TITLE PAGE: Geriatrics and Gerontology International submission

Full title:

Diagnostic test accuracy of self-reported screening instruments in identifying frailty in community-dwelling older people: A systematic review

Short running title:

Frailty screening instrument accuracy

Authors:

Rachel C. Ambagtsheer[1,2], Mark Q. Thompson[1], Mandy M. Archibald[1,3], Mavourneen G. Casey[2] and Timothy J. Schultz[1,4]

Affiliations:

1 National Health and Medical Research Council Centre of Research Excellence in Trans-Disciplinary Frailty Research to Achieve Healthy Ageing, Adelaide, Australia
2 Torrens University Australia, Adelaide, Australia
3 College of Nursing and Health Sciences, Flinders University, Adelaide, Australia
4 The Centre for Evidence-based Practice South Australia (CEPSA): a Joanna Briggs Institute Centre of Excellence, Adelaide, Australia

Corresponding author:

Rachel Ambagtsheer, Torrens University Australia, GPO Box 2025, Adelaide SA

5000 Australia. Telephone: (+61) 425 177 110.

Email: rambagtsheer@laureate.net.au.

Abstract

AIM:

Against a backdrop of ageing populations worldwide, it has become increasingly important to identify frailty screening instruments suitable for community settings. Self-reported and/or administered instruments may offer significant simplicity and efficiency advantages over clinician-administered instruments but their comparative diagnostic test accuracy has yet to be systematically examined.

The aim of this systematic review was to determine the diagnostic test accuracy of self-reported and/or self-administered frailty screening instruments against two widely accepted frailty reference standards (the Frailty Phenotype and the Frailty Index) within community-dwelling older adult populations.

METHODS:

We conducted a systematic search of the Embase, CINAHL, MEDLINE, PubMed, Web of Science, PEDro, PsycINFO, ProQuest Dissertations, Open Grey and GreyLit databases up to April 2017 (with an updated search conducted over May-July 2018) to identify studies reporting comparison of self-reported and/or self-administered frailty screening instruments against an appropriate reference standard, with a minimum sensitivity threshold of 80% and specificity threshold of 60%.

RESULTS:

We identified 24 studies that met our selection criteria. Four self-reported screening instruments across three studies met minimum sensitivity and specificity thresholds. However, in most cases, study design considerations limited the reliability and generalisability of the results. Additionally, meta-analysis was not conducted because no more than three studies were available for any of the unique combinations of index tests and reference standards.

CONCLUSIONS:

Although our study has demonstrated that a number of self-reported frailty screening instruments reported sensitivity and specificity within a desirable range for community application, additional diagnostic test accuracy studies are needed.

Keywords: (MESH): Aged, 80 and over; Frailty; Geriatric Assessment; Primary Health Care

1. Introduction

Frailty has been identified as a global public health priority in societies with ageing populations (1). Frailty is a state of decreased physiological reserve and increased vulnerability to stressor events, resulting in increased risk of adverse health outcomes such disability, hospitalisation, institutionalisation, and mortality (2–4). Frailty is a dynamic condition where improvement is possible (5,6) and interventions exist that can delay or reverse frailty (7). Identifying individuals who would benefit from timely identification and intervention, therefore, is a key priority in the management of frailty within the community (8,9).

There are currently a large number of different frailty screening instruments in existence, many of which have proven to be reliable and valid measures of frailty within different contexts (10). However, only some of these are suitable contenders to be considered within the scope of self-administered instruments, taking the form of either a postal survey or a self-completed questionnaire. Several systematic reviews have examined the utility of frailty screening within community settings, from the perspective of both self-report and test-based measurement (4,9,11–13).

Two systematic reviews have reported on the diagnostic test accuracy (DTA) of frailty screening instruments against a reference standard (14,15). A number of publications have examined the DTA of self-reported instruments for the identification of frailty since these reviews were published, hence the need for this review. One of the key complexities identified in these reviews regarding frailty screening is a lack of consensus on a definition of frailty, which is reflected in two separate reference standards and a large number of potential index tests (9).

The aim of this review was to identify the DTA of self-reported screening instruments against a frailty reference standard for community dwelling older adults. Specifically, our review questions were:

- How accurate are self-reported screening instruments against agreed reference standards?
- How does the accuracy of self-reported instruments vary according to whether the test is self-reported or self-administered?

2. Materials and Methods

We consulted the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement (16) and the JBI Critical Appraisal Checklist for Diagnostic Test Accuracy Studies tool (17) in developing the study design. The study protocol has been published previously (18) and the study is registered with both the PROSPERO (ID=CRD42017081379) and the JBI databases.

We diverged from our original protocol with respect to the following points:

- We focused on frailty alone rather than including pre-frailty for purposes of clarity.
- We excluded CGA (Comprehensive Geriatric Assessment) as a reference standard for frailty due to a lack of standardisation in terms of its administration between studies and the absence of a widely recognised threshold for frailty. We have, however included studies where FI was derived from CGA.
- We excluded studies conducted in hospital settings and emergency departments along with patients with specific conditions, e.g. cancer.
- We included studies from inception of the databases rather than studies published after Jan 1, 2001.
- Due to significant overlap between JBI and QUADAS and time limitations we critically appraised against the JBI framework only.
- Meta-analysis and sub-group analysis were not conducted due to substantial heterogeneity in the results.
- We excluded consideration of feasibility from our review given that our search strategy, which is focused on diagnostic accuracy, is likely to have omitted a high proportion of papers focused specifically on the feasibility of individual instruments.

