


RESEARCH NOTE

Open Access



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

P. Nabbe^{1*} , J. Y. Le Reste¹, M. Guillou-Landreat², E. Beck-Robert¹, R. Assenova³, D. Lazic⁴, S. Czachowski⁵, S. Stojanović-Špehar⁶, M. Hasanagic⁷, H. Lingner⁸, A. Clavería⁹, M. I. Fernandez San Martin¹⁰, A. Sowinska¹¹, S. Argiriadou¹², C. Lygidakis¹³, B. Le Floch¹, C. Doerr¹⁴, T. Montier¹⁵, H. Van Marwijk¹⁶ and P. Van Royen¹⁷

Abstract

Objective: From a systematic literature review (SLR), it became clear that a consensually validated tool was needed by European General Practitioner (GP) researchers 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. GP researchers from different European countries were selected. The SLR extracted tools were validated against the DSM. The Youden index was used as an effectiveness criterion and Cronbach's alpha as a reliability criterion. Ergonomics data were extracted from the literature. Ergonomics were 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 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.

Keywords: RAND/UCLA appropriateness method, Multicultural consensus, Delphi procedure, Depression diagnosis tool

Introduction

Primary care is a strategic place for depression diagnosis and treatment [1–5]. This led to a triple challenge:

- Improve early diagnosis.
- Provide a simple and effective diagnostic tool that allows medical research in daily practice.
- Gain consensus on the tool's use irrespective of nationality.

For medical research, there are common selection criteria: efficiency, reliability and ergonomics. 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, 7]. Under the auspices of the European General Practice Research Network (EGPRN), European GP researchers decided to find such a tool. Experts representing different cultures, languages and health systems sought consensus [6, 8].

Seven tools were found using a systematic literature review. They needed to be validated against a psychiatric examination using the DSM's major depression criteria, usable in primary care research and conceptually understandable by GPs and psychiatrists [9]. Consequently, this

*Correspondence: patrice.nabbe@univ-brest.fr

¹ EA 7479 SPURBO, Department of General Practice, Université de Bretagne Occidentale, Brest, France

Full list of author information is available at the end of the article

method of selection excluded tools such as PHQ, which are not validated against the DSM [10]. Then it was necessary to select the more reliable, efficient and ergonomic tool.

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

Methods

Criteria to compare

The psychometric properties, (sensitivity, specificity, positive and negative predictive values) of the tools were extracted [9]. They did not vary sufficiently to allow statistical comparison, as the study populations were different. Subsequently, a narrative review was undertaken to extract the reliability data (Cronbach's alpha, Cohen's kappa). The ergonomics were also important, but comparing this aspect of tools was complex due to the number of items, test duration, method of inquiry, score range, etc. A consensus, taking into account quantitative and qualitative criteria, based on an European expert panel, was the only alternative to ensure comparison [11].

Consensus procedure

The RAND/UCLA appropriateness method (RAM) is approved by major institutes, such as the NICE (National Institute for health and Clinical Excellence) in the United Kingdom or the HAS (Haute Autorité de Santé) in France. It was the most appropriate consensus method [12, 13].

Developed in the mid-1980s, it is an instrument to enable the measurement of the overuse and underuse of medical and surgical procedures. It allows a consensual choice in the comparison of complex processes [11].

RAND/UCLA is a "two-round modified Delphi process" which includes a nominal group. The Delphi rounds avoid leader opinion influence; the panel meeting creates the opportunity to discuss ratings and judgments face to face [14] (Fig 1).

Based on the result of a narrative review completed initially, the quality level of the RAM is increased when the results of a systematic review are used [11, 14].

The RAM is one of several methods that was developed to identify the collective opinion of experts [11]. With RAM, repeated assessment is used by all experts to rank relevance, objectivity and homogeneity [13]. The RAM produces appropriateness criteria and quality indicators with face, construct and predictive validity [15].

Experts' panel

The experts' panel was purposively selected from primary care, on research expertise, academic expertise, English level, gender, practice, native culture and language [16].

First step

The study started with a Delphi procedure to eliminate the less efficient and keep the more reliable tools. The comments took into account only validity data, not ergonomics.

