Un outil diagnostique consensuel de la dépression pour servir de nombreux pays : un défi !

Patrice Nabbe

UNIVERSITE BIOLOGIE BRETAGNE SANTE LOIRE

UBO Université de Bretagne Occidentale

Universiteit Antwerpen

THESE DE DOCTORAT DE

L'UNIVERSITE
DE BRETAGNE OCCIDENTALE
COMUE UNIVERSITE BRETAGNE LOIRE

ECOLE DOCTORALE N° 605
Biologie Santé
Spécialité : Epidémiologie, Analyse du Risque, Recherche Clinique

EN COTUTELLE AVEC L'UNIVERSITE DE ANVERS

Par

Patrice NABBE

Un outil diagnostique consensuel de la dépression pour servir de nombreux pays: un défi !
One consensual depression diagnosis tool to serve many countries: a challenge !

Thèse présentée et soutenue à Brest, le 5 juillet 2018

Unité de recherche : SPURBO E7479

Rapporteurs avant soutenance :

Lieve PEREMANS PU, Université de Anvers
Hilde BASTIAENS PU, Université de Anvers

Composition du Jury :

Présidente du jury :	Lieve PEREMANS	PU, Université de Anvers
Examinateurs :	Hilde BASTIAENS	PU, Université de Anvers
	Christian BERTHOU	PU, UFR Médecine, UBO
Directeurs de thèse :	Tristan MONTIER	Directeur de thèse, PU-PH, UFR Médecine, UBO
	Paul VAN ROYEN	Directeur de Thèse, PU, Université de Anvers
	Harm VAN MARWIJK	Co-Directeur, PU, Brighton and Sussex Medical School
	Jean Yves LE RESTE	Co-Directeur, PU-PA, UFR Médecine, UBO

Table of Contents

Je ne cherche pas à connaitre les réponses,
Je cherche à comprendre les questions.

Notre plus grande gloire n'est point de tomber,
Mais de savoir nous relever chaque fois que nous tombons.

Confucius

CHAPTER 1

Introduction

In this chapter, an overview will be given of the background and the research question.

Introduction
Depression and primary care
European General Practice Research Network implication in order to select a consensual tool
Constitution of the European team
Pitfalls and objectives, taxonomy of the diagnosis tools and research question
Ethics Approval and consent to participate

INTRODUCTION

Major depression affects 4.4% of the world's population [1][2][3]. Prevalence estimates vary in Europe but are around 10% for people attending general practice, and the prevalence is twice as high for women [4]. An increase of more than 18% was observed between 2005 and 2015 [5].

Depression is a disease comprising contextual distress, anxiety and somatoform disorders. This syndromic disorder is not easy to diagnose, however, due to the wide variety of ways in which it may be presented. Patients themselves experience difficulties to express their suffering and display their own form of illness expression [6]. Based in this inter-individual variability, the difficulties to diagnose and assess the severity of depression, may overestimate or underestimate the distress level of their patients by clinicians. Those difficulties may lead to inappropriate care and cause public health issues [7][8][9][10].

DEPRESSION AND PRIMARY CARE

Depression is the second most common chronic disorder in general practice. In most European countries, General practitioners (GPs) are the first and mostly only physicians to take care of depressed patients, but generally have little time to care. [7][11][12][13] GPs seem to be uncomfortable with depression definition and available diagnostic tools [14][15]. However, GPs have a high specificity but a low sensitivity to detect major depression in routine care. Nevertheless they also offer excellent follow-up and primary care is an efficient place to organize depression care [16][17]. A fast and efficient tool, with an excellent specificity and negative predictive value, would add value and improve performance management in general practice as it would save time.

EUROPEAN GENERAL PRACTICE RESEARCH NETWORK IMPLICATION IN ORDER TO SELECT A CONSENSUAL TOOL

Collaborative primary care mental health models can improve care and outcomes for European outpatients. The Challenge should be to reduce difference between incidence and prevalence rates of depression in General Practice across Europe, due to complex contextual variations with differences in health care systems, concepts, objectives and practices as well as cultural variations in the expression of the disease.

With this aim, the EGPRN (European General Practice Research Network) developed collaborative research in General Practice throughout Europe. The EGPRN requires a reliable, efficient and ergonomic tool, which will take into account cultural and linguistic differences [18][19]. International experts from different cultures, speaking different languages and with different health systems undertook a consensus, to identify such tools [20]. These tools had to be acceptable and informative for both GPs and for secondary care (Psychiatrist, Psychologist) and to improve their collaboration [21]. These diagnosis tool for depression would bring added value to the identification of the condition if have to be undertaken in a routine manner in the doctor's surgery' [22].

CONSTITUTION OF THE EUROPEAN TEAM

In EGPRN meeting of Zurich in October 2010, a first draft was proposed. Researchers from Belgium (University of Antwerp), France (university of Brest), Germany (University of Gottingen and Hannover), Greece (Association of Greek Gps – ELEGEIA), Italy (Association of Italian GPs), Poland (University of Torun), Spain (University of Barcelona and Vigo) and Netherland (VUmc of Amsterdam) composed a research team.

Two years and 4 meeting were necessaries to write the entire protocol. In October 2011 during the EGPRN meeting of Krakow (Poland), researchers from Bosnia (University of Sarajevo) and Croatia (University of Zagreb), interesting by preliminary results, joined the team. In EGPRN meeting of Ljubljana (Slovenia), in spring 2012, researchers from Bulgaria (University of Plovdiv) joined the team and closed the recruitment.

To diagnose depression could seem ridiculously simple, but in daily practice is not so easy, according to the reasons stated upper. The research team was build in order to identify a tool, acceptable by GPs, efficient to break the reluctance to use academics scales.

This tool:
- Had to serve firstly European collaborative research,
- Secondary could be proposed to GPs as an additional way to perform the depression diagnosis.

Researchers should be academics and composed a large multicultural group from different countries. So, an add of value was possible according:
- The linguistic abilities for literature review
- The linguistic abilities for translations
- The different resources of Universities and GPs associations
- The former research network already in use in each country
- The former practice network already in use in each country.

To manage heterogeneous multicultural groups create some difficulties as:
- Comprehension according the use of English language
- Comprehension related to the linguistic, health system, political system differences ...
- Motivation to participate in a long duration study.

To overcome those difficulties, to maintain the group cohesion, each EGPRN meeting was the time to reassure the group, using:
- The ask of the agreement of all teams for each step of research to ensure their participation
- The write of research protocols at each step, sent before meeting
- The checking of the global comprehension of research protocol during each meeting
- The physical and active help of each participant about understanding
- The physical presence to ensure collection and validation of data

- The help all teams for publication
- The assuming publication rank for each team.

According to all these benefits and harms, the European research group was motivated to follow up the entire study during 6 years.

PITFALLS AND OBJECTIVES, TAXONOMY OF THE DIAGNOSIS TOOLS AND RESEARCH QUESTION

To diagnose depression, besides the clinic approach, some tools and algorithms were available for use [23].

The American categorical tool: DSM (Diagnostic Statistical manual) is widely used. It is considered as a Gold Standard [24][25]. From the Beginning, the DSM endeavored to maintain a close relationship with the ICD of the WHO (Internal Classification of Disease of the World Health Organization). [26][27]. To improve its using in practice and research, some structural clinical interviews were built and validated against both DSM and ICD, as the CIDI (Composite International Composite Interview), the SCID (Structured Clinical Interview for DSM disorders) or more recently the MINI (Mini International Neuropsychiatric Interview) [28][29][30].

The DSM and structured clinical interviews are not so easy to use in practice [31]. Some screening tool, some self-report or interview diagnosis tools, more usable in primary care, were validated straight against the DSM or the structured clinical interview [29][32][33][34].

The English longitudinal tool: HDRS or HAM-D (Hamilton Depression rating Scale) [35][36][37][38][39] served at Gold Standard to the BDI (Beck Depression inventory), [40] and more recently to the MADRS (Montgomery-Asberg Depression Scale) [41][42].

An independent Structured clinical interview, centered on primary care, can also be used: the PRIME-MD (Primary Care Evaluation of Mental Disorders) diagnostic

instrument [43]. It evolved on a simple self-questionnaire: the PHQ-9 (Patient Health Questionnaire -9 items) [44][45].
Some recent tools were built independently and were validated, such as 4-DSQ [46]
.
This list was not exhaustive. Faced with this complexity, the researchers began by determining the gold standard selected. The gold standard had to be diagnostic and categorical to diagnose and assess severity of depression. The Gold Standard was to be accepted by the international community and by other specialties including psychiatrics.

About the diagnostic tool, it had to have a direct validation against the gold standard, and not against structured clinical interviews or tools conceived from it.
It had to be accepted by both GPs and psychiatrists.
It had to be embedded in primary care.
It had to be reliable and ergonomic:
- To enable its translation and cultural adaptation easily, on several countries simultaneously, taking part in the study.
- For its use in daily practice.

As requested by the Ethics Committee, and because there is a scientific debate to find out if there is a significant difference between the perinatal form and other forms of depression, perinatal depression has not been retained, like pregnant or postpartum patients [47][48].

Consequently, the research objective of this thesis was to select a tool that could be used and accepted by GPs to diagnose depression in adults' outpatients. It had to be efficient, reliable and easy to use in daily practice. This tool should be applicable in the European countries taking part in the study.

Several steps were necessaries to answer the research question, each step is developed in the chapter two: Method.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

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

BIBLIOGRAPHY

1. World Health Organization: **WORLD HEALTH STATISTICS - MONITORING HEALTH FOR THE SDGs**. *World Heal Organ* 2016:1.121.

2. Filipovic-Pierucci A, Samson S, Fagot J-P, Fagot-Campagna A: **Estimating the prevalence of depression associated with healthcare use in France using administrative databases**. *BMC Psychiatry* 2017, **17**:1.

3. 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.

4. Ayuso-Mateos JL, Vázques-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, Gaite 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.

5. Vos T, Allen C, Arora M, Barber RM, Brown A, Carter A, Casey DC, Charlson FJ, Chen AZ, Coggeshall M, Cornaby L, Dandona L, Dicker DJ, Dilegge T, Erskine HE, Ferrari AJ, Fitzmaurice C, Fleming T, Forouzanfar MH, Fullman N, Goldberg EM, Graetz N, Haagsma JA, Hay SI, Johnson CO, Kassebaum NJ, Kawashima T, Kemmer L, Khalil IA, Kyu HH, et al.: **Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015**. *Lancet* 2016, **388**:1545–1602.

6. Jorm AF: **Mental health literacy. Public knowledge and beliefs about mental disorders.** *Br J Psychiatry* 2000, **177**:396–401.

7. Ani C, Bazargan M, Hindman D, Bell D, Farooq MA, Akhanjee L, Yemofio F, Baker R, Rodriguez M: **Depression symptomatology and diagnosis:**

discordance between patients and physicians in primary care settings. *BMC Fam Pract* 2008, **9**:1.

8. Thomas-MacLean R, Stoppard J, Miedema BB, Tatemichi S: **Diagnosing depression: there is no blood test.** *Can Fam Physician* 2005, **51**:1102–1103.

9. Coventry PA, Hays R, Dickens C, Bundy C, Garrett C, Cherrington A, Chew-Graham C: **Talking about depression: a qualitative study of barriers to managing depression in people with long term conditions in primary care**. *BMC Fam Pract* 2011, **12**:10.

10. Nutting PA, Rost K, Dickinson M, Werner JJ, Dickinson P, Smith JL, Gallovic B: **Barriers to Initiating Depression Treatment in Primary Care Practice**. *J Gen Intern Med* 2002, **17**:103–111.

11. Torzsa P, Szeifert L, Dunai K, Kalabay L, Novák M: **[Diagnosis and therapy of depression in family practice].** *Orv Hetil* 2009, **150**:1684–93.