2.1 Selection Criteria

2.1.1. Types of Studies

We included observational studies published in English and conducted in community settings.

2.1.2 Participants

The participants in our included studies were community-dwelling older adults with a minimum mean age of 65 years, or where at least half of the study participants were aged 65 years and over. Studies of participants living in residential care settings were excluded. Studies that addressed a specific diagnosis or that were conducted in an acute setting (e.g. cancer patients, surgical patients or emergency department patients) were excluded.

2.1.3 Index Tests

Any index test purporting to measure frailty that was entirely self-reported (i.e., administered by an investigator but including no clinical or physical measurements) or that was self-administered was included. Tests that were partially self-reported were excluded unless test results for the self-reported items were presented independently of the non-self-reported items. We included

studies in which the self-report frailty instrument was completed by a proxy as well as studies where the older person self-completed the instrument. Studies using a self-reported FI were excluded as any FI (self-report, test-based, or combination) which meets the criteria of Searle et at (2008) may be considered to be a reference standard.

2.1.4 Reference Standards

Studies were included if they applied either of two frailty reference standards: the Frailty Phenotype (FP) (Fried et al., 2001) or the Frailty Index (Mitnitski et al., 2001). Studies applying no reference standard or a reference standard other than those specified above were excluded.

2.1.5 Diagnosis of interest

The diagnosis of interest was frailty.

2.2. Search Strategy

To identify published studies, we searched the databases MEDLINE/PubMed, PEDro, Embase, PsycINFO, CINAHL, Scopus, and Web of Science from inception. The initial search was conducted between March and April of 2017 and updated in July 2018. Our search strategy was developed in consultation with an academic librarian with a speciality in medicine.

The search strategy was developed in a scaffolded manner, commencing with a CINAHL and PubMed search to inform specified keyword analysis, including MeSH terms, for subsequent database searching. We then used truncated and expanded keyword variations of the terms relating to frailty, self-report, and screening, along with specific self-report screening tools (e.g., Kihon Checklist).

To identify unpublished and grey literature, we searched ProQuest, Open Grey, The Grey Literature Report database, and consulted websites of key gerontological research centres with a focus on frailty.

We also reviewed the reference lists of all included studies to identify additional studies of interest.

The search strategy syntax for individual databases is provided in Supporting Table 1.

2.3 Study selection

All studies of interest were exported from the respective databases and imported into Zotero Reference Manager version 4.0.29.17. Zotero was selected for this

process due to its compatibility with the various data extraction formats from the electronic databases, its low cost and an internal capability that made it relatively easy to identify and remove duplicates. Duplicates were identified using the inbuilt feature and manually checked before deletion by one researcher (RA).

The resulting unique records were exported into a Microsoft Excel 2016 worksheet before being assessed for title and abstract relevancy by two independent reviewers (RA and MT). We consulted a third reviewer (TS) over any differences in opinion regarding the inclusion of individual articles.

Agreed articles were extracted in full text format by one reviewer (RA) and reviewed independently by the same reviewers (RA and MT) with recourse to the third reviewer (TS) to achieve consensus. The reason for exclusion was retained for all articles reviewed after the initial title/abstract screen.

2.4 Quality review

Two reviewers (RA and MT) subjected the included full-text articles to an initial assessment against the inclusion criteria, and a number of studies were excluded at this point. The reasons for exclusion were retained.

Included studies were assessed against the Joanna Briggs Institute (JBI) Checklist for Diagnostic Test Accuracy Studies (17). This Checklist is based on the QUADAS 2 signaling questions (19) and relate to the design and conduct of the study (see Box 1).

Two reviewers (RA and MT) then assessed included studies independently against the JBI criteria. We recorded a value of "Yes", "No" or "Unclear" against each criterion and made additional notes where appropriate.

Box 1: JBI Criteria

1. RECRUITMENT: Was a consecutive or random sample of patients enrolled?

2. CASE CONTROL: Was a case control design avoided?

3. EXCLUSIONS: Did the study avoid inappropriate exclusions?

4. INDEX TEST INTERPRETATION: Were the index test results interpreted without knowledge of the results of the reference standard?

5. THRESHOLD: If a threshold was used, was it pre-specified?

6. REFERENCE STANDARD: Is the reference standard likely to correctly classify the target condition?

7. REFERENCE STANDARD INTERPRETATION: Were the reference standard results interpreted without knowledge of the results of the index test?

8. INTERVAL: Was there an appropriate interval between index test and reference standard?

9. SAME REFERENCE STANDARD: Did all patients receive the same reference standard?

10. ALL PATIENTS: Were all patients included in the analysis?

2.5 Data extraction

We developed an initial data extraction template, which was based on the JBI Data Extraction tool (17) and was finalised in consultation with the research team. We extracted country of origin, sample size, mean age of participants, % male/female, index test and reference standard, test thresholds, and self-report or self-administration status. Data was extracted independently by two reviewers (RA and MT) and then discussed to achieve consensus. Where necessary, we sought additional clarification or data from study authors, especially with respect to 2x2 data to allow calculation of diagnostic characteristics.

