

This item is the archived peer-reviewed author-version of:

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

Reference:

Nabbe P., Le Reste J.Y., Guillou-Landreat M., Munoz Perez M.A., Argyriadou S., Claveria A., Fernández San Martín M.I., Czachowski S., Lingner H., Lygidakis C., ...- Which DSM validated tools for diagnosing depression are usable in primary care research? A systematic literature review
European psychiatry - ISSN 0924-9338 - 39(2017), p. 99-105
Full text (Publisher's DOI): <https://doi.org/10.1016/J.EURPSY.2016.08.004>
To cite this reference: <http://hdl.handle.net/10067/1422180151162165141>

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

Abbreviated Title:

Which tools are usable for diagnosing depression in primary care?

Corresponding author: Nabbe P, patrice.nabbe@univ-brest.fr, phone: 00 33 6 07 63 14 90

Alternate author: Le Reste JY, lereste@univ-brest.fr

Author's list:

Nabbe P¹(MD), Le Reste JY²(MD, PhD), Guillou-Landreat M³(MD, PhD), Munoz Perez MA⁴(MD), Argyriadou S⁵(MD), Claveria A⁶(MD), Fernández San Martín MI⁷(MD), Czachowski S⁸(MD), Lingner H⁹(MD), Lygidakis C¹⁰(MD), Sowinska A¹¹(PhD), Chiron B¹²(MD), Derriennic J¹³(MD), Le Prielec A¹⁴(MD), Le Floch B¹⁵(MD), Montier T¹⁶ (MD, PhD) and Van Marwijk H¹⁷(MD, PhD) and Van Royen P¹⁸(MD, PhD).

1. Nabbe P. ERCCR SPURBO, Department of general practice, Université de Bretagne Occidentale, Brest, France. nabbe.patrice@univ-brest.fr
2. Le Reste JY. ERCCR SPURBO, Department of general practice, Université de Bretagne Occidentale, Brest, France. lereste@univ-brest.fr
3. Guillou-Landreat M., ERCCR SPURBO, Department of addictology, Université de Bretagne occidentale, Brest, France. Morgane.guillou@chu-brest.fr
4. Munoz M. IDIAP Jordi GOL unitat de suport a la recerca, Barcelona, Spain. mamunoz.bcn.ics@gencat.cat
5. Argyriadou S. The Greek Association of General Practitioners (ELEGEIA), Thessaloniki, Greece. starge@otenet.gr
6. Claveria A. Galician National Health Service, Universidad de Vigo, Rúa Oporta, 1, 36201 Vigo, Spain. Ana.Claveria.Fontan@sergas.es

7. Fernández San Martín MI. IDIAP Jordi Gol unitat de support a la recerca, Barcelona, Spain. mifsanmartin.bcn.ics@gencat.ca
8. Czachowski S. Department of Family Doctor, University Nicolaus Copernicus, Torun, Poland. s.czachowski@to.home.pl
9. Lingner H. Centre for public health and healthcare, Hannover Medical school, Hannover, Germany. Lingner.Heidrun@mh-hannover.de
10. Lygidakis C. Associazione Italiana Medici di Famiglia (AIMEF), Bologna, Italy. lygidakis@gmail.com
11. Sowinska A. Department of English, Nicolaus Copernicus University, Torun, Poland. sowinska@umk.pl
12. Chiron B. ERRCR SPURBO, Department of general practice, Université de Bretagne Occidentale, Brest, France. benoitichiron@yahoo.fr
13. Derriennic J. ERRCR SPURBO, Department of general practice, Université de Bretagne Occidentale, Brest, France. jeremy_derriennic@hotmail.com
14. Le Prielec A. ERRCR SPURBO, Department of general practice, Université de Bretagne Occidentale, Brest, France. anne.leprielec@wanadoo.fr
15. Le Floch B. ERRCR SPURBO, Department of general practice, Université de Bretagne Occidentale, Brest, France. blefloch1@univ-brest.fr
16. 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
17. Van Marwijk H. Primary care research center. Williamson building. Oxford road, Manchester, United Kingdom. Harm.vanmarwijk@manchester.ac.uk
18. Van Royen P. Department of Primary and Interdisciplinary Care. Faculty of Medicine and Health Sciences. Universiteit Antwerpen, Belgium. paul.vanroyen@uantwerpen.be

Abstract

Introduction: Depression occurs frequently in primary care. Its broad clinical variability makes it difficult to diagnose. This makes it essential that family practitioner (FP) researchers have validated tools to minimize bias in studies of everyday practice. Which tools validated against psychiatric examination, according to the major depression criteria of DSM-IV or 5, can be used for research purposes?

Method: An international FP team conducted a systematic review using the following databases: Pubmed, Cochrane and Embase, from 2000/01/01 to 2015/10/01.

Results: The three databases search identified 770 abstracts: 546 abstracts were analyzed after duplicates had been removed (224 duplicates); 50 of the validity studies were eligible and 4 studies were included. In 4 studies, the following tools were found: GDS-5, GDS-15, GDS-30, CESD-R, HADS, PSC-51 and HSCL-25. Sensitivity, Specificity, Positive Predictive Value, Negative Predictive Value were collected. The Youden index was calculated.

Discussion: Using efficiency data alone to compare these studies could be misleading. Additional reliability, reproducibility and ergonomic data will be essential for making comparisons.

