Which DSM validated tools for diagnosing depression are usable in primary care research? A systematic literature review

Article (Accepted Version)


This version is available from Sussex Research Online: http://sro.sussex.ac.uk/id/eprint/72747/

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher’s version. Please see the URL above for details on accessing the published version.

Copyright and reuse:
Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

http://sro.sussex.ac.uk
Which tool for diagnosing depression is usable in primary care clinical research? A RAND / UCLA consensus within the European General Practitioners Research Network.

The HSCL-25: a depression diagnosis tool usable in collaborative primary care research.

Nabbe P¹ (MD), Le Reste JY² (MD, PhD), Guillou-Landreat M³ (MD, PhD), Beck-Robert E⁴ (MD), Assenova R⁵ (MD), Lazic D⁶ (MD, PhD), Czachowski S⁷ (MD), Stojanović-Špehar S⁸ (MD), Hasanagic M⁹ (MD), Lingner H¹⁰ (MD), Clavería A¹¹ (MD, PhD), Fernandez San Martin MI¹² (MD, PhD), Sowinska A¹³ (MD, PhD), Argyriadou S¹⁴ (MD), Lygidakis C¹⁵ (MD), Le Floch B¹⁶ (MD), Doerr C¹⁷ (MD), Montier T¹⁸ (MD, PhD) and Van Marwijk H¹⁹ (MD, PhD) and Van Royen P²⁰ (MD, PhD).

1. Nabbe P. ERCR SPURBO, Department of general practice, Université de Bretagne Occidentale, Brest, France. nabbe.patrice@univ-brest.fr
2. Le Reste JY. ERCR SPURBO, Department of general practice, Université de Bretagne Occidentale, Brest, France. lereste@univ-brest.fr
3. Guillou-Landreat M. ERCR SPURBO, Department of addictology, Université de Bretagne Occidentale, Brest, France. Morgane.guillou@chu-brest.fr
4. Beck-Robert E. ERCR SPURBO, Department of general practice, Université de Bretagne Occidentale, Brest, France. lereste@univ-brest.fr robertemilie@free.fr
5. Assenova R. Department of General Practice, Medical University of Plovdiv, Faculty of Medicine, Plovdiv, Bulgaria. R_assenova@yahoo.com
6. Lazic D. Department of Family Medicine "Andrija Stampar" School of Public Health University of Zagreb School of Medicine, Rockefellerova 4, 10000 Zagreb, Croatia. djlazic@gmail.com
7. Czachowski S. Department of Family Doctor, University Nicolaus Copernicus, Torun, Poland. s.czachowski@to.home.pl
8. Stojanović-Špehar S. PhD/GP Department of Family Medicine "Andrija Štampar" School of Public Health Medical School, University of Zagreb Rockfellerova 4, 10000 Zagreb, Croatia. ssspehar1@inet.hr.
9. Hasaganic M. Department of General Practice, University of Sarajevo, Sarajevo, Bosnia. drmelidah@gmail.com
10. Lingner H. Allgemein Medizin Hochschule Hannover, Hannover, Germany. Lingner.Heidrun@mh-hannover.de
11. Clavería A. Galician Health Services, Instituto de Investigación Sanitaria Galicia Sur,
ABSTRACT

Background:
Few clinical diagnostic tools to rule in or rule out depression have been validated in primary care. From a previous systematic review study, it became clear that in terms of research, a validated clinical diagnostic tool (effective, reliable and usable) was needed by European General Practitioner (GP) investigators in order to include patients for research studies in daily practice care and for cross-sectional studies between psychiatrists and GPs.

Research question: Which clinical diagnostic tool for depression, validated against psychiatric examination according to DSM for primary care adult patients, do GPs select as the best for use in clinical research, taking into account the combination of efficiency, reliability and usability?
Method: A systematic literature review followed by a consensus procedure in two Delphi rounds with an expert panel meeting inserted (i.e. a RAND Appropriateness Method). For the experts’ group inclusion researchers and/or general practitioners (GPs) from different European countries were carefully included. The systematic literature review extracted tools validated against DSM as standard diagnostic criteria. The effectiveness criterion used was the Youden index was used as an effectiveness criterion and Cronbach’s alpha as the criterion of reliability. Usability data were extracted from the literature (structure, method of collection, duration...).

