One consensual depression diagnosis tool to serve many countries: a challenge! A RAND/UCLA methodology


This version is available from Sussex Research Online: http://sro.sussex.ac.uk/id/eprint/102014/

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.
One consensual depression diagnosis tool to serve many countries: a challenge!
A RAND / UCLA methodology.

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), Claveria A11 (MD, PhD), Fernandez San Martin MI12 (MD, PhD), Sowinska A13 (MD, PhD), Argyriadou S14 (MD), Lygidakis C15 (MD), Le Floc'h B16 (MD), Doerr C17 (MD), Montier T18 (MD, PhD) and Van Marwijk H19 (MD, PhD) and Van Royen P20 (MD, PhD).

1. Nabbe P. EA 7479 SPURBO, Department of general practice, Université de Bretagne Occidentale, Brest, France. patrice.nabbe@univ-brest.fr
2. Le Reste JY. EA 7479 SPURBO, Department of general practice, Université de Bretagne Occidentale, Brest, France. lereste@univ-brest.fr
3. Guillou-Landreat M. EA 7479 SPURBO, Department of addictology, Université de Bretagne Occidentale, Brest, France. Morgane.guillou@chu-brest.fr
4. Beck-Robert E. EA 7479 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, School of Medicine, University of 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. Department of Family Medicine "Andrija Štampar" School of Public Health, University of Zagreb, Croatia. ssspehar1@inet.hr
9. Hasaganic M. Department of General Practice, University of Sarajevo, Bosnia. drmelidah@gmail.com
10. Lingner H. Allgemein Medizin Hochschule Hannover, Germany. Lingner.Heidrun@mh-hannover.de
11. Claveria A. Galician Health Services, Instituto de Investigación Sanitaria Galicia Sur, Vigo, Spain. Ana.Claveria.Fontan@sergas.es
12. Fernández San Martin MI. IDIAP jordi GOL unitat de support a la recerca, Barcelona, mifsanmartin.bcn.ics@gencat.cat
13. Sowinska A. Department of English, Nicolaus Copernicus University, Torun, Poland.
14. Argyriadou S. The Greek Association of General Practitioners (ELEGEIA), Thessaloniki, Greece. starge@otenet.gr

15. Lygidakis C. Associazione Italiana Medici di Famiglia (AIMEF), Bologna, Italy. lygidakis@gmail.com

16. Le Floch B. EA 7479 SPURBO, Department of general practice, Université de Bretagne Occidentale, Brest, France. blefloch1@univ-brest.fr

17. Doerr C. Allgemein Medizin Hochschule Göttingen, Germany. christa@doedoerr.de

18. Montier T. Unite INSERM 1078, SFR 148 ScInBioS, Faculté de Médecine, Université de Bretagne Occidentale, Université Européenne de Bretagne, 22 Avenue Camille Desmoulins, 29238 Brest Cedex 2, France. Tristan.Montier@univ-brest.fr

19. Van Marwijk H. Division of Population Health, Health Services Research and Primary Care, School of Health Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Williamson Building, Oxford Road, Manchester M13 9PL, United Kingdom. Harm.vanmarwijk@manchester.ac.uk

20. Van Royen P. Department of Primary and Interdisciplinary Care. Faculty of Medicine and Health Sciences. Universiteit Antwerpen, Belgium. paul.vanroyen@uantwerpen.be
3 MESSAGES

- The HSCL-25 is a usable depression diagnosis tool in daily practice
- A primary care research network, a chance for collaborative research

ABSTRACT

Objective

From a systematic literature review (SLR), it became clear that a consensual validated tool was needed by European researchers General Practitioners (GPs) in order to allow multi-centred collaborative research, in daily practice throughout Europe. Which diagnostic tool for depression, validated against psychiatric examination according to the DSM, would GPs select as the best for use in clinical research, taking into account the combination of effectiveness, reliability and ergonomics? A RAND/UCLA, which combines the qualities of the Delphi process and of the nominal group, was used. Researchers GPs from different European countries were selected. The SLR extracted tools validated against the DSM. The Youden index was used as an effectiveness criterion, Cronbach’s alpha as a reliability criterion. Ergonomics data were extracted from the literature. Ergonomics was tested face-to-face.