Each expert received the study flow-chart; study method; efficiency, sample and reliability data and consent form. They had to rate the efficiency and reliability of each tool on a 9-point Likert scale [17]:

- Is this tool efficient for the diagnosis of depression in primary care?
- Is this tool reliable for the diagnosis of depression in primary care?

Consensus was defined as at least 70% of the experts rating questions at 7 or above [13]. A tool was considered appropriate if it scored higher than 70% on each question. Comments were collected in order to structure the experts' panel meeting.

Second step

The 2nd step (panel meeting) had to confirm the results of the 1st step and allow debate, without voting, resulting in a presentation of the selected tools. The following resources were provided to experts: methodology reminder, first-round results including all comments, ergonomic features, bibliography data and three 9-point Likert scale notation forms. The forms were completed at the beginning, after testing tools, and 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. If more than 70% of the experts agreed with the results, the first Delphi round was considered successful.

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".

Experts were invited to evaluate before and after testing the tools face-to-face in pairs. This was undertaken

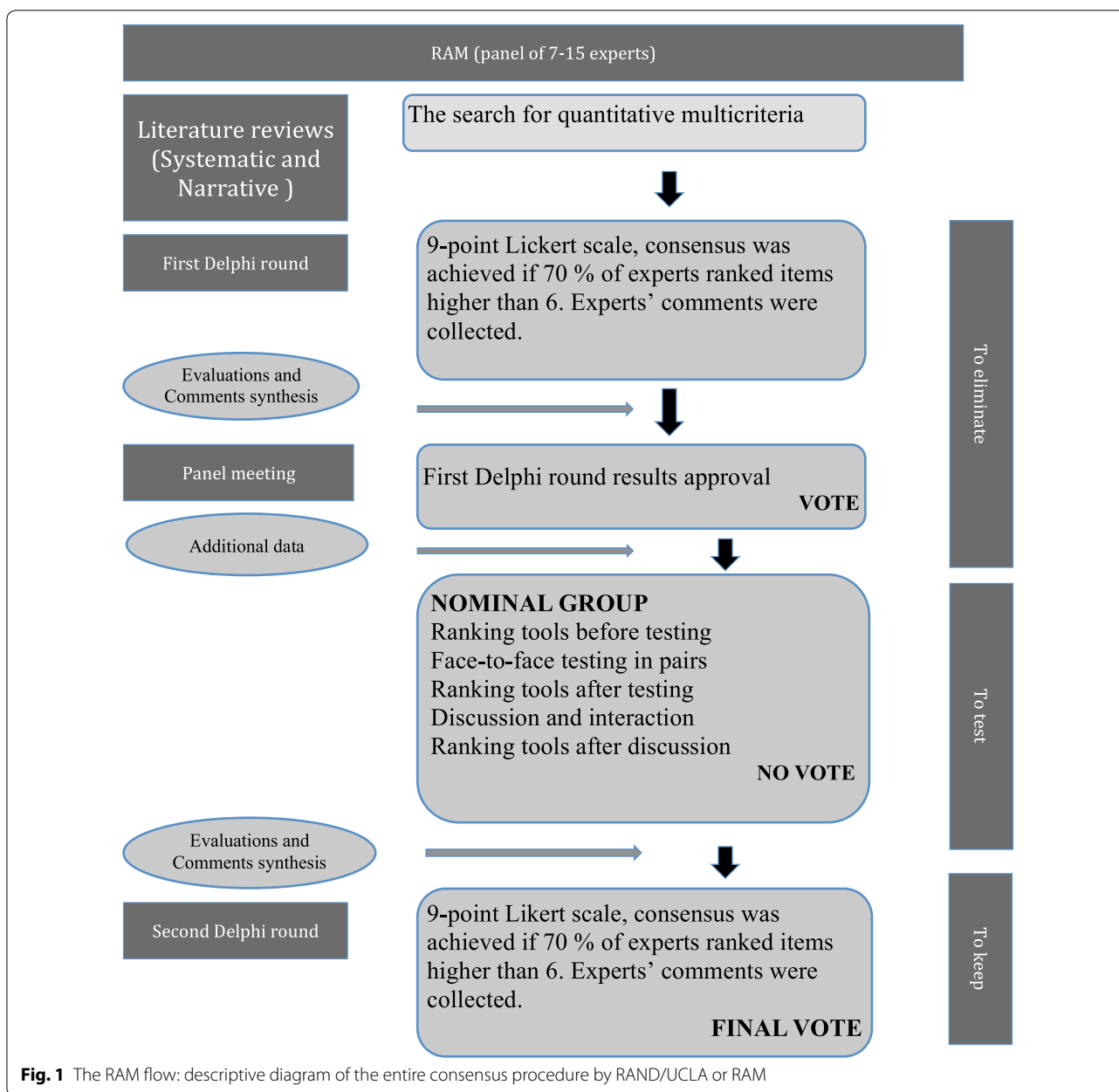


Fig. 1 The RAM flow: descriptive diagram of the entire consensus procedure by RAND/UCLA or RAM

to assess whether testing tools had modified their judgment. Then the ergonomics were discussed. The meeting ended with final evaluations. The entire meeting was recorded in both video and audio format for ultimate quality control.

No final consensus was required at the end of the meeting [11].

Third step

The goal was to select one 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 Efficiency, 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, 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 5000 inhabitants and

2 worked in urban areas with 2000–5000 inhabitants (Table 1).

The tools selected by the literature review were: GDS-5, 15 and 30 (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

The PSC-51, GDS-30 and CES-DR: eliminated for lack of efficiency.

The GDS-15 and GDS-5: eliminated for lack of reliability.

The HADS and the HSCL-25: considered efficient and reliable (Table 2).

Second step results

Eight experts participated and confirmed that HSCL-25 and HADS were the best-validated tools in terms of efficiency and reliability.

Before the ergonomics test, the experts had favoured HADS. Their individual opinions were modified after testing the HSCL-25 face-to-face (Table 3). Consensus was not sought at the end of the meeting.

Table 1 Expert panel-participants' characteristics

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

* PubMed database

Table 2 Results of the first Delphi round

