

Systematic review of accuracy of screening instruments for predicting fall risk among independently living older adults

Simon Gates, PhD;^{1*} Lesley A. Smith, PhD;² Joanne D. Fisher, PhD;³ Sarah E. Lamb, DPhil^{1,4}

¹Warwick Medical School Clinical Trials Unit, University of Warwick, Coventry, UK; ²Oxford Brookes University, School of Health and Social Care, Oxford, UK; ³Warwick Medical School, Gibbet Hill Campus, University of Warwick, Coventry, UK; ⁴Kadoorie Critical Care Research Centre, University of Oxford, Oxford, UK

Abstract—The objective of this study was to summarize the evidence on the accuracy of screening tools for predicting falling risk in community-living older adults. This study was designed as a systematic review. Prospective studies of clinical fall risk prediction tools that provided data on the number of participants who sustained falls during follow-up were included. We searched six electronic databases and reference lists of studies and review articles. Data were extracted by two reviewers independently, and methodological quality assessment was performed with a modified version of the Quality Assessment of Diagnostic Accuracy Studies checklist. Twenty-five studies were included. These studies evaluated 29 different screening tools, but only 6 of the tools were evaluated by more than one study. Methodological quality was variable, and many studies were small. No meta-analyses were performed because of heterogeneity. Most tools discriminated poorly between fallers and nonfallers. We found that existing studies are methodologically variable and the results are inconsistent. Insufficient evidence exists that any screening instrument is adequate for predicting falls.

Key words: accidental falls, assessment tool, clinical screening, fall risk, older adults, rehabilitation, risk assessment, screening, sensitivity, specificity, systematic review.

INTRODUCTION

Falls are a major health issue for older adults. About a third of people over the age of 65 will fall each year, and 5 to 10 percent of falls cause serious injury. Apart from the direct injuries resulting from falls, other long-term conse-

quences may include disability, fear of falling, and loss of independence, which can have serious effects on people's health and quality of life [1]. Systematic reviews of randomized controlled trials have concluded that several types of interventions may be effective in preventing falls, including strength and balance training, home hazard modification, and withdrawal of psychotropic medication [2]. Identification of older people at high risk of falling would theoretically allow targeting of fall-prevention interventions to those most likely to benefit from them. This screening process is distinct from the more intensive assessment procedures that are used to identify potentially modifiable risk factors in multifactorial fall prevention programs. For example, the American Geriatric Society/British Geriatric Society guideline [3] suggests that a simple screening algorithm, which incorporates a question about falls in the last year, and a timed performance test be

Abbreviations: CI = confidence interval, FRASE = Falls Risk Assessment Score for the Elderly, FRAT = Falls Risk Assessment Tool, IPD = individual patient data, NPV = negative predictive value, POMA = Performance Oriented Mobility Assessment, PPV = positive predictive value, ROC = receiver operating characteristic, TUG = Timed Up and Go (test).

*Address all correspondence to Simon Gates, PhD; Principal Research Fellow, Warwick Medical School Clinical Trials Unit, University of Warwick, Coventry, UK CV4 7AL; +44-(0)2476-575850; fax: +44-(0)2476-574657.

Email: s.gates@warwick.ac.uk

DOI: 10.1682/JRRD.2008.04.0057

administered and that those found by this screen to be at higher risk be given more intensive assessment and intervention.

Numerous studies have been conducted on risk factors for falls, and many factors related to future falls have been identified. The best predictors appear to be a history of falls and abnormalities of gait or balance [4]. Other factors such as visual impairment, medication use, and impaired cognition are less consistently associated with falls [4]. Numerous clinical screening instruments for identifying older people at high risk of falling have been proposed, and these vary in complexity from a single clinical test to scales involving 10 or more assessments. They may also include other questions such as whether a person has fallen in the past year. Screening tools have been developed for use in various populations, including hospitalized older adults, adults in residential care, and community-dwelling older people [5–7]. Different test attributes may be needed to predict falls successfully in different populations; for example, the timescale over which a prediction is needed varies from a few days or weeks in hospitalized patients to a year or more for community-living populations. Tools developed for one population may therefore be less accurate when used in a different setting. Many fall risk screening tools have been introduced into clinical practice in the United Kingdom in recent years [8] as components of clinics intended to reduce falls in community-living older people. However, their introduction has not been based on sound evidence that they are useful in discriminating between people who will fall and those who will not. In this review, we aim to assess and summarize the evidence for the accuracy of screening tests at predicting fallers in community-dwelling populations and to indicate where more research is needed.

METHODS

Search Strategy

We searched six electronic databases (MEDLINE [1966 to 22 May 2007], EMBASE [1980 to 22 May 2007], PsycInfo [1975 to 22 May 2007], CINAHL, and Social Science Citation Index [regular and expanded, all 1970 to 22 May 2007]). The search strategy was based on the Van der Weijden search strategy [9], which was found to be the most effective in a recent review [10]. We added extra terms to this search strategy (“risk assessment,” “assessment tool,” and the names of known screening

tools) because eligible studies for this review would not likely use terms usually associated with studies of diagnostic accuracy.

We assessed all studies included in six earlier reviews of related research [11–16] and inspected the reference lists of all eligible studies to further identify any potentially eligible studies that were cited by them. Studies that evaluated clinical tests for prediction of falls were included if they satisfied the following eligibility criteria:

1. Prospective cohort studies that evaluated the performance of one or more screening tests for predicting fallers.
2. The population was elderly people living in the community or substantially independently.
3. Falls were recorded prospectively; i.e., participants were followed up for falls occurring after the screening test was performed, with a follow-up duration of at least 3 months.
4. Data were presented (or could be calculated) on the number of fallers who were positive and negative for the screening test or summary statistics (sensitivity, specificity, positive predictive value [PPV], negative predictive value [NPV], or receiver operating characteristic [ROC] curves).

We included studies of older people living in residential care environments where they were substantially independent. We excluded hospital populations and other situations in which participants were not independent and populations of individuals with specific diagnoses such as stroke, Parkinson’s disease, or hip fracture. Studies that used a retrospective design relating performance on a screening test to a patient’s history of falls were excluded.