Conclusion: This study selected seven tools, usable in primary care research, for the diagnosis of depression. In order to define the best tools in terms of efficiency, reproducibility, reliability and ergonomics for research in primary care, and for care itself, further research will be essential.

Key Words: Depression, Systematic review of literature, Diagnostic tool, Primary Care Research.

BACKGROUND

Depression occurs frequently but it may be difficult to detect and acknowledge in primary care settings, where most patients present with physical symptoms [1–4]. The prevalence rates of depression differ across worldwide (from 2.2% to 10.4%), probably due to conceptual differences and different objectives when diagnosing [5][3, 6–8] and socio demographic factors [9]. There is a large overlap between depression and contextual distress, anxiety and somatoform disorders in primary care [10]. Family practitioners (FPs) experience problems

when diagnosing depression in their patients which may lead to over-prescription of antidepressant drugs. They are the first point of care in most European countries but they seem to be less comfortable with the use of formal diagnostic tools [11].

As FPs try to provide personal, contextual and integrated care, there may be a reluctance to diagnose and use psychiatric labels, such as depression, especially in the context of a somatic illness. These labels may 'separate' the patient with symptoms, such as fatigue, from his or her mental state [12]. Such normalization and diagnostic reluctance may frequently be beneficial for some patients with mild distress but not necessarily for others [13].

We therefore need better knowledge of the tools usable by FPs in the field of depression in adult patients [3]. Several tools exist that help FPs to diagnose depression in adult patients [14]. Identifying the ones that are validated, and evaluating them, will create an opportunity to enhance primary care depression diagnosis. In addition, it will ultimately reduce selection bias and misdiagnosis [15]. It could also improve communication among health professionals if the same tool could be used in primary care (by FPs) and secondary care (by psychiatrists), and improve antidepressant use.

However in accordance with this objective, the field of this research was focused on major depressive disorder according to the DSM. Bipolar depressive disorders, are not covered by the fields in this research, have not concerned the fields in this research [16, 17]. Minor or mood depressive disorders have not been taken into account because the diagnosis is not clearly designed especially in older patients [18].

The European General Practice Research Network (EGPRN) is committed to concepts that could advance research in primary care throughout Europe. The EGPRN has created a research agenda specifically designed for methodological and instrumental research, which includes the development of primary care epidemiology, focusing on patient-centered health. Therefore the EGPRN was specifically interested in the detection of a validated and feasible tool for depression diagnosis in Family Medicine, in order to support collaborative research throughout Europe. An international team, consisting of EGPRN members, was created by co-optation and willingness to participate in this study.

According to a meta-analysis, it seems that the specificity of FPs' depression diagnosis is high and is in accordance with DSM criteria for major depression, even where their sensitivity is low [3]. The choice of the best possible standard for diagnosis was the first stage for the

research team. The standard should be one which can be used by both psychiatry and primary care. It must also take into account a conceptual and cognitive approach which is common to both disciplines [19]. An interviewer-expert, using diagnostic criteria for major depression, according to the Diagnostic and Statistical Manual of mental disorders (DSM), satisfies these two criteria.

The DSM is a classification instrument, a standard categorical tool for research, designed to confirm depression. It describes a number of minor and major symptoms [20–22]. The DSM-5, once bereavement has been removed from the criteria for depression, is a further development from the DSM-IV [15, 16]. The DSM is mainly designed for research purposes and not for everyday practice [25]. Combining DSM with skilled professionals (psychiatrist, psychologists...) creates a robust standard or an external criterion with which to evaluate the tools for research purposes.

Subsequently, the research question was: Which diagnostic tools for depression, tested against a psychiatric examination using DSM IV-5, are usable in primary care research?

METHOD

Systematic review according to the PRISMA Guidelines [26, 27]

Research group constitution:

An international group of researchers in primary care, including a psychiatrist, from France, Spain, Portugal, Catalonia, Italy, Greece, Germany and Poland, was constituted during the EGPRN meetings in Zurich in late 2010. They met several times in order to construct the inclusion/exclusion criteria and research equation.

Step 1: Inclusion of articles

Identification:

The following electronic databases were screened: PubMed, Embase and Cochrane.

The following research equation was used for Pubmed: "Depression"[MeSH Major Topic] AND ("Physicians, Family"[All Fields] OR "General Practitioners"[All Fields] OR "Primary Health Care"[All Fields] OR "Family Practice"[All Fields]) AND ("Tool"[All Fields] OR "Scale"[All Fields] OR "questionnaire"[All Fields] OR "Criteria"[All Fields] OR "screening"[All Fields] OR "Diagnosis"[All Fields]) AND "adult"[MeSH Terms] AND ("2000/01/01"[PDAT]: "2015/10/01"[PDAT]):

This equation was adapted to the characteristics of each database.

A team of 2 international researchers undertook the database document search, working blind and pooling documents at the end of the identification process. They compiled a list of the articles which met the criteria. That list was sent to each national team, including the abstracts, in its own national language, along with a portion of the English abstracts, after duplicates had been removed. Then each national team undertook inclusion/exclusion procedures on these abstracts with 2 national researchers working blind. In addition, a team of two international researchers, working blind, completed the same process of inclusion / exclusion. The two teams of two researchers then compared their results to reach a consensus based on the qualitative criteria of inclusion / exclusion. All eligible abstracts were finally evaluated for identification.

Screening:

Inclusion criteria:

- Limited to the past 15 years (In order to have a comprehensive view of the most recent research).
- Adults and/or elderly patients
- English, Greek, Spanish, Italian, French, German, Polish languages.

