Diagnostic accuracy of screening tool for non-specialist health care settings: A summary of findings from ICOPE rapid reviews

Authors: Jotheeswaran Amuthavalli Thiyagarajan, Matteo Cesari, Suresh Kumar, Carolina Kralj, Finbarr C Martin, Shelly Chadha, John Beard, Islene Araujo de Carvalho

1Department of Ageing and Life Course, World Health Organization, Geneva, Switzerland.
2WHO Collaborating Centre on Frailty, Clinical Research and Geriatric Training, Gérontopôle - Toulouse University Hospital, Toulouse, France
3Department of Clinical Research, London School of Hygiene and Tropical Medicine, London, United Kingdom
4Centre for Global Mental Health Institute of Psychiatry, Psychology & Neuroscience, King's College London, United Kingdom
5Department of Medical Gerontology, King’s College London, United Kingdom
6Department of Management of Noncommunicable Diseases, Disability, Violence and Injury Prevention, World Health Organization, Geneva, Switzerland

BACKGROUND

The World Health Organization (WHO), Integrated Care for Older People (ICOPE) guidelines, provide guidance on evidence-based interventions to manage common impairments in older age, namely mobility, cognition, mood, vision and hearing, along with important geriatric syndromes such as urinary incontinence and falls. However, the guidelines provide very little guidance on potential measurements for detecting declines in the physical and mental capacity of older adults in the second half of life.

One of the cornerstones of the ICOPE strategy is to identify individuals who have a high risk of developing adverse outcomes over a specific period to target them for early preventative strategies and possibly treatment. For example, older people who are seemingly healthy but are found to have a high risk of experiencing significant loss to intrinsic capacity could be recommended to modify their lifestyle and behaviour (e.g. nutrition and exercise interventions) to avoid adverse health outcomes. They may also be prioritised for clinical investigation, which could lead to early diagnosis of an underlying condition (e.g. Alzheimer’s disease).

Existing risk assessment tools are ill-suited for ICOPE strategy because most are developed to predict outcomes for a specific condition or population and therefore have limited usefulness outside the scope of their intended target. Some, for example, focus on the prediction of risk for particular conditions (e.g. frailty, falls), whereas others focus on the prediction of health outcomes for specific cohorts (e.g. intensive care patients or nursing home). Moreover, the interventions recommended in the ICOPE guidelines are expected to produce synergistic effects on the intrinsic capacity and functional ability of the individual. Therefore, for optimal
benefit, interventions are best implemented with an integrated risk and needs assessment that can lead to a comprehensive care plan.

To progress implementation of ICOPE recommendations, a series of rapid systematic reviews were carried out to determine the validity and reliability of screening and diagnostic instruments for the following priority conditions: malnutrition, cognitive impairment, depressive symptoms, mobility impairment, vision impairment and hearing loss. The primary objective of the commissioned reviews was to assess the psychometric properties of the valid screening tests. The focus of this report is to summarise the diagnostic accuracy and capability of the screening tool to classify older persons correctly as having or not having an impairment, as well as to comment briefly on feasibility and acceptability.

**METHOD**

Two-stage review method was applied for assessing the accuracy of screening tools. First, three independent reviewers, conducted a rapid systematic review of screening tools for ICOPE priority conditions. Second, an independent reviewer assessed the quality of included reviews and studies, and meta-analysed the evidence for accuracy and validity of the test. All four rapid reviews were performed between August and October 2016.

Online databases PubMed and Cochrane were systematically searched in title and abstract. The search was limited to publications in English. Studies published from 2000 unto August 2016 were searched using terms selected from the literature, taking into account the three parameters of inclusion criteria, that is older people, primary care, and health conditions. Backward citation tracking was performed to identify additional relevant articles. The final selection of search terms was: (1) conditions of interest, (2) target population, (3) type of measurement instrument, (4) measurement properties.

**Study eligibility criteria**

Observational studies fulfilling the following criterias were considered for inclusion in the rapid reviews: a) reporting diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC), b) explore a target population of older individuals aged 60 years and over, c) published in a peer reviewed journal, d) published between the years 2000 and 2016 (August), and e) conducted in primary care settings or community. Also, a copyright information on each screening tool was also gathered and considered for drafting final considerations. Studies were excluded if they studied a specific patient population (e.g. older persons with dementia).

**Search strategy:**

A systematic literature search was conducted for each reviews using specific medical subject headings (MeSH) and words from “all fields” to identify studies in Pubmed and the Cochrane library. Broadly, search strategy included a combination of the following search terms: (screening) AND (conditions specific terms) and (older adults OR elderly OR ageing OR aged)
AND (Primary care OR community). Reference lists of eligible articles will be searched by hand to identify any additional, relevant articles, which will also be subject to the screening process.