Data Extraction

Data were extracted by two reviewers independently, and discrepancies were resolved by discussion or by reference to a third reviewer.

Quality Assessment

We used a modified version of the Quality Assessment of Diagnostic Accuracy Studies quality assessment tool (**Table 1**) [17]. Three items (numbers 4, 7, and 10) were omitted because they were not applicable to this review. Two additional questions about the statistical analysis were added.

Table 1.

Quality assessment criteria (Quality Assessment of Diagnostic Accuracy Studies [QUADAS] checklist).

QUADAS Criterion	Interpretation for this Review	Scoring
1. Was the spectrum of patients representative of the patients who will receive the test in practice?	Same as QUADAS criterion.	3: Consecutive series or random selection of eligible people. 2: Unclear. 1: Not consecutive or random series.
2. Were selection criteria clearly described?	Same as QUADAS criterion.	3: Sufficient detail that selection process could be replicated. 2: Incomplete information. 1: No information.
3. Is the reference standard likely to classify the target condition correctly?	3a. Did recording of falls use adequate follow-up period? 3b. Did recording of falls use an accurate method?	3: At least 12 months. 2: 6–12 months. 1: Less than 6 months. 3: Prospective data collection; falls diary or calendar. 2: Data collected at intervals. 1: Recall at end of follow-up only.
4. Is the period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	Same as QUADAS criterion.	Not used: Not relevant to this review.
5. Did the whole sample or a random selection of the sample receive verification using a reference standard of diagnosis?	5. Was falls data recorded for all participants?	3: If it was clear that there was no selection of patients to be followed up for falls. 2: Possibility of selection. 1: Clear selection.
6. Did patients receive the same reference standard regardless of the index test result?	6. Was falls data collected in the same way for all participants regardless of screening test result?	3: Clear that data collection methods did not differ. 2: Unclear or no information. 1: Clear differences.
7. Was the reference standard independent of the index test (i.e., the index test did not form part of the reference standard)?	Same as QUADAS criterion.	Not used: Not relevant to this review.
8. Was the execution of the index test described in sufficient detail to permit replication of the test?	8. Was the screening test described in sufficient detail to permit replication of the test?	3: Details of exactly how test was implemented. 2: Partial information. 1: Insufficient information.
9. Was the execution of the reference standard described in sufficient detail to permit its replication?	9. Was the duration of follow-up and method of ascertainment of falls status reported in sufficient detail to permit replication?	3: Sufficient detail. 2: Partial information. 1: Insufficient information.
10. Were the index test results interpreted without knowledge of the results of the reference standard?	Same as QUADAS criterion.	Not used: Not relevant to this review because falls data always collected after screening tests.
11. Were the reference standard results interpreted without knowledge of the results of the index test?	11. Was assessment of falls done without knowledge of the screening test results?	3: Reliable measures reported to ensure that participants and clinicians assessing falls were unaware of screening test results. 2: Possible blinding but insufficient detail. 1: Not done or not mentioned.

Table 1. (Continued)

Quality assessment criteria (Quality Assessment of Diagnostic Accuracy Studies [QUADAS] checklist).

QUADAS Criterion	Interpretation for this Review	Scoring
12. Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	12. Were data on age, sex, and diagnoses reported?	3: All reported. 2: Only one or two reported. 1: None reported.
13. Were uninterpretable/intermediate test results reported?	13. Were test results reported for all participants (including unclear/uninterpretable test results)?	3: All participants accounted for in screening test results. 2: Partial explanation. 1: Insufficient or no information.
14. Were withdrawals from the study explained?	14a. Were screening test and falls reported for all participants that entered the study? 14b. Were withdrawals from the study explained?	3: >80%. 2: 70%–80%. 1: <70%. 3. Complete statement of losses and withdrawals. 2. Partial explanation. 1. Insufficient or no information.
Additional (non-QUADAS) Criteria		
—	15. Were data presented for all screening tests performed?	3: Results for all screening tests. 1: One or more screening tests omitted from results.
—	16. Were methods of analysis adequately described and free from error?	3: Correct analysis and sufficient detail. 2: Appears correct but insufficient detail to be sure. 1: Errors in analytical method.

Statistical Analysis

For each study, the specificity, sensitivity, PPV, and NPV were either extracted from the article or calculated from the data extracted from the article (where possible). No meta-analyses were performed because of heterogeneity between studies. For each study, the sensitivity, specificity, PPV, and NPV were tabulated with their 95 percent confidence intervals (CIs).

RESULTS

Studies

The electronic search yielded 3,363 citations. Of these, 125 were selected for further consideration based on their title and abstract and full reports were obtained. The search of the reference lists of the review articles and all selected studies yielded an additional 15 potentially eligible studies, for which full reports were also obtained. We considered a total of 140 full reports (**Figure**); 25 studies were eligible and included in the review (**Appendix 1**, available online only at www.rehab.research.va.gov/jour/08/45/8/pdf/con-

[tents.pdf](#)). The most common reasons for exclusion were that the study included an ineligible population (usually hospitalized participants), falls data were not collected prospectively, or the required data could not be extracted. Some studies included screening test scores as independent variables in a multiple regression model for prediction of fallers and presented sensitivity and specificity for the whole model. These studies were not included. Many studies presented their results as differences in test scores between fallers and nonfallers. Such results are not clinically useful because fallers and nonfallers are unknown at the time the screening test is performed, and these studies have therefore not been included.

Methodological Quality

Methodological quality was variable (**Table 2**). Follow-up periods were 12 months or more in 13/25 studies and ranged from 3 months to 5 years. Only 11/25 studies used reliable, prospective methods of recording falls outcomes, such as diaries or calendars. The remainder used less reliable methods such as participant recall at pre-defined intervals or the end of the follow-up period only.