12. IFOP: **Les Français et le système de santé**. *La Vague* 2013, **FD/ AB N°1**.

13. Kovess-Masféty V, Saragoussi D, Sevilla-Dedieu C, Gilbert F, Suchocka A, Arveiller N, Gasquet I, Younes N, Hardy-Bayle M-C: **What makes people decide who to turn to when faced with a mental health problem? Results from a French survey.** *BMC Public Health* 2007, **7**:188.

14. Alonso J, Codony M, Kovess V, Angermeyer MC, Katz SJ, Haro JM, De Girolamo G, De Graaf R, Demyttenaere K, Vilagut G, Almansa J, Lépine JP, Brugha TS: **Population level of unmet need for mental healthcare in Europe.** *Br J Psychiatry* 2007, **190**:299–306.

15. Demyttenaere K, Bruffaerts R, Posada-Villa J, Gasquet I, Kovess V, Lepine JP, Angermeyer MC, Bernert S, de Girolamo G, Morosini P, Polidori G, Kikkawa T, Kawakami N, Ono Y, Takeshima T, Uda H, Karam EG, Fayyad JA, Karam AN, Mneimneh ZN, Medina-Mora ME, Borges G, Lara C, de Graaf R, Ormel J, Gureje O, Shen Y, Huang Y, Zhang M, Alonso J, et al.: **Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys.** *JAMA* 2004, **291**:2581–90.

16. Hérique A, Kahn J-P: **Réalités et recommandations dans la prescription et l'observance des antidépresseurs en médecine générale : évaluation des pratiques dans le traitement de la dépression en Lorraine et Champagne-Ardenne**. *Encephale* 2009, **35**:73–79.

17. OMS: **CIM-10 à usage PMSI**. 2015:888.
18. Lehti A, Hammarström A, Mattsson B: **Recognition of depression in people of different cultures: a qualitative study**. *BMC Fam Pract* 2009, **10**:53.
19. Kirmayer LJ: **Psychotherapy and the cultural concept of the person.** *Transcult Psychiatry* 2007, **44**:232–257.
20. 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.
21. Zhang J, Patel VL, Johnson TR, Shortliffe EH: **A cognitive taxonomy of medical errors**. *J Biomed Inform* 2004, **37**:193–204.
22. Mitchell AJ: **Clinical utility of screening for clinical depression and bipolar disorder.** *Curr Opin Psychiatry* 2012, **25**:24–31.
23. Mitchell AJ, Vaze A, Rao S: **Clinical diagnosis of depression in primary care: a meta-analysis.** *Lance* 2009, **374**:609–619.
24. Apa: *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV*. *Volume 4th*. American Psychiatric Association; 1994(VI).
25. American Psychiatric Association: **Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5)**. *Diagnostic Stat Man Ment Disord 4th Ed TR* 2013:280.
26. Dalal PK, Sivakumar T: **Moving towards ICD-11 and DSM-V: Concept and evolution of psychiatric classification**. *Indian J Psychiatry* 2009, **51**:310–319.
27. Dickmann CEB, Dickmann JRM, Broocks A, Square R, Klinik CF, Schwerin HK, College B: **Using ICD-10 Criteria ?** *Gen Pract* 2005:3–5.
28. Kessler RC, Andrews G, Mroczek D, Ustun B, Wittchen H-U: **The World Health Organization Composite International Diagnostic Interview short-form (CIDI-SF)**. *Int J Methods Psychiatr Res* 1998, **7**:171–185.
29. First MB, Spitzer RL, Gibbon M, Williams JBW: *Structured Clinical Interview for DSM-IV Axis I Disorders, Clinician Version (SCID-CV)*. American Psychiatric Press; 1997.
30. Sheehan D V, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, Hergueta T, Baker R, Dunbar GC: **The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic**

psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry* 1998, **59 Suppl 2**:22-33-57.

31. 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.

32. Teodorescu DS, Heir T, Hauff E, Wentzel-Larsen T, Lien L: **Mental health problems and post-migration stress among multi-traumatized refugees attending outpatient clinics upon resettlement to Norway**. *Scand J Psychol* 2012, **53**:316–332.

33. Jongenelis K, Pot AM, Eisses AMH, Gerritsen DL, Derksen M, Beekman ATF, Kluiter H, Ribbe MW: **Diagnostic accuracy of the original 30-item and shortened versions of the Geriatric Depression Scale in nursing home patients.** *Int J Geriatr Psychiatry* 2005, **20**:1067–1074.

34. Derogatis LR, Lipman RS, Rickels K, Uhlenhuth EH, Covi L: **The hopkins symptoms checklist (HSCL): A Self-report Sympotoms Inventory**. *Behav Sci* 1974, **19**:1–11.

35. Hamilton M: **A RATING SCALE FOR DEPRESSION**. *J Neurol Neurosurg Psychiatry* 1960, **23**:56–62.

36. Hamilton M: **Rating depressive patients.** *J Clin Psychiatry* 1980, **41**(12 Pt 2):21–4.

37. Grundy CT, Lunnen KM, Lambert MJ, Ashton JE, Tovey DR: **The Hamilton Rating Scale for Depression: One Scale or Many?** *Clin Psychol Sci Pract* 1994, **1**:197–205.

38. Paykel ES: **Use of the Hamilton Depression Scale in general practice.** *psychopharmacol Ser* 1990, **9**:40–7.

39. Bagby RM, Ryder AG, Schuller DR, Marshall MB: **The Hamilton Depression Rating Scale: has the gold standard become a lead weight?** *Am J Psychiatry* 2004, **161**:2163–2177.

40. Beck AT, Steer RA, Carbin MG: **Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation**. *Clin Psychol Rev* 1988, **8**:77–100.

41. Montgomery SA, Asberg M: **A new depression scale designed to be sensitive to change**. *Br J Psychiatry* 1979, **134**:382–389.

42. Williams JBW, Kobak KA: **Development and reliability of a structured interview guide for the Montgomery-Åsberg Depression Rating Scale (SIGMA)**. *Br J Psychiatry* 2008, **192**:52–58.

43. Spitzer RL, Williams JB, Kroenke K, Hornyak R, McMurray J: **Validity and utility of the PRIME-MD patient health questionnaire in assessment of 3000 obstetric-gynecologic patients: the PRIME-MD Patient Health Questionnaire Obstetrics-Gynecology Study.** *Am J Obstet Gynecol* 2000, **183**:759–69.

44. Kroenke K, Spitzer RL, Williams JB: **The PHQ-9: validity of a brief depression severity measure.** *J Gen Intern Med* 2001, **16**:606–613.

45. 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.

46. 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.

47. Grazioli R, Terry DJ: **The role of cognitive vulnerability and stress in the prediction of postpartum depressive symptomatology**. *Br J Clin Psychol* 2000, **39**:329–347.

48. STOWE ZN, HOSTETTER AL, NEWPORT D: **The onset of postpartum depression: Implications for clinical screening in obstetrical and primary care**. *Am J Obstet Gynecol* 2005, **192**:522–526.

CHAPTER TWO

Method

In this chapter, a brief overview will be given of the methodologies used in this thesis to answer the research questions

Overview
Research protocol
Systematic literature review
Consensus procedure: the RAND/UCLA
The translations, a forward-backward procedure supplemented by a cultural control check.
The French study validation

OVERVIEW

Each step of the thesis used a specific method. The first phase was an informal group consensus on the research protocol. The Second was a systematic literature review (SRL) to extract validated tool on depression diagnosis. The third was a consensus procedure, a RAND/UCLA, to select a single tool according efficiency criterion, reliability and ergonomics. The fourth was a translation of the selected tool, in each participating countries languages. A forward/backward translation followed by a cultural check was used to maintain linguistic and semantic stability. The fifth was the validation of the tool, at least in a country, in daily practice, with adults' outpatients.

RESEARCH PROTOCOL

The first goal of the thesis was to have an agreement of all participants on the research protocol. Il was important to have a strong validation of all participants. The main pitfall and barrier for this long research process and thesis should be the abandonment by not motivated participant. From the beginning, all the steps were explained and at each step recalled. The consensus support and active participation of all participants was maintained throughout the study. All members of the team were academic GPs, university members or research organisation member.

The strengths of the group were:
- A high motivation for research in general practice
- The necessity to have publications for their university or organisation
- The membership to the same network (EGPRN)
- The assurance of being independent of the pharmaceutical companies.
- The assurance to be active participants in research
- The follow up of the research group by an academic department of general practice.

The weaknesses were:
- The lack of financial support

- The limited time that every participants coul offer to the research process.

Systematic literature review

The objective of the systematic review was to extract validated diagnosis tools of depression usable in GP.
Upstream, an informal consensus was realised with an international research teams, come from Belgium, France, Germany, Italy, Netherland, Poland, Spain, Greece, composed by GPs' researchers and psychiatrist. The distribution of tasks was allocated to each researcher. Each researcher in his country had to find another researcher to work as a pair during the SLR process.

As describe in chapter one, the family of tools were numerous. The first problem was to choose just one Gold Standard in order to extracted tools related with it. Researchers met several times in order to select the Gold Standard, to construct the inclusion/exclusion criteria and the research equation, to select the indexed databases. The choice of a Gold Standard was crucial because it caused a strong orientation of the overall results of the full study.

To make sure not to make mistake in choosing the Gold Standard, the researchers decided it during the first phase of the SRL (Screening), when all validated tools would be extracted. This first phase gave taxonomy of validated tools for use in primary care. It was revealed that the DSM was the largest gold standard worldwide for validation studies. In addition DSM is a diagnostic tool and categorical, in that wa in line with the study.

At this time, in the second phase, using the DSM has been a major eligibility criterion.

The choice of exclusion and inclusion criterion was underpinned by the following requirements:
- The field of the research was embedded in GP and linked to psychiatrists; adult patients were concerned; the pregnancy and the post partum were not concerned

- Only validation studies with psychometrics efficiency data were concerned, correlation studies were not concerned
- The writing had to be in one of the languages of the experts.

The SRL was according to the PRISMA Guidelines. She respected each step of the flow PRISMA diagram, i.e. identification, screening, eligibility and inclusion [1, 2].

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 [3].

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 the post-partum experience. However, in accordance with the demands of our ethical committee, and because there is a scientific debate to discover whether there is a significant difference between perinatal and other forms of depression, perinatal depression was not retained.[4, 5].

The study was not in a 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.

At this step, the remaining articles were included. The tools validated against face-to-face psychiatric examination using DSM, according major depression criterion, in GP, were identified.
A team of two researchers analysed the included articles. The psychometrics properties were collected. Researchers ensured that validity data was calculated on the findings of each individual study and not extracted from elsewhere. These data will be essential for the next phase of the study.

Consensus procedure: RAND/UCLA

Which diagnostic tool for depression would GP researchers select as the most efficient, reliable and ergonomic for use in clinical research?

Criteria to compare

The psychometric properties, (sensitivity, specificity, positive and negative predictive values) of the tools were extracted [6]. 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. [7]

Consensus procedure

The RAND/UCLA (Research and development / University of California Los Angeles) appropriateness method (RAM) was selected.

It 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. [8][9]

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.[7].

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 [10].

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. [7][10]

The RAM is one of several methods that was developed to identify the collective opinion of experts. [7] With RAM repeated assessment is used by all experts, to rank relevance, objectivity and homogeneity [9]. The RAM produces appropriateness criteria and quality indicators with face, construct and predictive validity [11].

Experts’ panel

The experts’ panel was purposively selected from primary care, on research expertise, academic expertise, English level, gender, practice, native culture and language [12]. It was constituted by the research team, which had been enriched by two new countries, It included those follows countries: France, Germany, Italy, Netherland, Poland, Spain, Greece, English, Greek, German, Croatia, Bosnia.

First step

A Delphi procedure eliminated the less efficient and kept 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 [13]:

- 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 [9]. 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 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 [7].

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.

The translations, a forward-backward procedure supplemented by a cultural control check.

The objective was to translate the selected tool into the languages of the team members, without losing linguistic and semantic stability, and staying within the context of primary care [14][15].

A three-step standardized study was conducted among participating countries, including: (I) a forward translation (FT), (II) a backward translation (BT) and (III) a cultural check [16][17][18].