Exclusion criteria:

- Not in **IMRaD (Introduction, Methods, Results, and Discussion)** format [28].
- Depression was not the major topic.

- No diagnostic tool identified.
- The study was about children or pregnancy or post-partum depression. Depression is a common complication of post partum. But according to our ethical committee asking and as there is a scientific debate to know if there is a significant difference between depression and perinatal depression, perinatal depression were not retained [29, 30].
- The study was not in primary care setting.
- The tools were identified without validity data.

Eligibility:

A team of 2 researchers extracted the full text articles and sent each national team the articles in their own national language, as well as part of the English articles. Each national team undertook inclusion/exclusion for eligibility. In addition, a team of 2 members of the international research team undertook the same procedure, working blind. Then the two teams of two researchers merged their results to achieve greater reliability. The use of metric data comparison tools such as K-statistic was not possible; studies were not comparable in terms of population and sampling. All articles were finally assessed for eligibility using a qualitative group consensus among the four researchers.

Articles were excluded according to the following criteria:

Depression diagnosis was not the major topic of the study.

Efficiency data (Sensitivity, Specificity, Positive predictive value, Negative Predictive value) were absent or imported from another study.

Reliability was the only mentioned validity data in the article.

Language used in the study was not English, Greek, Spanish, Italian, French, German or Polish.

Researchers were not FPs.

Tool was only validated against another diagnostic tool without a face-to-face psychiatric examination using the DSM IV-5.

Tool was only a screening tool.

Step 2: Data extraction and Selection of tools

A team of two researchers analyzed the included articles. All validated diagnostic tools were extracted. The efficiency data (sensitivity, specificity, positive predictive value, negative predictive value, Youden Index [$Se + Sp - 1$]) were collected. Youden index is an index used for securing optimal thresholds for testing medical tools [31]. Researchers ensured that validity data was calculated on the findings of each individual study and not extracted from elsewhere.

RESULTS

The three databases search identified 770 abstracts: 546 abstracts were analyzed after duplicates had been removed (224 duplicates); 50 of the validity studies were eligible and 4 studies were finally included (Figure 1).

Table 1 shows the reasons for exclusion of abstracts and articles. Finally, seven tools were selected: the 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 (Table 2).

The 'entire, initial sample' of all the studies and the sensitivity and specificity data were collected. PPV and NPV were not always present in the articles. The calculation of the Youden index (sensitivity + specificity -1) enabled a comparison of the effectiveness of the tools to be made. GDS-5, GDS-15, HSCL-25 and HADS had a Youden Index greater than 0.6 (high effectiveness); while GDS-30, CESD-R and PSC-51 had less than 0.6 (Table 2).

Concerning the study of Sanchez-Garcia: this involved a Mexican population, aged 60 to 90 years old; 206 individuals participated among a random sample of 534 individuals from a population of 35,191 individuals. They benefited from a psychiatric interview conducted by a psychiatrist. All validity data sought by the research team were present in the article.

Concerning the study of De Waal: this involved a Dutch population, aged 25 to 80 years old; 473 individuals participated among a randomized sample of 589 individuals from a population of 1046 individuals. They have benefited from an interview with WHO-certified psychologist. PPV and NPV were not present in the article.

Concerning the study of Ortega-Orcos: this concerned a Spanish population, aged over 64 years old; 301 individuals participated and were randomized in a population served by a public primary care center. They were interviewed by trained doctors: a psychiatric interview based on the DSM. All valid data sought by the research team were present in the article.

Concerning the study of Fröjdh: this concerned a Swedish population over 65 years old; 37 individuals participated in sample 58 individuals for a population of 475 individuals. They were interviewed by trained doctors: a psychiatric interview based on the DSM. PPV and NPV were not present in the article.

DISCUSSION

The aim of the study was to find out which diagnostic tools, used for depression diagnosis in primary care, are validated against a psychiatric examination, using major depression criteria, according to DSM **IV-5**. Those tools were: GDS-30 and CESD-R [32], PSC-51 and HADS [33], the GDS-5 and GDS-15 [34], HSCL-25 [35]..

Comparison with existing literature:

Retaining a psychiatric examination based on the DSM was an effective means of comparing the efficiency of the tools [36]. The Youden index gave a robust comparison. It emerged subsequently that the pitfall of this study is that the use of the DSM by a psychiatrist, as comparison criteria, excluded very popular tools. The tools extracted by the literature review were not the tools commonly used in practice. This was intentional as our purpose was to select a tool for research. For example, the 4DSQ [37] is validated against a population-based mathematical model and not against clinical comparison criteria [38]. The PHQ-9, which is also very popular, is a follow-up tool [39], validated against the Hamilton Scale for follow-up

and often used as a diagnostic tool [40]. Nevertheless, the PHQ-9 was never validated against a psychiatric examination, using the DSM, in our team's languages. The PHQ-9 had been validated against DSM-4 in East Africa in 2009 [41] but the language was irrelevant as it did not fall within our criteria.

The research team made choices successively, throughout the entire process, in order to be as accurate as possible and to maintain the ability to communicate with other health professionals. These choices led to the elimination of some popular tools which had certain methodological restrictions preventing their validation according to our search method [19].