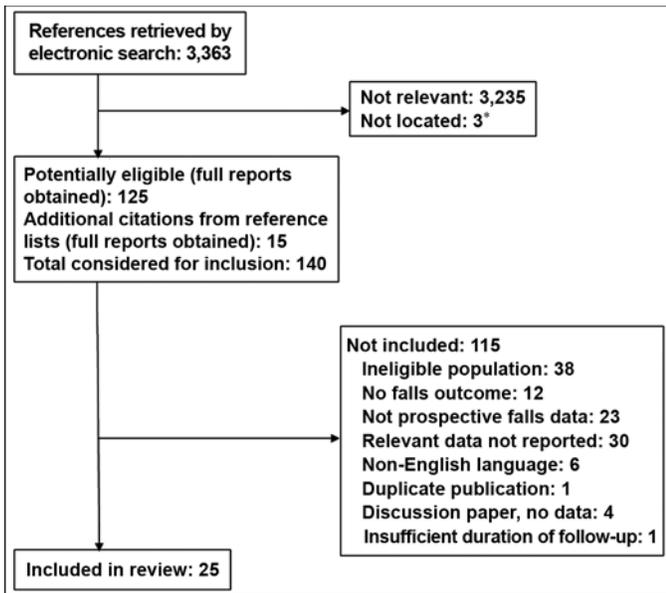


Figure. Flow chart of studies screened for use in review. *Studies not located: 1 conference abstract, reference apparently incorrect; 1 error in electronic database (wrong year; correct reference was considered for eligibility); 1 journal could not be located.

The majority of studies (21/25) reported the screening test procedures in sufficient detail or gave references to a description elsewhere. In some cases where several studies evaluated the same test, differences were noted in the performance or scoring of the test. For example, two studies evaluated a version of the Timed Up and Go (TUG) test, one including a 3 m walk and one a 5 m walk instead, and different versions of the Tinetti balance and gait scales were used by different studies.

Only four studies reported measures to ensure that assessment of falls outcomes was not influenced by knowledge of the screening test results. In these studies, clinicians who assessed falls and decided whether they qualified as outcomes were blinded. In most studies, falls were self-reported and whether participants knew whether their test result had classified them as high or low risk for falls was not clear.

Falls outcomes were reported in several different ways. The majority of studies (14/25) included all falls, but others reported recurrent falls, falls not due to an external hazard, falls not due to a medical event, indoor falls, or a combination of these.

In most studies (21/25), losses and exclusions were less than 20 percent; 3 studies had greater than 30 percent losses and exclusions, leading to a possibility of attrition bias in these studies. A substantial number of studies scored poorly for analysis; 8/25 did not report results for all of the screening tests that were performed, leading to a possibility of reporting bias, and 10/25 had errors in the reported analysis. These errors included unexplained discrepancies between the number of participants and the number of reported results [18], a clearly erroneous result for specificity [19], and incorrect exclusion of five participants from the results [20].

Test Performance

The included studies reported results for 29 different screening tests (Table 3 and Appendix 2; Appendix 2 is available online only at www.rehab.research.va.gov/jour/08/45/8/pdf/contents.pdf). Most tests were assessed by only one study, the main exceptions being the Tinetti gait, balance, and mobility scales (8 studies) and TUG test (4 studies). Generally, the screening tests had higher

Table 2. Quality assessment scores.

Study	Quality Assessment Criteria															
	1	2	3a	3b	5	6	8	9	11	12	13	14a	14b	15	16	
Bergland and Laake, 2005 [1]	3	3	3	3	3	3	3	3	2	2	3	3	3	3	3	
Bogle Thorbahn and Newton, 1996 [2]	3	2	2	2	3	3	3	3	1	3	2	3	3	3	3	
Cwikel et al., 1998 [3]	3	3	3	1	1	3	3	3	3	2	2	2	3	3	2	
Faber et al., 2006 [4]	2	2	2	3	2	3	3	3	2	2	3	3	1	3	1	
Flemming, 2006 [5]	2	2	1	1	3	1	3	1	2	1	3	3	3	3	1	
Hale et al., 1992 [6]	3	3	3	3	3	3	3	3	2	2	3	3	3	3	2	
Kario et al., 2001 [7]	2	3	3	3	1	3	3	3	2	3	1	3	1	3	2	
Killough, 2001 [8]	2	2	3	2	2	2	1	3	2	1	3	3	2	3	1	
Laessoe et al., 2007 [9]	2	2	3	3	3	3	3	3	2	2	2	3	3	1	3	
Lin et al., 2004 [10]	3	3	3	2	1	3	3	3	2	2	1	1	1	2	1	
Lundin-Olsson et al., 1997 [11]	2	1	2	1	3	3	2	1	2	3	3	3	3	3	3	
Lundin-Olsson et al., 2000 [12]	3	3	2	2	3	3	3	2	2	3	3	3	3	3	3	

Table 2. (Continued)

Quality assessment scores.

Study	Quality Assessment Criteria															
	1	2	3a	3b	5	6	8	9	11	12	13	14a	14b	15	16	
Lundin-Olsson et al., 2003 [13]	3	3	2	3	3	3	3	3	2	3	2	3	3	3	3	
Morris et al., 2007 [14]	2	3	3	2	2	3	3	3	2	2	3	3	3	1	3	
Murphy et al., 2003 [15]	3	3	3	2	1	3	3	2	2	3	1	3	3	1	1	
Nandy et al., 2004 [16]	3	3	2	1	3	3	3	3	2	2	1	3	3	3	2	
Okumiya et al., 1998 [17]	2	1	3	1	2	3	2	3	2	1	2	3	1	1	2	
Raiche et al., 2000 [18]	3	1	3	3	2	3	2	3	2	2	3	3	3	3	3	
Rosendahl et al., 2003 [19]	3	2	3	3	3	3	3	3	2	3	3	3	2	3	3	
Stel et al., 2003 [20]	1	2	3	3	3	3	3	3	2	2	1	1	3	2	3	
Studenski et al., 1994 [21]	2	3	2	3	3	3	3	3	3	3	3	3	3	1	3	
Tinetti et al., 1986 [22]	3	2	1	3	3	3	3	3	3	2	3	3	3	3	3	
Trueblood et al., 2001 [23]	3	3	2	1	3	3	3	3	2	2	2	3	3	3	2	
Vellas et al., 1997 [24]	3	3	3	2	3	3	3	3	2	2	1	1	3	2	3	
Verghese et al., 2002 [25]	3	3	3	2	3	3	3	3	3	3	3	3	3	3	3	