Results: The literature review compared 7 diagnostic tools validated against psychiatric examination according to DSM. In the first Delphi round, two instruments were considered sufficiently effective and reliable for use: the Hospital Anxiety and Depression Scale (HADS) and the Hopkins Symptoms Checklist-25 (HSCL-25). Usability was tested during the expert panel meeting. In the second Delphi round, experts selected the HSCL-25 according to 3 criteria: efficiency, reliability and usability and it was this last criterion that was considered particularly effective for HSCL-25 when compared with the HADS.

Conclusion: A multicultural consensus on one diagnostic tool for depression has been obtained: the clinical tool HSCL-25. This tool will provide the opportunity to select homogeneous populations across Europe to undertake collaborative clinical research in daily practice, between GPs and between GPs and psychiatrists.

BACKGROUND

Depression is the leading cause of disability throughout the world, affecting 350 million people. Primary care is a strategic place for screening, diagnosis, and treatment of depressive disorders [1]. Depressive disorders will increase its already high burden on OCDE countries by 2030 [2]. For the patient, depression causes a significant decline in his/her ability to cope and aggravates his/her psychological distress [3]. Risk of suicide, closely linked to depression, is a dramatic potential consequence of depressive disorders [4][5]. Depression is a recurrent problem for public health care, because of its social and economic consequences [6]. Several studies stress that an early diagnosis is very important; it may decrease the duration and impact of the depression [7].

We have a triple challenge:
- try to improve in daily practice this early detection of depression
- to provide simple and effective means that allow medical research in daily practice.
- agreeing to use this tool, regardless of the countries.

In general practice:

General Practitioners (GPs) are the first port of call in the health care system in most European countries [8][9]. Diagnosing major depression has a high specificity, but a low
sensitivity in Primary Care Medicine [10]. For 25 years, academics in psychiatry and primary care assume that only 10–50% of depressed patients are adequately treated, primarily because of the failure to recognize depression. Those facts could be related to the patient’s reluctance to express their symptoms and the variability of the symptomatology of depression [11]. However, another barrier to recognition is that the traditional psychiatric interview is impractical in primary care settings because it takes too much time [12]. In this context, some tools can support physicians in their diagnosis of depression but most tools are not suitable for primary care use. In primary care, the time of medical consultation is short. An efficient tool has to be sensible and usable to facilitate depression diagnosis. Moreover, a good tool is highly needed for pragmatic research in primary care. [12]

In medical research:

There are common selection criteria: efficiency, stability and usability. Moreover, the tool must be consensually accepted by researchers and have face validity. It must be validated to predict psychiatric referrals well and should be accepted by the two professional communities [13]. These are important conditions for pragmatic research in primary care. [12]

In collaborative research:

The purpose was also to develop collaborative research about depression, in primary care throughout Europe. Research networks such as EGPRN need reliable, efficient and usable tools which take into account cultural and linguistic differences [14]. With this common goal, European primary care researchers decided to work together to find such a tool. International experts from different cultures, speaking different languages and with different health systems have to seek consensus [15]. These tools have to be acceptable and informative for both GPs and for secondary care (Psychiatrist, Psychologist) and to improve their collaboration [13][16].

In our recent systematic review of literature covering the last fifteen years, we found seven tools for psychiatric examination, which were validated according to the major depression criteria in DSM IV-5, usable in primary care research, and conceptually understandable by GPs and psychiatrists [17].

Based on these criteria, our research question was what consensual diagnostic tool for depression, GPs researchers would select for use in clinical research as the most efficient, reliable and feasible.?
METHOD

Design: a RAND/UCLA process to define the best depression tool for general practice research.

