Diagnostic test accuracy of simple instruments for identifying frailty in community-dwelling older people: a systematic review

ANDREW CLEGG¹, LUKE ROGERS², JOHN YOUNG¹

¹Academic Unit of Elderly Care and Rehabilitation, University of Leeds, Bradford Institute for Health Research, Bradford Teaching Hospitals NHS Foundation Trust, Duckworth Lane, Bradford, West Yorkshire, UK
²Department of Elderly Medicine, Bradford Teaching Hospitals NHS Foundation Trust, Bradford, West Yorkshire, UK

Address correspondence to: Tel: (+44) 01274 383440; Fax: (+44) 01274 382766. A. Clegg. Email: a.p.clegg@leeds.ac.uk

Abstract

Background: frailty is a state of vulnerability to adverse outcomes. Routine identification of frailty is recommended in international guidance. This systematic review investigates the diagnostic test accuracy (DTA) of simple instruments for identifying frailty in community-dwelling older people.

Methods: the review methodology followed Cochrane procedures. Databases were searched from January 1990 to October 2013. Prospective studies assessing the DTA of simple instruments for identifying frailty in community-dwelling older people (aged ≥65 years) as index tests against a reference standard phenotype model, cumulative deficit frailty index or comprehensive geriatric assessment were eligible for inclusion. Sensitivity, specificity, positive predictive value, negative predictive value and likelihood ratios were calculated for index tests. Risk of bias was assessed using the QUADAS-2 checklist.

Results: three studies involving 3,261 participants were included. Median frailty prevalence was 10.5%. Seven index tests were assessed: gait speed, timed-up-and-go test, PRISMA 7 questionnaire, self-reported health, general practitioner clinical assessment, polypharmacy and Groningen Frailty Index. For a gait speed of <0.8 m/s, the sensitivity = 0.99 and specificity = 0.64. For the PRISMA 7, the sensitivity = 0.83 and specificity = 0.83. For the timed get-up-and-go test of 10 s, the sensitivity = 0.93 and specificity = 0.62. DTA was notably lower for all other index tests. All three studies were judged at unclear risk of bias.

Discussion: slow gait speed, PRISMA 7 and the timed get-up-and-go test have high sensitivity for identifying frailty. However, limited specificity implies many false-positive results which means that these instruments cannot be used as accurate single tests to identify frailty.

Keywords: frailty, diagnostic test accuracy, instruments, sensitivity, specificity, older people

Background

Frailty is a state of vulnerability to poor resolution of homeostasis following a stressor event and is independently associated with important adverse outcomes [1]. Routine identification of frailty to enable improved, evidence-based care pathways for older people has been advocated in international consensus guidance [2].

The two established models of frailty are the Cardiovascular Health Study (CHS) phenotype model [3] and the Canadian Study of Health and Aging (CSHA) cumulative deficit model [4] which demonstrate overlap in identification of frailty and statistical convergence [1]. Although validated in large epidemiological studies, and used in research studies, these models of frailty are not yet suitably developed for use in routine care. The usual method of identifying frailty in clinical practice is comprehensive geriatric assessment (CGA), but this is a time- and resource-intensive specialist approach. Ideally, a frailty identification process should be based on a simple test that is quick to use and interpret by non-specialist staff.

Previous systematic reviews of frailty instruments have identified and classified the instruments [5], or investigated
the reproducibility, validity and responsiveness to change [6].

Although this is important for frailty researchers, information on the diagnostic test accuracy (DTA) of instruments for identifying frailty is more useful for clinicians.

The objective for this review was to provide information on the DTA of simple instruments for identifying frailty.

Methods

The full protocol is described on the PROSPERO database (registration number CRD42013005940). The review methodology followed Cochrane DTA systematic review procedures [7].

Criteria for considering studies for this review

Prospective studies assessing the DTA of one or more simple instruments for identifying frailty in community-dwelling older people (index tests) against a reference standard were considered for inclusion.

Participants

Participants were community-dwelling older people, defined for this review as mean age in the study population of 65 years and over. DTA studies of hospitalised patients, people in intermediate care and care home residents were excluded.