- Bergland A, Laake K. Concurrent and predictive validity of "getting up from lying on the floor." *Aging Clin Exp Res.* 2005;17(3):181–85. [PMID: 16110729]
- Bogle Thorbahn LD, Newton RA. Use of the Berg Balance Test to predict falls in elderly persons. *Phys Ther.* 1996;76(6):576–83. [PMID: 8650273]
- Cwikel JG, Fried A, Biderman A, Galinsky D. Validation of a fall-risk screening test, the Elderly Fall Screening Test (EFST), for community-dwelling elderly. *Disabil Rehabil.* 1998;20(5):161–67. [PMID: 9622261]
- Faber M, Bosscher RJ, Van Wieringen PC. Clinimetric properties of the performance-oriented mobility assessment. *Phys Ther.* 2006;86(7):944–54. [PMID: 16813475]
- Flemming PJ. Utilization of a screening tool to identify homebound older adults at risk for falls: Validity and reliability. *Home Health Care Serv Q.* 2006;25(3–4):1–22. [PMID: 17062508]
- Hale WA, Delaney MJ, McGaghie WC. Characteristics and predictors of falls in elderly patients. *J Fam Pract.* 1992;34(5):577–81. [PMID: 1578207]
- Kario K, Tobin JN, Wolfson LI, Whipple R, Derby CA, Singh D, Marantz PR, Wassertheil-Smoller S. Lower standing systolic blood pressure as a predictor of falls in the elderly: a community-based prospective study. *J Am Coll Cardiol.* 2001;38(1):246–52. [PMID: 11451282]
- Killough J. Validity of a new fall risk screen [abstract]. *J Geriatr Phys Ther.* 2001;28(3):119.
- Laessle U, Hoeck HC, Simonsen O, Sinkjaer T, Voigt M. Fall risk in an active elderly population—Can it be assessed? *J Negat Results Biomed.* 2007;6:2. [PMID: 17257414]
- Lin MR, Hwang HF, Hu MH, Wu HD, Wang YW, Huang FC. Psychometric comparisons of the timed up and go, one-leg stand, functional reach, and Tinetti balance measures in community-dwelling older people. *J Am Geriatr Soc.* 2004;52(8):1343–48. [PMID: 15271124]
- Lundin-Olsson L, Nyberg L, Gustafson Y. "Stops walking when talking" as a predictor of falls in elderly people. *Lancet.* 1997;349(9052):617. [PMID: 9057736]
- Lundin-Olsson L, Nyberg L, Gustafson Y. The Mobility Interaction Fall chart. *Physiother Res Int.* 2000;5(3):190–201. [PMID: 10998775]
- Lundin-Olsson L, Jensen J, Nyberg L, Gustafson Y. Predicting falls in residential care by a risk assessment tool, staff judgement, and history of falls. *Aging Clin Exp Res.* 2003;15(1):51–59. [PMID: 12841419]
- Morris R, Harwood RH, Baker R, Sahota O, Armstrong S, Masud T. A comparison of different balance tests in the prediction of falls in older women with vertebral fractures: A cohort study. *Age Ageing.* 2007;36(1):78–83. [PMID: 17264139]
- Murphy MA, Olson SL, Protas EJ, Overby AR. Screening for falls in community-dwelling elderly. *J Aging Phys Activity.* 2003;11:66–80.
- Nandy S, Parsons S, Cryer C, Underwood M, Rashbrook E, Carter Y, Eldridge S, Close J, Skelton D, Taylor S, Feder G; Falls Prevention Pilot Steering Group. Development and preliminary examination of the predictive validity of the Falls Risk Assessment Tool (FRAT) for use in primary care. *J Public Health (Oxf).* 2004;26(2):138–43. [PMID: 15284315] Erratum in: *J Public Health (Oxf).* 2005;27(1):129–30.
- Okumiya K, Matsubayashi K, Nakamura T, Fujisawa M, Osaki Y, Doi Y, Ozawa T. The timed "up & go" test is a useful predictor of falls in community-dwelling older people. *J Am Geriatr Soc.* 1998;46(7):928–30. [PMID: 9670889]
- Raiche M, Hebert R, Prince F, Coriveau H. Screening older adults at risk of falling with the Tinetti balance scale. *Lancet.* 2000;356(9234):1001–2. [PMID: 11041405]
- Rosendahl E, Lundin-Olsson L, Kallin K, Jensen J, Gustafson Y, Nyberg L. Prediction of falls among older people in residential care facilities by the Downton index. *Aging Clin Exp Res.* 2003;15(2):142–47. [PMID: 12889846]
- Stel VS, Pluijm SM, Deeg DJ, Smit JH, Bouter M, Lips P. A classification tree for predicting recurrent falling in community-dwelling older persons. *J Am Geriatr Soc.* 2003;51(10):1356–64. [PMID: 14511154]
- Studenski S, Duncan PW, Chandler J, Samsa G, Prescott B, Hogue C, Bearon LB. Predicting falls: The role of mobility and nonphysical factors. *J Am Geriatr Soc.* 1994;42(3):297–302. [PMID: 8120315]
- Tinetti ME, Williams TF, Mayewski R. Fall risk index for elderly patients based on number of chronic disabilities. *Am J Med.* 1986;80(3):429–34. [PMID: 3953620]
- Trueblood PR, Hodson-Chennault N, McCubbin A, Youngclarke D. Performance and impairment-based assessments among community dwelling elderly: Sensitivity and specificity. *Issues Aging.* 2001;24(1):2–6.
- Vellas BJ, Wayne SJ, Romero L, Baumgartner RN, Rubenstein LZ, Garry PJ. One-leg balance is an important predictor of injurious falls in older persons. *J Am Geriatr Soc.* 1997;45(6):735–38. [PMID: 9180669]
- Verghese J, Buschke H, Viola L, Katz M, Hall C, Kuslansky G, Lipton R. Validity of divided attention tasks in predicting falls in older individuals: A preliminary study. *J Am Geriatr Soc.* 2002;50(9):1572–76. [PMID: 12383157]

Table 3.