The FT was conducted with an incorporated Delphi procedure [19][8][20]. It is a rigorous way to reach consensus [21][13][22]. It is a systematic, interactive method which involves a panel of experts using iterative procedures [23].
This process requires:
- Anonymity of participants, which ensures response reliability and avoids contamination,
- Iteration, which allows participants to refine their views in the light of the progress of the group's work,
- Feedback control under the responsibility of the investigator,
- Statistical aggregation of the group’s responses to allow a quantitative and qualitative analysis of the data [9][24][25][26].

The EGPRN French team ensured that the whole process followed the protocol. The FT had to be validated by the daily board of the study, composed of members of the EGPRN, all-active within the research process.

The NIs selected translators to set up two translation teams, which worked blind for both FT and BT. Translators, had to be knowledgeable about healthcare terminology. The FT team involved one member of the FP research group and one

official translator from every country involved. The BT team involved one (or two) FPs and one official translator [14].

The NIs recruited a panel of experts in their own countries, anonymized the experts' responses and allocated an identification number for later identification [9]. Initially, 20 to 30 experts were recruited per country in order to maintain at least 15 participants until the end of the last round. The selection criteria for each FP expert were: being a native of his/her country of residence and speaking his/her native language; being an English speaker [15]. Over half had to be involved in teaching and/or research activities. In order to assess the representativeness of the panel by its maximum variation, the experts provided the following information: their gender, practice setting, years of practice and publications [27].

According to Brislin's Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures, once the FT had been completed, a BT was performed with two goals: (I) to ensure that linguistic translation problems were identified, (II) to collect translation problems which were independent of the linguistic translation itself. Nevertheless, a linguistic translation was insufficient as translation biases related to cultural aspects in each country were possible; this required cultural control to ensure homogeneity [17][18][28][29] [30].

An FP researcher and a PhD linguist analyzed all backward translations and compared them with the original version to establish whether there were any significant difference in terms of meaning. Their report was submitted to a consensus group whose task was to clarify the nature of the anomaly from three problem areas:
A Backward Translation Problem was eliminated if the difference was explained by an incorrect back translation.
A Forward Translation Problem was defined as an anomaly in transcribing the original English.
A Cultural Effect was considered validated if there were no linguistic problems with the translation but where the item needed a modification to be understood by the patients in their own "everyday" language.

At the end of the forward/backward translation and the cultural effect control check, a linguistically stable, definitive translation, which maintained meaning, was produced for each country. This encompassed the structure and order of the questions, item by item, as well as the method of use.

The French study validation

The selected tool has been translated into the languages of team members. If the guarantees have been taken to ensure the mean stability of the tool during the translation process, we must also ensure that the psychometric qualities have little varied. A validation study of the translated tool should be done in daily practice, within GP surgeries, to outpatients.
The overall design of the study imposed from the beginning a close correspondence between GPs and psychiatrists. It was therefore necessary that GPs and researchers psychiatrists work together to drive the study.

A quantitative cross-validation study in an adult French general practice population was carried by the research team of the Soins primaires, Santé Publique, Registre des tumeurs de Bretagne Occidentale (EA 7479 SPURBO). It was a comparative, non-inferiority, multi-centred, survey. The study team constituted of two physician researchers, three GP trainees trained in psychiatric assessment using a structured clinical interview, a psychiatrist, a statistician, 20 GPs, a Data Manager and a Research Coordinator.

The study was carried out in northern Finistère (Brittany, France). The population was a mix of patients from urban, semi-rural and rural environments. In the waiting-room, patients were given a leaflet explaining the study, a questionnaire and a consent form. The participants made the recruitment spontaneously after reading the explanatory notice.

<u>Inclusion criteria</u>

The patients needed to be adults (over 18 years of age). Patients had to give their written informed consent to participate. They completed the questionnaire and submitted it to the study team.

Exclusion criteria

To avoid possible cases of puerperal depression, which requires specific management, women with a reported pregnancy were not included in the study [31][32][33]. Also excluded were adults consulting for a medical certificate, patients with schizophrenia or related disorders and patients requiring emergency care.

The sample size was calculated according depression prevalence in general population. A structured clinical interview was use as reference. Two groups of outpatients should be compared, depress and non depress and allow the calculation of psychometrics efficiency properties (sensitivity, specicity, negative and positive predictive values). For logistical reasons, it was decided to have a not similar ratio at the randomisation sampling stage between each group. The inclusion period was 20 weeks. The duration of participation for each patient was 1 week. The study was conducted between June 2015 and February 2016. Delays could furthermore generate loss to follow up of patients; therefore, to include 1100 patients was necessary.

The final data analysis was carried out after the database freezes at the end of the data review meeting. The data was analysed by the Data Management Unit of the Brest University Hospital (Brest CHRU).

PPV and NPV were calculated based on a contingency table. The sensitivity and specificity values could not be obtained directly, as the samples of structured clinical interview positive/negative patients were not similar due to the different ratios at the randomization sampling stage. They had to be calculated from the predictive values and the patient frequency at positive depress patient according to the following formulas:

$$\boldsymbol{Se} = \frac{\boldsymbol{PPV} * \boldsymbol{P}(\boldsymbol{HSCL}+)}{\boldsymbol{P}(\boldsymbol{HSCL}+) * \boldsymbol{PPV} + \boldsymbol{P}(\boldsymbol{HSCL}-) * (\mathbf{1} - \boldsymbol{NPV})}$$

$$Sp = \frac{NPV * P(HSCL-)}{1 - [P(HSCL+) * PPV + P(HSCL-) * (1 - NPV)]}$$

P: Prevalence; PPV = Positive Predictive Value; NPV = Negative Predictive Value
P(HSCL+) = Patient HSCL+ frequency; P(HSCL-) = Patient HSCL- frequency
The confidence intervals were then obtained by a Bootstrap method. (Bootstrap Percentile).
At the end of this study, with the design of this whole method, a tool for depression diagnosis would be selected. It would provide the efficiency and reliability qualities necessary for research in primary care and dailypractice. It would have a design adapted to GP. It would be translate in several languages. Translated and original tool will be closed linguistcally and semantically.
A standardised protocol validation would be created to allow its validation in different European countries.

BIBLIOGRAPHY

1. Counsell C: **Formulating questions and locating primary studies for inclusion in systematic reviews.** *Ann Intern Med* 1997, **127**:380–387.
2. 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.
3. 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.
4. 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.
5. Whiffen VE: **The comparison of postpartum with non-postpartum depression: a rose by any other name.** *Journal of psychiatry & neuroscience : JPN* 1991:160–165.
6. Nabbe P, Le Reste JY, Guillou-Landreat M, Munoz Perez MA, Argyriadou S, Claveria A, Fernandez San Martín M, 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.

7. 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.

8. HAS, Haute Autorité Santé: **Bases Méthodologiques Pour L'élaboration de Recommandations Professionnelles Par Consensus Formalisé.** 2006.

9. 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.

10. Letrilliart L, Vanmeerbeek M: **À la recherche du consensus : quelle méthode utiliser ?** *exercer* 2011, **99**:170–177.

11. McGory ML, Shekelle PG, Ko CY: **Development of quality indicators for patients undergoing colorectal cancer surgery.** *J Natl Cancer Inst* 2006, **98**:1623–1633.

12. Skulmoski GJ, Hartman FT, Krahn J: **The Delphi Method for Graduate Research**. *J Inf Technol Educ* 2007, **6**:1.

13. 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.

14. Jones PS, Lee JW, Phillips LR, Zhang XE, Jaceldo KB: **An adaptation of Brislin's translation model for cross-cultural research**. *Nurs Res* 2001, **50**:300–304.

15. SPOONER D, PACHANA N: **Ecological validity in neuropsychological assessment: A case for greater consideration in research with neurologically intact populations**. *Arch Clin Neuropsychol* 2006, **21**:327–337.

16. Cuéllar I, Paniagua FA: *Handbook of Multicultural Mental Health : Assessment and Treatment of Diverse Populations*. Academic Press; 2000.

17. Sousa VD, Rojjanasrirat W: **Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline.** *J Eval Clin Pract* 2011, **17**:268–74.

18. Maneesriwongul W, Dixon JK: **Instrument translation process: a methods review.** *J Adv Nurs* 2004, **48**:175–86.
19. Jones J, Hunter D: **Consensus methods for medical and health services research**. *Bmj Clin Res Ed* 1995, **311**:376–380.
20. Linstone HA, Turoff M: *The Delphi Method : Techniques and Applications*. Addison-Wesley Pub. Co., Advanced Book Program; 1975.
21. Graham B, Regehr G, Wright JG: **Delphi as a method to establish consensus for diagnostic criteria.** *J Clin Epidemiol* 2003, **56**:1150–6.
22. De Villiers MR, De Villiers PJT, Kent AP: **The Delphi technique in health sciences education research.** *Med Teach* 2005, **27**:639–643.
23. Hasson F, Keeney S, McKenna H: **Research guidelines for the Delphi survey technique.** *J Adv Nurs* 2000, **32**:1008–1015.
24. Powell C: **The Delphi technique: myths and realities.** *J Adv Nurs* 2003, **41**:376–82.
25. Romm FJ, Hulka BS: **Developing criteria for quality of assessment: effect of the Delphi technique.** *Health Serv Res* 1979, **14**:309–312.
26. Jamieson S: **Likert scales: how to (ab)use them.** *Med Educ* 2004, **38**:1217–8.
27. Anadón M, Guillemette F, P: **La recherche qualitative est-elle nécessairement inductive ?** *Rech Qual* 2007, **Hors Série**:26–37.
28. Bullinger M, Anderson R, Cella D, Aaronson N: **Developing and evaluating cross-cultural instruments from minimum requirements to optimal models**. *Qual Life Res* 1993, **2**:451–459.
29. Beaton DE, Bombardier C, Guillemin F, Ferraz MB: **Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures**. *Spine (Phila Pa 1976)* 2000, **25**:3186–3191.
30. Herdman M, Fox-Rushby J, Badia X: **"Equivalence" and the translation and adaptation of health-related quality of life questionnaires.** *Qual Life Res* 1997, **6**:237–47.
31. Grazioli R, Terry DJ: **The role of cognitive vulnerability and stress in the prediction of postpartum depressive symptomatology**. *Br J Clin Psychol* 2000, **39**:329–347.

32. Uher R, Payne JL, Pavlova B, Perlis RH: **MAJOR DEPRESSIVE DISORDER IN DSM-5: IMPLICATIONS FOR CLINICAL PRACTICE AND RESEARCH OF CHANGES FROM DSM-IV**. *Depress Anxiety* 2014, **31**:459–471.

33. Stowe ZN, Hostetter AL, Newport D: **The onset of postpartum depression: Implications for clinical screening in obstetrical and primary care**. *Am J Obstet Gynecol* 2005, **192**:522–526.

34. Beck F, Guignard R: **La dépression en France (2005-2010) : prévalence, recours au soin et sentiment d'information de la population**. *Sante Homme* 2012, **421**:43–45.

CHAPTER 3

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?

Nabbe P
Le Reste JY
Guillou-Landreat M
Munoz M
Argyriadou S
Claveria A
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

Published in the Eur Psychiatry. 2017 Jan;39:99-105.

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 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 anti depressant use.

However in accordance with this objective, the field of this research focused on major depressive disorder according the DSM. Bipolar depressive disorders, are not covered by the fields in this research and have not been the concern of this research [16, 17]. Minor depressive disorders or mood disorders have not been taken into

account because the diagnosis is not clearly defined, particularly where older patients are concerned [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.
Consequently, 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 (Spain), 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 the post-partum experience. However, in accordance with the demands of our ethical committee, and because there is a scientific debate to discover whether there is a significant difference between perinatal and other forms of depression, perinatal depression was not retained. [29, 30]

The study was not in a 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 analysed 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 analysed after duplicates had been removed (224 duplicates); 50 of the validity studies were eligible and 4 studies were finally included (Figure 1).