EGPRN has developed a research agenda in 2010 to facilitate a common structured approach. Through the EGPRN, European primary care researchers have been recruited from national experts. A network of researchers was created with this objective: to find a cross-sectional diagnostic tool usable by both primary care and psychiatry, allowing collaborative research between countries of different cultures and languages. Experts had to be academic researchers or teachers, FPs and fluent in English.

A systematic literature review covering the last fifteen years, found seven tools usable in primary care research, conceptually understandable by GPs and psychiatrists [17]. Their psychometric properties, sensitivity (Se), specificity (Sp), positive and negative predictive values (PPV, NPV) do not vary sufficiently to allow statistical comparison, however. The study populations were different, and the reliability and usable qualities of each tool had to be considered as insufficient researched.

To come up with the best possible tool, we needed an evaluation and selection process which would ensure experts can work independently but with opportunities for discussion. [18]. The RAND/UCLA process is as a consensus method which combines independence and debate among experts and was selected by the study’s scientific committee as the best possible research process for this study [19].

The RAND/UCLA Appropriateness Method (RAM) combines the qualities of the Delphi process [20] and the nominal group [21]. Developed in the eighties, it allows a consensual choice in comparing complex processes. In medical research it has been used to measure the quality of care procedure [19]. It combines, after a literature review, two Delphi rounds with a meeting of experts. This process combines the advantages of the focus group (communication, face-to-face debate and interaction about ratings) and the quality of Delphi (independent evaluation by experts, working blind, in order to remove any leader effect). Seven to fifteen experts are necessary to achieve RAM [22][19]. The quality level of RAM is increased when they use in their procedure the results of a systematic review instead of a simple review.

We followed 3 steps.

First step:

The study started by a Delphi: (i) in order to eliminate the least efficient tool, (ii) in order to keep the more reliable.
Each expert received: introduction letter, study flow chart, study method, efficiency and sample and reliability data, consent form. They had to rate the efficiency and reliability of each tool on a 9 point Likert scale [23]:

- Is this tool an effective aid to diagnose depression in primary care?
- Is this tool a reliable aid to diagnose depression in primary care?

Consensus was defined as at least 70% of the experts rating questions at 7 or above [22]. A tool was selected if it had a result greater than 70% of the two questions. Commentaries were collected in order to structure the experts’ meeting discussion.

**Second step:**
The 2^nd^ step (experts’ meeting) ensured the results of the 1^st^ step and debate, without voting, the usable features of selected tools.

Experts were equipped with the following resources:

- a reminder about the methodology of the study
- the results of the first round, including all the experts’ comments.
- details about the usable features of each tool
- bibliography data
- a list of additional information to be completed by the experts.
- the individual results from each expert and his/her comments.
- 3 notation grids showing the usable features of each tool. One had been filled in at the beginning, another after testing tools and the last one at the end of the experts’ meeting.

All grids were collected and kept by the study committee at the end of the experts’ meeting.

The experts were invited to discuss the results of the first round on efficiency and reliability and whether they agreed with them or not. If more than 70% of the experts agreed with the results of the first round Delphi, consensus on selected tools was achieved.

The tools selected were then presented. The experts were invited to read all documents and to rate the following statements:

- "This tool is easy to use in family practice".
- "This tool could easily be introduced during a consultation".
- "This tool could be understood by patients".
- "I like this tool".
- "Patients could be surprised by this tool".

Then experts were invited, working in pairs, to discuss and test the tools face-to-face. Afterwards each expert was asked individually to rate on a 9 point Likert scale the tools using the same grid, in order to assess whether testing the tools had modified their judgment. Then a discussion about the feasibility of the selected tool was held among all the experts. The moderator focused on improving the experts’ personal expression. The meeting ended with final quotations using the same grid for each tool. The entire meeting was recorded in both video format and as an audio file for ultimate quality control.
No final consensus was required at the end of the meeting, which is in line with the RAM instructions.

Third step:
The goal of the 3rd step was to vote, in the light of the results of the 1st and 2nd steps, for the best tool in terms of efficiency, reliability and usability.