Results reported as summary receiver operating characteristic (ROC) curves.

Study	Screening Test	Outcome	ROC AUC	95% CI
Lin et al., 2004 [1]	TUG	Falls	0.61	Not given
	One-leg standing	Falls	0.53	Not given
	Functional reach	Falls	0.51	Not given
	Tinetti balance	Falls	0.56	Not given
Stel et al., 2003 [2]	Mediolateral sway (<i>n</i> = 161)	Recurrent falls	0.67	(0.57–0.77)
	Tandem stand (<i>n</i> = 161)	Recurrent falls	0.61	(0.49–0.73)
	Leg extension strength (<i>n</i> = 419)	Recurrent falls	0.58	(0.51–0.64)
	Handgrip strength (<i>n</i> = 419)	Recurrent falls	0.57	(0.51–0.64)

1. Lin MR, Hwang HF, Hu MH, Wu HD, Wang YW, Huang FC. Psychometric comparisons of the timed up and go, one-leg stand, functional reach, and Tinetti balance measures in community-dwelling older people. *J Am Geriatr Soc.* 2004;52(8):1343–48. [PMID: 15271124]

2. Stel VS, Pluijm SM, Deeg DJ, Smit JH, Bouter M, Lips P. A classification tree for predicting recurrent falling in community-dwelling older persons. *J Am Geriatr Soc.* 2003;51(10):1356–64. [PMID: 14511154]

AUC = area under curve, CI = confidence interval, TUG = Timed Up and Go (test).

specificity than sensitivity, indicating that a higher proportion of nonfallers than fallers were correctly identified. Specificity of at least 80 percent was reported 22 times, compared with only 8 reports of sensitivity of 80 percent or more. Only two tests had any result for which sensitivity and specificity both exceeded 80 percent [20–21], although a larger study of one of these tests (Mobility Interaction Fall chart) did not confirm this result [22]. Almost no replication of test evaluation was reported in independent studies. Most tests were assessed by only one included study. Where several studies did evaluate the same test, differences in the conduct of the test, scoring, cutoff points, or outcome measures meant that the results could not be compared or combined in meta-analyses.

Tinetti Gait, Balance, and Mobility Scales

Eight studies evaluated the Tinetti gait, balance, or mobility scales (also referred to as the Performance Oriented Mobility Assessment [POMA]) [18,20,23–28]. Five studies provided data on the balance scale, three on the gait scale, and five on the combined mobility scale. The assessments performed, scoring systems, and outcomes varied between studies, so we judged it inappropriate to perform a meta-analysis. For the overall mobility scale, the sensitivity found in different studies varied between 0.27 and 0.76, with specificity between 0.52 and 0.83. The PPV was between 0.31 and 0.68, and the NPV was between 0.67 and 0.88. For the balance scale alone, sensitivity and PPV were similarly variable (0.23–0.80 and 0.33–0.86, respectively). Specificity was

0.66 or higher for all studies and cutoffs examined, and NPV was always at least 0.78. The gait scale alone had similar results to the balance scale, with high NPV (0.80–0.86) and specificity (0.63–0.95) but lower or more variable sensitivity (0.20–0.68) and PPV (0.43–0.46). One study reported results for the balance scale using ROC curve statistics [28]. The area under the curve was 0.559 (no CI given), showing poor discriminatory ability.

Timed Up and Go Test

The four studies that assessed the TUG test [24,28–30] used different versions of the test and different cutoff values, hence no meta-analysis was performed. In common with most other tests, specificity was higher than sensitivity and NPV generally higher than PPV. One study reported the area under the ROC curve as 0.61 (no CI given), but no information was provided on the best-performing cutoff value [28]. The TUG test was the best-performing test of the four screening tests in this study.

Mobility Interaction Fall Chart

This test was evaluated in two studies by Lundin-Olsson et al. [21–22]. Considerable heterogeneity was found between the results (I^2 values of 95% for sensitivity and 65% for specificity), so meta-analysis was not appropriate. The earlier study found high values for sensitivity and specificity (0.85 and 0.82, respectively), which were not confirmed by the later and larger study (0.43 and 0.69, respectively). PPV and NPV were similarly higher in the earlier study.

Functional Reach

Murphy et al. found relatively high values for sensitivity and specificity (0.73 and 0.88, respectively), although the CIs were wide because of this study's small size [20]. However, the larger study by Lin et al. suggested that this test had almost no discriminatory ability between fallers and nonfallers (area under ROC curve was 0.51 with no CI given) [28].

Tandem Stance

Stel et al. found an area under the ROC curve of 0.61 (95% CI = 0.49–0.73) for this test, suggesting that its discriminatory ability was poor [31]. Murphy et al. found poor sensitivity (55%) but good specificity (94%) [20].

Walking Tests

The two studies that assessed walking tests differed in their methods of assessment; one assessed the predictive ability of a 5-minute walk (i.e., the distance walked in 5 minutes) [20], whereas the other evaluated a timed gait test (i.e., the time taken to walk a total of 40 feet) [25]. The 5-minute walk, using a cutoff of 1,000 feet, had high sensitivity and NPV (0.93 and 0.82, respectively) but lower specificity and PPV (0.44 and 0.21, respectively). Results of the timed 40-foot walk varied with the cutoff used.

DISCUSSION

Adequate evidence for us to determine with any confidence how good any screening test was at predicting fallers was not available. Most tests have been evaluated by only one study, and where multiple studies exist, they are incompatible in important ways. Moreover, many studies have small sample sizes or suffer from methodological problems or poor reporting. With current evidence, therefore, providing a quantitative summary of the accuracy of any fall risk screening tool is not possible. For robust determination of the accuracy of screening tests, further high-quality studies are needed. These should seek to use compatible study designs to allow pooling of their results in future systematic reviews and a large enough sample size to estimate sensitivity and specificity with sufficient precision.

