol*. 2006;59: 1155–1161.

61. O'Connell B, Myers H. The sensitivity and specificity of the Morse Fall Scale in an acute care setting. *J Clin Nurs*. 2002;11:134–136.

62. Nyberg L, Gustafson Y. Using the Downton Index to predict those prone to falls in stroke rehabilitation. *Stroke*. 1996;27:1821–1824.

63. Hill K, Vrantsidis F, Jessup R, McGann A, Pearce J, Collins T. Validation of a falls risk assessment tool in the sub-acute hospital setting: a pilot study. *Australas J Podiatric Med*. 2004;38:99–108.

64. Prosser L, Canby A. Further validation of the Elderly Mobility Scale for measurement of mobility of hospitalized elderly people. *Clin Rehabil*. 1997;11:338–343.

65. Moore T, Martin J, Stonehouse J. Predicting falls: risk assessment tool versus clinical judgement. *Perspectives*. 1996;20:8–11.

66. Price C, Suddes M, Wood B, Harrison S, O'Shea D. Validation of a visual analogue scale for assessment of falls risk in elderly inpatients. *Age Ageing*. 1999;28(Suppl 1):15.

67. Eagle J, Salama S, Whitman D, Evans L, Ho E, Olde J. Comparison of three instruments in predicting accidental falls in selected inpatients in a general teaching hospital. *J Gerontol Nurs*. 1999;25:40–45.

68. Schwendimann R, De Geest S, Milisen K. Evaluation of the Morse Fall Scale in hospitalised patients. *Age Ageing*. 2006;35:311–313.

69. Feld J, Rabadi M, Blau A, Jordan B. Berg balance scale and outcome measures in acquired brain injury. *Neurorehabil Neural Repair*. 2001;15:239–244.

70. Smith J, Forster A, Young J. Use of the 'STRATIFY' falls risk assessment in patients recovering from acute stroke. *Age Ageing*. 2006;35:138–143.