

TITLE

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

ABREVIATED TITLE

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. ERRCR SPURBO, Department of general practice, Université de Bretagne Occidentale, Brest, France. nabbe.patrice@univ-brest.fr
2. Le Reste JY. ERRCR SPURBO, Department of general practice, Université de Bretagne Occidentale, Brest, France. lereste@univ-brest.fr
3. Guillou-Landreat M. ERRCR SPURBO, Department of addictology, Université de Bretagne Occidentale, Brest, France. Morgane.guillou@chu-brest.fr
4. Beck-Robert E. ERRCR 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 Rockefellerova 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,

- Rosalía Castro 21, 3601 Vigo, Spain. Ana.Claveria.Fontan@sergas.es
12. Fernández San Martín MI. IDIAP Jordi Gol unitat de support a la recerca, Barcelona, Spain. mifsanmartin.bcn.ics@gencat.ca
 13. Sowinska A. Department of English, Nicolaus Copernicus University, Torun, Poland. sowinska@umk.pl
 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. ERCCR SPURBO, Department of general practice, Université de Bretagne Occidentale, Brest, France. blefloch1@univ-brest.fr
 17. Doerr C. Allgemein Medizin Hochschule Göttingen, Göttingen, Germany. christa@docdoerr.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, 1 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

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 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 2nd step (experts' meeting) ensured the results of the 1st 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 et 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

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

* 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

	Efficiency		Reliability		Conclusion
	Median (average)	Scores >6 as percentage	Median (average)	Scores >6 as percentage	
PSC 51	5 (5)	0	7 (6.9)	80	<i>Eliminated Tools: reliable but not efficient</i>
GDS 30	4 (3.6)	0	7 (7.3)	90	
CES DR	4 (3.8)	0	8 (8.1)	90	
GDS 15	8 (7.7)	100	6 (6.6)	0	<i>Eliminated tools: efficient but not reliable</i>
GDS 5	7 (7.4)	91	2 (1.8)	0	
HADS	7 (7.2)	90.9	7 (7.4)	100	Selected tools: considered both efficient and reliable
HSCL 25	7.5 (7.3)	82	9 (8.5)	100	

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	Statements put to the experts	Scores >6 as percentage on a 9 point Likert scale		
		First quotation: After reading only usable data	Second quotation: After testing and discussion of the questionnaires in pairs	Third quotation: After discussion among all the experts
HADS	This tool is easy to use in FP's practice	50	12,5	12,5
	This tool could easily be introduced during a consultation	25	12,5	12,5
	This tool could be understood by patients	37,5	12,5	12,5
	I like this tool	25	12,5	12,5
	Patients could be surprised by this tool	75	62,5	62,5
HSCL-25	This tool is easy to use in FP's practice	87,5	100	100
	This tool could easily be introduced during a consultation	87,5	75	75
	This tool could be understood by patients	87,5	62,5	75
	I like this tool	87,5	87,5	87,5
	Patients could be surprised by this tool	25	0	0

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 a 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

1. Sharp LK, Lipsky MS: **Screening for depression across the lifespan: a review of measures for use in primary care settings.** *Am Fam Physician* 2002, **66**:1001–1008.
2. Mathers CD, Loncar D: **Projections of global mortality and burden of disease from 2002 to 2030.** *PLoS Med* 2006, **3**:2011–2030.
3. Sartorius N, Ustün TB, Costa e Silva J a, Goldberg D, Lecrubier Y, Ormel J, Von Korff M, Wittchen HU: **An international study of psychological problems in primary care. Preliminary report from the World Health Organization Collaborative Project on “Psychological Problems in General Health Care”.** *Arch Gen Psychiatry* 1993, **50**:819–24.
4. Reynolds CF, Kupfer DJ: **Depression and aging: a look to the future.** *Psychiatr Serv Washingt DC* 1999, **50**:1167–1172.
5. Morvan Y, Prieto A, Briffault X, Blanchet A, Dardennes R, Rouillon F, Lamboy B: **La dépression dépression en France: Prévalence, facteurs associés et consommation de soins.** *Baromètre santé 2005* 2007:459–485.
6. 2001 TWHR: *A Public Health Approach to Mental Health.* 2001.