2.6 Data analysis

We used Microsoft Excel and SPSS v.25 to perform descriptive analysis of the data. We calculated sensitivity (Se), specificity (Sp), positive predictive value (PPV), negative predictive value (NPV) and likelihood ratios along with their associated 95% confidence intervals within the Review Manager (RevMan) v.5.3 software (The Nordic Cochrane Centre, Cochrane Collaboration, Copenhagen). We used Revman to construct forest plots of the data to display heterogeneity.

For the purposes of this study, we adopted 80% as a minimum sensitivity threshold (9,20,21) and 60% as a minimum specificity threshold (20) as being acceptable. Although good screening tests have both high sensitivity and specificity (i.e., results close to 1), in practice there is often a trade-off in favour of one over the other (22). In the case of frailty screening within community settings, high sensitivity and low specificity tends to be the more preferable scenario where identification of as many frail individuals as possible is a priority (23). Following this logic, it may be more desirable to have a higher number of false positives than false negatives, signifying that there is a greater chance of potentially identifying as frail people who are not frail rather than missing those who are.

Additionally, we used Microsoft Excel to construct the Youden Index (sensitivity + specificity-1). The Youden Index is a single summary measure reflecting how closely the DTA result matches the ideal of no false positives or false negatives (24). Although the Youden Index adds additional interpretive information (as a

kind of balancing mechanism across the sensitivity and specificity values), it is not appropriate to consider it in isolation, especially in the context of our study. For example, the Youden Index, as a single statistic, does not take into consideration decisions such as prioritisation of high sensitivity over high specificity into account.

The 'metandi' command in Stata requires a minimum of four studies for meta-analysis of DTA studies (25). However, this review did not identify more than three studies comparing the same index test and reference standard. Consequently, we did not conduct meta-analysis and have synthesised results in tabular and narrative formats.

3. Results

All results presented below are descriptive in nature. Across all possible comparisons of index and reference tests, the highest number of included studies per comparison was only three. In this single instance, a comparison of Tilburg Frailty Indicator (TFI) (index test) versus FP, the bivariate model was not calculated as the number of studies was below the minimum required (n=4). Therefore, as meta-analysis was not possible, all comparisons are presented in tabular and graphical format, along with a narrative synthesis.

3.1 Search results and study characteristics

The search results (including the updated search) returned 18,034 results in total. The large number of results returned was due to a number of factors including conducting the search from the inception of the databases, the large number of databases searched, extensive duplication between databases and the need to allow for numerous combinations of the search term "self" in relation to screening. Study results are shown in the flow diagram in Figure 1. Of the total records identified, 10,579 duplicates were identified and removed, leaving 7,455 records to be screened by title and abstract. Initial agreement between reviewers was 77.8%; where assessments differed, these were resolved through discussion or occasionally referred to the third reviewer.

After title and abstract screening, 7,164 articles were excluded due to lack of relevance or because they could not be sourced in full text format. In all, 291 studies were assessed in full-text format against the inclusion/exclusion criteria. Of these, 267 were subsequently excluded. Key reasons for exclusion included not having or an inappropriate reference standard applied (28.1%), focus on a specific disease (18.1%), being a non-DTA or observational study (16.9%), using non-self-reported instruments (9.4%) and not meeting the age criteria (7.9%). Ultimately, 24 studies were deemed to have met the inclusion/exclusion criteria and underwent JBI critical appraisal.

<Insert Figure 1 about here please>

3.2 Included studies

Across the 24 included studies, the sample size ranged from 52 to 27,527, with a total of 84,984 participants. Characteristics of included studies are summarised in Table 1. The mean age of participants, where stated, ranged from 65.3 years up to 85.7 years.

In all, there were 31 instances of screening instruments compared against a frailty reference standard across the included studies, with some studies having more than one combination of index test and reference standard. The most frequently implemented instruments were the FRAIL scale (22.6% of instances), the Groningen Frailty Indicator (12.9%) and Self-Reported Health (12.9%), representing a mix of multi-dimensional and uni-dimensional indicators.

Almost two-thirds (64.5%) of all comparisons were made against the FP alone as a reference standard, 19.4% against both the FP and the FI, and 16.1% against the FI alone.

The most common mode of administration was self-reporting to an external interviewer (54.8% of instances). Approximately one-quarter of cases involved self-administration of the instrument (25.8%). Lastly, 19.4% of instances did not specify mode of administration.

3.3 Methodological quality

We assessed the 24 included articles against the JBI quality criteria. The results of the methodological quality assessment against the JBI criteria for included studies are shown in Table 2. Methodological quality of studies ranged from a low of 30% up to 90%, with almost 90% of included studies meeting ≥ 50% of quality criteria. The JBI criteria were designed for DTA-specific studies, however in this review we have also applied them to population studies. Quality should therefore be interpreted in light of the study design. We encourage readers to refer to Table 1 for context-specific information regarding study setting and design when interpreting the DTA results.

3.4 Test accuracy

We were able to obtain sufficient data to calculate diagnostic test accuracy from 14 of the included studies (Table 3, Supporting Table 2, Supporting Table 3, Supporting Figure 1). Sensitivity and specificity of frailty screening instruments against the two reference standards varied widely between included studies that provided DTA data (Table 3).

Self-reported screening instruments meeting the minimum sensitivity and specificity thresholds included the PRISMA-7 against the FP (Two studies; Se:100.0%, Sp: 80.0% (26) and Se:93.3%, Sp: 78.2% (27)), the GFI against the FP (Se:100.0%, Sp: 80.0% (26)); Self-Rated Health against the FP (Se: 85%, Sp:73% (27)); and Self-Reported Physical Activity against the FP (Se:80.6%, Sp:84.2% (28)) (Table 4). All instruments scoring high sensitivity also returned a Youden index value above 0.5.