**Synthesis Methods:**

The search hits were inserted in reference library (EndNote and reference manager), and duplicates were removed. Two independent reviewers reviewed the review of the construct of interest. All titles and abstracts were independently screened by the review and scored as relevant or not relevant based on the inclusion and exclusion criteria aforementioned. Identified studies were circulated to WHO ICOPE expert panel for review and comments. At this stage, additional studies recommended were also included in the review. Subsequently, full texts were assessed for inclusion by the reviewer responsible for individual construct, according to the eligibility criteria. Note: No review was undertaken for assessment of malnutrition, existing systematic review was used to synthesise the evidence.

**Assessment of methodological quality**

One independent reviewer (ATJ), masked to other reviewer’s judgement, performed the risk of bias assessment. Quality of included systematic reviews was assessed using AMSTAR tool. Individual observational studies were for quality using the QUADAS-2 tool as recommended by The Cochrane Collaboration. Key areas critical to quality assessment are participant selection, blinding, and missing data. Instead of the overall risk score, a narrative summary description was produced for the numbers of studies that were found to have a high, low, or unclear risk of bias as well as concerns regarding applicability.

**Statistical analysis:**

We were principally interested in the test accuracy of screening for the diagnosis of impairments using a dichotomous variable, ‘cases’ or ‘non-cases’. Therefore, we applied the Cochrane diagnostic accuracy framework for the analysis of a single test and fitted the extracted data to a standard two-by-two data table showing binary test results cross-classified with a binary reference standard. Where data allowed, we used Review Manager 5.3 to calculate sensitivity, specificity, and their 95% confidence intervals (CIs) from the two-by-two tables abstracted from the included studies. We presented data graphically; using forest plots to allow necessary visual inspection and comparison of individual studies. If data allowed, we had calculated summary estimates of test accuracy. We used the bivariate approach to give summary estimates of test accuracy at common thresholds and common time points and to use the hierarchical summary receiver operating characteristics (HSROC) model to explore differing thresholds across studies.
RESULT

Narrative summary of the rapid reviews

Depressive symptoms

Seven studies assessing ten different screening tools for depressive symptoms in primary care settings were included in the scoping review. Four recent systematic reviews and three individual studies were analysed in detail. The latter (non-reviews) were also reported since they examined the diagnostic accuracy of tools that had not been previously depicted.

The Geriatric Depression Scale (GDS) was found to be the most frequently explored screening tool in this review. Two recent systematic reviews explored this tool exclusively, assessing the diagnostic accuracy of its different versions (GDS-30 and GDS-15) using different cut-off scores and across different settings [1-3]. Similarly, different cut-off scores for the Center for Epidemiologic Studies Depression scale (CES-D) were also examined across 33 studies included in two other reviews [4]. Other tools reported in only a few studies included the EURO-D, the Cornell Scale for Depression in Dementia (CSDD), the Caribbean Culture-Specific Screen (CCSS), the Patient Health Questionnaire, the 10-item Kessler Psychological Distress Scale (K10), the WHO’s Well-being 5-item Index (WHO-5) and SelfCARE(D), amongst others. Several diagnostic instruments have been used in the general population to identify depression. However, surprisingly, only two have been specially designed for older people: the Geriatric Depression Scale [5] and the EURO-D scale [6].

Malnutrition

Two identified systematic reviews examined the most appropriate nutritional screening tools, regarding validity and reliability, for identifying malnutrition risk in older adults living in the community [7, 8]. The review identified ten screening tools for use in community-dwelling older adults: Mini Nutritional Assessment- Short Form (MNA-SF), Malnutrition Universal Screening Tool (MUST), Nutrition Screening Initiative (NSI), which includes the DETERMINE Checklist and Level I and II Screen, Australian Nutritional Screening Initiative (ANSI), Seniors in the Community: Risk Evaluation for Eating and Nutrition (SCREEN I and SCREEN II), Short Nutritional Assessment Questionnaire (SNAQ), Simplified Nutritional Appetite Questionnaire (SNAQ), and two unnamed tools. MNA-SF appears to be the most appropriate nutrition-screening tool for use in community-dwelling older adults although MUST and SCREEN II also have evidence to support their use. Although MNA was widely applied screening tool, small proportion of studies assessed the accuracy of the test in primary care and community settings [9].
Cognitive impairment (but not fulfilling the criteria for dementia diagnosis)

The rapid review identified a systematic review that examined the diagnostic accuracy of cognitive screening instruments in older adults aged 50 years and over[10]. The review included 41 studies that addressed the diagnostic accuracy of very brief and brief screening instruments that could be administered in primary care and seven studies that addressed instruments that could be self-administered. For this review identified 10 instruments, including (most to least frequent) MMSE, VF, IQCODE, CDT, SPMSQ, AMT, MIS, Mini-Cog, TICS-m and FCSRT testing for dementia; 5 (CDT, MMSE, Mini-Cog, VF, IQCODE) testing for MCI and 2 (MMSE and Mini-Cog) for dementia and MCI combined were meta-analysed individually.