The selected tools are categorical and have been little used in everyday family medicine up until now, although this may change rapidly with the introduction of the new primary care mental health nurses in several European countries. On one hand, psychiatrists argue that the difficulty of having to combine validity, utility and disease status [42] in one tool prevents clinicians from using them. On the other hand, FPs are dubious about the validity of DSM for primary care and, therefore, will not use scales [43]. In addition, these tools were mainly developed for research, and not for (general) practice purposes, which might explain their limited use. The GDS-30 was developed in 1982 [44] to diagnose and quantify depression in elderly patients. It was designed with 30 items, using binary response, centered on the previous week's symptoms. It is widely used for research purposes [45, 46]. The GDS-5 and GDS-15 are short versions of the GDS-30 [47, 48] designed for better ergonomic use. The CESD-R was developed in 1977 [49] to diagnose and quantify depression. It was designed with 20 items, using a 4-point Likert scale, centered on the previous week's symptoms. It is also widely used [50] in research. The PSC-51 is a 51-item physical symptoms list. PSC-51 is little used [33]. The HADS was developed in 1983 to diagnose and quantify depression in hospital [51]. It is designed with 14 items, using a 4-point Likert scale, and is centered on the previous week's symptoms. It is a widely used tool [52] in research. The HSCL-25 was developed in 1974 to diagnose and quantify depression [53]. It is designed with 25 items, using a 4-point Likert scale, centered on the previous week's symptoms. It is widely used and specifically used with refugees [54, 55].

Strengths and limitations of the study:

This collaborative work followed a well-defined and rigorous methodology. The broad-based research team consisted of primary care providers or researchers from several countries and cultures. However, not all European countries were represented. Nevertheless, members of this literature review cover a broadly based linguistic range: Romance, Greek, Germanic and Slavic languages. Through a stepwise process, a list of diagnostic tools, usable for depression diagnosis in primary care research and based on the DSM, could be determined.

Selection bias may have occurred but it is limited by the use of a multilingual team, two pairs of two researchers, working blind, at all stages of the selection and inclusion process and also by the wide range of the search equation.

Information bias was possible but limited by the thoroughness of the search. A complete collection of all the summaries and all the full-text articles was assembled. No documents were omitted. The relevant outcomes, such as PPV and NPV, were not always present. The choice of database is debatable the team oriented the search towards a primary care setting.

Confusion bias was limited by using a group consensus procedure to establish the final list at each step (identification, screening, eligibility and inclusion).

Teaching implications:

In family medicine medical education, students are often faced with the question of how to make a depression diagnosis. Many trainees feel that they have difficulty in detecting depression and consequently they do not know whom to treat, whom to follow up and whom to discharge. Even though this study is mainly focused on research, the use of categorical tools will be of great help to those young physicians. They will be able to assess their practice with these tools and to establish robust professional methods for handling depression diagnosis. As always, a tool is only an entry point for the diagnosis and for the conversation with the patient about the labelling of their symptoms. Students have to be taught how to introduce a tool into the consultation; how to stimulate patients to use a tool; how to interpret, discuss and record the results, and subsequently, how to follow up their patients with that help.

Research implications:

The studies collected by the systematic literature review involved adult patients. Only 1 study of 4 has a wide range of include age, between 20 and 80 years. The other 3 studies included a population superior to 60 or 65 years. It may be difficult to extend the results of this study to the entire adult population. However, can we consider that age is the only discriminating factor, given that in Europe the working population from 50 to 64 years represents 1/3 of the active population aged 20 to 64 years [56]? Discrimination of population of future studies should be taken into account other factors and not only the age, such as the ability to cope for example [57]. Future research will need to ensure that these tools have their place in the treatment of adult patients, *inter alia* regardless of their age.

Within the perspective of collaborative studies about depression in primary care, FPs show a good level of specificity in diagnosing depression according to DSM criteria [3] but choosing tools to be shared by FPs and psychiatrists will be a challenge. The choice of a common tool could be based on statistical criteria but the choice could also be influenced by clinical criteria of usefulness [42]. Further research, which applies a standardized methodology, will be necessary to choose the best possible tool, in terms of reliability, efficiency and ergonomics, for undertaking Europe-wide collaborative studies between GPs and psychiatrists [58].

Using only efficiency data could be misleading in the comparison of tools. Therefore, completing this study by researching the reliability data for these tools would have added value, whether this were achieved through the use of the COSMIN statement [59] [48] or by finding additional data on Cronbach's Alpha or Cohen's Kappa in the literature for each tool. It would also be useful to find ergonomics (easy to use) data and that could also be undertaken by means of a literature search. Ergonomics must be taken into account. This is particularly important in primary care because of the importance of the usually limited consultation time.

After collecting these data and analyzing the results, the research team will undertake an expert consensus, using the RAND/UCLA methodology, to find which one of the 7 funded tools is the best, in terms of reproducibility, reliability and ergonomics, for research in primary care.

CONCLUSION

This study selected seven validated tools, usable in primary care, for the diagnosis of depression: GDS-30, CESD-R, PSC-51, HADS, GDS-5, GDS-15, and HSCL-25.

There is need for further research on reliability and ergonomic data for these tools in order to define the best tools in terms of efficiency, reproducibility, reliability and ergonomics for collaborative research in primary care and psychiatry.