Of the results reported above, most (5 of 6) were based on self-reported rather than self-administered instruments. Only the GFI against the FP (Se:100.0%, Sp: 80.0%) as reported in Braun et al's study (26) was self-administered. However, as the sample size was small (n=52), this result should be interpreted with caution.

4. Discussion

4.1 Summary of main findings

We did not find strong and reliable evidence in support of the DTA of self-reported instruments for the identification of frailty included within our study. Two candidates (the PRISMA-7 and the Groningen Frailty Indicator) developed specifically for the identification of frailty met our minimum sensitivity and specificity requirements against the Frailty Phenotype (26,27), as did Self-Reported Health (Hoogendijk et al., 2012). However, the studies from which they were drawn were characterised by wide confidence intervals and relatively small sample sizes, and in the case of one study (Hoogendijk et al., 2012), a higher prevalence of frailty than would be expected to be found within the community due to study design. Only Self-Reported Physical Activity (28), not a frailty screening instrument per se, but rather a single self-reported criterion of the FP, simultaneously met our sensitivity and specificity criteria and was based on a relatively large sample size (n=4000). However, none of the studies described above met the JBI criterion for random or consecutive recruitment and/or were deliberately structured to achieve an over-representation of frail individuals, limiting their reliability and generalisability.

Our review found very high heterogeneity with respect to DTA, study design, index test, and reference standard between the included studies. This is an important consideration because study methodology, sample size, setting, and selection of both index test and reference standard are all likely to have influenced DTA results (Leeflang et al., 2008). Very few studies compared the same index test and reference standard, making a meta-analysis statistically unviable. In this respect, our results were consistent with other studies in finding substantial variability for the DTA of frailty screening instruments (11,14,15).

In the included studies the reference standard was commonly modified, which complicates interpretation of our results. The majority of studies which used the

FP as a reference standard in this review included variables which differed from the original formulation specified by Fried and colleagues (49). The FP is commonly modified due to the availability of variables or ease of data collection across studies, and the modification of variables has implications for frailty prevalence (50), and thus also DTA findings. Furthermore, we observed considerable variation in terms of the threshold for frailty used for the FI across a number of studies, which ranged from 0.08 up to 0.35. These factors impacted on our ability to make definitive recommendations regarding the DTA of various self-reported instruments in their ability to identify frailty, and this also impacts health care providers and policy makers in determining whether a particular instrument is sufficiently accurate to apply for population level or clinic in frailty screening.

No conclusive statement can be made regarding the accuracy of self-reported versus self-administered tests. In this review, the screening tests were interviewer-administered in the majority of instances (54.8%), a much smaller proportion were self-administered (25.8%), and 19.4% had an unknown mode of administration. In a number of studies, screening tests were either interviewer or self-administered concurrent to a broader assessment which included a range of questions covering geriatric syndromes, health conditions and disability. Different methods of questionnaire administration have been identified as impacting the quality of data collected, with differences most marked between interview and self-administration (51). Factors such as cognitive burden of questionnaires, control over pace of interview, rapport between interviewer and respondent, and social desirability bias are recognised as contributing to potential differences in responses (51). Therefore, the accuracy of using self-reported but not self-administered frailty screening tools within community-based frailty screening, particularly in populations with low levels of education or literacy, is unclear.

4.2 Strengths and limitations

The key strength of this review is to deliver the first (to our knowledge) comprehensive appraisal of the DTA of self-reported and/or self-administered screening instruments for the identification of frailty. Consequently, we anticipate that the data presented within this review will be particularly relevant to those seeking to implement surveys conducted via post, online or in waiting-room type environments. It has also greatly expanded the number of studies and instruments included in previous reviews. Further, we have calculated and reported comprehensive DTA statistics for each of the included studies. For a number of studies, this data was not reported in the original publications and has been sourced through direct communication with authors.

Our review has also identified a number of novel measures that might be further explored in future DTA studies of instruments for the identification of frailty; for example, the inclusion of instruments not specific to frailty such as self-reported

health and self-reported physical exhaustion, both of which returned higher sensitivity results than some of the instruments designed specifically to identify frailty. In addition, our study makes a range of new contextual and diagnostic information available, including the Youden Index, which is of potential clinical relevance in making decisions about screening.

There were a number of limitations associated with our study. The most significant of these is the heterogeneity characterising our included studies (particularly with regard to the Frailty Index, where multiple thresholds have been applied), making interpretation challenging. However, in the absence of consensus within the field on many aspects of frailty screening, we believe that it remains important to present the full range of results so that policy makers and practitioners can come to their own conclusions about the appropriateness of various instruments based on their intended context of use. Additionally, we acknowledge that a number of the included studies were not explicitly designed as DTA studies but, rather, may have been designed for another purpose, such as population-level cohorts. A further limitation is that we did not include self-reported FIs as index tests within our study. Despite the fact that a self-reported FI can be used for frailty screening, it also meets the criteria of being a reference standard for frailty (52), and hence was outside the scope of this review. Lastly, although we focused this review on older adults aged 65 years and over, we acknowledge the possibility that a potential source of the heterogeneity we observed in the results may be due to differences in functional ability between younger and older age groups within this cohort.