A recent survey identified the screening tools most commonly used by falls clinics (i.e., clinics that perform multifactorial risk assessment and intervene to reduce

risk factors for falling) in the United Kingdom in 2006 [8]. The Falls Risk Assessment Tool (FRAT) was by far the commonest, with smaller numbers using the Tinetti POMA, TUG test, Falls Risk Assessment Score for the Elderly (FRASE) [32], and Berg Balance Scale. Strong evidence of the predictive ability of any of these measures does not exist, and one of them, the FRASE, was not evaluated by any studies included in this review. In the one study that evaluated the FRAT, specificity was 80 to 97 percent depending on the cutoff point used but sensitivity was only 15 to 59 percent [33]. Therefore, these measures may not give good predictions of which people are likely to fall and who might benefit from detailed multifactorial risk assessment. Inaccurate selection of people for assessment by falls clinics may be an important constraint on their effectiveness; inclusion of a high proportion of low-risk people in the clinic population will lead to waste of resources, whereas omission of a large proportion of people at high risk would limit the reduction in falls that the clinic could achieve.

The Tinetti balance, gait, and mobility scales were evaluated by the most studies, but clearly, different versions of the test were used by different studies [34]. Tinetti described two versions of the screening tool in different articles: one comprising 13 balance and 9 gait assessments [35] and one 8 balance tests (maximum score 15) and 8 gait tests (maximum score 13) [23]. Of the four studies that used both balance and gait assessments, two had a maximum score of 40 (with different cutoffs) [26–27] and two had a maximum of 28 (with the same positive test criterion [19/28]) [18,23]. For the balance scale, four studies reported a total score of 16, one reported 15, and one reported 26, and for the gait scale, two reported a maximum score of 12 and one reported 13. The number of items assessed in the balance scale varied from 9 to 14. The differences among the cutoff points, scoring systems, and outcomes used precluded any combination of results for these tests.

We excluded 23 studies from this review because they did not use a prospective design. Instead, they performed the screening tests and evaluated their relationship with historical falls. This study design may give misleading results because past falls may affect test performance, which may overestimate the accuracy of screening instruments. Moreover, it does not represent a real-life clinical situation, as clinicians need to use a tool that can predict future falls. For these reasons, retrospective studies were excluded.

Many of the excluded studies could potentially provide relevant data for a future individual patient data (IPD) review because the relevant data were probably collected but the results were not presented in a way that allowed extraction of the numbers of people who were predicted by the test to fall and those who actually fell during follow-up. Access to the original IPD for these studies could yield more data on the performance of screening instruments and may be able to give a more accurate picture of their predictive ability without the need for conducting new studies. Attempting to collect IPD may be worth considering in a future review but would be a major undertaking.

Some excluded studies assessed the relationship of a screening test to a gold standard (usually another screening test) rather than measuring falls. For example, Whitney et al. analyzed correlations of the TUG test with the Physiological Profile Assessment, a comprehensive falls risk assessment [36]. This type of analysis may give a result much more quickly, without the need for long-term follow-up, but is potentially misleading. No screening test is perfect, and at present, no screening test is known to be accurate enough to be regarded as a gold standard. A good level of agreement between two tests, therefore, does not necessarily mean that prediction of falls will be good.

The studies in this review assessed the ability of screening tools to predict fallers. However, whether participants were counted as “fallers” was measured in various ways. Some studies included all falls reported by the participants, whereas others excluded falls likely to have been caused by internal events (such as stroke), drugs, or unusual hazards. Some studies limited “fallers” to individuals who experienced two or more falls during the follow-up periods (recurrent fallers) on the basis that this result may be more clinically relevant; those who experience repeated falls are more likely to suffer injury or other adverse consequences of falling. However, data on fall-related injuries, one of the most clinically important consequences of falls, could be extracted from only two studies [37–38]. Therefore, very little information is available on whether a screening tool for predicting fallers will also be able to predict those who will suffer fall-related injuries or other outcomes relevant to long-term health. The disparity among outcome measures is likely to contribute to statistical heterogeneity among the results, and future studies should ensure that they use standardized and clinically relevant outcomes [39].

Studies relevant to this review may have been difficult to locate because they may not have included search terms for diagnostic studies in their titles and abstracts. We tested the accuracy of our search strategy against two earlier reviews that included studies relevant to this review [12–13]. Our strategy located 16/21 studies in one review and 15/23 in the other. Examination of the studies not located revealed that none of them was eligible for this review. The titles and abstracts of these studies lacked any terms describing their population or methodology, and hence would have been difficult to locate with any electronic search strategy. This finding emphasized the importance of additional searching of reference lists to locate relevant studies.

What the most important attributes of a screening test are and how good it needs to be to be useful will depend on what it is used for. Probably the commonest use would be as an initial screen for an unselected population to identify the individuals that may benefit from further assessment. For this purpose, a high NPV might be the most useful attribute, since it would mean that those not referred for further assessment would only include a few people at high risk of falls. However, the predictive value of a test depends on the prevalence of the condition in the population; for a given sensitivity and specificity, the NPV will increase as the prevalence declines. The incidence of falls in the general population over age 65 is generally thought to be around 30 percent a year. If 30 percent of the population were fallers, achieving an NPV of 90 percent with a specificity of 70 percent would require a sensitivity of 82 percent. It seems unlikely that any of the tests included in this review would perform this well. The NPV is relatively insensitive to the value of specificity, and if specificity were 50 percent rather than 70 percent, the required sensitivity increases to only 88 percent. However, an NPV of 90 percent would still mean that 18 percent of fallers would not be identified on the screening test and that 46 percent of those who screened positive would not fall. If less than 30 percent of the population were fallers, the NPV would be improved at the expense of a reduction in PPV; this would mean that more of those referred for further assessment would actually not be fallers.