Index tests

Simple instruments for identifying frailty included questionnaires and brief assessments and were defined as the index tests for this review.

Reference standards

The CHS phenotype model, CSHA FI and specialist CGA were the reference standard tests. Studies that assessed DTA of one or more index tests against one or more of the three reference standards were considered for inclusion.

Search methods for identification of studies

We searched Medline, EMBASE, CINAHL, Web of Science, Cochrane database of systematic reviews, Cochrane database of abstracts of reviews, AMED, PsychInfo, Scopus and Pedro to identify studies for inclusion. Databases were searched from January 1990 to October 2013.

Selection of studies

Two independent reviewers assessed all titles and full-text reports for inclusion. Disagreements were resolved by consensus.

Data extraction and management

Two independent reviewers extracted all data, and disagreements were resolved by consensus. The reference standard frailty model and index tests, including cut points for diagnosis, were recorded. Data on true-positive (TP), true-negative (TN), false-positive (FP) and false-negative (FN) rates were extracted or calculated from reported data using RevMan 5.2 software.

Assessment of methodological quality

Two independent reviewers independently assessed methodological quality using the QUADAS-2 tool [8].

Statistical analysis and data synthesis

Sensitivity and specificity, with 95% confidence intervals, were calculated for each index test by constructing 2 × 2 tables for TP, TN, FP and FN rates using RevMan 5.2 software. A summary receiver operating characteristic (SROC) curve comparing all included index tests was plotted.

Other assessments of test accuracy that provide additional information for practitioners were also calculated, including positive predictive validity (PPV), negative predictive validity (NPV), positive likelihood ratio (LR+) and negative likelihood ratio (LR−). PPVs, NPVs, LR+ and LR− values were estimated from TP, TN, FP and FN data.

Table 1. Summary of diagnostic test accuracy results for seven simple instruments for identifying frailty

<table>
<thead>
<tr>
<th>Index test (units)</th>
<th>Cut-off</th>
<th>Reference standard</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Positive predictive value</th>
<th>Negative predictive value</th>
<th>Positive likelihood ratio</th>
<th>Negative likelihood ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gait speed (m/s)</td>
<td>&lt;0.7</td>
<td>Phenotype model</td>
<td>0.93</td>
<td>0.77</td>
<td>0.35</td>
<td>0.98</td>
<td>4.19</td>
<td>0.09</td>
</tr>
<tr>
<td>Gait speed (m/s)</td>
<td>&lt;0.8</td>
<td>Phenotype model</td>
<td>0.99</td>
<td>0.64</td>
<td>0.26</td>
<td>0.99</td>
<td>2.80</td>
<td>0.01</td>
</tr>
<tr>
<td>Gait speed (m/s)</td>
<td>&lt;0.9</td>
<td>Phenotype model</td>
<td>1.00</td>
<td>0.56</td>
<td>0.22</td>
<td>1.00</td>
<td>2.28</td>
<td>0.00</td>
</tr>
<tr>
<td>PRISMA 7</td>
<td>≥3</td>
<td>Phenotype model</td>
<td>0.83</td>
<td>0.83</td>
<td>0.40</td>
<td>0.97</td>
<td>5.00</td>
<td>0.20</td>
</tr>
<tr>
<td>Timed-up-and-go test (s)</td>
<td>&gt;10</td>
<td>Phenotype model</td>
<td>0.93</td>
<td>0.62</td>
<td>0.17</td>
<td>0.99</td>
<td>2.46</td>
<td>0.11</td>
</tr>
<tr>
<td>Self-rated health</td>
<td>≤6</td>
<td>Phenotype model</td>
<td>0.83</td>
<td>0.72</td>
<td>0.29</td>
<td>0.97</td>
<td>3.00</td>
<td>0.23</td>
</tr>
<tr>
<td>General Practitioner assessment</td>
<td>Dichotomous</td>
<td>Phenotype model</td>
<td>0.67</td>
<td>0.76</td>
<td>0.28</td>
<td>0.94</td>
<td>2.86</td>
<td>0.43</td>
</tr>
<tr>
<td>Polypharmacy</td>
<td>≥5 medications</td>
<td>Phenotype model</td>
<td>0.67</td>
<td>0.72</td>
<td>0.24</td>
<td>0.94</td>
<td>2.40</td>
<td>0.46</td>
</tr>
<tr>
<td>Groningen Frailty Indicator</td>
<td>≥4</td>
<td>Phenotype model</td>
<td>0.58</td>
<td>0.72</td>
<td>0.22</td>
<td>0.93</td>
<td>2.10</td>
<td>0.58</td>
</tr>
</tbody>
</table>
Results of the search