In order to focus the review, we deliberately excluded studies focusing on certain populations (cancer and surgical patients) and settings (residential care, acute care, and emergency departments). Therefore, our findings do not extend beyond community settings. Further, studies that focused on the feasibility of tool administration and predictive, rather than diagnostic, accuracy were outside the scope of this review. Where the DTA information on the accuracy of screening instruments is limited, it may be optimal to also consider the predictive accuracy of these instruments (mortality, hospitalisation, institutionalisation, etc.) before implementing them at a population level or in a clinical setting. The ability of self-reported instruments to predict adverse outcomes is an important feature of screening instruments that should be considered where DTA findings are inconclusive. Lastly, despite the comprehensiveness of our search strategy, it may be possible that we have inadvertently omitted studies that were relevant to our review.

4.3 Implications

There are a number of implications for frailty screening using self-reported or self-administered instruments that can be drawn from our results. Firstly, the relatively low accuracy of many of the formal screening instruments currently in wide use potentially restricts the field of choice; however, this decision is largely dependent on the purpose for screening. The appropriate sensitivity for a self-reported instrument in the identification of frailty may vary according to the context in which such an instrument is used. While a lower sensitivity may be appropriate in a primary care setting where follow-up investigation can more readily occur, the presence of a large number of false positive results may be problematic in larger scale population level screening of frailty. The ethics of frailty screening require follow-up consultation with a health professional in the event of a positive result (53), therefore, the DTA of a screening instrument has implications in terms of health resource utilisation. Other outcomes, such as predictive validity, may need to be considered alongside DTA when considering a self-report instrument for the identification of frailty.

Alternatively, another option is to consider whether a self-reported reference standard could be practically applied in preference to a frailty screening instrument. Frailty screening instruments are commonly developed as alternatives to reference standards for purposes of efficiency, but this may come at the expense of clinical relevance and accuracy (54). A fully self-reported FI has been identified as having similar characteristics to a test-based FI (50). Requiring a minimum of 30 variables, this could be considered as a viable alternative. Otherwise, in contexts where self-reporting is not possible, and where equipment and space allow, the Frailty Phenotype, a combined test-based and self-report FI or a non-self-reported instrument (such as Gait Speed) meeting accuracy requirements may be able to be administered.

Regardless, any decision on a frailty screening instrument depends on the purpose for selecting the instrument, and which approach to frailty best fits the requirements of the health organisation or practitioner recommending the test. It is widely recognised within the frailty literature that the FP and FI approaches to measuring frailty are essentially different (although complementary) (55). For example, the FP has been proposed as more amenable to a ‘first contact’ with an individual, as it relies on general signs and symptoms, and does not readily signify what should be done by way of therapeutic follow-up. In contrast, the FI, as a multi-dimensional frailty assessment (often based on a comprehensive assessment), can indicate where clinicians or health service providers might need to focus their intervention (55). This difference can influence the motivation for selecting a screening instrument (10); for example, if the instrument is to be used within a large, population-based study, further investigation will be a priority. Conversely, if it is to be used for diagnostic screening and assessment, for example within a primary care context, the ability to intervene based on information collected will be important.

Although the body of literature on frailty is expanding rapidly (56), it appears that there remains an insufficient number of high-quality, sufficiently-powered DTA

studies to enable meaningful conclusions to be drawn about the performance of individual frailty screening instruments. More studies are needed examining the DTA of self-reported frailty screening instruments for the identification of frailty in community settings. In this review, we have combined large population-based studies and smaller clinical studies of community-dwelling participants. The justification for this choice was to maximise the available evidence on screening instruments and centralise the results to inform research and practice. However, as the frailty evidence base grows, it may be useful to narrow the inclusion criteria to DTA studies only in a future review.

5. Conclusion

This study has identified several self-reported instruments with potential for application within community settings. However, despite four screening instruments across three studies reporting sensitivity and specificity within a desirable range, diagnostic accuracy was clouded by study design and sampling issues, in particular participant selection. The current evidence for the DTA of many screening instruments does not support their widespread use to identify frailty in community dwelling adults. Predictive validity, which was outside the scope of this review, may be an alternative outcome to inform health policy and practice decision-making regarding instrument selection for this population. Further well-designed DTA studies of self-reported screening instruments to identify frailty are required.

Acknowledgements

Our thanks to Michael Draper, Research Librarian, University of Adelaide, for assistance with developing the search strategy.

This work was supported by the National Health and Medical Research Council of Australia (NHMRC) via funding provided for the Centre of Research Excellence in Frailty and Healthy Ageing (#1102208). The sponsors had no role in the design or preparation of the paper.

MA acknowledges the fellowship support provided by the Canadian Institutes of Health Research and the NHMRC. This research was supported by an Australian Government Research Training Program Scholarship to MT. This review will contribute towards the requirements towards a PhD through Torrens University (RA) and the University of Adelaide (MT).

Disclosure Statement

No potential conflicts of interest were disclosed.