Tools assessed in single study (including General Practitioner Assessment of Cognition (GPCOG), 3-Word Memory Test, 6-item screener, Ascertain Dementia 8 (AD8), Benton's Orientation Test, Brief IADL (4IADL), Cognitive Assessment Screening Test (CAST), Computer Assessment of Mild Cognitive Impairment (CAMCI), Fluid Object Memory Evaluation, abbreviated, Functional Activities Questionnaire (FAQ), Immediate recall, Immediate Recall (Logical Memory I), Informant Report of Memory Problems (IRMP), Kendrick Cognitive tests, Labyrinth Test, Lawton ADL, Function 2 (MF-2), Memory Impairment Screen by Telephone (MIS-T), Minimum Data Set Cognition Scale (MDS-Cog), MMblind, Orientation Memory Concentration (OMC), Oral trails, Rey figure copy, Self-Administered Gerocognitive Examination (SAGE), Short Blessed Test (SBT), Short Concord Informant Dementia Scale and the Single-item informant report) were not reported[10]. MMSE was shown to be the most utilised screening test for dementia, MCI and dementia+MCI combined, being explored in 25, six, and five studies respectively.

Hearing impairment

Two reviewers independently identified 39 studies in the rapid review that examined screening and diagnostic instrument for hearing impairment. Studies included were identified and included if they were published in the past twenty years (1996 – 2016). Out of 39 records, 15 records were identified as suitable for the review, but the full-text was only available for 12 studies. Hence the full text of all the 12 records was retrieved and included in this review. There were several differences among the studies included regarding the study design, outcome of interest and methods used. Hence a meta-analysis was not possible. We reviewed all the included studies narratively. All the studies included in the review were published articles in peer-reviewed journals. The review identified four studies from LMICs namely Brazil [11, 12], China [13] and South Africa [14].

Includes studies examined the accuracy of several screening tools for primary care or community settings, including clinical testing methods (whispered voice test, pure tone audiometry, hand-held audioscope, smartphone-based PTA), single-question screening (asking, “Do you have difficulty with your hearing?”) or multiple-item patient questionnaires (HHIE-S), and handheld audiometers. Relatively simple tests, such as the whispered voice at 2 feet and a single question regarding perceived hearing loss, appear to be nearly as accurate
compared with a more detailed hearing loss questionnaire or a handheld audiometric device for detecting hearing loss. Compared to the single subjective question, the review found that the whispered voice test was an accurate test for detecting hearing impairment and compares favourably with the portable audioscope. Despite some variations in the methodology of studies and the populations sampled, findings are relatively consistent.

**Vision impairment**

Two reviewers independently identified 52 records and screened all 52 during the process. Out of 52 records, 12 records were identified as suitable for the review and hence the full text of all the 12 records were retrieved and included in the review. There were several differences among the studies included regarding the study design, outcome of interest and methods used. Hence a meta-analysis was not possible. We reviewed and summarised all the included studies narratively. The review examined several diagnostic tests, suitable for primary care and community settings. They include Snellen’s Visual Acuity Chart (2 and 6 meters), Portable Eye Examination Kit (PEEK), Smart hone-based Snellen’s Visual Acuity, Log MAR Vision Charts, Early Treatment Diabetic Retinopathy Study (ETDRS), Pinhole eye test, Amsler Grid. The accuracy of E-Test Card and National Eye Institute – Vision Functioning Questionnaire (NEI-VFQ) was also investigated.

The review found that screening questionnaires had low accuracy for identifying persons with impaired visual acuity compared with the Snellen eye chart or an ophthalmologic examination. Included studies also found no visual acuity test had both high sensitivity and specificity compared with a detailed ophthalmologic examination. One study tested the validity of the Amsler grid had reported poor accuracy for identifying visual conditions. However, the review found convincing evidence visual acuity test cards and portable eye examination kit.

**Mobility impairment**

The review identified 2026 studies on single domain and summary measure of mobility impairment, which include 125 on muscle mass, 254 on hand grip strength, 88 on chair stand test, 19 on stair climb test, 168 on Berg balance scale, 257 on Timed up and go test, 184 on short physical performance battery, 174 on one-leg-stand, 86 on functional reach test, and 671 on usual gait speed. Of the ten measurements evaluated, the SPPB was the measurement with most favourable characteristics (e.g. highest score in reliability and validity) and has been extensively investigated in different populations ranging from vigorous to ADL limited or frail with more than 30 studies investigating at least one psychometric property. A recent systematic review investigated predictive ability of SPPB for all-cause mortality was also included[15]. The review found strong construct validity for SPPB among older persons living in the community.
MOVING FROM EVIDENCE TO CONSIDERATION:

Screening tool for consideration:

Based on the scoping reviews, following screening and or diagnostic assessments are considered for the ICOPE risk assessment tool.