The review process is summarised in Figure 1 using the PRISMA guidelines [9].

The searches identified 5,677 citations. Of these, we considered 19 as potentially relevant and retrieved the full-text articles. Sixteen articles were excluded: 10 did not apply a reference standard test [10–19], and 6 did not report DTA data [20–25].

Three studies involving 3,261 participants were included and are summarised in Supplementary data, Appendix 1 available in Age and Ageing online [26–28]. Seven simple instruments were assessed as index tests. The reported mean age in study participants was 74.7 years (range: 70.0–78.6 years), 47.5% were male (range: 43.1–49%), and the median frailty prevalence was 10.5% (range: 7.7–11.6%).

The first study assessed the accuracy of gait speed, measured in m/s over a 4 m distance, using a range of cut-off values compared with the phenotype model as the reference standard [26].

The second study assessed the timed-up-and-go test (TUGT) at a range of cut-off values compared with the phenotype model as the reference standard [27]. The TUGT measures, in seconds, the time taken to stand up from a standard chair, walk a distance of 3 m, turn, walk back to the chair and sit down.

The third study assessed the accuracy of five index tests compared with both the phenotype model and specialist CGA as reference standards [28]. The first index test was clinical judgment in which a general practitioner made a judgment based on the question ‘Would you consider this patient to be frail if frailty is defined as a loss of resources in several domains of functioning (physical, psychological, social), increasing the risk of adverse outcomes?’ The second index test was polypharmacy defined as five or more medications. The third was the Groningen Frailty Indicator (GFI) which is a 15-item questionnaire that is suitable for postal completion [29]. A score of ≥4 indicates the possible presence of moderate to severe frailty. The fourth was the PRISMA 7 tool which is a 7-item questionnaire to identify disability that is also suitable for postal completion [30]. A score of ≥3 is considered to identify frailty. The fifth index test was self-rated health assessed using the question ‘How would you rate your health on a scale of 0–10?’. A cut-off value of ≤6 was used to identify frailty.

Findings

The main findings are summarised in Table 1 and the summary forest plot (Supplementary data, Appendix 2 available in Age and Ageing online). The DTA of all seven instruments is further compared in the SROC curve (Supplementary data, Appendix 3 available in Age and Ageing online).

Gait speed, using a cut-off value of <0.8 m/s, has high sensitivity (0.99) and moderate specificity (0.64) for identifying frailty, with narrow confidence limits. However, although NPV is high at 0.99, PPV is relatively low with a value of 0.26. Similar results are obtained when a cut-off value of <0.9 m/s is used. A cut-off value of <0.7 m/s has lower sensitivity (0.93) but higher specificity (0.78).

The PRISMA 7 tool has relatively high sensitivity (0.83) and specificity (0.83) for identifying frailty, but wide confidence limits indicate considerable uncertainty. A TUGT score of >10 s has relatively high sensitivity (0.93) and specificity (0.62) for identifying frailty.

The test accuracy was lower for self-reported health (sensitivity = 0.83; specificity = 0.72), general practitioner clinical assessment (sensitivity = 0.67; specificity = 0.77), polypharmacy (sensitivity = 0.67; specificity = 0.72) and GFI (sensitivity = 0.58; specificity = 0.72). Confidence intervals were notably wide for these instruments.

Methodological quality of included studies

All three studies were judged at unclear risk of methodological bias.