Results

The SLR extracted 7 tools. 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). After testing face-to-face, HSCL-25 was selected. A multicultural consensus on one diagnostic tool for depression was obtained for the HSCL-25. This tool will provide the opportunity to select homogeneous populations for European collaborative research in daily practice.
INTRODUCTION

Primary care is a strategic place for screening, diagnosis, and treatment of depression [1][2][3][4][5]. This led to a triple challenge:
- Trying to improve early detection of depression.
- Providing a simple and effective diagnostic tool that allows medical research in daily practice.
- Gaining consensus on the use of this tool, irrespective of nationality.

For medical research, there are common selection criteria: effectiveness, reliability and ergonomics. Moreover, the tool must be consensually accepted by researchers and have face validity. It must be validated to indicate when psychiatric referral is required and should be accepted by both psychiatrists and General Practitioners (GPs) [6]. Under the auspices of the European General Practice Research Network (EGPRN), European GPs researchers have decided to work to find such a tool. Experts from different cultures, speaking different languages and with different health systems, looked for consensus [7]. These tools have to be acceptable and informative for both GPs and for Psychiatrist and to improve their collaboration [6][8].

Seven tools were found, using a systematic literature review. They were validated against a psychiatric examination using DSM's major depression criteria, usable in primary care research, conceptually understandable by GPs and psychiatrists [9]. However, this method of tool selection has excluded tools such as PHQ [10]. Nevertheless, the object of the research was to have a consensual tool between psychiatrists and GPs. The research team had to reject tools that were not jointly validated by GPs and psychiatrists.

Based on these criteria, the research question was: which diagnostic tool for depression would GP researchers select as the most efficient, reliable and ergonomic for use in clinical research?

MAIN TEXT

Method

European primary care researchers were recruited from EGPRN national experts. Experts had to be academic researchers or teachers, GPs and fluent in English.

The psychometric properties, among the seven tools found in the SRL were extracted [9]. However their properties, sensitivity (Se), specificity (Sp), positive and negative predictive values (PPV, NPV) did not vary sufficiently to allow statistical comparison. The study populations were different. Consequently the reliability and usable qualities of each tool had to be considered.
A selection process was needed to ensure that experts could work independently but with opportunities for discussion [11]. Accordingly a RAND / UCLA process was selected by the study’s scientific committee as the best possible research process for this study [12].

The RAND/UCLA Appropriateness Method (RAM) following a literature review, combines the qualities of a Delphi process [13] with a nominal group [14]. It allows a consensual choice in comparing complex processes [12]. Seven to fifteen experts are necessary [15][12]. The quality level is increased when the results of a systematic review are used in the procedure instead of a narrative review.

First step:

The study started with a Delphi procedure in order to eliminate the less efficient and to keep the more reliable tools. Each expert received: the introductory letter; study flow chart; study method; effectiveness, sample and reliability data; a consent form. They had to rate the effectiveness and reliability of each tool on a 9-point Likert scale [16]:
- Is this tool an effective aid for the diagnosis of depression in primary care?
- Is this tool a reliable aid for the diagnosis of depression in primary care?

Consensus was defined as at least 70% of the experts rating questions at 7 or above [15]. A tool was selected if it scored higher than 70% on each question. Comments were collected in order to structure the experts’ discussion meeting.

Second step:

The 2nd step (experts’ meeting) confirmed the results of the 1st step and allowed debate, without voting, on the usable features of selected tools. Experts were equipped with the following resources: a reminder about the methodology, the results of the first round including all comments, details about the usable features, bibliography data, notation grids on a 9-point Likert scale. The first was filled in at the beginning, another after testing tools and the last one at the end of the experts’ meeting.

The experts were invited to discuss the results of the first round 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 was considered achieved.

The tools selected were then presented. The experts were invited to rate the following statements:
- "This tool is easy to use in general 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 to test the tools face-to-face, working in pairs. Afterwards, each expert was asked individually to rate the tools, in order to assess whether testing them had modified his/her judgment. Then a discussion was held about the ergonomics. The meeting ended with final quotations. 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, in accordance with RAM instructions.

Third step:

The goal was to vote for the best tool. At the end of the experts’ meeting, all discussions were transcribed. Each expert received the transcript independently.

The final question was: “Which is the most appropriate tool for the diagnosis of depression in adult patients, in General Practice, in Europe, in terms of its Effectiveness, Reproducibility and Ergonomics”. The experts were asked to vote on each tool and to comment on their responses.

Results

Eleven experts from 8 European countries participated. They were all GPs 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. (Table 1).

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 Symptom 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:

The PSC-51, GDS-30 and CES-DR were eliminated for lack of effectiveness. The GDS-15 and GDS-5 were eliminated for lack of reliability. The HADS and the HSCL-25 were considered efficient and reliable (Table 2).