Geriatric depression scale:

The Geriatric Depression Scale (GDS) was found to be the most widely examined screening tool for late-life depression in primary care settings. This tool is available in public domain; hence there is no user fee or copyright issue. This instrument is considered to be a useful evaluation tool in this population due to its more straightforward format (Yes/No) and due to its focus assessing issues related to loss, cognitive function and self-image as opposed to somatic symptoms, which may be attributed to comorbid physical conditions and the ageing process.

A 2016 meta-analytic review conducted [2] examined the diagnostic accuracy of the shorter versions (15-items or less) of the Geriatric Depression Scale (GDS). Authors identified 32 studies (n=13141), ten based in primary care settings and 14 in the community. The mean age of the participants ranged between 66.4 and 87 years and the prevalence of depression reported across studies varied from 3.2% to 64.1%. Major depressive disorder was diagnosed using different gold standard tools including GMS AGECAT, ICD-1 checklist, PRIME-MD, SCID, MINI, CIDI and DIS. DSM-IV and ICD-10 criteria were also included in all of the studies. The verbal administration was the most utilised method (17 studies), followed by self-administration observed in 5 studies. Information about administration mode was unavailable for two studies, and even though the majority of the assessments were conducted in English, other languages were used, including Spanish, Malay, Dutch, Korean, Farsi and Tai.

Due to the small number of primary studies examining shorter versions, a meta-analysis could only be conducted for the 15-item version of GDS. Subgroup analysis of primary care setting showed a pooled sensitivity of 0.92 (95% CI 0.83-0.96) and specificity of 0.63 (95% CI 0.42-0.80) for a cut-off score of 5. For the community setting, pooled sensitivity was lower (0.78, 95% CI 0.42-0.80) and specificity was higher (0.90, 95% CI 0.74-0.96) (figure 1 & 2). Comparably, when the cut-off score was 6, sensitivity for GDS-15 in primary care settings reduced to 0.77 (95% CI 0.69-0.84) while the specificity slightly increased 0.74 (95% CI 0.47-0.90). Contrarily, in community settings, performing GDS-15 at cut-off score 6, showed slightly higher sensitivity (0.80, 95% CI 0.54-0.93) and similar specificity (0.90, 95% CI 0.74-0.96). When comparing Youden’s Indexes for both settings, GDS-15 appears to perform better in community settings (0.68) compared to when applied in primary care practice (0.55). Meta-regression performed to explore heterogeneity between studies, revealed that diagnostic accuracy was predicted by language (p=0.05) and study country (p=0.005) but not by setting (p=0.66), administration mode (p=0.80), age group (p=0.11) or gender (p=0.54).
Figure 1: Summary HROC plot: GDS-15 accuracy by different cut-points in primary care settings

![Summary HROC plot: GDS-15 accuracy by different cut-points in primary care settings](image)

Legend:
- GDS15 (cutpoint 5)
- GDS15 (cutpoint 6)
- GDS15 (cutpoint 7)
- GDS15 (cutpoint 4)
- GDS15 (cutpoint 3)

Figure 2: Forest plot of sensitivity and specificity of GDS 15 in primary care settings

<table>
<thead>
<tr>
<th>Study</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Cut-point</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Van Marwijk, 1995</td>
<td>24</td>
<td>149</td>
<td>11</td>
<td>402</td>
<td>3</td>
<td>0.69 [0.51, 0.83]</td>
<td>0.73 [0.69, 0.77]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kalt, 1999</td>
<td>12</td>
<td>33</td>
<td>1</td>
<td>84</td>
<td>4</td>
<td>0.92 [0.64, 1.00]</td>
<td>0.72 [0.63, 0.80]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arthur, 1999</td>
<td>10</td>
<td>34</td>
<td>2</td>
<td>155</td>
<td>4</td>
<td>0.83 [0.52, 0.98]</td>
<td>0.82 [0.76, 0.87]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abbas, 1998</td>
<td>28</td>
<td>9</td>
<td>6</td>
<td>39</td>
<td>5</td>
<td>0.82 [0.65, 0.93]</td>
<td>0.81 [0.67, 0.91]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lyness, 1997</td>
<td>146</td>
<td>24</td>
<td>13</td>
<td>106</td>
<td>5</td>
<td>0.92 [0.86, 0.96]</td>
<td>0.82 [0.74, 0.88]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grosso, 2007</td>
<td>36</td>
<td>6</td>
<td>8</td>
<td>251</td>
<td>6</td>
<td>0.82 [0.67, 0.92]</td>
<td>0.98 [0.95, 0.99]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blank, 2004</td>
<td>51</td>
<td>15</td>
<td>13</td>
<td>46</td>
<td>6</td>
<td>0.80 [0.68, 0.89]</td>
<td>0.75 [0.63, 0.86]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friedman, 2005</td>
<td>102</td>
<td>206</td>
<td>23</td>
<td>629</td>
<td>6</td>
<td>0.82 [0.74, 0.88]</td>
<td>0.75 [0.72, 0.78]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D'Ath, 1994</td>
<td>61</td>
<td>33</td>
<td>6</td>
<td>94</td>
<td>6</td>
<td>0.91 [0.82, 0.97]</td>
<td>0.74 [0.65, 0.81]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Robinson, 2002</td>
<td>51</td>
<td>85</td>
<td>16</td>
<td>151</td>
<td>7</td>
<td>0.76 [0.64, 0.86]</td>
<td>0.64 [0.58, 0.70]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Another systematic review conducted a meta-analysis to assess the diagnostic and clinical utility of the 30 and 15 item versions of the GDS in primary care (n=4869)[1]. The majority of the studies were conducted in high-income countries. After adjustment for meta-analysis, GDS-30 and GDS-15 showed modest to adequate pooled sensitivity (77.4%, 95% CI 66.3-86.8 and 81.3%, 95% CI 77.2-85.2) and specificity (65.4%, 95% CI 44.2-83.8 and 78.4%, 95% CI 71.2-84.8) respectively. However, both tools displayed poor case-finding performance, with low positive predictive values (39% and 37% respectively). Overall, GDS-15 showed better screening accuracy (Youden’s Index 0.59), with a higher fraction of correctly identified cases than GDS-30 (77.6% and 71.2% Chi2 =24.8 p<0.001). Nevertheless, these results should be interpreted with caution as selective reporting of cut-off scores may influence pooled results. Moreover, future research should prioritise on briefer versions of the GDS, which is more appealing for less resourced settings, and time constraints in current primary care practice [2]. Overall, compared to other screening instruments, GDS-15 was cross-culturally validated screening tool, which can be administered by non-specialist health professionals in primary care or community settings.