7. Tiemens BG, Ormel J, Jenner J a, van der Meer K, Van Os TW, van den Brink RH, Smit a, van den Brink W: **Training primary-care physicians to recognize, diagnose and manage depression: does it improve patient outcomes?** *Psychol Med* 1999, **29**:833–845.
8. Verhaak PFM, van den Brink-Muinen A, Bensing JM, Gask L: **Demand and supply for psychological help in general practice in different European countries: access to primary mental health care in six European countries.** *Eur J Public Health* 2004, **14**:134–40.
9. Kringos D, Boerma W, Bourgueil Y, Cartier T, Dedeu T, Hasvold T, Hutchinson A, Lember M, Oleszczyk M, Pavlic DR, Svab I, Tedeschi P, Wilm S, Wilson A, Windak A, Van der Zee J, Groenewegen P: **The strength of primary care in Europe: an international comparative study.** *Br J Gen Pract* 2013, **63**.
10. Mitchell AJ, Vaze A, Rao S: **Clinical diagnosis of depression in primary care: a meta-analysis.** *Lancet* 2009, **374**:609–619.
11. Jorm AF: **Mental health literacy. Public knowledge and beliefs about mental disorders.** *Br J Psychiatry* 2000, **177**:396–401.
12. Lake CR, Baumer J: **Academic psychiatry's responsibility for increasing the recognition of mood disorders and risk for suicide in primary care.** *Curr Opin Psychiatry* 2010, **23**:157–66.
13. Zhang J, Patel VL, Johnson TR, Shortliffe EH: **A cognitive taxonomy of medical errors.** *J Biomed Inform* 2004, **37**:193–204.
14. Kirmayer LJ: **Psychotherapy and the cultural concept of the person.** *Transcult Psychiatry* 2007, **44**:232–257.
15. Steinert C, Hofmann M, Kruse J, Leichsenring F: **The Prospective Long-Term Course of Adult Depression in General Practice and the Community. A Systematic Literature Review.** *J Affect Disord* 2013.
16. Dezetter A, Briffault X, Bruffaerts R, De Graaf R, Alonso J, König HH, Haro JM, de Girolamo G, Vilagut G, Kovess-Masféty V: **Use of general practitioners versus mental health professionals in six European countries: the decisive role of the organization of mental health-care systems.** *Soc Psychiatry Psychiatr Epidemiol* 2013, **48**:137–149.
17. Nabbe P, Le Reste JY, Guillou-Landreat M, Munoz Perez MA, Argyriadou S, Claveria A, Fernandez San Martín MI, Czachowski S, Lingner H, Lygidakis C, Sowinska A, Chiron B, Derriennic J, Prielec A Le, Floch B Le, Montier T, Van Marwijk H, Van Royen P: **Which DSM validated tools for diagnosing depression are usable in primary care research? A systematic literature review.** *Eur Psychiatry* 2016.
18. Letrilliart L, Vanmeerbeek M: **À la recherche du consensus : quelle méthode utiliser ?** *exercer* 2011, **99**:170–177.
19. Fitch K, Bernstein SJ, Aguilar MD, Burnand B, LaCalle JR, Lazaro P, Loo M van het, McDonnell J, Vader JP, Kahan JP: *The RAND/UCLA Appropriateness Method User's Manual.* 2001.
20. Skulmoski GJ, Hartman FT, Krahn J: **The Delphi Method for Graduate Research.** *J Inf Technol Educ* 2007, **6**:1.
21. Jones J, Hunter D: **Consensus methods for medical and health services research.** *Bmj Clin Res Ed* 1995, **311**:376–380.