List of abbreviations and definitions:

COSMIN - Consensus-based Standards for the selection of health Measurement Instruments

DSM – Diagnostic and Statistical Manual of Mental Disorders

EGPRN - European General Practice Research Network

IMRAD - IMRaD - Introduction, Methods, Results, and Discussion

FP - Family Practitioner

NPV - Negative Predictive Value

PPV - Predictive Positive Value

PRISMA - Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Se - Sensibility

Sp - Specificity

Competing interests:

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

The authors have no non-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, led meetings and reviewed the article. MPMA collected data and reviewed the article. AS collected data and reviewed the article. CA collected data and reviewed the article. FSMMI collected data and reviewed the article. CS collected data and reviewed the article. LH collected data and reviewed the article. LC collected data and reviewed the article. SA collected data and reviewed the article. CB contributed to conception and design. DJ contributed to conception and design. LPA collected data. LFB collected data and reviewed the article. VMH designed the study and reviewed the article, gave final approval of the version to be published. VRP designed the study and reviewed the article, gave final approval of 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.

Funding for this study was provided by the EGPRN.

Bibliography

1. Licht-Strunk E, Windt D Van Der, Van Der Windt DAWM, Van Marwijk HWJ, De Haan M, Beekman ATF: **The prognosis of depression in older patients in general practice and the community. A systematic review.** *Fam Pract* 2007, **24**(December 2006):168–180.
2. Licht-strunk E, Beekman ATF, Haan M De, Marwijk HWJ Van: **The prognosis of undetected depression in older general practice patients . A one year follow-up study.** *J Affect Disord* 2009, **114**:310–315.
3. Mitchell AJ, Vaze A, Rao S: **Clinical diagnosis of depression in primary care: a meta-analysis.** *Lance* 2009, **374**:609–619.
4. Licht-Strunk E, Van Marwijk HWJ, Hoekstra T, Twisk JWR, De Haan M, Beekman ATF: **Outcome of depression in later life in primary care: longitudinal cohort study with three**

years' follow-up. *BMJ Br Med J* 2009, **338**:a3079.

5. Ayuso-Mateos JL, Vázquez-Barquero JL, Dowrick C, Lehtinen V, Dalgard OS, Casey P, Wilkinson C, Lasa L, Page H, Dunn G, Wilkinson G, Ballesteros J, Birkbeck G, Børve T, Costello M, Cuijpers P, Davies I, Diez-Manrique JF, Fenlon N, Finne M, Ford F, Gaité L, Gomez del Barrio A, Hayes C, Herrán A, Horgan A, Koffert T, Jones N, Lehtilä M, McDonough C, et al.: **Depressive disorders in Europe: Prevalence figures from the ODIN study.** *Br J Psychiatry* 2001, **179**(OCT.):308–316.

6. Alonso J, Angermeyer MC, Bernert S, Bruffaerts R, Brugha TS, Bryson H, de Girolamo G, Graaf R, Demyttenaere K, Gasquet I, Haro JM, Katz SJ, Kessler RC, Kovess V, Lépine JP, Ormel J, Polidori G, Russo LJ, Vilagut G, Almansa J, Arbabzadeh-Bouchez S, Autonell J, Bernal M, Buist-Bouwman M a, Codony M, Domingo-Salvany a, Ferrer M, Joo SS, Martínez-Alonso M, Matschinger H, et al.: **Prevalence of mental disorders in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project.** *Acta Psychiatr Scand Suppl* 2004, **109**:21–27.

7. Paykel ES, Brugha T, Fryers T: **Size and burden of depressive disorders in Europe.** *European Neuropsychopharmacology* 2005:411–423.

8. Ustün TB, Ayuso-Mateos JL, Chatterji S, Mathers C, Murray CJL: **Global burden of depressive disorders in the year 2000.** *Br J Psychiatry* 2004, **184**:386–92.

9. Bromet E, Andrade LH, Hwang I, Sampson NA, Alonso J, Girolamo G de, Graaf R de, Demyttenaere K, Hu C, Iwata N, Karam AN, Kaur J, Kostyuchenko S, Lépine J-P, Levinson D, Matschinger H, Mora ME, Browne MO, Posada-Villa J, Viana MC, Williams DR, Kessler RC: **Cross-national epidemiology of DSM-IV major depressive episode.** *BMC Med* 2011, **9**:90.

10. Farvolden P, McBride C, Bagby RM, Ravitz P: **A Web-based screening instrument for depression and anxiety disorders in primary care.** *J Med Internet Res* 2003, **5**:e23.

11. Lehti A, Hammarström A, Mattsson B: **Recognition of depression in people of different cultures: a qualitative study.** *BMC Fam Pract* 2009, **10**:53.

12. Wittkampf KA, Zwieten M Van, Smits FT, Schene AH: **Patients' view on screening for depression in general practice.** *Can J Psychiatry* 2008(August):438–444.