	Efficiency		Reliability		Conclusions
	Median (average)	Scores > 6 as percentage	Median (average)	Scores > 6 as percentage	
PSC 51	5 (5)	0	7 (6.9)	80	Eliminated tools: reliable but not efficient
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	Eliminated tools: efficient but not reliable
GDS 5	7 (7.4)	91	2 (1.8)	0	
HADS	7 (7.2)	91	7 (7.4)	100	Selected tools: considered both efficient and reliable
HSCL 25	7.5 (7.3)	82	9 (8.5)	100	

Table 3 Evaluation progression during the experts' meeting

Tools	Statements put to the experts	Scores > 6 as percentage on a 9-point Likert scale		
		First evaluation: after reading only usable data	Second evaluation: after testing and discussion of the questionnaires in pairs	Third evaluation: after discussion among all the experts
HADS	This tool is easy to use in GP'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 GP'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

All comments were collected and were returned to the experts in the document they were sent 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 efficiency, 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 the most interesting tool for diagnosing depression in terms of the combination of its efficiency, 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 distress [18–20]. It has robust efficiency and reliability scores [21–23].

This RAM study was based on a systematic literature review [9], of higher quality than the original RAM with a non-systematic literature review. The ergonomic factor was an important criterion in maintaining a relationship between patients and GPs. Researchers demonstrated by this process how ergonomics were decisive in choosing a tool suitable for future research [24].

HSCL 25 has been widely used for evaluation among traumatised populations and used many times in primary care [25–29]. HADS has been widely used over a long period for clinical and research purposes [30]; has been

translated into several languages [31] and validated for use in primary care. Nevertheless, HADS seemed complicated for research purposes in daily practice [32–34].

The PSC-51, the CES-DR [35] and the GDS (GDS-30) were considered but efficiency was too low. The GDS was developed specifically to detect depression in elderly patients [36]. It was rejected in the 2 shorter versions: GDS-15 and GDS-5 as reliability was too low [37–41].

In conclusion, the HSCL-25 best combined efficiency, reliability and ergonomics for diagnosis of depression within European primary care practice from a research perspective. It will allow multi-centred collaborative research throughout Europe. HSCL-25 could allow transverse research between psychiatrists and GPs. The group will be vigilant as a self-administered questionnaire must be easily understood by the general population. Its translation into several European languages allows collaborative research. Application in practice must be demonstrated for each national translation.