22. Bourrée F, Michel P, Salmi LR: **Consensus methods: Review of original methods and their main alternatives used in public health.** *Rev Epidemiol Sante Publique* 2008, **56**:e13–e21.
23. Hassan T, Barnett D: **Delphi type methodology to develop consensus on the future design of EMS systems in the United Kingdom.** *Emerg Med J EMJ* 2002, **19**:155–159.
24. Ganann R, Ciliska D, Thomas H: **Expediting systematic reviews: methods and implications of rapid reviews.** *Implement Sci* 2010, **5**:56.
25. Hignett S, Carayon P, Buckle P, Catchpole K: **State of science: human factors and ergonomics in healthcare.** *Ergonomics* 2013, **56**:1491–503.
26. Oruc L, Kapetanovic A, Pojskic N, Miley K, Forstbauer S: **Screening for PTSD and depression in Bosnia and Herzegovina : validating the Harvard Trauma Questionnaire and the Hopkins Symptom Checklist.** *Int J* 2008, **1**.
27. Tinghög P, Al-Saffar S, Carstensen J, Nordenfelt L: **The association of immigrant- and non-immigrant-specific factors with mental ill health among immigrants in Sweden.** *Int J Soc Psychiatry* 2010, **56**:74–93.
28. Tinghög P, Carstensen J: **Cross-cultural equivalence of HSCL-25 and WHO (ten) wellbeing index: Findings from a population-based survey of immigrants and non-immigrants in Sweden.** *Community Ment Health J* 2010, **46**:65–76.
29. Nettelblatt P, Hansson L, Stefansson CG, Borgquist L, Nordström G: **Test characteristics of the Hopkins Symptom Check List-25 (HSCL-25) in Sweden, using the Present State Examination (PSE-9) as a caseness criterion.** *Soc Psychiatry Psychiatr Epidemiol* 1993, **28**:130–3.
30. Munk-Jørgensen P, Fink P, Brevik JI, Dalgard OS, Engberg M, Hansson L, Holm M, Joukamaa M, Karlsson H, Lehtinen V, Nettelblatt P, Stefansson C, Sørensen L, Jensen J, Borgquist L, Sandager I, Nordström G: **Psychiatric morbidity in primary public health care: a multicentre investigation. Part II. Hidden morbidity and choice of treatment.** *Acta Psychiatr Scand* 1997, **95**:6–12.
31. Sandanger I, Moum T, Ingebrigtsen G, Dalgard OS, Sørensen T, Bruusgaard D: **Concordance between symptom screening and diagnostic procedure: the Hopkins Symptom Checklist-25 and the Composite International Diagnostic Interview I.** *Soc Psychiatry Psychiatr Epidemiol* 1998, **33**:345–54.
32. Strand BH, Dalgard OS, Tambs K, Rognerud M: **Measuring the mental health status of the Norwegian population: a comparison of the instruments SCL-25, SCL-10, SCL-5 and MHI-5 (SF-36).** *Nord J Psychiatry* 2003, **57**:113–118.
33. Veijola J, Jokelainen J, Läksy K, Kantojärvi L, Kokkonen P, Järvelin M-R, Joukamaa M: **The Hopkins Symptom Checklist-25 in screening DSM-III-R axis-I disorders.** *Nord J Psychiatry* 2003, **57**:119–123.
34. Zigmond AS, Snaith RP: **Hospital Anxiety and Depression Scale (HADS):** *Ann Gen Psychiatry* 1983, **67**:361–70.
35. Reda AA: **Reliability and Validity of the Ethiopian Version of the Hospital Anxiety and Depression Scale (HADS) in HIV Infected Patients.** *PLoS One* 2011, **6**:6.
36. Bjelland I, Dahl AA, Haug TT, Neckelmann D: **The validity of the Hospital Anxiety and Depression Scale. An updated literature review.** *J Psychosom Res* 2002, **52**:69–77.

37. Andrews B, Hejdenberg J, Wilding J: **Student anxiety and depression: comparison of questionnaire and interview assessments.** *J Affect Disord* 2006, **95**:29–34.
38. Spinhoven P, Ormel J, Sloekers PP, Kempen GI, Speckens AE, Van Hemert AM: **A validation study of the Hospital Anxiety and Depression Scale (HADS) in different groups of Dutch subjects.** *Psychol Med* 1997, **27**:363–370.
39. Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, Leirer VO: **Development and validation of a geriatric depression screening scale: a preliminary report.** *J Psychiatr Res* 1983, **17**:37–49.
40. Friedman bruce, Heisel M j, Delavan R: **Psychometric Properties of the 15-Item Geriatric Depression.** *J Am Geriatr Soc* 2005, **53**:1570–1576.
41. Chattat R, Ellena L, Cucinotta D, Savorani G, Mucciarelli G: **A study of the validity of different short versions of the geriatric depression scale.** *Arch Gerontol Geriatr* 2001, **Suppl 7**:81–86.
42. D’Ath P, Katona P, Mullan E, Evans S, Katona C: **Screening, detection and management of depression in elderly primary care attenders: the acceptability and performance of the GDS15 and the development of shorter versions.** *Fam Pract* 1994, **11**:260–266.
43. Incalzi RA, Cesari M, Pedone C, Carbonin PU: **Construct validity of the 15-item Geriatric Depression Scale in older medical inpatients.** *J Geriatr Psychiatry Neurol* 2003, **16**:23–28.
44. Van Marwijk HWJ, Wallace P, De Bock GH, Hermans J, Kaptein AA, Mulder JD: **Evaluation of the feasibility, reliability and diagnostic value of shortened versions of the geriatric depression scale.** *Br J Gen Pract* 1995, **45**:195–199.
45. De Waal MWM, Arnold IA, Spinhoven P, Eekhof JAH, Assendelft WJJ, Van Hemert AM, Waal MWM De, Arnold ÆIA, Spinhoven ÆP, Eekhof ÆJAH, Hemert ÆAM Van: **The role of comorbidity in the detection of psychiatric disorders with checklists for mental and physical symptoms in primary.** *Soc Psychiatry Psychiatr Epidemiol* 2009, **44**:78–85.
46. Zhang Z, Shi Z, Wang L, Liu M: **One year later: Mental health problems among survivors in hard-hit areas of the Wenchuan earthquake.** *Public Health* 2011, **125**:293–300.
47. Hinton DE, Pham T, Tran M, Safren SA, Otto MW, Pollack MH: **CBT for vietnamese refugees with treatment-resistant PTSD and panic attacks: A pilot study.** *J Trauma Stress* 2004, **17**:429–433.
48. Elmer F, Seifert I, Kreibich H, Thieken AH: **Delphi method.** *Innovation* 2010, **30**:93–113.