CONCLUSIONS

At present, recommending any screening test for routine clinical use is not possible. Despite the number of

studies that have been conducted, no strong evidence exists that any screening test is useful for identifying fallers. If screening tests are to be used in clinical practice, further studies are needed to provide reliable estimates of their performance and give a sound evidence base for their use. Future studies should use a sufficiently large sample size to estimate sensitivity and specificity with high precision, be conducted in a clinically relevant population, include a sufficient duration of follow-up, and have reliable methods of recording of falls. However, some evidence exists that simple screening questions may perform as well as more complex screening tests in predicting who will fall. A history of falls and reported abnormalities of gait or balance are consistently found to be the best predictors of future falls [4], and little or no additional value may be gained by performing a complex screening test. For example, in two studies of fall risk factors, a history of falls had sensitivity of 0.93 and 0.95 and specificity of 0.20 and 0.21 for prediction of single falls in the subsequent year [40–41]. For prediction of recurrent falls, two further studies found sensitivity of 0.77 and 0.78 and specificity of 0.52 and 0.54 [42–43]. These values are comparable to those of many screening tests and suggest that simply asking about past falls may give a reasonably accurate prediction with minimal effort. However, these data are from only a few independent cohorts and the accuracy of this screening question has not yet been adequately reviewed. A recent systematic review was able to extract data to calculate likelihood ratios for only these four studies [4]. Further research, possibly involving IPD from the many existing risk factor studies, is necessary to establish the accuracy of simple screening questions and whether complex screening tests offer any additional predictive accuracy.

ACKNOWLEDGMENTS

This material was based on work supported by the UK National Health Service, Service Delivery and Organisation Programme, project number SDO/139/2006.

Authors' contributions—Concept and design: SEL, SG; Literature searching: SG, JDF; Data extraction: SG, LAS, JDF; Analysis and interpretation: SG, LAS, SEL, JDF; Drafting of manuscript: SG, SEL; Revision of manuscript: SG, SEL, LAS, JDF.

Conduct and reporting of the study were independent of the funder.

The authors have declared that no competing interests exist.

REFERENCES

1. Kannus P, Sievanen H, Palvanen M, Jarvinen T, Parkkari J. Prevention of falls and consequent injuries in elderly people. *Lancet*. 2005;366(9500):1885–93. [PMID: 16310556]
2. Gillespie LD, Gillespie WJ, Robertson MC, Lamb SE, Cumming RG, Rowe BH. Interventions for preventing falls in elderly people. *Cochrane Database Syst Rev*. 2003;(4): CD000340. [PMID: 14583918]
3. Guideline for the prevention of falls in older persons. American Geriatrics Society, British Geriatrics Society, American Academy of Orthopaedic Surgeons Panel on Falls Prevention. *J Am Geriatr Soc*. 2001;49(5):664–72. [PMID: 11380764]
4. Ganz D, Bao Y, Shekelle PG, Rubenstein LZ. Will my patient fall? *JAMA*. 2007;297(1):77–86. [PMID: 17200478]
5. Oliver D, Britton M, Seed P, Martin FC, Hopper AH. Development and evaluation of evidence based risk assessment tool (STRATIFY) to predict which elderly inpatients will fall: Case-control and cohort studies. *BMJ*. 1997; 315(7115):1049–53. [PMID: 9366729]
6. Lord SR, Menz HB, Tiedemann A. A physiological profile approach to falls risk assessment and prevention. *Phys Ther*. 2003;83(3):237–52. [PMID: 12620088]
7. Hill K, Vratsidis 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. *Aust J Pod Med*. 2004; 38(4):99–108.
8. Lamb S, Gates S, Fisher J, Cooke M, Carter Y, McCabe C. Scoping exercise on fallers' clinics: Report to the National Coordinating Centre for NHS Service Delivery and Organisation R & D (NCCSDO). Coventry (England): NCCSDO; 2007.
9. Van der Weijden T, Ijzermans CJ, Dinant GJ, Van Duijn NP, De Vet R, Buntinx F. Identifying relevant diagnostic studies in Medline. The diagnostic value of the erythrocyte sedimentation rate (ESR) and dipstick as an example. *Fam Pract*. 1997;14(3):204–8. [PMID: 9201493]
10. Leeflang MM, Scholten RJ, Rutjes AW, Reitsma JB, Bossuyt PM. Use of methodological filters to identify diagnostic accuracy studies can lead to the omission of relevant studies. *J Clin Epidemiol*. 2006;59(3):234–40. [PMID: 16488353]
11. Myers H. Hospital fall risk assessment tools: A critique of the literature. *Int J Nurs Pract*. 2003;9(4):223–35. [PMID: 12887374]
12. Perell KL, Nelson A, Goldman RL, Luther SL, Prieto-Lewis N, Rubenstein LZ. Fall risk assessment measures: An analytic review. *J Gerontol A Biol Sci Med Sci*. 2001; 56(12):M761–66. [PMID: 11723150]