**Mini-Nutritional Assessment (MNA)**

The MNA is an inexpensive screening tool requiring no laboratory investigations[9]. In the included review, out of 13 studies, only two studies found a sensitivity below 70% compared to a detailed nutritional assessment or BMI <19 and <21; but both studies showed high specificity. These results suggest the validity of the full MNA to screen for malnutrition in the elderly is very good. The assessment validity of the MNA using specificity suggests a broader variation, ranging from 13 - 98%, due to the lack of a gold standard for nutritional assessment. The standard used for specificity ranges from simple anthropometry to a comprehensive nutritional assessment. However, the results from 6 studies that obtained specificity >70% demonstrate an excellent reliability for the MNA assessment[9].

Meta-analysis for MNA-SF was performed to estimate the diagnostic accuracy. The validity of the MNA-SF is nearly as good as the MNA full form, with a sensitivity of 98-73% and specificity of 63-100% in 6 different studies (Figure 4). The MNA-SF pattern of sensitivity and specificity was illustrated by hierarchical summary receiver operating characteristic curve. Good tests have lines that rise steeply and pass close to the top left-hand corner, where both the sensitivity and specificity are 1. The areas under the ROC curves represent the overall summary accuracy of the MNA-SF total score as a test for malnutrition (Figure 3). The summary estimate for sensitivity was 0.90 (95% CI, 0.82–0.95). The summary estimate for specificity was 0.87 (95% CI, 0.72–0.94). These results indicate that MNA–SF is highly sensitivity screening tool compared to a variety of nutritional parameters (biochemical, anthropometry, clinical examination, and full MNA test). This review suggests that compare to full MNA assessment; the 6-item MNA-SF is more suitable and effective nutritional screening tool for primary care and community settings.
Figure 3: Hierarchical summary ROC curve showing diagnostic accuracy of MNA-short form

Figure 4: MNA-SF sensitivity and specificity in primary care studies

<table>
<thead>
<tr>
<th>Study</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rubenstein 2001</td>
<td>87</td>
<td>0</td>
<td>2</td>
<td>53</td>
<td>0.98 [0.92, 1.00]</td>
<td>1.00 [0.93, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aid Malek, 2015</td>
<td>90</td>
<td>42</td>
<td>3</td>
<td>70</td>
<td>0.97 [0.91, 0.99]</td>
<td>0.63 [0.53, 0.71]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stsek 2014</td>
<td>174</td>
<td>77</td>
<td>20</td>
<td>362</td>
<td>0.90 [0.85, 0.94]</td>
<td>0.82 [0.79, 0.86]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kostka 2014</td>
<td>583</td>
<td>229</td>
<td>74</td>
<td>762</td>
<td>0.89 [0.86, 0.91]</td>
<td>0.77 [0.74, 0.79]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kostka 2014</td>
<td>768</td>
<td>164</td>
<td>175</td>
<td>1107</td>
<td>0.81 [0.79, 0.84]</td>
<td>0.87 [0.85, 0.89]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kaiser 2011</td>
<td>22</td>
<td>18</td>
<td>8</td>
<td>224</td>
<td>0.73 [0.54, 0.88]</td>
<td>0.93 [0.88, 0.96]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Mini-mental status examination