Figure 1: Articles Inclusion (Related to PRISMA guideline)

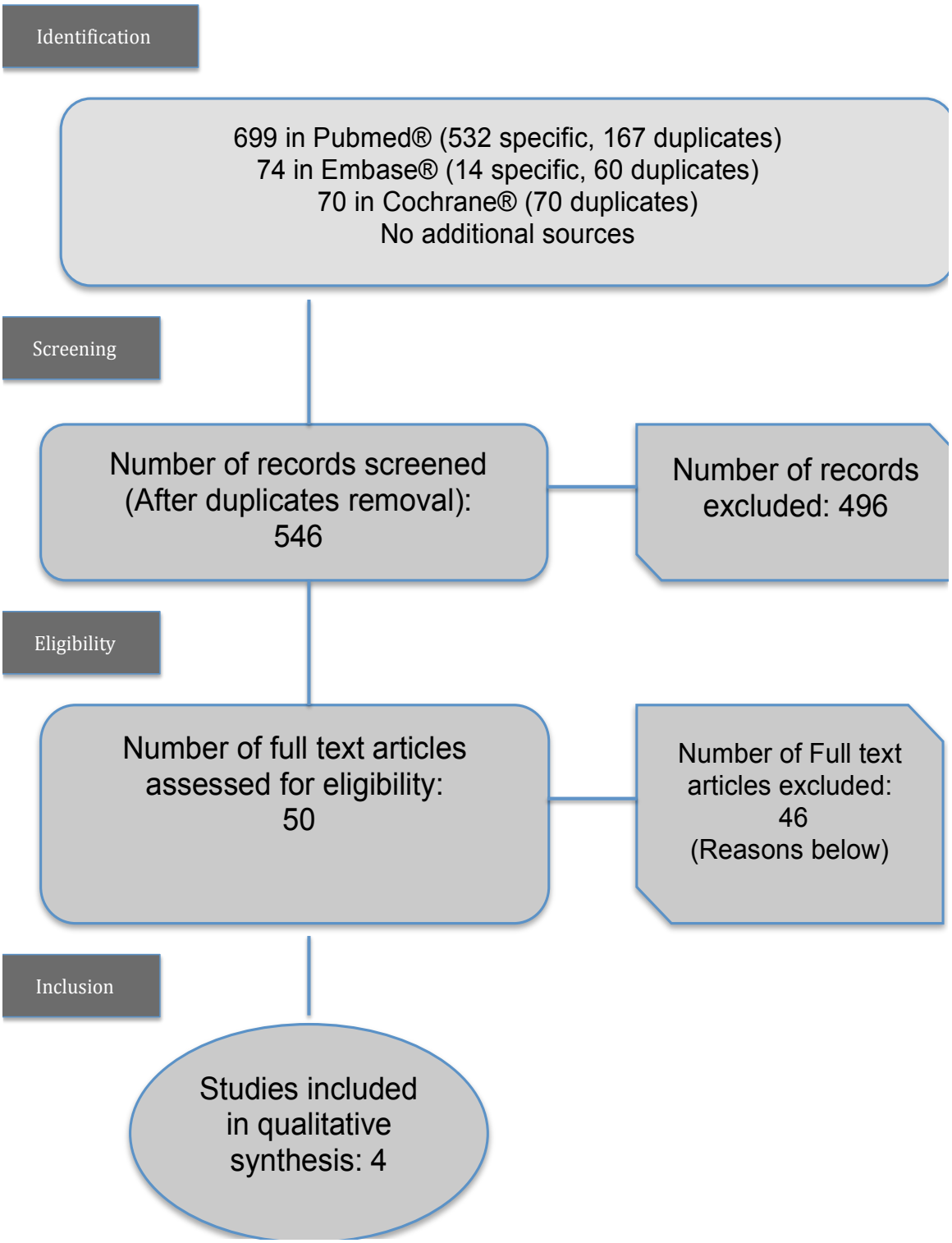


Table 1 shows the reasons for exclusion of abstracts and articles. Finally, seven tools were selected: the GDS-5, 15 and 30 items (Geriatric Depression Scale with 5, 15 and 30 items), the HSCL-25 (Hopkins Symptoms Checklist with 25 items), the HADS (Hospital Anxiety Depression Scale), the PSC-51 (physical symptom checklist in 51 items), and the CES-DR (Center for Epidemiologic Studies Depression Scale-Revised (Table 2).

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	Mean Age
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.	Fröjdh K, Hakansson NA & al.	2004	HSCL-25	MADRS based on DSM-IV	Yes	74	0.94	0.94			0.88	78.5

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 a sample 58 individuals out of 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, PSC-51 and HADS, the GDS-5 and GDS-15, HSCL-25 [32][33][34][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 is validated against a population-based mathematical model and not against clinical comparison criteria [37][38]. The PHQ-9, which is also very popular, is a follow-up tool, validated against the Hamilton Scale for follow-up and often used as a diagnostic tool [39][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 but the language was irrelevant as it did not fall within our criteria [41].

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 in one tool prevents clinicians from using them [42]. 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 to diagnose and quantify depression in elderly patients [44]. 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 designed for better ergonomic use [47,48]. The CESD-R was developed in 1977 to diagnose and quantify depression [49]. It was designed with 20 items, using a 4-point Likert scale, centered on the previous week's symptoms. It is also widely used in research [50]. 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 in research [52]. 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 in 4 has a wide age range, between 20 and 80 years. The other 3 studies included a population over 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]? In future studies, when discriminating within a population, there are other factors which should be taken into account, apart from age, for example, the ability to cope, which is not only age-related. [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 but choosing tools to be shared by FPs and psychiatrists will be a challenge. [3] 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 or by finding additional data on Cronbach's Alpha or Cohen's Kappa in the literature for each tool. [59][48] 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 analysing 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

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, Vázques-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, Gaite 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.

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. Fröjdh K, Håkansson A, Karlsson I, Frojdh 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 JJL, 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.** *J Nerv Ment Dis* 1997, **185**:102–107.

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: **Vieillissement 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.

CHAPTER 4

One consensual depression diagnosis tool to serve many countries: a challenge!

A RAND / UCLA methodology.

Nabbe P
Le Reste JY
Guillou-Landreat M
Beck-Robert E
Assenova R
Lazic D
Czachowski S
Stojanović-Špehar S
Hasaganic M
Lingner H
Clavería A
Fernández San Martín MI
Sowinska A
Argyriadou S
Lygidakis C
Le Floch B
Doerr C
Montier T
Van Marwijk H
Van Royen P

Published in BMC Res Notes (2018) 11:4

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.

Key Words

RAND/ UCLA Appropriateness Method, Multicultural Consensus, Delphi Procedure, Depression Diagnosis Tool

INTRODUCTION

Primary care is a strategic place for depression diagnosis and treatment. [1][2][3][4][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 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

METHOD

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]. (Figure 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

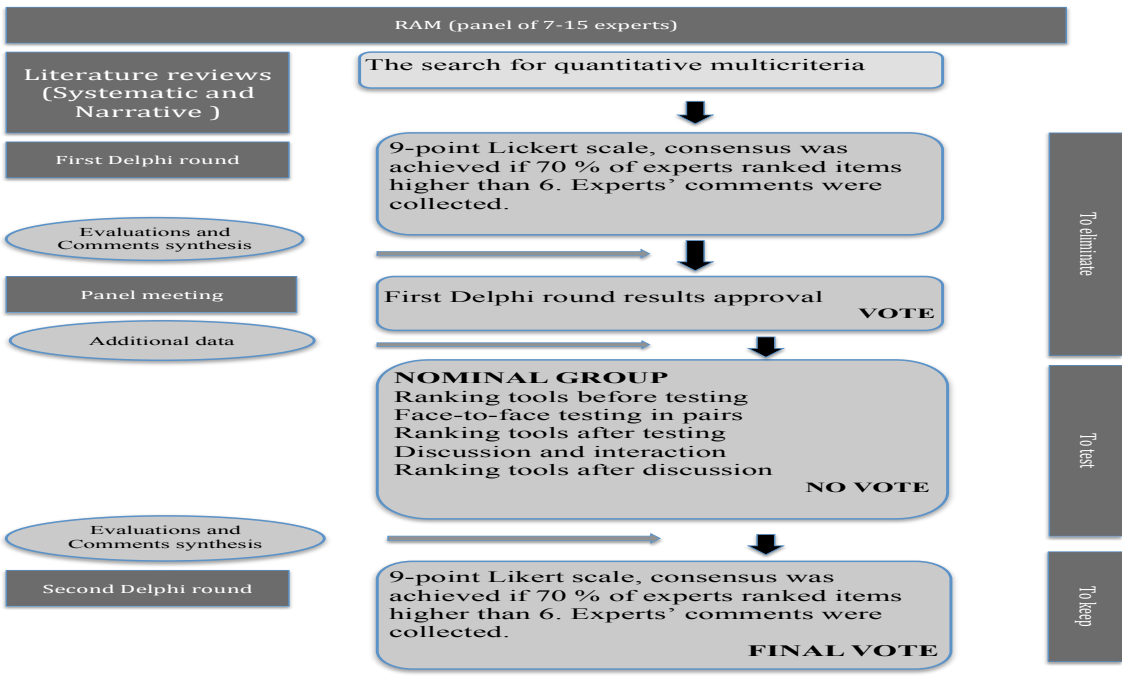
The study started with a Delphi procedure to eliminate the less efficient and keep the more reliable tools. The comments took into account only validity data, not ergonomics.

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.

Figure 1: Flow diagram RAND/UCLA



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 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 5,000 inhabitants and 2 worked in urban areas with 2,000 to 5,000 inhabitants. (Table1)

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 to 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 to 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*

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

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)

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	*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**	

Second step

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.

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.

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	25	12.5	12.5

	consultation			
	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

Third step

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][19][20]. It has robust efficiency and reliability scores [21][22][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][26][27][28][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][33][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][38][39][40][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 transversal 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 to 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.

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

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. World Health Organization: **The World Health Report 2001 : Mental Health : New Understanding, New Hope.** World Health Organization; 2001.
4. 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.
5. 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**.

6. Zhang J, Patel VL, Johnson TR, Shortliffe EH: **A cognitive taxonomy of medical errors**. *J Biomed Inform* 2004, **37**:193–204.
7. 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.
8. 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.
9. Nabbe P, Le Reste JY, Guillou-Landreat M, Munoz Perez MA, Argyriadou S, [illegible]
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.
10. 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**:1533–1543.
11. 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.
12. HAS, Haute Autorité Santé: **Bases Méthodologiques Pour L'élaboration de Recommandations Professionnelles Par Consensus Formalisé.** 2006.
13. 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.
14. Letrilliart L, Vanmeerbeek M: **À la recherche du consensus : quelle méthode utiliser ?** *exercer* 2011, **99**:170–177.
15. McGory ML, Shekelle PG, Ko CY: **Development of quality indicators for patients undergoing colorectal cancer surgery.** *J Natl Cancer Inst* 2006, **98**:1623–1633.

16. Skulmoski GJ, Hartman FT, Krahn J: **The Delphi Method for Graduate Research**. *J Inf Technol Educ* 2007, **6**:1.

17. 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.

18. 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.

19. Derogatis LR, Unger R, Derogatis LR, Unger R: **Symptom Checklist-90-Revised**. In *The Corsini Encyclopedia of Psychology*. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2010.

20. Lipman RS, Covi L, Shapiro AK: **The Hopkins Symptom Checklist (HSCL)--factors derived from the HSCL-90.** *J Affect Disord* 1979, **1**:9–24.

21. 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.

22. 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.

23. 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.

24. Hignett S, Carayon P, Buckle P, Catchpole K: **State of science: human factors and ergonomics in healthcare.** *Ergonomics* 2013, **56**:1491–503.

25. 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**.

26. 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.