13. Jarnlo GB. Functional balance tests related to falls among elderly people living in the community. *Eur J Geriatr*. 2003;5(1):7–14.
14. Oliver D, Daly F, Martin FC, McMurdo ME. Risk factors and risk assessment tool for falls in hospital in-patients: A systematic review. *Age Ageing*. 2004;33(2):122–30. [\[PMID: 14960426\]](#)
15. Piirtola M, Era P. Force platform measurements as predictors of falls among older people—A review. *Gerontology*. 2006;52(1):1–16. [\[PMID: 16439819\]](#)
16. Scott V, Votova K, Scanlan A, Close J. Multifactorial and functional mobility assessment tools for fall risk among older adults in community, home-support, long-term and acute care settings. *Age Ageing*. 2007;36(2):130–39. [\[PMID: 17293604\]](#)
17. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: A tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol*. 2003;3:25. [\[PMID: 14606960\]](#)
18. Faber M, Bosscher RJ, Van Wieringen PC. Clinimetric properties of the performance-oriented mobility assessment. *Phys Ther*. 2006;86(7):944–54. [\[PMID: 16813475\]](#)
19. Flemming PJ. Utilization of a screening tool to identify homebound older adults at risk for falls: Validity and reliability. *Home Health Care Serv Q*. 2006;25(3–4):1–22. [\[PMID: 17062508\]](#)
20. Murphy MA, Olson SL, Protas EJ, Overby AR. Screening for falls in community-dwelling elderly. *J Aging Phys Activity*. 2003;11:66–80.
21. Lundin-Olsson L, Nyberg L, Gustafson Y. The Mobility Interaction Fall chart. *Physiother Res Int*. 2000;5(3):190–201. [\[PMID: 10998775\]](#)
22. Lundin-Olsson L, Jensen J, Nyberg L, Gustafson Y. Predicting falls in residential care by a risk assessment tool, staff judgement, and history of falls. *Aging Clin Exp Res*. 2003;15(1):51–59. [\[PMID: 12841419\]](#)
23. Tinetti ME, Williams TF, Mayewski R. Fall risk index for elderly patients based on number of chronic disabilities. *Am J Med*. 1986;80(3):429–34. [\[PMID: 3953620\]](#)
24. Trueblood PR, Hodson-Chennault N, McCubbin A, Young-clarke D. Performance and impairment-based assessments among community dwelling elderly: Sensitivity and specificity. *Issues Aging*. 2001;24(1):2–6.
25. Verghese J, Buschke H, Viola L, Katz M, Hall C, Kuslansky G, Lipton R. Validity of divided attention tasks in predicting falls in older individuals: A preliminary study. *J Am Geriatr Soc*. 2002;50(9):1572–76. [\[PMID: 12383157\]](#)
26. Hale WA, Delaney MJ, McGaghie WC. Characteristics and predictors of falls in elderly patients. *J Fam Pract*. 1992;34(5):577–81. [\[PMID: 1578207\]](#)
27. Raiche M, Hebert R, Prince F, Corriveau H. Screening older adults at risk of falling with the Tinetti balance scale. *Lancet*. 2000;356(9234):1001–2. [\[PMID: 11041405\]](#)
28. Lin MR, Hwang HF, Hu MH, Wu HD, Wang YW, Huang FC. Psychometric comparisons of the timed up and go, one-leg stand, functional reach, and Tinetti balance measures in community-dwelling older people. *J Am Geriatr Soc*. 2004;52(8):1343–48. [\[PMID: 15271124\]](#)
29. Morris R, Harwood RH, Baker R, Sahota O, Armstrong S, Masud T. A comparison of different balance tests in the prediction of falls in older women with vertebral fractures: A cohort study. *Age Ageing*. 2007;36(1):78–83. [\[PMID: 17264139\]](#)
30. Okumiya K, Matsubayashi K, Nakamura T, Fujisawa M, Osaki Y, Doi Y, Ozawa T. The timed “up & go” test is a useful predictor of falls in community-dwelling older people. *J Am Geriatr Soc*. 1998;46(7):928–30. [\[PMID: 9670889\]](#)
31. Stel VS, Pluijms SM, Deeg DJ, Smit JH, Bouter M, Lips P. A classification tree for predicting recurrent falling in community-dwelling older persons. *J Am Geriatr Soc*. 2003;51(10):1356–64. [\[PMID: 14511154\]](#)
32. Cannard G. Falling trend. *Nurs Times*. 1996;92(2):36–37. [\[PMID: 8577589\]](#)
33. Nandy S, Parsons S, Cryer C, Underwood M, Rashbrook E, Carter Y, Eldridge S, Close J, Skelton D, Taylor S, Feder G; Falls Prevention Pilot Steering Group. Development and preliminary examination of the predictive validity of the Falls Risk Assessment Tool (FRAT) for use in primary care. *J Public Health (Oxf)*. 2004;26(2):138–43. [\[PMID: 15284315\]](#)
Erratum in: *J Public Health (Oxf)*. 2005;27(1):129–30.
34. Kopke S, Meyer G. The Tinetti test: Babylon in geriatric assessment. *Z Gerontol Geriatr*. 2006;39(4):288–91. [\[PMID: 16900448\]](#)
35. Tinetti ME. Performance-oriented assessment of mobility problems in elderly patients. *J Am Geriatr Soc*. 1986;34(2):119–26. [\[PMID: 3944402\]](#)
36. Whitney JC, Lord SR, Close JC. Streamlining assessment and intervention in a falls clinic using the Timed Up and Go Test and Physiological Profile Assessment. *Age Ageing*. 2005;34(6):567–71. [\[PMID: 16267180\]](#)
37. Bergland A, Laake K. Concurrent and predictive validity of “getting up from lying on the floor.” *Aging Clin Exp Res*. 2005;17(3):181–85. [\[PMID: 16110729\]](#)
38. Vellas BJ, Wayne SJ, Romero L, Baumgartner RN, Rubenstein LZ, Garry PJ. One-leg balance is an important predictor of injurious falls in older persons. *J Am Geriatr Soc*. 1997;45(6):735–38. [\[PMID: 9180669\]](#)
39. Lamb SE, Jørstad-Stein EC, Hauer K, Becker C; Prevention of Falls Network Europe and Outcomes Consensus Group. Development of a common outcome data set for fall injury prevention trials: The Prevention of Falls Network Europe consensus. *J Am Geriatr Soc*. 2005;53(9):1618–22. [\[PMID: 16137297\]](#)

40. Chu LW, Chi I, Chiu AY. Incidence and predictors of falls in the Chinese elderly. *Ann Acad Med Singapore*. 2005; 34(1):60–72. [\[PMID: 15726221\]](#) Erratum in: *Ann Acad Med Singapore*. 2005;34(7):469.
41. Teno J, Kiel DP, Mor V. Multiple stumbles: A risk factor for falls in community-dwelling elderly. A prospective study. *J Am Geriatr Soc*. 1990;38(12):1321–25. [\[PMID: 2254571\]](#)
42. Luukinen H, Koski K, Laippala P, Kivela SL. Predictors for recurrent falls among the home-dwelling elderly. *Scand J Prim Health Care*. 1995;13(4):294–99. [\[PMID: 8693215\]](#)
43. Luukinen H, Koski K, Kivela SL, Laippala P. Social status, life changes, housing conditions, health, functional abilities and life-style as risk factors for recurrent falls among the home-dwelling elderly. *Public Health*. 1996;110(2):115–18. [\[PMID: 8901255\]](#)

Submitted for publication April 24, 2008. Accepted in revised form August 4, 2008.