13. Licht CMM, De Geus EJC, Seldenrijk A, Van Hout HPJ, Zitman FG, Van Dyck R, Penninx BWJH: **Depression is associated with decreased blood pressure, but antidepressant use increases the risk for hypertension.** *Hypertension* 2009, **53**:631–638.
14. Williams JW, Pignone M, Ramirez G, Perez Stellato C: **Identifying depression in primary care: a literature synthesis of case-finding instruments.** *Gen Hosp Psychiatry* 2002, **24**:225–37.
15. Murphy JM, Berwick DM, Weinstein MC, Borus JF, Budman SH, Klerman GL: **Performance of screening and diagnostic tests. Application of receiver operating characteristic analysis.** *Arch Gen Psychiatry* 1987, **44**:550–555.
16. Smith DJ, Griffiths E, Kelly M, Hood K, Craddock N, Simpson S a: **Unrecognised bipolar disorder in primary care patients with depression.** *Br J Psychiatry* 2011, **199**:49–56.
17. Phillips ML, Kupfer DJ: **Bipolar disorder diagnosis: challenges and future directions.** *Lancet* 2013, **381**:1663–1671.
18. Gallarda T, Lôo H: **Dépression et personnes âgées.** *Encephale* 2009:269–280.
19. Zhang J, Patel VL, Johnson TR, Shortliffe EH: **A cognitive taxonomy of medical errors.** *J Biomed Inform* 2004, **37**:193–204.
20. APA: *Diagnostic and Statistical Manual of Mental Disorders (4th Ed.)*. 1994:69–81.
21. Mittal VA, Walker EF: **Diagnostic and statistical manual of mental disorders.** *Psychiatry Research* 2011:158–159.
22. Van Weel-Baumgarten EM, Van Den Bosch WJ, Van Den Hoogen HJ, Zitman FG: **The validity of the diagnosis of depression in general practice: is using criteria for diagnosis as a routine the answer?** *Br J Gen Pract J R Coll Gen Pract* 2000, **50**:284–287.
23. Wakefield JC, First MB: **Validity of the bereavement exclusion to major depression: does the empirical evidence support the proposal to eliminate the exclusion in DSM-5?** *World psychiatry Off J World Psychiatr Assoc WPA* 2012, **11**:3–10.
24. Maercker A: **When grief becomes a disorder.** *Eur Arch Psychiatry Clin Neurosci* 2007, **257**:435–436.

25. Zimmerman M, Galione J: **Psychiatrists' and nonpsychiatrist physicians' reported use of the DSM-IV criteria for major depressive disorder.** *J Clin Psychiatry* 2010, **71**:235–238.
26. Counsell C: **Formulating questions and locating primary studies for inclusion in systematic reviews.** *Ann Intern Med* 1997, **127**:380–387.
27. Bland CJ, Meurer LN, Maldonado G: **A systematic approach to conducting a non-statistical meta-analysis of research literature.** *AAMC Acad Med J Assoc Am Med Coll* 1995, **70**:642–653.
28. Sollaci LB, Pereira MG: **The introduction, methods, results, and discussion (IMRAD) structure: a fifty-year survey.** *J Med Libr Assoc* 2004, **92**(July):364–367.
29. O'Hara MW, Zekoski EM, Philipps LH, Wright EJ: **Controlled prospective study of postpartum mood disorders: comparison of childbearing and nonchildbearing women.** *J Abnorm Psychol* 1990, **99**:3–15.
30. Whiffen VE: **The comparison of postpartum with non-postpartum depression: a rose by any other name.** *Journal of psychiatry & neuroscience : JPN* 1991:160–165.
31. Smits N: **A note on Youden's J and its cost ratio.** *BMC Med Res Methodol* 2010, **10**:89.
32. Sánchez-garcía S, Juárez-cedillo T, García-gonzález JJ, Espinel-bermúdez C, Gallo JJ, Wagner FA, Vázquez-estupiñán F, García-peña C: **Usefulness of two instruments in assessing depression among elderly Mexicans in population studies and for primary care.** 2008, **50**:447–456.
33. De Waal MWM, Arnold IA, Spinhoven P, Eekhof JAH, Assendelft WJJ, Van Hemert AM: **The role of comorbidity in the detection of psychiatric disorders with checklists for mental and physical symptoms in primary care.** *Soc Psychiatry Psychiatr Epidemiol* 2009, **44**:78–85.
34. Ortega Orcos R, Salinero Fort MA, Kazemzadeh Khajoui A, Vidal Aparicio S, de Dios del Valle R: **[Validation of 5 and 15 items Spanish version of the geriatric depression scale in elderly subjects in primary health care setting].** *Rev Clin Esp* 2007, **207**:559–562.
35. Fröjd K, Håkansson A, Karlsson I, Fröjd K, Hakansson A: **The Hopkins Symptom Checklist-25 is a sensitive case-finder of clinically important depressive states in elderly**

- people in primary care. *Int J Geriatr Psychiatry* 2004, **19**(August 2003):386–390.
36. Perkins NJ, Schisterman EF: **The Youden Index and the optimal cut-point corrected for measurement error.** *Biometrical J Biometrische Zeitschrift* 2005, **47**:428–441.
37. Terluin B, Van Marwijk HWJ, Adèr HJ, de Vet HCW, Penninx BWJH, Hermens MLM, Van Boeijen CA, Van Balkom AJLM, van der Klink JJJ, Stalman WAB: **The Four-Dimensional Symptom Questionnaire (4DSQ): a validation study of a multidimensional self-report questionnaire to assess distress, depression, anxiety and somatization.** *BMC Psychiatry* 2006, **6**:34.
38. Chambe J, Le Reste J, Maisonneuve H, Sanselme A, Oho-Mpondo J, Nabbe P, Terluin B: **Evaluating the validity of the French version of the Four-Dimensional Symptom Questionnaire with differential item functioning analysis.** *Fam Pract* 2015, **32**:474–9.
39. Kroenke K, Spitzer RL, Williams JBW, L?we B: **The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: A systematic review.** *Gen Hosp Psychiatry* 2010, **32**:345–359.
40. Hamilton M: **A RATING SCALE FOR DEPRESSION.** *J Neurol Neurosurg Psychiatry* 1960, **23**:56–62.
41. Gelaye B, Williams MA, Lemma S, Deyessa N, Bahretibeb Y, Shibre T, Wondimagegn D, Lemenhe A, Fann JR, Vander Stoep A, Andrew Zhou X-H: **Validity of the Patient Health Questionnaire-9 for depression screening and diagnosis in East Africa.** *Psychiatry Res* 2013, **210**:653–61.
42. Kendell R, Jablensky A: **Distinguishing between the validity and utility of psychiatric diagnoses.** *Am J Psychiatry* 2003, **160**:4–12.
43. van Rijswijk E, van Hout H, van de Lisdonk E, Zitman F, van Weel C: **Barriers in recognising, diagnosing and managing depressive and anxiety disorders as experienced by Family Physicians; a focus group study.** *BMC Fam Pract* 2009, **10**:52.
44. 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.
45. Li D, Zhang D, Shao J, Qi X, Tian L: **A meta-analysis of the prevalence of depressive**