27. 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.
28. 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.
29. Munk-Jørgensen P, Fink P, Brevik JI, Dalgard OS, Engberg M, Hansson L, Holm M, 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.
30. Zigmond AS, Snaith RP: **Hospital Anxiety and Depression Scale (HADS):** *Ann Gen Psychiatry* 1983, **67**:361–70.
31. 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.
32. 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.
33. Andrews B, Hejdenberg J, Wilding J: **Student anxiety and depression: comparison of questionnaire and interview assessments.** *J Affect Disord* 2006, **95**:29–34.
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**:363–370.
35. 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.

36. 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.

37. Friedman bruce, Heisel M j, Delavan R: **Psychometric Properties of the 15-Item Geriatric Depression**. *J Am Geriatr Soc* 2005, **53**:1570–1576.

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, **Suppl 7**:81–86.

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**:260–266.

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**:23–28.

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**:195–199.

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.

CHAPTER 5

Nine HSCL-25 translations, a forward-backward procedure supplemented by a cultural control check.

Abreviated title

HSCL- 25: translation into nine European languages

Nabbe P
Le Reste JY
Guillou-Landreat M
Assenova R
Lazic D
Czachowski S
Stojanović-Špehar S
Hasaganic M
Lingner H
Clavería A
Fernández San Martín MI
Sowinska A
Argyriadou S
Lygidakis C
Le Floch B
Doerr C
Montier T
Van Marwijk H
Van Royen P

This article is awaiting review on the journal: BMC psychology.

ABSTRACT

Introduction: The Hopkins Symptom Checklist-25 (HSCL-25) is a depression diagnosis tool. Effective, reliable and ergonomic, it can be used in daily practice. To allow family practitioners (FPs) to use it, it had to be translated into various European languages. The entire translation process had to ensure homogeneity.

Method: Forward-backward translation with two translators (an academic and an FP researcher) recruited for the forward translation (FT). A panel of English-speaking FPs was set up in each country. A minimum size of 15 experts was requested. The panel of experts finalized the FT using a Delphi procedure. Then, a different translator, who did not know the original version, undertook an English backward translation. Linguists compared the two English versions. Differences were listed to analyze the cultural impact of translation according to a multicultural consensus group.

Results: Translations into 9 languages were completed. The composition of each panel had to include a maximum variation of researchers, teachers and practitioners. One to two Delphi rounds by country were sufficient. To ensure the original meaning, all versions were subjected to a cultural check.

Conclusion: Translations into Greek, Polish, Bulgarian, Croatian, Catalan, Galician, Spanish, Italian and French have been finalized without altering the meaning.

INTRODUCTION

How to manage people with depression in primary care is a growing challenge within Europe. Family Practitioners (FPs) are at the frontline and, at the same time, secondary care services are increasingly under threat [1][2][3][4]. Depression manifests itself in various ways: (I) as a syndromic ‘disorder’ in which contextual distress, anxiety and somatoform disorders overlap; (II) as a difficulty many patients experience in expressing, acknowledging and discussing their suffering; (III) as a long-term condition with both a subjective and an objective aspect which can be measured [5]. Based on these three inter-individual variabilities, FPs may experience difficulties in diagnosis and may easily misjudge the symptom levels if they do not use formal instruments to guide the discussion [6][7]. These difficulties may lead to inappropriate care and cause public health issues [8][9][10]. A short discussion of the results on a relevant questionnaire is often the first step towards an open dialogue focused on the patient.

The incidence and prevalence rates of depression therefore differ widely in family practice, due to complex contextual variations, differences in health care systems, concepts of disorder, objectives and practices, as well as cultural variations in the expression of the disorder [11][12].

Collaborative primary care mental health models can improve care and outcome for patients. With the aim of supporting them, the European General Practice Research Network (EGPRN) developed a collaborative research agenda. [13] FPs, whether within or outside this network, require a reliable, standardized, efficient and ergonomic tool which should take into account cultural and linguistic differences. [14][15] The EGPRN adopted a standardized methodology including European FPs experts from different cultures, who speak different languages, within different healthcare systems, to set up an established consensus procedure to identify such tools [16][17].

These tools had to be acceptable to both FPs and psychiatrists, and informative for both, to improve collaboration [18]. They must be routinely feasible in the physician's

surgery, in either primary or psychiatric care, and to be extremely practical for research purposes [19]. These tools had to be validated and reliable.

A handbook was developed consensually in order to select a single tool and then translated into different languages, using a forward and backward translation (inspired by Brislin's model). It was a consensual procedure that has been used internationally in other cross-cultural studies [20][21][22]. At each step, the key points and purposes were debated and chosen by consensus among European experts.

Initially, a systematic review of literature in the indexed databases, according to PRISMA criteria, was produced. Seven tools validated against a psychiatric examination using the DSM's major depression criteria were collected [23]. A Consensus procedure (RAND/UCLA) made it possible to select one tool according to its effectiveness, reliability and ergonomics [24]. European researchers selected the Hopkins Symptom Checklist-25 items (HSCL-25) [23][25][26]. It is a highly validated, reliable diagnostic tool for ranking levels of depression [27][28]. It is a self-rating scale on the existence and severity of both anxiety and depression symptoms experienced during the preceding week [29][30].

The objective of this study was to translate the HSCL-25 into the languages of the team members, without losing homogeneity, and staying within the context of primary care [22][31].

METHOD

A three-step standardized study was conducted among participating countries, including: (I) a forward translation (FT), (II) a backward translation (BT) and (III) a cultural check [11][32][33]. (Figure 1)

The FT was conducted with an incorporated Delphi procedure [34][35][36]. It is a rigorous way to reach consensus [37][38][39]. It is a systematic, interactive method which involves a panel of experts using iterative procedures [40].
This process requires:

- Anonymity of participants, which ensures response reliability and avoids contamination,
- Iteration, which allows participants to refine their views in the light of the progress of the group's work,
- Feedback control under the responsibility of the investigator,
- Statistical aggregation of the group's responses to allow a quantitative and qualitative analysis of the data [41][42][43][44].

The EGPRN French team ensured that the whole process followed the protocol. The FT had to be validated by the daily board of the study, composed of members of the EGPRN, all-active within the research process.

The NIs selected translators to set up two translation teams which worked blind, for both FT and BT. Translators had to be knowledgeable about healthcare terminology. The FT team involved one member of the FP research group and one official translator from every country involved. The BT team involved one (or two) FPs and one official translator [22].

The NIs recruited a panel of experts in their own countries, anonymized the experts' responses and allocated an identification number for later identification [41]. Initially, 20 to 30 experts were recruited per country in order to maintain at least 15 participants until the end of the last round. The selection criteria for each FP expert were: being a native of his/her country of residence and speaking his/her native language; being an English speaker [31]. Over half had to be involved in teaching and/or research activities. In order to assess the representativeness of the panel by its maximum variation, the experts provided the following information: their gender, practice setting, years of practice and publications [45].

According to Brislin's Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures, once the FT had been completed, a BT was performed with two goals: (I) to ensure that linguistic translation problems were identified, (II) to collect translation problems which were independent of the linguistic translation itself. Nevertheless, a linguistic translation was insufficient as translation biases related to

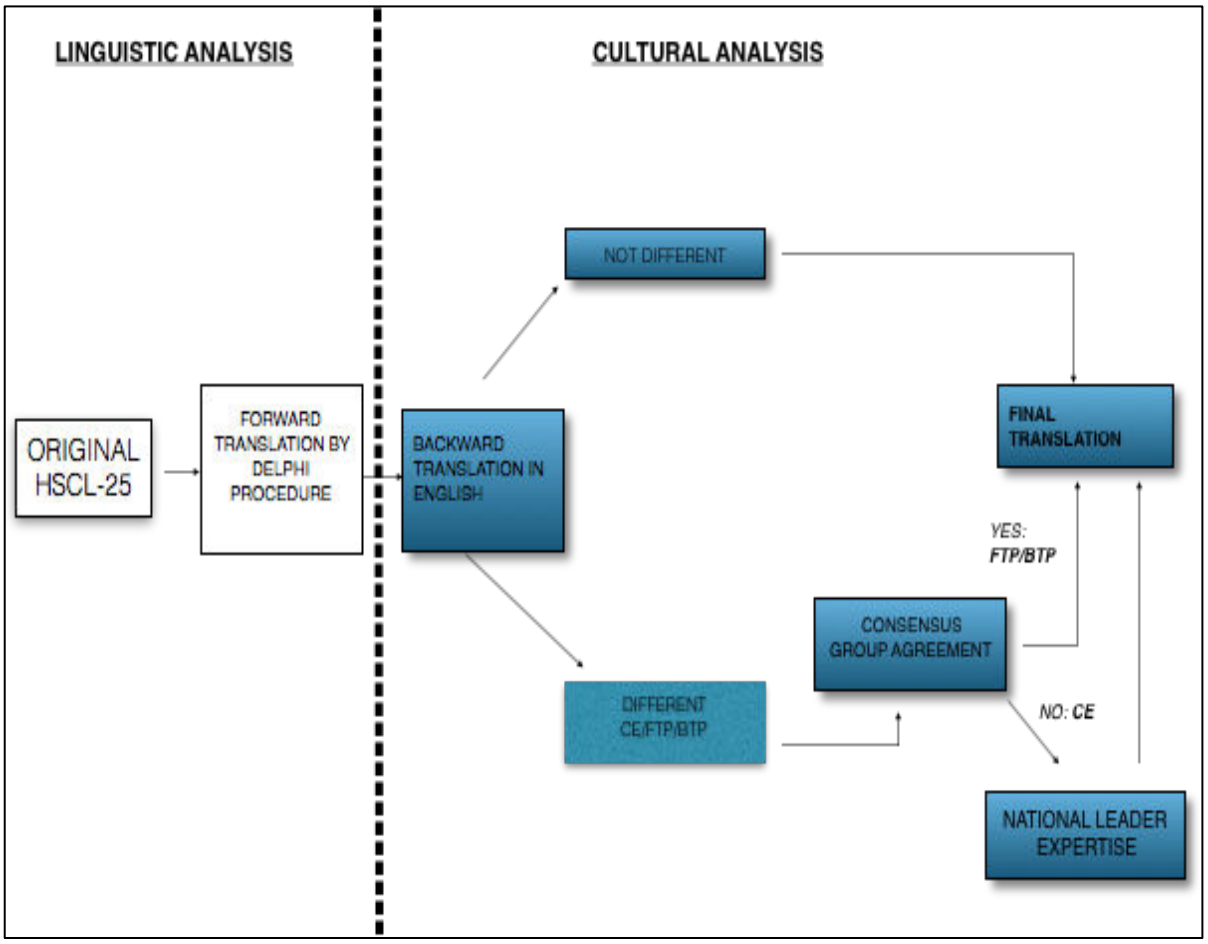
cultural aspects in each country were possible; this required cultural control to ensure homogeneity [17][20][32][33][46].

An FP researcher and a PhD linguist analyzed all backward translations and compared them with the HSCL-25 original version to establish whether there were any significant difference in terms of meaning. Their report was submitted to a consensus group whose task was to clarify the nature of the anomaly from three problem areas:
A Backward Translation Problem was eliminated if the difference was explained by an incorrect back translation.
A Forward Translation Problem was defined as an anomaly in transcribing the original English.
A Cultural Effect was considered validated if there were no linguistic problems with the translation but where the item needed a modification to be understood by the patients in their own “everyday” language.

At the end of the forward/backward translation and the cultural effect control check, a linguistically stable, definitive translation, which maintained meaning, was produced for each country. This encompassed the structure and order of the questions, item by item, as well as the method of use of the HSCL-25.

Ethical request: The EGPRN French team was in charge of checking the volunteering process and confirming there were no potential risks or benefits related to participants
The Comité d’Ethique of the Université de Bretagne Occidentale gave its approval for the whole process.
The EGPRN French team recruited National Investigators (NIs) and requested their consent, arranged voluntary participation in the study and produced an absence of conflict of interest statement.
Each NI asked the participants for their signed consent.