References

1. Cesari M, Prince M, Thiyagarajan JA, De Carvalho IA, Bernabei R, Chan P, et al. Frailty: An Emerging Public Health Priority. J Am Med Dir Assoc [Internet]. 2016;17(3):188–92. Available from: http://dx.doi.org/10.1016/j.jamda.2015.12.016

2. Clegg A, Young J, Iliffe S, Rikkert MO, Rockwood K. Frailty in elderly people. Lancet. 2013;381(9868):752–62.

3. Shamliyan T, Talley KMC, Ramakrishnan R, Kane RL. Association of frailty with survival : A systematic literature review. Ageing Res Rev [Internet]. 2017;12(2):719–36. Available from: http://dx.doi.org/10.1016/j.arr.2012.03.001

4. Vermeulen J, Neyens JCL, van Rossum E, Spreeuwenberg MD, de Witte LP. Predicting ADL disability in community-dwelling elderly people using physical frailty indicators: a systematic review. BMC Geriatr. 2011;11:33.

5. Espinoza SE, Jung I, Hazuda H. Frailty transitions in the San Antonio longitudinal study of aging. J Am Geriatr Soc. 2012;60(4):652–60.

6. Gill TM, Gahbauer EA, Allore HG, Han L. Transitions Between Frailty States Among Community-Living Older Persons. Arch Intern Med. 2006;166:418–23.

7. Puts MTE, Toubasi S, Andrew MK, Ashe MC, Ploeg J, Atkinson E, et al. Interventions to prevent or reduce the level of frailty in community-dwelling older adults: A scoping review of the literature and international policies. Age Ageing. 2017;46(3):383–92.

8. Clegg A, Bates C, Young J, Ryan R, Nichols L, Ann Teale E, et al. Development and validation of an electronic frailty index using routine primary care electronic health record data. Age Ageing [Internet]. 2016;45(3):353–60. Available from: http://ageing.oxfordjournals.org/content/45/3/353.short

9. Apóstolo J, Cooke, Richard; Bobrowicz-Campos E, Santana S, Marcucci M, Cano, Antonio; Vollenbroek-Hutten M, Germini F, et al. Predicting risk and outcomes for frail older adults: an umbrella review of frailty screening tools. JBI Database Syst Rev Implement Reports. 2017;15(4):1154–208.

10. Dent E, Kowal P, Hoogendijk EO. Frailty measurement in research and clinical practice: A review. Eur J Intern Med [Internet]. 2016;31:3–10. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27039014

11. Pijpers E, Ferreira I, Stehouwer CDA, Nieuwenhuijzen Kruseman AC. The frailty dilemma. Review of the predictive accuracy of major frailty scores. Eur J Intern Med [Internet]. 2012;23(2):118–23. Available from: http://dx.doi.org/10.1016/j.ejim.2011.09.003

12. Sutton JL, Gould RL, Daley S, Coulson MC, Ward E V, Butler AM, et al. Psychometric properties of multicomponent tools designed to assess frailty in older adults: A systematic review. BMC Geriatr [Internet]. 2016 Feb;16(1):55. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4772336&tool=pmcentrez&rendertype=abstract

13. Drubbel I, Numans ME, Kranenburg G, Bleijenberg N, de Wit NJ, Schuurmans MJ. Screening for frailty in primary care: a systematic review of the psychometric properties of the frailty index in community-dwelling older people. BMC Geriatr [Internet]. 2014;14(1):27. Available from: http://bmcgeriatr.biomedcentral.com/articles/10.1186/1471-2318-14-27

14. Pialoux T, Goyard J, Lesourd B. Screening tools for frailty in primary health care: a systematic review. Geriatr Gerontol Int [Internet]. 2012;12(2):189–97. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22233158

15. Clegg A, Rogers L, Young J. Diagnostic test accuracy of simple instruments for identifying frailty in community-dwelling older people: a systematic review. Age Ageing. 2015;44:148–52.

16. Moher D, Liberati A, Tetzlaff J, Altman DG. Academia and Clinic Annals of Internal Medicine Preferred Reporting Items for Systematic Reviews and Meta-Analyses : Ann Intern Med. 2009;151(4):264–9.

17. The Joanna Briggs Institute (JBI). Joanna Briggs Institute Reviewers' Manual: 2015 edition / Supplement: The systematic review of studies of diagnostic test accuracy. Adelaide: The Joanna Briggs Institute; 2015. p. 46.

18. Ambagtsheer RC, Thompson MQ, Archibald MM, Casey MG, Schultz TJ. Diagnostic test accuracy of self-reported frailty screening instruments in identifying community-dwelling older people at risk of frailty and pre-frailty. JBI Database Syst Rev Implement Reports [Internet]. 2017;15(10):2464–8. Available from: http://insights.ovid.com/crossref?an=01938924-201710000-00009

19. Whiting P, Rutjes A, Westwood M, Mallett S, Deeks J, Reitsma J, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(529e36).

20. Forti P, Rietti E, Pisacane N, Olivelli V, Maltoni B, Ravaglia G. A comparison of frailty indexes for prediction of adverse health outcomes in an elderly cohort. Arch Gerontol Geriatr [Internet]. 2012;54(1):16–20. Available from: http://dx.doi.org/10.1016/j.archger.2011.01.007

21. Dent E, Perez-Zepeda M. Comparison of five indices for prediction of adverse outcomes in hospitalised Mexican older adults: A cohort study. Arch Gerontol Geriatr [Internet]. 2015;60(1):89–95. Available from: http://dx.doi.org/10.1016/j.archger.2014.09.011

22. Leeflang MMG, Deeks JJ, Gatsonis C, Bossuyt PMM. Annals of Internal

Medicine Academia and Clinic. Ann Intern Med. 2008;889–98.