This review identified numerous brief instruments that primary care providers can use to screen for cognitive impairment. The MMSE was, by far, the most studied screening tool for dementia and MCI. This 11-item tool has become the most often used method for assessing cognitive impairment in multiple settings [16]. It was originally developed by Folstein in 1975 [17] and, since then, has been studied in different languages across numerous countries. One of this tool’s characteristics is its administration time, usually ranging between 6 and 10 minutes. However, in current time-restricted health practices, the latter may still result lengthy, thus, encouraging the employment of other shorter, yet effective, screening tests. Several cut-points have been described. However, studies suggest 23/24 and 24/25 as appropriate cut-off scores for primary care settings[10]. When assessing MMSE’s diagnostic accuracy, meta-analysis results of 26 studies showed high-pooled sensitivity (0.93 95% CI 0.91-0.94) and specificity (0.88 95% CI 0.87-0.89), respectively. A Youden’s Index of 0.81, confirms its high effectiveness for detecting dementia in primary care settings.

Compared to dementia, evidence examining test performance of screening instruments for MCI appears to be more limited. A total of 15 studies assessing MCI exclusively were included in the USPSTF guidelines, most published during the last decade[10]. According to the gold standard, the prevalence of MCI observed across studies ranged from 15% to 83.5%. Moreover, even though, where reported, educational levels of participants tended to be low, the fact that educational level was not reported in most of the studies suggests possible bias. There was considerable variation regarding MCI diagnostic criteria, with most studies implementing either Petersen’s criteria (which mainly focuses on amnestic MCI), Winblad criteria (including amnestic and non-amnestic MCI) or suboptimal performance on cognitive testing (using CDR score). This heterogeneity could affect the comparability of diagnostic estimates[10].

MMSE was the most examined instrument to screen for MCI in primary care settings. However, six studies in the included review [10] assessed MCI exclusively. One of the studies evaluated the accuracy of MMSE stratified by ethnicity (white, black and Latino populations) [18] and another based on educational level (more or less than high school) [19]. Pooled meta-analysed estimates disclosed a low sensitivity (0.61 95% CI 0.57-0.64) yet slightly higher specificity (0.78 95% CI 0.76-0.80) for MMSE to identify MCI. Although Youden’s index was low (0.39), when comparing diagnostic accuracies and body of literature available for each of the other methods, MMSE appears to be the most suitable method for screening for MCI in primary care settings.
**Whisper voice test**

Four included studies examined the accuracy of the whispered voice test in older adults. These studies evaluated the diagnostic accuracy of a whispered voice at 2 feet for identification of >25 or >30 dB hearing loss. The prevalence of hearing impairment ranged from 26% to 61%. Three studies used similar techniques for the whispered voice test and a 30 dB positivity threshold for hearing impairment by audiometry[20-22]. The fourth study used a different technique for the test and a 25 dB positivity threshold, and its results were reported in such a way that it was not possible to calculate an overall sensitivity and specificity, although specificities for sensitivities of 40% and 82% were provided[23]. In addition, the distance from the examiner to the person’s ear was less than half the distance in the other studies (11 inches (28 cm) v 24 inches (61 cm)). In the three comparable studies the sensitivity of the whispered voice test was either 90% or 100% and specificity ranged from 80% to 87%. Overall, meta-analysed sensitivity was 83% (95% CI 69% to 97%) (Figure 5). Positive likelihood ratios ranged from 4.6 to 7.7, showing that a positive test is moderately strong in ruling in hearing impairment. Negative likelihood ratios were zero or close to it, showing no hearing impairment when the test is negative. The evidence above suggests that WVT is comparatively better than the single or few non-standardised questions for screening hearing impairment.

An earlier review that compared accuracy of WVT with subjective and structured standard questionnaire also found similar results[24].

**Figure 5 : Diagnostic accuracy of whispered voice test**
**Visual Acuity Card and Portable Eye Examination Kit**

The test card consists of two sizes of optotypes on a small square plastic card. Either the illiterate, directional E or the Landolt C was used. The Peek Acuity application was written in Android and was used on a Galaxy SIII GT-I9300 (Samsung C&T Corp) running Android 4.0. The application was directly installed onto the test devices. Screen brightness was set to 100% within the application. Peek Acuity follows the standard ETDRS chart design with a 5 × 5 grid optotypes letter E displayed in 1 of 4 orientations (90°, 180°, 270°, and 0°). The participant points in the direction they perceive the arms of the E to be pointing and the tester uses the touch screen to swipe accordingly, translating the gestures from the patient. The tester is masked to the presented optotype and is unaware whether the participant is providing the correct response. These two tests have been validated in low resourced health care settings and non-specialist health professionals can be trained to administer in primary care or community settings.