Figure 1: The translation procedure



FTP: Forward translation problem; BTP: Backward translation problem; CE: Cultural Effect

RESULTS

11 NIs from 8 European countries participated. They were all FPs, EGPRN members, and fluent in English. The NIs panel was composed of 8 women. 10 of the NIs practiced in urban areas of more than 5000 inhabitants and 1 worked in an urban area of between 2000 and 5000 inhabitants. 8 were teachers and researchers, 3 were solely researchers, with a total of 152 publications. The average number of years of practice was 21.3 years and 12.4 years of research. Among the 11 NIs, 2 NIs were from two distinct cultural regions of coastal Spain: Catalonia and Galicia; 2 NIs were Croats; other countries were each represented by a single NI. (Table 1)

Table 1: National investigators panel

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

F: female; M: male; FPs: family practitioners

Forward translation

14 experts (from Germany) to 31 (from Spain) were recruited for the Delphi procedure. In compliance with the selection criteria, they were all FPs, all English speakers. The European panel consisted of 215 FPs (111 male and 104 female).
20 of the experts worked in a city of <2000 inhabitants, 36 in a city of between 2000 and 5000 inhabitants, 159 in a city of >5000 inhabitants. Their clinical experience was analyzed according to years of practice: an average of 16.4 years' experience.
In Poland, Bulgaria, Germany, Spain and the Catalonia region of Spain, there was only one Delphi round, and two rounds in the other countries..
Almost all proposals were accepted in one round (273/320: 85.3%). The one where consensus was not reached entered the second round; the NI and the Forward official translator synthesized the experts' comments to produce a new translation proposition for a second round. (Table 2)

Table 2: Characteristics of the experts' panels per country

	N (Female)	Practice (mean years)	Area of Practice according to number of inhabitants			Academic researcher and/or teacher		Number of publications	2nd round participants
			< 2000	2000 to 5000	> 5000	Number	Experience (mean, years)		
Bulgaria	22 (13)	20,5	1	5	16	5	5,4	8	No second round
Catalonia	22 (9)	15,7	0	2	20	20	10,5	22	No second round
Croatia	16(13)	19,2	1	1	14	16	11,5	15	15
France	16 (7)	12,5	1	7	8	15	6,3	11	15
Galicia	20 (6)	22,3	0	0	20	17	13,1	19	20
Germany	14 (8)	16,7	0	3	11	9	10	6	No second round
Greece	26 (13)	10.9	10	9	7	24	5,1	26	15
Italy	18 (6)	17,2	3	2	13	13	14	12	No second round
Poland	30 (18)	11,9	4	6	20	26	13,1	10	No second round
Spain	31 (11)	19,5	0	1	30	27	12	30	No second round
Total	215 (104)	15,55	20	36	178	172	10,1	159	4 second round

Some translation issues required a second proposal and another Delphi round

In Croatian, eleven proposals were rejected in the first round.
For example, for Item-17 (feeling blue) the first proposal was "Bili ste tužni", which was considered to be too focused on melancholia, so, it was modified to "Bili ste sjetni", closer to the concept of sadness. All new proposals were accepted during the second round.

In the French translation, consensus was not reached on eighteen proposals in the first round and needed further specification in the second round. For example, for Item-25 (sleep disturbance), the first proposal was "vous n'arrivez pas à dormir" which was modified to "votre sommeil était perturbé ", closer to the English word: 'disturbed'. All new proposals were accepted during the second round.

As a German version of the HCL-25 already existed, the German NIs proposed that their expert panel would discuss the official version anyway, but without the forward translation process. All items were accepted in the first Delphi round. At this step, the Germans NIs stopped the procedure. No cultural check was performed.

Nine Greek proposals were rejected in the first round.
For example, for Item-1 (Being scared for no reason): the first proposal was "Είμαι τρομοκρατημένος χωρίς αιτία". This proposal was considered "too strong". Consensus was reached on the second proposal: "Είμαι τρομαγμένος χωρίς αιτία". All new proposals were accepted during the second round.

In the Italian translation, consensus was not reached on five proposals in the first round.
For example, for Item-5 (heart racing), the first proposal was "avere tachicardia", which was considered to be too focused on medicine, therefore it was modified to "sentire le cuore battere veloce", which was more familiar to the reviewers. All new proposals were accepted during the second round.

In the Spanish Galician translation, consensus was not reached on three proposals in the first round. For example: for Item-6 (trembling), the first proposal was “trema”, the present indicative of the verb “tremar”. The second proposal was “ten tremores”, which was accepted in the second round. All new proposals were accepted during the second round. (Table 3)

Table 3: Results of the first Delphi round

Item/Country	Galicia	Spain	Catalonia	France	Italy	Bulgaria	Croatia	Greece	Germany	Poland
1 **Being scared for no reason**	C	C	C	C	C	C	C	NC	C	C
2 **Feeling fearful**	C	C	C	C	C	C	NC	C	C	C
3 **Faintness**	C	C	C	NC	NC	C	NC	NC	C	C
4 **Nervousness**	C	C	C	C	C	C	C	C	C	C
5 **Heart racing**	C	C	C	NC	C	C	C	C	C	C
6 **Trembling**	NC	C	C	NC	NC	C	C	C	C	C
7 **Feeling tense**	C	C	C	C	C	C	C	C	C	C
8 **Headache**	C	C	C	C	C	C	C	C	C	C
9 **Feeling panic**	C	C	C	NC	C	C	NC	C	C	C
10 **Feeling restless**	NC	C	C	NC	C	C	NC	C	C	C
11 **Feeling low in energy**	C	C	C	C	C	C	NC	NC	C	C
12 **Blaming oneself**	C	C	C	NC	NC	C	C	C	C	C
13 **Crying easily**	C	C	C	C	C	C	C	NC	C	C
14 **Losing sexual interest**	C	C	C	NC	C	C	NC	C	C	C
15 **Feeling lonely**	C	C	C	NC	C	C	NC	C	C	C
16 **Feeling hopeless**	C	C	C	C	C	C	NC	C	C	C
17 **Feeling blue**	C	C	C	NC	C	C	NC	C	C	C
18 **Thinking of ending one’s life**	C	C	C	C	C	C	C	NC	C	C
19 **Feeling trapped**	C	C	C	NC	C	C	C	C	C	C
20 **Worrying too much**	C	C	C	NC	C	C	NC	NC	C	C
21 **Feeling no interest**	C	C	C	NC	C	C	NC	NC	C	C
22 **Feeling that everything is an effort**	C	C	C	C	C	C	C	C	C	C
23 **feelings of Worthlessness**	C	C	C	NC	C	C	C	NC	C	C
24 **Poor appetite**	C	C	C	C	C	C	C	NC	C	C
25 **Sleep disturbance**	NC	C	C	NC	C	C	C	C	C	C
26 Choose the best answer for how you felt over the past week	C	C	C	NC	C	C	C	C	C	C
27 Not at all	C	C	C	C	NC	C	C	C	C	C
28 A little	C	C	C	NC	C	C	C	C	C	C

29 Quite a bit	C	C	C	C	C	C	C	C	C	C
30 Extremely	C	C	C	C	C	C	C	C	C	C
31 The HSCL-25 score is calculated by dividing the total score (sum score of items) by the number of items answered (ranging between 1.00 and 4.00). It is often used as the measure of distress. The patient is considered as a "probable psychiatric case" if the mean rating on the HSCL-25 is ≥ 1.55.	C	C	C	NC	NC	C	C	C	C	C
32 A cut-off value of ≥ 1.75 is generally used for diagnosis of major depression defined as "a case in need of treatment". This cut-off point is recommended as a valid predictor of mental disorder as assessed independently by clinical interview, somewhat depending on diagnosis and gender.	C	C	C	NC	C	C	C	C	C	C
The administration time of HSCL 25 is 5 to10 minutes										
C: Consensus, NC: No Consensus										

Backward translation and cultural check

The initial instructions, the 25 items, the quotation and the explanatory sentences were back-translated. A total of 36 propositions was subjected to analysis. All backward translations were compared linguistically to the original. Differences were noted for submission to the NIs and the consensus group. Three consensus group meetings were necessary with national feedback between each.

For the Hellenic language

The translation was mainly based on an adaptation according to gender. The experts concluded that there was a general cultural effect affecting all parts of the scale. However, there were no real differences in meaning, and so the Greek HSCL-25 scale remained stable in comparison with the original.

For the Slavic languages

Poland: 13 items were different. 7 due to a BTP, 6 required a cultural adaptation.
Bulgaria: 3 items were different. 2 due to a BTP, 1 required a cultural adaptation.
Croatia: 8 items were different. 2 due to a BTP, 8 required a cultural adaptation.

Most items resulted from a conceptual issue: therefore « Heart racing » became « Palpitations », « Trembling » « Tremors », and « an effort » « a burden » in Polish; moreover « Feeling low in energy » became « Feeling low in energy » in Bulgarian; « Feeling restless » was translated by « Anxiety » in all three languages, for there were no equivalent words to express these ideas. A word-by-word translation, in that case, was impossible. « Headache » turned into the plural form « Headaches » in Polish for grammatical reasons.

For Croatia, the main cultural aspect was the use of the present perfect, which is a tense of state and not of action, commonly expressed in all daily life: in Items 2, 7, 9, 10, « feeling » was logically replaced by « you have been ». Only one item appeared to be stronger, « Faintness », where the term was replaced by « weakness », but the word weakness in Croatian was in fact equivalent. Finally, translation into Bulgarian showed it to be the most stable of those three languages.

For Romance languages

Italy: 7 items were different. 6 due to a BTP, 1 required a cultural adaptation
France: 5 items were different. 4 due to a BTP, 1 required a cultural adaptation.
Spain: 6 items were different. 1 due to a BTP, 5 required a cultural adaptation
Catalonia: 7 items were different. 4 due to a BTP, 3 required a cultural adaptation.
Galicia: 5 items were different. 1 due to a BTP, 4 required a cultural adaptation.

For the French scale, the tense used in everyday language should be the present. However, the past tense was the tense used in the forward version. The past tense, in everyday French, is considered an older, upper-class use of the language. All

tenses were modified: e.g. « Tout était un effort pour vous" was modified by « Tout est un effort pour vous » in the final definition.

For the Hispanic languages (e.g. standard Spanish, Catalan and Galician) and Italian, the translation had to be modified according to gender, and more precisely on the male plural form for the Italian scale, which was the usual way of speaking/writing in that country.

The item « Faintness » was translated into « Weakness » in all three Hispanic languages (e.g. « Debilidad », « Debilitat » and « Debilidade » in Standard Spanish, Catalan and Galician respectively).

The same was observed for the item « Heart racing », translated into « Palpitations » (i.e. « Palpitaciones » and « Palpitacions » in both Standard Spanish and Galician versions).

From a Galician and Catalan point of view, « Blame oneself » turned into « Blame yourself » in backward translation, as the term « oneself » wasn't commonly employed.

To finish, « Feeling no interest » was translated by « No siente interes por nada » in Standard Spanish to be understood by the patient, and « Worthless feeling » became « Feeling useless ». But in Standard Spanish, « inutil » meant « worthless » as well.

As far as the Galician scale was concerned, item 14 « losing sexual interest », was translated into « Loss of sexual interest » which expressed a state, not an action, as in the original English version: but the local experts considered it a normal way of speaking/writing in that language.

For all of languages

Item-17 « Feeling Blue » coming from the Afro-American culture, would come from a contraction of "having the blue devils" or having the blues (blows or hits) to the soul. This expression induced a cultural effect in 6 out of the 9 languages. A word-by-word rendition was then impossible and required a cultural adaptation.

The items-15 « Feeling lonely », 18 « Thinking of ending one's life », 19 « Feeling trapped » and 25 « Sleep disturbance » remained stable after BT.

As regards the 10 scale instructions and the quotation question, the BT was different from the original version of 9 items except the explanation concerning the administration time required to use the tool. Many translation problems were related to ‘cultural’ effects specific to the languages. For example: in French, some terms were replaced by typical expressions commonly employed in questionnaires: e.g. « pencil-and-paper » was translated by « auto questionnaire » and « Not at all » by « Pas du tout d’accord ».