At the end of the experts’ meeting, all discussions were transcribed. Each expert received the transcript independently, incorporated into a synthesis document.

This synthesis document included:
- a reminder on the purpose of the study.
- the results of the systematic literature review
- a reminder of the methodology and the results of the first Delphi round.
- a reminder of efficiency, reliability and sample data of the tools selected at the end of the meeting.
- the full transcript of the discussion among experts with a synthesis of the key points of the discussion.
- the distribution of the results of the latest listing of five statements for each of the selected tools at the end of the discussion.

The final question was: “Which is the most appropriate tool for the diagnosis of depression in adult patients, in Family practice, in Europe, in terms of its Efficiency, Reproducibility and Usability ”. The experts were asked to vote on each tool and to comment on their response.

RESULTS

Eleven experts from 8 European countries participated in the RAND / UCLA. They were all GPs, EGPRN members and fluent in English. The panel was composed of 9 women and 2 men. Of the 11 experts 9 practised in urban areas of more than 5,000 inhabitants and 2 worked in urban areas with 2,000 to 5,000 inhabitants. (Table1).
Table 1: Expert panel- characteristics of participants

<table>
<thead>
<tr>
<th>Experts</th>
<th>Gender</th>
<th>Country</th>
<th>University statement</th>
<th>Number of inhabitants</th>
<th>Office type</th>
<th>International publication number*</th>
<th>Years of practice</th>
<th>Years of research</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F</td>
<td>Spain (Galicia)</td>
<td>Teacher/ Researcher</td>
<td>&gt;5000</td>
<td>FPs group office</td>
<td>15</td>
<td>20</td>
<td>14</td>
</tr>
<tr>
<td>2</td>
<td>F</td>
<td>Spain (Cataluña)</td>
<td>Teacher/ Researcher</td>
<td>&gt;5000</td>
<td>FPs group office</td>
<td>13</td>
<td>22</td>
<td>25</td>
</tr>
<tr>
<td>3</td>
<td>F</td>
<td>Greece</td>
<td>Teacher/ Researcher</td>
<td>&gt;5000</td>
<td>FPs and paramedic group office</td>
<td>14</td>
<td>30</td>
<td>18</td>
</tr>
<tr>
<td>4</td>
<td>M</td>
<td>Italy</td>
<td>Researcher</td>
<td>&gt;5000</td>
<td>FPs group office</td>
<td>23</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>Germany</td>
<td>Researcher</td>
<td>2000 to 5000</td>
<td>Practice stop since 2 years</td>
<td>19</td>
<td>23</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>M</td>
<td>Poland</td>
<td>Teacher/ Researcher</td>
<td>&gt;5000</td>
<td>FPs group office</td>
<td>20</td>
<td>30</td>
<td>12</td>
</tr>
<tr>
<td>7</td>
<td>F</td>
<td>Croatia</td>
<td>Teacher/ Researcher</td>
<td>&gt;5000</td>
<td>Alone</td>
<td>6</td>
<td>20</td>
<td>12</td>
</tr>
<tr>
<td>8</td>
<td>F</td>
<td>Bosnia</td>
<td>Teacher/ Researcher</td>
<td>2000 to 5000</td>
<td>FPs group office</td>
<td>2</td>
<td>22</td>
<td>12</td>
</tr>
<tr>
<td>9</td>
<td>F</td>
<td>Croatia</td>
<td>Teacher/ Researcher</td>
<td>&gt;5000</td>
<td>FPs group office</td>
<td>18</td>
<td>30</td>
<td>20</td>
</tr>
<tr>
<td>10</td>
<td>F</td>
<td>Bulgaria</td>
<td>Teacher/ Researcher</td>
<td>&gt;5000</td>
<td>FPs group office</td>
<td>9</td>
<td>14</td>
<td>12</td>
</tr>
<tr>
<td>11</td>
<td>F</td>
<td>Germany</td>
<td>Researcher</td>
<td>&gt;5000</td>
<td>FPs group office</td>
<td>4</td>
<td>18</td>
<td>7</td>
</tr>
</tbody>
</table>