Two included studies investigated the accuracy of Visual acuity card and portable eye examination kit in diagnosing vision impairment[25, 26]. The results showed good sensitivity and specificity when used in community-based settings. The VA test cards had a sensitivity of 85% for distant vision and 100% for near vision. The specificity was 96% and 84% for distant and near vision. The Portable eye examination kit is a versatile instrument that has alternatives for vision assessment through LED lights, hand movements, finger counting etc. The sensitivity and specificity of this tool was 85% and 98% respectively. This test could be performed within 70-80 seconds on an average. The PEEK application is calibrated with letter ‘E’ optotypes and with ETDRS chart format. It is also designed to be uploaded and used in any android phones. Another advantage of using this tool inbuilt with smartphone is its ability to have standard luminosity even at settings where it is poor. The only concern with this use of the tool is the distance. The distance from which the examiner does the assessment has to be validated in future research.

**Short physical performance battery**

The SPPB is based on three timed tasks: standing balance, walking speed, and chair stand tests, all which require advanced sequential movements or activities that require strength, balance, dexterity, and cognitive control. The timed results of each subtest are rescaled according to predefined cut-points for obtaining a score ranging from 0 (worst performance) to 12 (best performance). As there is no universal consensus on definition of mobility impairment, the SPPB was evaluated in terms of its reliability and concurrent validity.

Three included studies found that older people with SPPB scores 10 at baseline had significantly higher odds of mobility impairment at follow-up [27-29]. Two of them had a 4-year follow-up while the third had follow-ups at 6, 12, and 18 months. Two included studies, conducted in United States, showed that a baseline score of 4–6 on the SPPB indicated that the individual had a 2.9–4.9 increase in risk of developing a future mobility impairment [27, 28].
When the baseline score was 7–9 on the SPPB, one study found a 1.9–2.1 relative risk of future mobility impairment whereas another study reported a 1.6–1.8 relative risk of developing a mobility limitation. A baseline score of 10–12 on the SPPB was predictive of no future mobility impairment. Another study compared the SPPB (0.74 odds ratio on a 95% confidence interval of 0.53–1.03) to gait speed (0.57 odds ratio on a 95% CI of 0.02–16.0) at the 12-month follow-up[29] and found increased risk of mobility impairment in subsequent follow-up assessed in community dwelling older adults. Overall, pooled results suggest that community dwelling older persons with lower SPPB score are three times more likely to develop mobility impairment than their counterpart (3.3 relative risk on a 95% CI of 1.4-5.3)(figure 6).

This review also found that walking speed, chair rises, and standing balance (components of the SPPB) were all associated with mortality. Included twelve studies focused on community living older people and examined the relationship between SPPB and all cause mortality. The length of follow-up ranged from a minimum of 1 year to a maximum of 11 years. As compared to an SPPB score of 10–12, scores of 7–9, 4–6, and 0–3 were associated with greater and progressively increasing risks of all-cause mortality. After adjustment for age, sex, and body mass index, pooled odds ratio for mortality is three times higher among older persons scored 0-3 compared to 10-12 in SPPB (figure 7). Heterogeneity, measured as I²%, was insignificant for the comparison between SPPB scores lower score versus higher score. Based on the concurrent validity evidence summarized above, this review strongly supports the role of SPPB scores as a marker for risk stratification for mobility impairment.

Figure 6: Predictive ability of SPPB in older persons living in the community. Outcome: Mobility impairment
DISCUSSION:

For a screening instrument, high sensitivity (i.e. a very high proportion of individual screen positives) is a key factor, whereas high specificity is less important. This is due to the fact that the second step of the diagnostic process, after an initial positive screening with the test, consists of a diagnostic interview performed by trained clinicians or health workers, during which false positives (patients screening positive on the initial screening test but not meeting criteria for the impairment or condition) will be detected.

Unlike many other disorders, impairments have no universally accepted criterion standard. Several screening and diagnostic instruments included in this review have been used to define the presence or absence of the conditions. In this report, multiple objective and self-reported screening instruments of vision impairment, hearing impairment, cognitive impairment, depressive symptoms and mobility impairment have been identified and described. Preliminary rapid reviews were performed to identify the instruments that have been studied more than others so far. This first step was necessary to better focus a second in-depth analysis of the validation studies of commonly used screening instruments at primary care or community settings. Such a two-step approach to the analysis of the literature was adopted given the huge number of instruments that have been designed, developed and adapted over
the years for older persons. This review identified a final set of six screening instruments that showed adequate accuracy and predictive ability for measuring ICOPE priority conditions (see the table below for details).

<table>
<thead>
<tr>
<th>Table 1: Psychometric properties of screening tool</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Screening tool for ICOPE priority condition</strong></td>
</tr>
<tr>
<td><strong>Conditions</strong></td>
</tr>
<tr>
<td>Screening tool</td>
</tr>
<tr>
<td>Diagnosis</td>
</tr>
<tr>
<td>Measurement type</td>
</tr>
<tr>
<td>Time taken in minutes</td>
</tr>
<tr>
<td>Method of administration</td>
</tr>
<tr>
<td>Administered by (specialist or non-specialist)</td>
</tr>
<tr>
<td>Educational bias</td>
</tr>
<tr>
<td>Validated population</td>
</tr>
<tr>
<td>Cultural bias</td>
</tr>
<tr>
<td>Setting</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
</tr>
<tr>
<td>Specificity (%)</td>
</tr>
<tr>
<td>PPV (95% CI)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><em>LAMIC</em> - Low and Middle Income Countries</td>
</tr>
</tbody>
</table>