Second step results:

Eight experts participated and confirmed that the HSCL-25 and the HADS were the best-validated tools in terms of effectiveness and reliability.
Before the ergonomics were tested, the experts had favoured HADS. Their individual points of view were modified after testing the HSCL-25 face-to-face (Table 3). Consensus was not sought at the end of the meeting.

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

**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.

**HSCL-25**: The answers are on a 1 to 4 Likert scale; the responses are recorded by checking on a table; the answers are simpler.

**Third step results:**

The 8 experts who participated in the whole procedure were asked to vote:
"Which is the most appropriate tool to diagnose depression in adult patients in General Practice, in Europe, in terms of its effectiveness, its reliability and its ease of use?"
- 6 answered, "In my opinion, the HSCL-25 is the most appropriate tool to diagnose depression in Primary Care practice."
- 2 answered, "In my opinion, the HADS is the most appropriate tool to diagnose depression in Primary Care practice."

The experts gave final comments (for example):
- "After analysing all the psychometric properties, the most useful test in primary care in many countries in Europe, with numerous cultural variations, is the HSCL-25."
- "In terms of effectiveness, reliability and ergonomics, the HSCL-25 is my first choice. However, I must add that the 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 the HSCL-25."
- "The HSCL-25: Simple, detailed enough for the diagnosis, short administration time, easy to understand."

**Discussion**

The HSCL-25 appeared to be the most interesting tool for diagnosing depression in terms of the combination of its effectiveness, reliability and ergonomics. It is a self-rating scale derived from the SCL-90 which is a multidimensional psychological test instrument for the assessment of psychological symptoms and psychological distress [17][18][19]. It has robust effectiveness and reliability scores [20][21][22].

With a RAND/UCLA methodology, based on a systematic literature review [9], which is of higher quality than the original RAM with a non-systematic literature review. The ergonomic factor of a tool was an important criterion in maintaining a relationship between patients and
GPs. Researchers demonstrated by this process, how ergonomics was a decisive factor in choosing a tool with a view to future research [23].

On one hand, HSCL 25 has been widely used for evaluation among traumatised populations and used many times in primary care [24][25][26][27][28]. On the other hand, HADS has been widely used over a long period for clinical and research purposes [29]. It has been translated into several languages [30] and validated for use in primary care. Nevertheless, HADS seemed complicated for research purposes in daily practice [31][32][33].

The PSC-51, the CES-DR [34] and the GDS (GDS-30) were considered but the effectiveness was too low. The GDS was developed specifically to detect depression in elderly patients [35]. It was rejected in the 2 shorter versions: GDS-15 and GDS-5 as their reliability was too low [36][37][38][39][40].

As an efficient, reliable and usable tool, the HSCL-25 will allow multi-centred collaborative research, in daily practice throughout Europe. HSCL-25 could allow transversal research between psychiatrist and GPs. The group will take great care as the process involves a self-administered questionnaire that the general population can easily understand.

In Conclusion, the HSCL-25 was the best combination in terms of effectiveness, reliability and ergonomics, for the diagnosis of depression within European primary care practice in a research perspective. Its translation into several European languages will allow collaborative research. Subsequently, application in practice must be demonstrated for each translation in the countries concerned.

**LIMITATIONS**

The quality of the panel was important for the overall quality level. The panel conformed to the requirements of variability in culture, language and practice. 4 language families were represented: Germanic, Slavic, Hellenic and Romance. The panel size was sufficient (7 to 15 experts) [12]. The deadlines for the Delphi rounds were short. Each judgment was performed blind [41]. In order to reduce information bias, each expert received a record of all the 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 review [42].

The tools found in literature were not anonymised. The judgment of each expert could possibly take his knowledge into account. Nevertheless, the experts’ opportunity for debate during meetings controlled this confusion bias.

An original selection bias could be discussed, given the systematic review of literature. From the outset, the gold standard was the psychiatric examination based on DSM’s major depression criteria. Tools with a high level of validity, such as PHQ [43], but which did not use this gold standard as its starting point, could not be selected. The objective of the SRL was to focus on
the tools and the list did not have to be exhaustive. It would be interesting to do another study using another gold standard, such as the Hamilton test [44] and compare the results.

**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

**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. 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.

**Consent to publish**

All authors of the manuscript have read and agreed to its content and are accountable for all aspects of the accuracy and integrity of the manuscript in accordance with ICMJE criteria. That the article is original, has not already been published in a journal, and is not currently under consideration by another journal.