Discussion

An ideal diagnostic test has high sensitivity and high specificity for detecting the condition of interest. High sensitivity means that there will be few false-negative results. High specificity means that there will be few false-positive results.

Gait speeds of ≤0.9, ≤0.8 and ≤0.7 m/s; the PRISMA 7 tool and the TUGT demonstrate high sensitivities and
DTA of simple instruments for identifying frailty in older people

moderate specificities for identifying frailty, whereas self-reported health, general practitioner clinical assessment, polypharmacy and GFI are notably less accurate.

A gait speed of \( \leq 0.8 \text{ m/s} \) (taking longer than 5 s to walk 4 m) has a sensitivity of 0.99 for identifying frailty which means that virtually all older people with frailty would be expected to have a positive test. Furthermore, the high NPV of 0.99 indicates that older people with a gait speed >0.8 m/s are extremely unlikely to have frailty. Interestingly, a gait speed of \( \leq 0.8 \text{ m/s} \) is also the chosen cut-off in the European consensus definition of sarcopenia, which is considered a key component of frailty [31]. However, the corresponding sensitivity of 0.64 and PPV of 0.26 mean that only one person in four who have a gait speed of \( \leq 0.8 \text{ m/s} \) will be frail on the basis of a reference standard. Applying these rules to the other instruments identifies a relatively high proportion of error rates, but particularly so for self-reported health, general practitioner clinical assessment, polypharmacy and GFI. This means that the utility of any of the instruments as a single test for identifying frailty in routine care is limited.

It is possible to improve diagnostic accuracy through two approaches. Firstly, predictive values are proportionate to population prevalence, so applying the test in an older population, in which the prevalence of frailty is likely to be higher, will potentially improve accuracy. Secondly, a two-stage assessment process could be used, whereby a test with high sensitivity is followed by a reference standard test, or second index test with high specificity, to identify frailty with greater accuracy. However, how this might be achieved reliably in routine care would need to be investigated in a future research study.

**Strengths of the review**

This review used a comprehensive search strategy and followed Cochrane procedures. Two independent reviewers screened all potential studies for inclusion and extracted data, reducing potential risk of bias. Reported frailty prevalence estimates for each included study were similar to those provided in a recent systematic review [32]. It is therefore likely that the DTA estimates are broadly generalisable to older people in community settings.

**Weaknesses of the review**

The review identified only three studies testing seven index tests. Five of the instruments were tested in a relatively small number of participants, resulting in considerable uncertainty regarding diagnostic accuracy.

A DTA approach uses binary cut points to identify frailty. However, frailty may be better viewed as a continuous health state rather than present or absent. This has proved possible with using the cumulative deficit frailty model [4] and gait speed [33]. A meta-analysis of nine cohort studies investigating gait speed identified a pooled mortality hazard ratio of 0.88 (95% CI 0.87 to 0.90) for each 0.1 m/s increment in gait speed [33].

**Conclusions**

Slow gait speed, PRISMA 7 and TUGT all have high sensitivity but limited specificity as simple instruments for identifying frailty. This means that there are many false-positive test results which limit their DTA.

Use of these tools in older populations with higher baseline prevalence of frailty is likely to improve test accuracy. Use of a simple instrument with a high sensitivity followed by either a reference standard test or second simple instrument in a two-stage approach to diagnosis would potentially improve accuracy but requires further investigation.

**Key points**

- Routine identification of frailty has been advocated in international consensus guidance.
- The frailty phenotype model, cumulative deficit frailty index and CGA are time consuming.
- This systematic review reports evidence on the accuracy of simple frailty instruments.
- Slow gait speed, PRISMA 7 questionnaire and TUGT have high sensitivity but limited specificity for identifying frailty.
- This means that these instruments cannot be used as accurate single tests to identify frailty.

**Conflicts of interest**

None declared.

**Supplementary data**

Supplementary data mentioned in the text are available to subscribers in *Age and Ageing* online.

**References**

A. Clegg et al.


Received 10 January 2014; accepted in revised form 1 July 2014