symptoms in Chinese older adults. *Arch Gerontol Geriatr* 2014, **58**:1–9.

46. Goring H, Baldwin R, Marriott A, Pratt H, Roberts C: **Validation of short screening tests for depression and cognitive impairment in older medically ill inpatients.** *Int J Geriatr Psychiatry* 2004, **19**:465–471.

47. Marc LG, Raue PJ, Bruce ML: **Screening performance of the 15-item geriatric depression scale in a diverse elderly home care population.** *Am J Geriatr Psychiatry* 2008, **16**:914–921.

48. Weeks SK, McGann PE, Michaels TK, Penninx BWJH: **Comparing various short-form Geriatric Depression Scales leads to the GDS-5/15.** *J Nurs Scholarsh* 2003, **35**:133–137.

49. Radloff LS: **The CES-D Scale: A Self-Report Depression Scale for Research in the General Population.** *Appl Psychol Meas* 1977, **1**:385–401.

50. Van Dam NT, Earleywine M: **Validation of the Center for Epidemiologic Studies Depression Scale--Revised (CESD-R): pragmatic depression assessment in the general population.** *Psychiatry Res* 2011, **186**:128–132.

51. Zigmond AS, Snaith RP: **Hospital Anxiety and Depression Scale (HADS):** *Ann Gen Psychiatry* 1983, **67**:361–70.

52. 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.

53. 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.

54. Ekblad S, Roth G: *Diagnosing Posttraumatic Stress Disorder in Multicultural Patients in a Stockholm Psychiatric Clinic. Volume 185;* 1997.

55. Jones L: **Exposure to Political Violence and Psychological Well-being in Bosnian Adolescents: A Mixed Method Approach.** *Clinical Child Psychology and Psychiatry* 2005:157–176.

56. Levasseur S: **Vieillessement de la population active.** *Rev l'OFCE* 2015, **6**:339–370.

57. Le Reste JY, Nabbe P, Manceau B, Lygidakis C, Doerr C, Lingner H, Czachowski S, Munoz M, Argyriadou S, Claveria A, Le Floch B, Barais M, Bower P, Van Marwijk H, Van

Royen P, Lietard C: **The European General Practice Research Network Presents a Comprehensive Definition of Multimorbidity in Family Medicine and Long Term Care, Following a Systematic Review of Relevant Literature.** *Journal of the American Medical Directors Association* 2013:319–325.

58. 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.

59. Mokkink LB, Terwee CB, Gibbons E, Stratford PW, Alonso J, Patrick DL, Knol DL, Bouter LM, de Vet HCW: **Inter-rater agreement and reliability of the COSMIN (Consensus-based Standards for the selection of health status Measurement Instruments) checklist.** *BMC Med Res Methodol* 2010, **10**(box C):82.

Figure 1: Articles Inclusion (Related to PRISMA guideline)