First, it is important to underline that all of the identified instruments (GDS, SPPB, MMSE) have to be considered as screening tools rather than diagnostic tools. Such a characteristic is related to the fact that 1) physical and mental capacity results from the interaction of multiple systems and apparatuses beyond the measured domain, and 2) a nosological condition focused on the impairments is not yet objectively operationalised. For example, the ICD-10-CM diagnosis code has been released for sarcopenia. However, sarcopenia is a clinical condition requiring the simultaneous assessment of muscle quantity (i.e. muscle mass) and quality (i.e. muscle strength). Therefore, no objective instrument can singularly and independently serve as the diagnostic tool.

None of the studies included in this review considered the potential disadvantages of screening, in addition to the extra costs of implementing a screening programme. There is a risk that older persons who are screened ‘out’ and are not considered ‘at risk’ (false negatives) could suffer from neglect as a result of screening. For example, nutritional screening might potentially result in ineffective or harmful interventions for some patients, but we found no study that assessed such potential outcomes of the screening programmes.

Future studies should consider the different implications of ‘opportunistic screening’ versus ‘screening as part of ongoing care’. While both approaches may result in changes in referrals to concerned health professionals, the direction of effect for the latter approach might be different. Screening as part of ongoing primacy care service might be more likely to improve staff awareness and might indirectly result in a reduced need for referrals. Although the included studies do not provide enough evidence on this issue, other direct evidence suggests that routine screening improves health professionals’ awareness of unmet care needs [31].

The rationale for promoting the screening tool is to detect targeted priority conditions in an early—pre-symptomatic—phase and to realise better treatment outcomes by early intervention compared to the usual care in a later—symptomatic—phase. Screening programmes (like all healthcare services) must prove not only the effectiveness of clinical intervention but also the economic cost-effectiveness. Over the past 20 years, decisions in the healthcare sector have focused increasingly on health economic goals. Due to demographic
ageing and technological progress, the provision of healthcare services is characterised by an increasing shortage of resources.

Several feasibility issues arise from this review. Many studies highlighted limitations in the interpretation of their outcomes mainly due to selective reporting of cut-off scores that might, on occasion, have inflated the accuracy of the tools. Moreover, several studies examining screening tools were focused on the general adult population with no specific subgroup analysis for older people. Studies exploring shorter versions of screening tools, such as GDS, are currently limited, and future research should include multiple cut-off point reporting and summary indicators of discriminability, such as the area under the ROC curve or Youden’s Index. Similarly, a significant gap was identified in the amount of available evidence assessing screenings for subsyndromal conditions (e.g. MCI and subthreshold depression). However, these findings could be expected considering the scales are not specifically designed for these conditions.

Despite a large body of well-conducted diagnostic accuracy studies, only a handful of instruments have been studied in more than one study applicable to primary care or community care settings. Further, the majority of studies were conducted in high-income countries. Therefore, future research should validate the screening tool in a non-specialist healthcare setting.

This review also found that very few screening tools are available in the public domain for clinicians and researchers. Many validated screening tools are pulled behind a wall of active copyright enforcement by the authors. In fact, clinical tools we take for granted, such as the Katz Index of Independence in Activities of Daily Living, fall into the same ‘benign neglect’ copyright category as the MMSE did before 2000. We recommend that authors of widely used clinical tools provide explicit permissive licensing, ideally with a form of copyleft. Any new tool developed with public funds can be required to use a copyleft or similar license to guarantee the freedom to distribute and improve it, similar to the requirement for open-access publication of research.

Annex:

1. Glossary

**Cognitive impairment:** A loss or abnormality in attention functions, memory functions or higher-level cognitive functions.

- Attention functions are special mental functions that focus on an external stimulus or internal experience for a specific period of time.
- Memory functions are special mental functions that register and store information and retrieve it as needed.
- Higher-level cognitive functions are special mental functions that involve the frontal lobes of the brain. They include complex goal-directed behaviours such as decision-making, abstract thinking, planning and carrying out.
plans, mental flexibility, and deciding which behaviours are appropriate under specific circumstances. These are often called executive functions.

**Depressive symptoms**: The presence of distress or some degree of impaired functioning in the absence of depressive episode/disorder.

**Hearing impairment**: The loss or abnormality in sensory functions relating to sensing the presence of sounds and discriminating the location, pitch, loudness or quality of sounds.

**Mobility impairment**: A loss or abnormality in any form of moving by changing body position or location or by transferring from one place to another, by carrying, moving or manipulating objects, by walking, running or climbing, and by using various forms of transportation.

**Vision impairment**: A loss or abnormality in sensory functions relating to sensing the presence of light and sensing the form, size, shape or colour of the visual stimuli.

**Reference**