* PubMed Database
The tools selected by the literature review were: GDS-5, 15 and 30 items (Geriatric Depression Scale with 5, 15 and 30 items), the HSCL-25 (Hopkins Symptoms Checklist with 25 items), the HADS (Hospital Anxiety Depression Scale), the PSC-51 (physical symptom checklist in 51 items), and the CES-DR (Center for Epidemiologic Studies Depression Scale-Revised).
First step Results:

Table 2: Results of the first Delphi round

<table>
<thead>
<tr>
<th></th>
<th>Efficiency</th>
<th>Reliability</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median (average)</td>
<td>Scores &gt;6 as percentage</td>
<td>Median (average)</td>
</tr>
<tr>
<td>PSC 51</td>
<td>5 (5)</td>
<td>0</td>
<td>7 (6.9)</td>
</tr>
<tr>
<td>GDS 30</td>
<td>4 (3.6)</td>
<td>0</td>
<td>7 (7.3)</td>
</tr>
<tr>
<td>CES DR</td>
<td>4 (3.8)</td>
<td>0</td>
<td>8 (8.1)</td>
</tr>
<tr>
<td>GDS 15</td>
<td>8 (7.7)</td>
<td>100</td>
<td>6 (6.6)</td>
</tr>
<tr>
<td>GDS 5</td>
<td>7 (7.4)</td>
<td>91</td>
<td>2 (1.8)</td>
</tr>
<tr>
<td>HADS</td>
<td>7 (7.2)</td>
<td>90.9</td>
<td>7 (7.4)</td>
</tr>
<tr>
<td>HSCL 25</td>
<td>7.5 (7.3)</td>
<td>82</td>
<td>9 (8.5)</td>
</tr>
</tbody>
</table>

PSC-51 and GDS-30 and CES-DR were eliminated for lack of efficiency. The main reasons were low Youden index, selection bias in studies, too many lost during follow up. GDS-15 and GDS-5 were considered efficient but unreliable. The main reasons were: disparity in Cronbach’s alpha value for GDS-15, only a reliability study for GDS-5. HADS and HSCL-25 were considered efficient and reliable. Population and sample were considered of lower quality for HSCL-25 compared with HADS.

Second step results:

8 experts participated actively in the whole process and had the opportunity to express their opinions. 6/8 of the experts agreed with the results of the first Delphi round and confirmed that HSCL-25 and HADS were the best-validated tools in terms of efficiency and reliability.

Table 3: The development of quotations during the experts’ meeting
### Tools

<table>
<thead>
<tr>
<th>Tools</th>
<th>Statements put to the experts</th>
<th>Scores &gt;6 as percentage on a 9 point Likert scale</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>First quotation: After reading only usable data</td>
</tr>
<tr>
<td>HADS</td>
<td>This tool is easy to use in FP's practice</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>This tool could easily be introduced during a consultation</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>This tool could be understood by patients</td>
<td>37,5</td>
</tr>
<tr>
<td></td>
<td>I like this tool</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Patients could be surprised by this tool</td>
<td>75</td>
</tr>
<tr>
<td>HSCL-25</td>
<td>This tool is easy to use in FP’s practice</td>
<td>87,5</td>
</tr>
<tr>
<td></td>
<td>This tool could easily be introduced during a consultation</td>
<td>87,5</td>
</tr>
<tr>
<td></td>
<td>This tool could be understood by patients</td>
<td>87,5</td>
</tr>
<tr>
<td></td>
<td>I like this tool</td>
<td>87,5</td>
</tr>
<tr>
<td></td>
<td>Patients could be surprised by this tool</td>
<td>25</td>
</tr>
</tbody>
</table>

Before the usability was tested, the experts favoured HADS, but their individual points of view were modified after testing HSCL-25 face-to-face (Table 3). No consensus was collected at the end of the meeting on any single tool, in line with the RAM instructions.