An interesting point to note: not only were there translation similarities (often with stronger meanings or medical connotations) between languages belonging to the same linguistic group, but also similarities between different groups: the best example concerns Slavic and Hispanic languages about Item 3 « Faintness », which was translated by « Weakness » in Catalan, Standard Spanish, Galician, and also in Croatian, having a more prosaic than medical connotation.

At the end of the cultural analysis, the consensus group finally concluded that there were no changes of meaning, and the translation was finalized in all 9 languages.

ITEM	HSCL-25 ORIGINAL VERSION	GREECE	POLAND	BULGARIA	CROATIA
A	Choose the best answer for how you felt over the past week	Επιλέξτε την καλύτερη απάντηση για το πώς αισθανθήκατε την τελευταία εβδομάδα	Wybierz najlepszą odpowiedź	Изберете отговора, който най-добре описва как сте се чувствали през изминалата седмица	Izaberite jedan odgovor koji najbolje opisuje kako ste se osjećali tijekom prošlog tjedna:
1	Being scared for no reason	Είμαι τρομαγμένος/η χωρίς αιτία	Bać się bez powodu	Чувство за уплаха без причина	Bili ste bezrazložno uplašeni
2	Feeling fearful	Αισθάνομαι φοβισμένος /η	Poczucie strachu	Чувство за страх	Bojali ste se
3	Faintness	Αίσθημα λιποθυμιάς	Omdlenia	Отпадналост	Bili ste slabi
4	Nervousness	Νευρικότητα	Nerwowość	Нервност	Bili ste nervozni
5	Heart racing	Ταχυπαλμία	Kołatanie serca	Сърцебиене	Ubrzano vam je lupalo srce
6	Trembling	Τρεμούλα	Drżenia	Треперене	Drhtali ste
7	Feeling tense	Αισθάνομαι υπερένταση	Poczucie napięcia	Чувство за напрежение	Bili ste napeti
8	Headache	Πονοκέφαλος	Bóle głowy	Главоболие	Boljela vas glava
9	Feeling panic	Αισθάνομαι πανικό	Uczucie paniki	Чувство за паника	Bili ste u panici
10	Feeling restless	Αισθάνομαι ταραχή	Uczucie niepokoju	Чувство на безпокойство	Bili ste uznemireni
11	Feeling low in energy	Αισθάνομαι ότι δεν έχω ενέργεια	Poczucie braku energii	Усещане за понижена енергия	Niste imali dovoljno energije
12	Blaming oneself	Κατηγορώ τον εαυτό μου	Obwinianie samego siebie	Самообвинение	Okrivljavali ste se
13	Crying easily	Εύκολο κλάμα	Płaczliwość	Плачливост	Bili ste plačljivi
14	Losing sexual interest	Απώλεια σεξουαλικού ενδιαφέροντος	Utrata zainteresowań sferą seksualną	Загубата на сексуален интерес	Niste bili zainteresirani za spolni odnos
15	Feeling lonely	Αισθάνομαι μοναξιά	Poczucie osamotnienia	Чувство за самотност	Bili ste usamljem
16	Feeling hopeless	Αισθάνομαι απελπισμένος/η	Poczucie beznadziejności	Чувство за безнадежност	Osjećali ste sebeznadno
17	Feeling blue	Νοιώθω πεσμένος/η	Poczucie przygnębienia	Чувстам се нещастен	Bili ste sjetni
18	Thinking of ending one's life	Σκέφτομαι να δώσω τέλος στη ζωή	Myśli samobójcze	Мисли за самоубийство	Razmišljali ste da si oduzmete život
19	Feeling trapped	Αισθάνομαι παγιδευμένος /η	Poczucie uwięzienia	Чувстам се като в капан	Osjećali ste sekao da ste u klopci
20	Worrying too much	Ανησυχώ υπερβολικά	Zamartwianie się	Притеснявам се твърде много	Bili ste previše zabrinuti
21	Feeling no	Αισθάνομαι ότι τίποτε	Poczucie braku	Чувство за загуба	Bez interesa za bilo što

	interest	δεν είναι ενδιαφέρον	zainteresowań	на интерест	
22	Feeling that everything is an effort	Αισθάνομαι ότι για το καθε τί χρειάζεται να κάνω προσπάθεια	Poczucie, że wszystko jest ciężarem	Чувство, че всичко изисква усилие	Sve vam je bilo naporno
23	**feelings of Worthlessness**	Αισθάνομαι ότι δεν αξίζω τίποτε	Poczucie bezwartościowości	Чувство за безполезност	Osjećali ste se bezvrijedno
24	Poor appetite	Μείωση της όρεξης	Słaby apetyt	Лош апетит	Imali ste slab apetit
25	Sleep disturbance	Διαταραχές ύπνου	Zaburzenia snu	Нарушения на съня	Imali ste problema sa spavanjem

ITEM	HSCL-25 ORIGINAL VERSION	CASTILE	CATALONIA	GALICIA	ITALY	FRANCE
A	Choose the best answer for how you felt over the past week	Elija la respuesta que mejor describa cómo se ha sentido durante la semana pasada	Triï la millor resposta per indicar com s'ha sentit en la darrera setmana	Escolla a resposta que mellor describa como se sentiu durante a semana pasada	Scegliere la risposta più adatta su come ti sei sentito/a nell'ultima settimana	Veuillez choisir la réponse qui décrit le mieux comment globalement vous vous sentiez toute la semaine dernière
1	Being scared for no reason	Se asusta sin motivo	Estar espantat/espantada sense motiu aparent	Asústase sen motivo	Avere paura senza motivo	Vous avez peur sans raison
2	Feeling fearful	Siente miedo	Sentir por	Ten medo	Sentirsi impauriti	Vous vous sentez effrayé
3	Faintness	Debilidad	Debilitat	Debilidade	Sensazione di mancamento	Vous avez une sensation d'étourdissement
4	Nervousness	Nerviosismo	Nerviosisme	Nerviosismo	Esseri nervosi	Vous vous sentez nerveux
5	Heart racing	Palpitaciones	Cor accelerat	Palpitacións	Sentire il cuore battere veloce	Vous avez l'impression que votre cœur bat anormalement vite
6	Trembling	Tiembla	Tremola	Ten tremores	Tremore	Vous avez la sensation de trembler
7	Feeling tense	Se siente tenso/a	Sentir-se tens/a	Séntese tenso/a	Sensazione di tensione	Vous vous sentez tendu
8	Headache	Dolor de cabeza	Mal de cap	Dor de cabeza	Avere mal di testa	Vous avez des maux de tête
9	Feeling panic	Siente pánico	Sensació de pànic	Sente pánico	Sensazione di panico	Vous vous sentez paniqué
10	Feeling restless	Siente inquietud	Sensació d'inquietud	Séntese inquedo/a	Sensazione di irrequietezza	Vous vous sentez agité
11	Feeling low in energy	Siente que le falta energía	Sensació de manca d'energia	Sente que lle falta enerxía	Sentirsi senza energia	Vous manquez d'énergie
12	Blaming oneself	Se culpa a sí mismo/a	Culpar-se un/a mateix/a	Cúlpase a si mesmo/a	Avere sensi di colpa	Vous ressentez une sensation de culpabilité
13	Crying easily	Llora con facilidad	Plora fàcilment	Chora con facilidade	Piangere facilmente	Vous pleurez facilement
14	Losing sexual interest	Pierde el interés sexual	Pèrdua de l'interès sexual	Perda do interese sexual	Perdere l'interesse sessuale	Vous ressentez un désintérêt pour la vie sexuelle
15	Feeling lonely	Se siente solo/a	Sentir-se sol/a	Séntese só/soa	Sentirsi soli	Vous avez une sensation de solitude

16	Feeling hopeless	Se siente sin esperanza	Sentiment de desesperança	Séntese sen esperanza	Sentirsi senza speranza	Vous vous sentez désespéré
17	Feeling blue	Se siente triste	Sentir-se trist/a	Séntese triste	Sentirsi tristi	Vous avez le cafard
18	Thinking of ending one's life	Piensa en acabar con su vida	Pensa en treure's la vida	Pensa en acabar coa súa vida	Avere pensieri di togliersi la vita	Vous avez pensé à mettre fin à votre vie
19	Feeling trapped	Se siente atrapado/a	Sentir-se atrapat/atrapada	Séntese atrapado/a	Sentirsi intrappolati	Vous vous sentez pris au piège
20	Worrying too much	Se preocupa en exceso	Preocupar-se en excés	Preocúpase en exceso	Preoccuparsi troppo	Vous vous inquiétez trop
21	Feeling no interest	No siente interés por nada	Sentiment de manca d'interès	Non sente interese por nada	Non avere alcun interesse	Plus rien ne vous intéresse
22	Feeling that everything is an effort	Siente que todo le cuesta un esfuerzo	Sentir que tot és un esforç	Sente que todo lle supón un esforzo	Sentire che tutto è uno sforzo	Tout est un effort pour vous
23	**feelings of Worthlessness**	Se siente inútil	Sentir-se inútil	Séntese inútil	Sentirsi inutili	Vous avez le sentiment d'être bon à rien
24	Poor appetite	poco apetito	Pèrdua de la gana	Poco apetito	Avere poco appetito	Vous avez perdu l'appétit
25	Sleep disturbance	Problemas para dormir	Alteració de la son	Alteracións do sono	Disturbi del sonno	Votre sommeil est perturbé

SCALE INSTRUCTIONS ORIGINAL VERSION	GREECE	POLAND	BULGARIA	CROATIA
The HSCL-25 score is based on pencil-and-paper self-report of 25 questions about the presence and intensity of anxiety and depression symptoms over the last week.	Η βαθμολογία του HSCL-25 βασίζεται σε γραπτό ερωτηματολόγιο αυτοαξιολόγησης 25 ερωτήσεων σχετικά με την παρουσία και την ένταση των συμπτωμάτων άγχους και κατάθλιψης κατά την τελευταία εβδομάδα. Οι συμμετέχοντες απαντούν σε μία από τις τέσσερις κατηγορίες για κάθε ερώτημα σε μια κλίμακα εύρους τεσσάρων βαθμών με τιμές από 1 μέχρι 4.	Ocena testu HSCL-25 oparta jest na kwestionariuszu 25 pytań, w którym zakreśla się na papierze obecność i nasilenie objawów lęku i depresji w ciągu ostatniego tygodnia.	Резултатът от HSCL-25 се основава на самостоятелно попълнен инструмент на хартиен носител, включващ 25 въпроса за наличието и интензивността на симптоми на тревожност и депресия през последната седмица.	HSCL-25 skor sastoji se od 25 pitanja koja se rješavaju jednostavno olovkom i papirom, a temelji se na samoprocjeni prisutnosti i intenzitetu ansksioznih i depresivnih simptoma tijekom prošlog tjedna.
Participants answer to one of four categories for each item on a four-point scale ranging from 1 to 4		Badani odpowiadają na jedno z czterech możliwych kategorii na skali mierzącej wartości od 1 do 4.	Участниците избират една от категориите за всяка позиция по скала от четири точки от 1.00 до 4.00.	Ispitanici odgovaraju jednom od četiri kategorija za svako pitanje na skali od 1-4.
1.“Not at all”	Καθόλου	Wcale	Съвсем не	Nimalo
2.“A little”	Λίγο	Trochę	Незначително	Malo
3.”Quite a bit”	Αρκετά	Znacznie	Съвсем малко	Dosta
4.“Extremely”	Πάρα πολύ	Bardzo mocno	Извънредно	Jako