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 (7–15 experts) [11]. The deadlines for the Delphi rounds were short. Each judgment was performed blind [42]. 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 [43].

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

A systematic literature review creates the possibility of original selection bias. From the outset, the gold standard was the psychiatric examination based on the DSM's major depression criteria. Tools with a high level of validity but which did not use this gold standard as their starting point, such as PHQ [44], could not be selected. The objective of the SRL was to focus on the tools; the list was not exhaustive. It could be worthwhile to initiate a study using another gold standard, such as the Hamilton test [45], and compare results.

Abbreviations

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

NP made substantial contributions to conception and design, acquisition of data, analysis and interpretation of data. He has been involved in drafting the manuscript and also agreed to be accountable for all aspects of the work by ensuring that questions related to the accuracy and integrity of any part of the work were appropriately investigated and resolved. LRJY made substantial contributions to conception and design, acquisition of data, analysis and interpretation of data. He has been involved in drafting the manuscript and revising it critically for important intellectual content. GLM made substantial contributions to conception and design and has been involved in revising it critically for important intellectual content. LD, SSS, HM, LH, CA, FSMMI, SA, SA, LC, CS and DC made substantial contributions to acquisition, analysis and interpretation of data and have been involved in revising it critically for important intellectual content. LFB made substantial contributions to conception and design and has been involved in drafting the manuscript. MT has been involved in revising it critically for important intellectual content and has given final approval for the version to be published. VMH and VRP made substantial contributions to conception and design, have been involved in revising it critically for important intellectual content and have given final approval for the version to be published. All authors read and approved the final manuscript.

Author details

¹ EA 7479 SPURBO, Department of General Practice, Université de Bretagne Occidentale, Brest, France. ² EA 7479 SPURBO, Department of Addictology, Université de Bretagne Occidentale, Brest, France. ³ Department of General Practice, Medical University of Plovdiv, Faculty of Medicine, Plovdiv, Bulgaria. ⁴ Department of Family Medicine "Andrija Štampar", School of Public Health, School of Medicine, University of Zagreb, Zagreb, Croatia. ⁵ Department of Family Doctor, University Nicolaus Copernicus, Torun, Poland. ⁶ Department of Family Medicine "Andrija Štampar" School of Public Health, University of Zagreb, Zagreb, Croatia. ⁷ Department of General Practice, University of Sarajevo, Sarajevo, Bosnia and Herzegovina. ⁸ Allgemein Medizin Hochschule Hannover, Hannover, Germany. ⁹ Galician Health Services, Instituto de Investigación Sanitaria Galicia Sur, Vigo, Spain. ¹⁰ IDIAP Jordi GOL Unitat de Support a la Recerca, Barcelona, Spain. ¹¹ Department of English, Nicolaus Copernicus University, Torun, Poland. ¹² The Greek Association of General Practitioners (ELEGIA), Thessaloniki, Greece. ¹³ Institute for Health and Behaviour, Research Unit INSIDE, University of Luxembourg, Luxembourg, Luxembourg. ¹⁴ Allgemein Medizin Hochschule Göttingen, Göttingen, Germany. ¹⁵ 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. ¹⁶ 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, UK. ¹⁷ Department of Primary and Interdisciplinary Care, Faculty of Medicine and Health Sciences, Universiteit Antwerpen, Antwerp, Belgium.

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 declare that they have no competing interests.

Availability of data and materials

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

Consent for publication

Not applicable.

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: No.

2014-A01790-47; Référence CPP: CPP Ouest VI 872; No. enregistrement Clinical Trial.gov: NCT02414711). All study participants signed a consent form.

Funding

The study had a Grant of 8000 Euros from the European General Practitioner Research Network.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 20 October 2017 Accepted: 21 December 2017