The whole comments were collected and they were returned to the experts in the document that were sent to them for the 3rd phase (as example):

*For HADS:*
- The questions are difficult for patients to understand; the answers are difficult for patients because they correspond to positive and negative choices; this tool is too long.
- The answers require explanation and reformulation for patients.
- HADS seems best for psychotherapists and HSCL-25 looks better for patients.
For HSCL-25:
- It’s a self-administered, easy tool and seems to be the most transferable to DSM; the results are detailed.
- The answers are on 1 to 4 Likert scale; the responses are recorded by checking on a table; the answers are simpler than HADS.
- HSCL-25 has already been tested across Europe, specifically among refugee populations; it is usable for both screening and diagnosis.

Third step results:

To the 8 experts who participated in the whole procedure, they were asked to vote by answering the following question:
"Which is the most appropriate tool to diagnose depression in adult patients in Family practice, in Europe, in terms of its effectiveness, its reliability and its ease of use?"

- 6/8 of the experts answered, "In my opinion, HSCL-25 is the most appropriate tool to diagnose depression in Primary Care practice."
- 2/8 of the experts answered, "In my opinion, HADS is the most appropriate tool to diagnose depression in Primary Care practice."

No comments were received for this final step, however, the experts gave, as example:
- "After analysing all the psychometric properties, the most useful test in primary care in many countries in Europe, with numerous cultural variations is HSCL-25."
- "In my view, HSCL-25 is the easiest to use in practice and the easiest for patients to understand."
- "In terms of efficiency, reliability and usability, HSCL-25 is my first choice. However, I must add that HADS is the best known and most commonly applied tool in clinical practice as well as in scientific discussions between different medical and non-medical professionals. In communication and discussion with our colleagues, it is crucial for the monitoring of depressed patients; we have to think about this if we choose HSCL-25."
- "HSCL-25: Simple, detailed enough for the diagnosis, administration time low, easy to understand."

DISCUSSION

In our study, the HSCL 25 appeared to be the most interesting tool for diagnosing depression in primary care. Most international experts stated that HSCL-25 was the most appropriate tool, in terms of its efficiency, reliability and feasibility. We used a RAND/UCLA methodology, based on an systematic literature review [17] and a rapid review [24]. This method is consequently a modified RAM of higher quality than the original one that requires a non-systematic literature review. In the health care and especially in primary care, feasibility
of a tool is an important criterion in establishing or maintaining a relationship between patient and practitioner. Researchers demonstrated, by this consensus process, how feasibility was a decisive factor in choosing a tool with a view to future research in primary care [25].

The present study confirmed the usefulness and relevance of the HSCL-25 in primary care research. Originally developed for non-traumatized populations, it was later adapted and widely used for evaluation among traumatized populations [26]. A 2009 study in Sweden showed that HSCL-25 had comparable results between subjects from a Scandinavian population and subjects from a population of immigrants from the Middle East [27]. This study suggested that HSCL-25 could be used for population surveys within multicultural groups in western Europe [28]. HSCL-25 has been used and tested many times in primary care situations, compared with the diagnosis of psychiatric disorder carried out by Primary Care practitioners [29][30]. It has robust efficiency and reliability scores [31–33].

The HADS is a tool that has been widely used over a long time for clinical purposes and for research [34]. It is widely used and has been translated into several languages [35]. It is a usable and validated for use in primary care [36, 37]. It has proved its value in the communication it allows between medical specialities [36]. Nevertheless this tool seems complicated for research purposes in primary care daily practice. Its psychometric qualities are more suitable for screening than as ‘a case-identifier’ tool for depression [38].

The GDS or GDS-30 was developed specifically to detect depression in elderly patients [39]. GDS had been rejected in 2 shorter versions of the tool: GDS-15 and GDS-5. The efficiency of those versions was acceptable but the reliability was not. The shorter the version, the lower the reliability [40–44]. Experts did not select these scales because their efficiency and reliability data were inadequate. The Youden index for the GDS-30, validated against DSM-IV, was poor. The Cronbach’s alpha was too low to allow GDS-15 and GDS-5 to be selected. PSC-51 and CES-DR were considered but the Youden Index score was too low and they were quickly discarded [45].