23. Nunes DP, Duarte YA de O, Santos JLF, Lebrão ML. Screening for frailty in older adults using a self-reported instrument. Rev Saude Publica. 2015;49(1).

24. Youden WJ. Index for rating diagnostic tests. Cancer. 1950;32–5.

25. Harbord R. METANDI: Stata module to perform meta-analysis of diagnostic accuracy. [Internet]. 2008. Available from: https://econpapers.repec.org/software/bocbocode/s456932.htm

26. Braun T, Grüneberg C, Thiel C. German translation, cross-cultural adaptation and diagnostic test accuracy of three frailty screening tools: PRISMA-7, FRAIL scale and Groningen Frailty Indicator . Z Gerontol Geriatr [Internet]. 2018;51(3):282–92. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85027176649&doi=10.1007%2Fs00391-017-1295-2&partnerID=40&md5=2bb381e8c1c32a498ef49a10cdc727c0

27. Hoogendijk EO, Van Der Horst H, Deeg DJH, Frijters DHM, Prins BAH, Jansen AAP, et al. The identification of frail older adults in primary care: Comparing the accuracy of five simple instruments. Age Ageing. 2012;0:1–4.

28. Auyeung TW, Lee JSW, Leung J, Kwok T, Woo J. The selection of a screening test for frailty identification in community-dwelling older adults. J Nutr Health Aging [Internet]. 2014;18(2):199–203. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med8&NEWS=N&AN=24522474

29. Bongue B, Buisson A, Dupre C, Beland F, Gonthier R, Crawford-Achour É. Predictive performance of four frailty screening tools in community-dwelling elderly. BMC Geriatr [Internet]. 2017;17(1):262. Available from: http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L622198737

30. Cawthon P, Marshall L, Michael Y, Dam T, Ensrud K, Barrett-Connor E, et al. Frailty in older men: prevalence, progression, and relationship with mortality. J Am Geriatr Soc [Internet]. 2007 Aug;55(8):1216–23. Available from: http://proxy.library.adelaide.edu.au/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=c8h&AN=106183014&site=ehost-live&scope=site

31. Pereira de Llano PM, Lange C, Pires Nunes D, Pastore CA, Pinto AH, Pilotto Casagranda L, et al. Frailty in rural older adults: Development of a care algorithm . ACTA Paul Enferm [Internet]. 2017;30(5):520–30. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85041201385&doi=10.1590%2F1982-0194201700075&partnerID=40&md5=67ab007fd70948480e48f725cf65

df22%0Ahttp://proxy.library.adelaide.edu.au/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=c8h&A

32. Jouanny PD. Evaluation of frailty in primary care: Which tools for which prediction? Eur Geriatr Med. 2014;5(September 2014):S119.

33. Mijnarends DM, Schols JMGA, Meijers JMM, Tan FES, Verlaan S, Luiking YC, et al. Instruments to Assess Sarcopenia and Physical Frailty in Older People Living in a Community (Care) Setting: Similarities and Discrepancies. J Am Med Dir Assoc [Internet]. 2015;16(4):301–8. Available from: http://dx.doi.org/10.1016/j.jamda.2014.11.011

34. Mossello E, Profili F, Di Bari M, Bandinelli S, Razzanelli M, Salvioni A, et al. Postal screening can identify frailty and predict poor outcomes in older adults: longitudinal data from INTER-FRAIL study. Age Ageing [Internet]. 2016;45(4):469–74. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27013497

35. Ng TP, Feng L, Nyunt MSZ, Larbi A, Yap KB. Frailty in older persons: Multisystem risk factors and the Frailty Risk Index (FRI). J Am Med Dir Assoc [Internet]. 2014;15(9):635–42. Available from: http://dx.doi.org/10.1016/j.jamda.2014.03.008

36. Roppolo M, Mulasso A, Gobbens RJ, Mosso CO, Rabaglietti E. A comparison between uni- and multidimensional frailty measures: prevalence, functional status, and relationships with disability. Clin Interv Aging. 2015;10:1669–78.

37. Satake S, Senda K, Hong Y-J, Miura H, Endo H, Sakurai T, et al. Validity of the Kihon Checklist for assessing frailty status. Geriatr Gerontol Int [Internet]. 2016 Jun [cited 2017 Mar 15];16(6):709–15. Available from: http://doi.wiley.com/10.1111/ggi.12543

38. Sternberg SA, Levin R, Dkaidek S, Edelman S, Resnick T, Menczel J. Frailty and osteoporosis in older women--a prospective study. Osteoporos Int [Internet]. 2014;25(2):763–8. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med8&NEWS=N&AN=24002542

39. Szlejf C, Parra-Rodríguez L, Rosas-Carrasco O. Osteosarcopenic Obesity: Prevalence and Relation With Frailty and Physical Performance in Middle-Aged and Older Women. J Am Med Dir Assoc [Internet]. 2017;18(8):733.e1-733.e5. Available from: http://dx.doi.org/10.1016/j.jamda.2017.02.023

40. Yamada Y, Nanri H, Watanabe Y, Yoshida T, Yokoyama K, Itoi A, et al. Prevalence of Frailty Assessed by Fried and Kihon Checklist Indexes in a Prospective Cohort Study: Design and Demographics of the Kyoto-Kameoka Longitudinal Study. J Am Med Dir Assoc [Internet]. 2017;18(8):733.e7-733.e15. Available from: http://dx.doi.org/10.1016/j.jamda.2017.02.022