Published online: 03 January 2018

References

- Sharp LK, Lipsky MS. Screening for depression across the lifespan: a review of measures for use in primary care settings. *Am Fam Phys.* 2002;66:1001–8.
- Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med.* 2006;3:2011–30.
- World Health Organization. The World Health Report 2001: Mental Health: new understanding. New Hope: World Health Organization; 2001.
- 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.
- 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:e742.
- Zhang J, Patel VL, Johnson TR, Shortliffe EH. A cognitive taxonomy of medical errors. *J Biomed Inform.* 2004;37:193–204.
- 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;48:137–49.
- 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;152:65–75.
- Nabbe P, Le Reste JY, Guillou-Landreau M, MunozPerez MA, Argyriadou S, Claveria A, Fernandez San Martín MI, Czachowski S, Lingner H, Lygidakis C, Sowinska A, Chiron B, Derriennic J, Le Prielec A, Le Floch B, 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;39:99–105.
- Santos I, Tavares B. Sensitivity and specificity of the Patient Health Questionnaire-9 (PHQ-9) among adults from the general population. *Cad Saúde.* 2013;9:1533–43.
- Fitch K, Bernstein SJ, Aguilar MD, Burnand B, LaCalle JR, Lazaro P, van het Loo M, McDonnell J, Vader JP, Kahan JP. The RAND/UCLA appropriateness method user's manual. Santa monica: Rand corp; 2001.
- HAS, Haute Autorité Santé. Bases méthodologiques pour l'élaboration de recommandations professionnelles par consensus formalisé. Saint-Denis La Plaine: HAS; 2006.
- 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:e13–21.
- Letrilliart L, Vanmeerbeek M. À la recherche du consensus : quelle méthode utiliser? *Exercer.* 2011;99:170–7.
- McGory ML, Shekelle PG, Ko CY. Development of quality indicators for patients undergoing colorectal cancer surgery. *J Natl Cancer Inst.* 2006;98:1623–33.
- Skulmoski GJ, Hartman FT, Krahn J. The delphi method for graduate research. *J Inf Technol Educ.* 2007;6:1.
- 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;19:155–9.
- Derogatis LR, Lipman RS, Rickels K, Uhlenhuth EH, Covi L. The Hopkins Symptom Checklist (HSCL): a self-report symptom inventory. *Behav Sci.* 1974;19:1–15.
- Derogatis LR, Unger R, Derogatis LR, Unger R. Symptom checklist-90-revised. The corsini encyclopedia of psychology. Hoboken: Wiley, Inc.; 2010.
- Lipman RS, Covi L, Shapiro AK. The Hopkins Symptom Checklist (HSCL)—factors derived from the HSCL-90. *J Affect Disord.* 1979;1:9–24.
- 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.
- 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;57:113–8.
- Veijola J, Jokelainen J, Läsky 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;57:119–23.
- Hignett S, Carayon P, Buckle P, Catchpole K. State of science: human factors and ergonomics in healthcare. *Ergonomics.* 2013;56:1491–503.
- Oruc L, Kapetanovic A, Pojskic N, Milek 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;1:105–16.
- 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.
- 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. *Commun Ment Health J.* 2010;46:65–76.
- Nettelbladt 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.
- Munk-Jørgensen P, Fink P, Brevik JJ, Dalgard OS, Engberg M, Hansson L, Holm H, Joukamaa M, Karlsson H, Lehtinen V, Nettelbladt 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.
- Zigmond AS, Snaith RP. Hospital Anxiety and Depression Scale (HADS). *Ann Gen Psychiatry.* 1983;6:361–70.
- 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.
- 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.
- Andrews B, Hejdenberg J, Wilding J. Student anxiety and depression: comparison of questionnaire and interview assessments. *J Affect Disord.* 2006;95:29–34.
- 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;27:363–70.
- De Waal MWM, Arnold IA, Spinhoven P, Eekhof JAH, Assendelft WJJ, Van Hemert AM, De Waal MWM, Arnold AEA, Spinhoven AEP, Eekhof AEA, Van Hemert AEA. 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.
- 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.
- Friedman B, Heisel MJ, Delavan R. Psychometric properties of the 15-item geriatric depression. *J Am Geriatr Soc.* 2005;53:1570–6.

38. 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;33(Suppl 7):81–6.
39. 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;11:260–6.
40. 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;16:23–8.
41. 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;45:195–9.
42. Elmer F, Seifert I, Kreibich H, Thieken AH. Delphi method. *Innovation.* 2010;30:93–113.
43. Ganann R, Ciliska D, Thomas H. Expediting systematic reviews: methods and implications of rapid reviews. *Implement Sci.* 2010;5:56.
44. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA.* 1999;282:1737–44.
45. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry.* 1960;23:56–62.

Submit your next manuscript to BioMed Central
and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