Through this consensus procedure, we suggest the HSCL-25 as an efficient, reliable and easy-to-use tool with good psychometric qualities. This confirms its value in clinical use and for research. Today in Europe, the movement of refugee populations is important. HSCL-25 has already been used in clinical settings both for refugee and general populations, [46, 47]. This has already allowed English-speaking populations and practitioners to use it. For non-English speaking populations, translations will be of great importance. The usability of HSCL-25 will reassure students of the feasibility of such tests in primary care. Teachers could use this validated tool as the basis of their courses for depression diagnosis in FM. They should have enough confidence in the reliability and efficiency data to teach it to their students.
An efficient, reliable and usable tool such as HSCL-25 allows collaborative research, multi-centered in daily practice throughout Europe. This cross-sectional tool could allow transversal research between psychiatry and primary care. Nevertheless its translation into most European languages should be carefully undertaken, which is now the main task of the research group. The group will take great care, as the process involves a self-administrated questionnaire, that the general population can easily understand the tool language.

Our methodology had limitations that we took into account and controlled. The experts’ panel quality was important for the quality level of the RAM. All experts were researchers, FPs and fluent in English. The panel conformed to the requirements of variability in culture, language and practice. 4 families of language were represented: Germanic, Slavic, Hellenic and Romance. The panel size was sufficient (7 to 15 experts) [19]. The deadlines for the Delphi rounds were short in order to block communication between experts. Each judgment was performed blind, as far as possible[48]. In order to reduce information bias, each expert received a record combining all bibliographic sources of the data provided. The reliability data were mainly based on Cronbach's alpha values. Those values were extracted using an additional literature rapid review inducing an information bias. A systematic literature review could have been more discriminating but that was not necessary for the RAM.

The tools found in literature were not anonymized during any of the procedure. The judgment of each expert could possibly take into account his knowledge or practice. The experts’ opportunity for debate during meetings controlled this confusion bias, in line with the RAM instructions.

Conclusion:

We recommend the HSCL-25 as the best possible tool, in terms of efficiency, reliability and usability, to diagnose depression within European primary care practice in a research perspective. Its translation into many European languages will allow collaborative research. The research group will carefully undertake those translations. Its application in practice is possible in English but will have to be demonstrated (following translation) in the translated languages.

List of abbreviations and definitions:

DSM – Diagnostic and Statistical Manual of Mental Disorders
EGPRN - European General Practice Research Network
SRL – Systematic Review of literature
RAND – Research And Development
RAM – RAND Appropriateness Method
RAND/UCLA – Research and Development / University of California Los Angeles
NPV – Negative Predictive Value
PPV – Positive Predictive Value
Se – Sensitivity
Sp - Specificity

Competing interests:
The study had a Grant of 8000 Euros from the EGPRN. The authors have no financial competing interests to declare.

Authors’ contributions section:
NP designed the study, collected data, led meetings, drafted the article and submitted it for publication. LRJY designed the study, collected data, attended meetings and reviewed the article. GLM reviewed the article. BRE collected data and moderated meetings. AR participated in RAND/UCLA. LD participated in RAND/UCLA. CS participated in RAND/UCLA and reviewed the article. SSS participated in RAND/UCLA. HM participated in RAND/UCLA. LH participated in RAND/UCLA. CA participated in RAND/UCLA. FSMMI participated in RAND/UCLA. SA participated in RAND/UCLA. AS participated in RAND/UCLA. LC participated in RAND/UCLA. LFB reviewed the article. DC participated in RAND/UCLA. MT reviewed the article and gave final approval for the version to be published. VMH designed the study, reviewed the article, and gave final approval for the version to be published. VRP designed the study, reviewed the article, and gave final approval for the version to be published.

Acknowledgements:
We would like to thank all FPs who participated in the research process throughout Europe and all trainees in Family Medicine from Brest University who participated in the research process and Mrs. Alex Gillman for her accurate translations.

Bibliography