41. Drubbel I, Bleijenberg N, Kranenburg G, Eijkemans RJCRJ, Schuurmans

MJ, de Wit NJ, et al. Identifying frailty: do the Frailty Index and Groningen Frailty Indicator cover different clinical perspectives? a cross-sectional study. BMC Fam Pract [Internet]. 2013;14(1):64. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3665587&tool=pmcentrez&rendertype=abstract

42. Jung H-W, Yoo H-J, Park S-Y, Kim S-W, Choi J-Y, Yoon S-J, et al. The Korean version of the FRAIL scale: clinical feasibility and validity of assessing the frailty status of Korean elderly. Korean J Intern Med [Internet]. 2016 May 1 [cited 2017 Mar 15];31(3):594–600. Available from: http://kjim.org/journal/view.php?doi=10.3904/kjim.2014.331

43. McCaul KA, Almeida OP, Norman PE, Yeap BB, Hankey GJ, Golledge J, et al. How many older people are frail? Using multiple imputation to investigate frailty in the population. J Am Med Dir Assoc [Internet]. 2015;16(5):439.e1-7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25783623

44. Orkaby AR, Hshieh TT, Gaziano JM, Djousse L, Driver JA. Comparison of two frailty indices in the physicians' health study. Arch Gerontol Geriatr [Internet]. 2017 Jul [cited 2017 Mar 15];71:21–7. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0167494317301309

45. Qiao X, Wang C, Tian X, Dong L, Liu N, Jin Y, et al. Cross-cultural adaptation and validation of the Comprehensive Frailty Assessment Instrument in Chinese community-dwelling older adults. Geriatr Gerontol Int. 2018;18(2):301–7.

46. Dong L, Liu N, Tian X, Qiao X, Gobbens RJJ, Kane RL, et al. Reliability and validity of the Tilburg Frailty Indicator (TFI) among Chinese community-dwelling older people. Arch Gerontol Geriatr [Internet]. 2017;73(December 2016):21–8. Available from: http://dx.doi.org/10.1016/j.archger.2017.07.001

47. Ntanasi E, Yannakoulia M, Kosmidis M-HM-H, Anastasiou CACA, Dardiotis E, Hadjigeorgiou G, et al. Adherence to Mediterranean Diet and Frailty. J Am Med Dir Assoc [Internet]. 2018 Apr;19(4):315–322.e2. Available from: http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L619975676

48. Theou O, Brothers TD, Mitnitski A, Rockwood K. Operationalization of Frailty Using Eight Commonly Used Scales and Comparison of Their Ability to Predict All-Cause Mortality. J Am Geriatr Soc [Internet]. 2013 Sep;61(9):1537–51. Available from: http://proxy.library.adelaide.edu.au/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=c8h&AN=104222764&site=ehost-live&scope=site

49. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, et al. Frailty in older adults: Evidence for a phenotype. Journals Gerontol Ser a-Biological Sci Med Sci. 2001;56(3):M146–56.

50. Theou O, Cann L, Blodgett J, Wallace LMK, Brothers TD, Rockwood K. Modifications to the frailty phenotype criteria: Systematic review of the current literature and investigation of 262 frailty phenotypes in the survey of health, ageing, and retirement in Europe. Ageing Res Rev [Internet]. 2015 May;21:78–94. Available from: http://dx.doi.org/10.1016/j.arr.2015.04.001

51. Bowling A. Mode of questionnaire administration can have serious effects on data quality. J Public Health (Bangkok). 2005;27(3):281–91.

52. Searle SD, Mitnitski A, Gahbauer EA, Gill TM, Rockwood K. A standard procedure for creating a frailty index. BMC Geriatr [Internet]. 2008;8. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-54949131669&doi=10.1186%2F1471-2318-8-24&partnerID=40&md5=2e5d258ea720d2778ff0a5ce1026e430

53. Ambagtsheer RC, Beilby JJ, Visvanathan R, Dent E, Yu S, Braunack-Mayer AJ. Should we screen for frailty in primary care settings? A fresh perspective on the frailty evidence base: A narrative review. Prev Med (Baltim) [Internet]. 2019;119(May 2018):63–9. Available from: https://doi.org/10.1016/j.ypmed.2018.12.020

54. Rockwood K, Theou O, Mitnitski A. What are frailty instruments for? Age Ageing [Internet]. 2015;44(4):545–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25824236

55. Cesari M, Gambassi G, Van Kan GA, Vellas B. The frailty phenotype and the frailty index: Different instruments for different purposes. Age Ageing [Internet]. 2014;43(1):10–2. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84890382740&doi=10.1093%2Fageing%2Faft160&partnerID=40&md5=c0b194a8a1659134a466d436836ae02c

56. Cruz-Jentoft AJ, Kiesswetter E, Drey M, Sieber CC. Nutrition , frailty , and sarcopenia. Aging Clin Exp Res. 2017;29(1):43–8.

Figure Legends:

Figure 1: Study Flow Diagram

Supporting Information Legends:

Supporting Table 1: Search Syntax by Database

Supporting Table 2: 2 x 2 Table Data for Selected Studies

Supporting Table 3: Additional Diagnostic Test Accuracy Results

Supporting Figure 1: Forest Plots, Diagnostic Test Accuracy Statistics