**Acknowledgements**

We would like to thank all GPs who participated in the research process throughout Europe and all trainees in General Practice from Brest University who participated in the research process and Mrs. Alex Gillman our proof-reader for her accurate translations.

**Competing interests**

The authors have no financial competing interests to declare.

**Availability of data and materials**

The datasets used and analyzed during the current study are available from the corresponding
author on reasonable request.

Ethics Approval and consent to participate
The entire study obtained the ethical agreement of the CPP (Protection of Persons Committee) of the University Hospital of Brest; (ID RCB: n°2014-A01790-47; Référence CPP: CPP Ouest VI 872; N° enregistrement Clinical Trial.gov: NCT02414711). All study participants signed a consent form.

Funding
The study had a Grant of 8000 Euros from the EGPRN.

Bibliography


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>8</td>
<td>F</td>
<td>Bosnia</td>
<td>Teacher/ Researcher</td>
<td>2000 to 5000</td>
<td>GPs group office</td>
<td>2</td>
<td>22</td>
<td>12</td>
</tr>
<tr>
<td>10</td>
<td>F</td>
<td>Bulgaria</td>
<td>Teacher/ Researcher</td>
<td>&gt;5000</td>
<td>GPs group office</td>
<td>9</td>
<td>14</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>9</td>
<td>F</td>
<td>Croatia</td>
<td>Teacher/ Researcher</td>
<td>&gt;5000</td>
<td>GPs group office</td>
<td>18</td>
<td>30</td>
<td>20</td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>Germany</td>
<td>Researcher</td>
<td>2000 to 5000</td>
<td>Stopped practising 2 years ago</td>
<td>19</td>
<td>23</td>
<td>5</td>
</tr>
<tr>
<td>11</td>
<td>F</td>
<td>Germany</td>
<td>Researcher</td>
<td>&gt;5000</td>
<td>GPs group office</td>
<td>4</td>
<td>18</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>F</td>
<td>Greece</td>
<td>Teacher/ Researcher</td>
<td>&gt;5000</td>
<td>GPs 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>GPs group office</td>
<td>23</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>M</td>
<td>Poland</td>
<td>Teacher/ Researcher</td>
<td>&gt;5000</td>
<td>GPs group office</td>
<td>20</td>
<td>30</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>F</td>
<td>Spain (Cataluña)</td>
<td>Teacher/ Researcher</td>
<td>&gt;5000</td>
<td>GPs group office</td>
<td>13</td>
<td>22</td>
<td>25</td>
</tr>
<tr>
<td>1</td>
<td>F</td>
<td>Spain (Galicia)</td>
<td>Teacher/ Researcher</td>
<td>&gt;5000</td>
<td>GPs group office</td>
<td>15</td>
<td>20</td>
<td>14</td>
</tr>
</tbody>
</table>

* PubMed Database
Table 2: Results of the first Delphi round

<table>
<thead>
<tr>
<th></th>
<th>Effectiveness</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>91</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>
<tr>
<td>Tools</td>
<td>Statements put to the experts</td>
<td>Scores &gt;6 as percentage on a 9-point Likert scale</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>--------------------------------</td>
<td>-----------------------------------------------</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td></td>
<td>First quotation: After reading only usable data</td>
<td>Second quotation: After testing and discussion of the questionnaires in pairs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HADS</td>
<td>This tool is easy to use in GP’s practice</td>
<td>50</td>
<td>12.5</td>
</tr>
<tr>
<td></td>
<td>This tool could easily be introduced during a consultation</td>
<td>25</td>
<td>12.5</td>
</tr>
<tr>
<td></td>
<td>This tool could be understood by patients</td>
<td>37.5</td>
<td>12.5</td>
</tr>
<tr>
<td></td>
<td>I like this tool</td>
<td>25</td>
<td>12.5</td>
</tr>
<tr>
<td></td>
<td>Patients could be surprised by this tool</td>
<td>75</td>
<td>62.5</td>
</tr>
<tr>
<td>HSCL-25</td>
<td>This tool is easy to use in GP’s practice</td>
<td>87.5</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>This tool could easily be introduced during a consultation</td>
<td>87.5</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>This tool could be understood by patients</td>
<td>87.5</td>
<td>62.5</td>
</tr>
<tr>
<td></td>
<td>I like this tool</td>
<td>87.5</td>
<td>87.5</td>
</tr>
<tr>
<td></td>
<td>Patients could be surprised by this tool</td>
<td>25</td>
<td>0</td>
</tr>
</tbody>
</table>