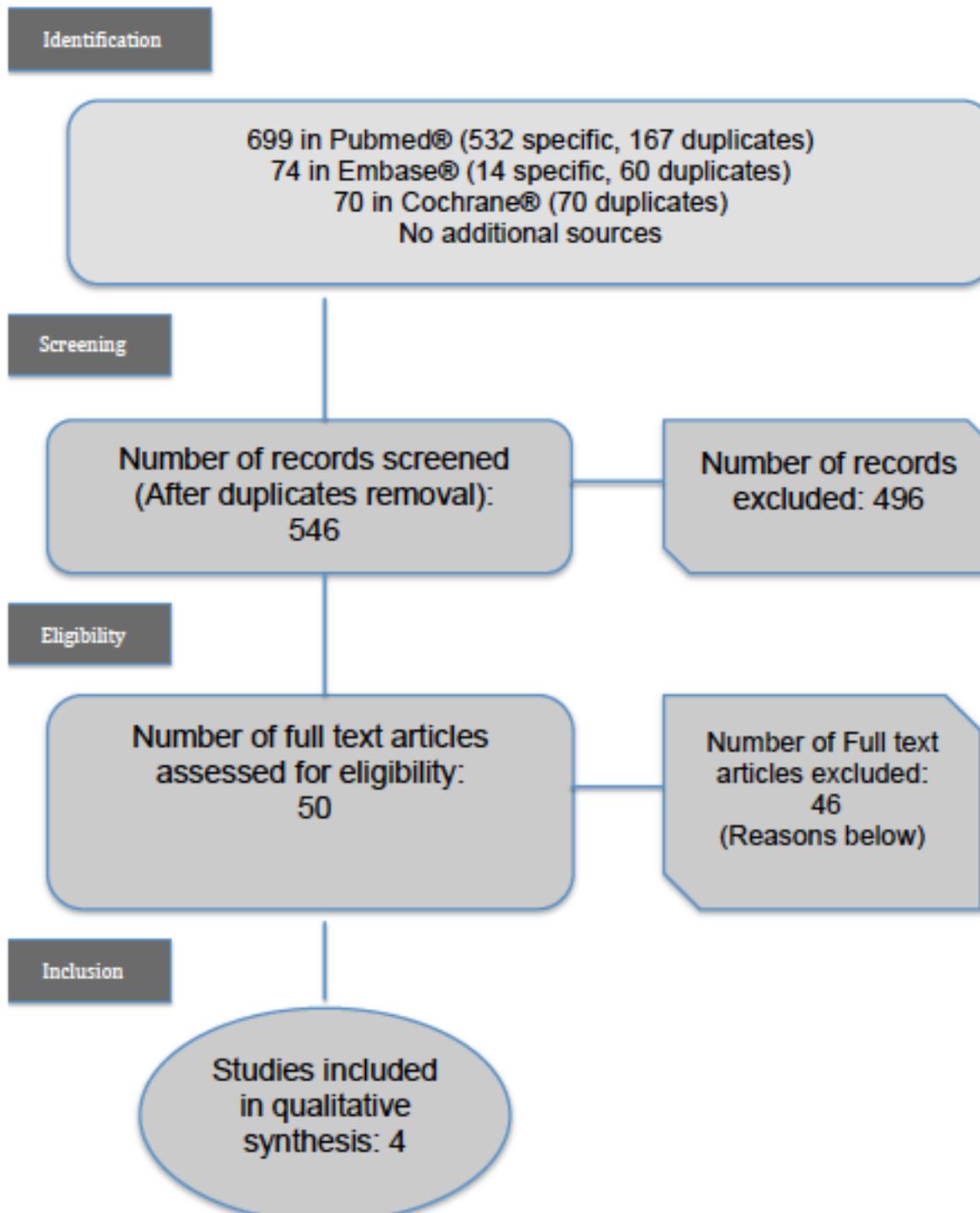


Table 1: Reasons for Abstracts, Articles and Tools Exclusion

Reasons for non eligibility	
Not in IMRAD format	67
Depression was not the major topic	41
No diagnostic tool identified	233
The study was about children or pregnancy or post-partum depression	21
Irrelevant: not pen and pencil and free of charge and not free	1
The study was not in primary care setting	32
The tools were identified without validity data.	458
Reasons for non inclusion	
Depression diagnosis was not the major topic of study	1
Efficiency data were absent or imported from another study,	8
Reliability was the only mentioned validity data in the article	2
Language used in the study is not English, Dutch, German, Polish, Greek, Italian, Spanish, French or Portuguese	8
Researchers are not FPs	0
Tool was exclusively tested against another tool	14
Tool was only validated against another diagnostic tool without a face-to-face psychiatric examination using the DSM IV-5	49
Tool was only a screening tool.	4

Results of exclusion have been summarized. The addition of results showing the number of eligible or included articles is not provided here. This information is given in a flow PRISMA diagram.

Table 2: Validated tools: Bibliographic and validity data

TITLE	First two Authors	Publication Year	Tool	Tool used in interview	Interviewer using DSM-IV or 5 criteria	Individuals	Se	Sp	PPV	NPV	YI	Age range
Usefulness of two instruments in assessing depression among elderly Mexicans in population studies and for primary care.	Sánchez-García S, Juárez-Cedillo T & al.	2008	CES-DR	Semi-structured tool based on the DMS-IV	Yes	206	0.82	0.49	0.50	0.88	0.31	71.2
			GDS				0.54	0.79	0.61	0.74	0.33	
The role of comorbidity in the detection of psychiatric disorders with checklists for mental and physical symptoms in primary care.	De Waal MWM, Arnold IA & al.	2009	HADS	SCAN 2.1 based on DSM-IV	Yes	473	0.65	0.79			0.44	48.8
			PSC-51				0.90	0.59			0.49	
Validation of 5 and 15 items Spanish version of the geriatric depression scale in elderly subjects in primary health care setting.	Ortega Orcos R, Salinero Fort MA & al.	2007	GDS-5	Clinical Diagnosis of Depression (using DSM-IV criteria)	Yes	301	0.86	0.87	0.51	0.97	0.72	74.3
			GDS-15				0.82	0.98	0.86	0.97	0.79	
The Hopkins Symptom Checklist-25 is a sensitive case-finder of clinically important depressive states in elderly people in primary care.	Fröjd K, Hakansso NA & al.	2004	HSCL-25	MADRS based on DSM-IV	Yes	74	0.94	0.94			0.88	78.5

Abbreviation: CES-DR: Center for Epidemiologic Studies-Depression scale; GDS, GDS-15, GDS-5: Geriatric Depression Scale in 30 or 15 or 5 items; HADS: Hospital Anxiety Depression Scale; PSC-51: Physical symptom Checklist in 21 items; SCAN 2.1: schedules for clinical assessment in neuropsychiatry based on DSM-IV diagnoses; HSCL-25: Hopkins Symptom Checklist in 25 items; MADRS: The Montgomery-Asberg-Depression-Rating Scale according DSM-IV criteria; YI: Youden Index; GP: General Practitioner; PPV: Positive Predictive Value; NPV: Negative Predictive Value; Se sensibility; Sp: Specificity.