SCALE INSTRUCTIONS ORIGINAL VERSION	CASTILE	CATALONIA	GALICIA	ITALY	FRANCE
The HSCL-25 score is based on pencil-and-paper self-report of 25 questions about the presence and intensity of anxiety and depression symptoms over the last week.	La puntuación HSCL-25 se basa en un cuestionario auto cumplimentado con lápiz y papel, de 25 preguntas sobre la presencia y la intensidad de ansiedad y síntomas depresivos en la última semana.	L'escala HSCL-25 es basa en un qüestionari auto administrat de 25 preguntes, sobre la presència i la intensitat de símptomes d'ansietat i depressió en la darrera setmana.	A puntuación HSCL-25 baséase nun cuestionario cumprimentado con lapis e papel, de 25 preguntas sobre a presenza e a intensidade de ansiedade e síntomas depresivos na última semana.	Il punteggio dell'HSCL-25 si basa sulla compilazione di un questionario di autovalutazione in cartaceo ("carta/penna") di 25 domande sulla presenza e intensità di sintomi di ansia e depressione nel corso dell'ultima settimana.	La HSCL-25 est un auto-questionnaire en 25 questions relatives à la présence et à l'intensité des symptômes d'anxiété et de dépression durant toute la semaine dernière.
Participants answer to one of four categories for each item on a four-point scale ranging from 1 to 4	Los/ las participantes responden una de cuatro categorías para cada ítem, en una escala de cuatro puntos que van desde 1 a 4.	Els/les participants responen a una de les quatre categories per a cada ítem en una escala de quatre punts que va de l'1 al 4.	Os participantes responden unha de catro categorías para cada ítem, nunha escala de catro puntos que van desde 1 a 4.	I partecipanti rispondono a una delle quattro categorie per ciascun sintomo su una scala di punteggio che va da 1 a 4.	Les participants cotent chaque proposition, sur une échelle en quatre points, cotée de 1 à 4.
1."Not at all"	En absoluto	Gens	En absoluto	Per niente	Pas du tout d'accord
2."A little"	Un poco	Una mica	Un pouco	Poco	Un peu d'accord
3."Quite a bit"	Bastante	Bastant	Bastante	Abbastanza	Plutôt d'accord
4."Extremely"	Mucho	Molt	Moito	Moltissimo	Complètement d'accord

SCALE INSTRUCTIONS ORIGINAL VERSION	GREECE	POLAND	BULGARIA	CROATIA
The HSCL-25 score is calculated by dividing the total score (sum score of items) by the number of items answered (ranging between 1.00 and 4.00). It is often used as the measure of distress.	Η βαθμολογία του HSCL-25 υπολογίζεται διαιρώντας τη συνολική βαθμολογία (αθροιστική βαθμολογία των ερωτημάτων), διά του αριθμού των ερωτημάτων που απαντήθηκαν (κυμαινόμενο μεταξύ του 1,00 έως 4,00). Συχνά χρησιμοποιείται για τη μέτρηση της δυσφορίας.	Wynik testu HSCL-25 jest obliczany poprzez podzielenie całkowitej liczby punktów (suma punktów z każdej pozycji testu) przez liczbę pozycji na które udzielono odpowiedzi (w skali od 1 do 4). Często służy on do pomiaru dystresu.	HSCL-25 резултатът се изчислява, като се раздели общият брой точки (сбор точки по критерий) на броя на отговорените критерии (вариращи между 1,00 и 4,00). Той често се използва като мярка за страдание.	Skor HSCL-25 se izračunava dijeljenjem ukupnog zbroja (zbroj skora pojedinih pitanja) s brojem odgovorenih pitanje (raspon od 1,00 do 4,00). Obično se koristi za mjerenje distresa.
The patient is considered as a "*probable psychiatric case"* if the mean rating on the HSCL-25 is ≥1.55.	Ο ασθενής θεωρείται σαν "πιθανό ψυχιατρικό περιστατικό" εάν η μέση βαθμολογία του HSCL-25 είναι >=1,55	Pacjenta uważamy za "prawdopodobny przypadek psychiatryczny" jeśli średnia ocena w teście HSCL-25 jest >/ (większa lub równa) 1,55.	Пациентът се приема като "вероятно психиатричен случай", ако средната оценка по HSCL-25 е ³ 1,55.	Pacijent se smatra « vjerojatno psihijatrijskim slučajem » ako je srednja vrijednost na HSCL-25 ≥ 1,55.
A cut-off value of ≥1.75 is generally used for diagnosis of major depression defined as "*a case, in need of treatment"*. This cut-off point is recommended as a valid predictor of mental disorder as assessed independently by clinical interview, somewhat depending on diagnosis and gender.	Το όριο του >= 1,75 γενικώς χρησιμοποιείται για τη διάγνωση της μείζονος κατάθλιψης που ορίζεται ως "περίπτωση που χρήζει θεραπείας". Αυτό το όριο συνίσταται σαν ένας έγκυρος προγνωστικός δείκτης ψυχικής διαταραχής, όπως εκτιμάται ανεξάρτητα από την κλινική εικόνα, η οποία εξαρτάται κάπως από τη διάγνωση και το φύλο.	Wartość graniczną>/ (większą lub równą) 1,75 ogólnie przyjmuje się w diagnozowaniu ciężkiej depresji, definiowanej jako „przypadek wymagający leczenia." Wartość ta jest zalecana jako istotny czynnik w przewidywaniu obecności choroby psychicznej, wymagającej jednak niezależnego wywiadu klinicznego i w pewnym sensie zależy od rozpoznania i płci.	Гранична стойност от ³ 1,75 обикновено се използва за диагностициране на тежка депресия и определя случая като "случай, нуждаещ се от лечение". Тази гранична стойност, получена независимо от клиничното интервю и зависеща до определена степен от диагнозата и пола, се препоръчва като валиден предиктор за психично разстройство.	Razdjelna točka (cut-off) ≥1,75 se koristi za dijagnozu velikog depresivnog poremećaja i to kao „slučaj koji zahtjeva liječenje". Razdjelna točka se preporuča kao validni prediktor mentalnog poremećaja podjednako kao i sama procjena neovisnim kliničkim intervjuom, dijelom ovisan o dijagnozi i spolu.
The administration time of HSCL 25 is 5 to 10 minutes.	Ο χρόνος χορήγησης του HSCL 25 είναι 5 έως 10 λεπτά.	Czas na wykonanie testu HSCL 25 wynosi od 5 do 10 minut.	Времето за провеждане HSCL-25 е от 5 до 10 минути.	Vrijeme za ispunjavanje HSCL-25 je 5-10 minuta.

SCALE INSTRUCTIONS ORIGINAL VERSION	CASTILE	CATALONIA	GALICIA	ITALY	FRANCE
The HSCL-25 score is calculated by dividing the total score (sum score of items) by the number of items answered (ranging between 1.00 and 4.00). It is often used as the measure of distress.	La puntuación del HSCL-25 se calcula dividiendo la puntuación total (sumando la puntuación de todos las preguntas) entre el número de respuestas (varía entre 1,00 y 4,00). Se usa habitualmente para medir el malestar psicológico.	La puntuació total del HSCL-25 es calcula dividint la suma de la puntuació dels diferents ítems pel número d'ítems contestats. El resultat total oscil·la entre 1,00 i 4,00. Aquesta escala sovint s'utilitza com a mesura del malestar psicològic.	A puntuación do HSCL-25 calcúlase dividindo a puntuación total (a suma de todas as preguntas) entre o número de respostas (cuxa puntuación oscila entre 1,00 e 4,00). Úsase de forma habitual para medir o nivel del malestar psicológico.	Il punteggio dell' HSCL-25 si calcola dividendo il punteggio totale (somma dei punteggi degli elementi) con il numero di elementi risposti (che variano da 1,00 a 4,00). Spesso si usa come misura di ansietà	Le score du HSCL- 25 se calcule en divisant la somme des cotations des propositions par le nombre de réponses reçues. Le résultat final est compris entre 1,00 à 4,00. Il est couramment utilisé pour mesurer la souffrance psychologi-que.
The patient is considered as a "*probable psychiatric case*" if the mean rating on the HSCL-25 is ≥1.55.	El/la paciente se considera un "probable caso psiquiátrico" si el valor medio del HSCL-25 es ≥1,55.	El/la pacient és considerat/considerada com a " probable cas psiquiàtric " si la qualificació mitjana del HSCL-25 és ≥ 1,55.	Considérase que o/a paciente é un "caso psiquiátrico probable" se o valor medio do HSCL-25 é ≥ 1,55.	Il paziente è considerato come un "probabile caso psichiatrico" se il punteggio medio dell'HSCL-25 è ≥1,55.	Le patient est considéré comme « probablement atteint d'un trouble psychiatrique » si le score moyen du HSCL-25 est supérieur ou égal à 1,55.
A cut-off value of ≥1.75 is generally used for diagnosis of major depression defined as "*a case, in need of treatment*". This cut-off point is recommended as a valid predictor of mental disorder as assessed independently by clinical	Por lo general se usa un valor de corte de ≥1,75 para el diagnóstico de depresión mayor, definida como "un caso que necesita tratamiento". Este valor de corte se considera un predictor válido de un trastorno mental, evaluado de forma independiente mediante	Generalment s'utilitza un punt de tall ≥1,75 per al diagnòstic de la depressió major i es defineix com " cas que precisa de tractament". Es recomana aquest punt de tall com un predictor vàlid de trastorn mental com ho seria l'avaluació independent per entrevista clínica, depenent en part del diagnòstic i del gènere.	Polo xeral, úsase un valor de corte ≥ 1,75 para diagnosticar a depresión maior, definida como "un caso que precisa tratamento". Este valor de corte recoméndase como un predictor válido dun trastorno mental, avaliado independentemente por medio de entrevistas clínicas, aínda que depende	Un cut-off che sia >=1,75 è normalmente usato per la diagnosi di depressione maggiore definita come "un caso che necessita di trattamento". Questo cut-off è raccomandato come un valido predittore di disordine mentale come valutato in modo	Un score supérieur ou égal à 1,75 diagnostique généralement une dépression caractérisée et définit « un patient nécessitant un traitement ». Ce seuil est considéré comme un score prédictif validé des troubles mentaux. Il a été évalué de manière

interview, somewhat depending on diagnosis and gender.	entrevista clínica, aunque depende en parte del diagnóstico y el género.		en parte do diagnóstico e do xénero.	indipendente da un colloquio clinico, dipendente in qualche modo dalla diagnosi e dal genere	indépendante par des études cliniques. Il varie peu quelles que soient les situations diagnostiques et le sexe.
The administration time of HSCL 25 is 5 to 10 minutes.	El tiempo de administración del HSCL-25 es de 5 a 10. minutos.	El temps d'administració del HSCL 25 és de 5 a 10 minuts.	O tempo de realización do HSCL-25 é de 5 a 10 minutos.	Il tempo di somministrazione dell'HSCL-25 è da 5 a 10 minuti.	Remplir le questionnaire HSCL-25 prend entre 5 et 10 minutes.

DISCUSSION

Using a three step qualitative procedure, ecologically embedded in primary care, nine consensual translations of the HSCL-25 were obtained which were homogeneous to the original version, in 3 language families, e.g. Hellenic, Slavic and Romance. A German version already existed.

They aimed to meticulously track inconsistencies between any local translations, which could lead to misinterpretation. This methodical and transcultural validation ensured the transfer of the same content from one language to another as well as its reliability. [17][46]

The Greek translation remained the most stable, followed by Bulgarian. Item 17, « Feeling blue » was the most challenging to translate, followed by Item 3 « Faintness » and Item 5 « Heart racing ».
Some scales needed an ultimate adaptation in terms of tense (French, Croatian) and others in terms of gender (Greek, Italian, and Hispanic languages).

Research and teaching implications

Translation remains the most crucial step in the adoption of a well-developed instrument by another nation using a different language. Errors in translation may distort the original intent of this instrument and compromise the validity and reliability of the resulting instrument. [47] There are what may be called semantic issues affecting comparability in international studies since the same word is interpreted differently across countries and cultures. [48][49] Moreover, certain terms and concepts may not exist in other languages, or may have additional connotations that back-translations do not always reveal. Challenges arise, not only because of the content of word-to-word, literal translation, but also because of the linguistic form of the language, such as tone, and syntax. [50]

The translations of the HSCL, compared to the original version, are now linguistically similar, in terms of meaning. However, the scales need further testing as this first step is not sufficient to complete the task of translating them and supporting their cross-cultural validity. The external and internal validity of each version has to be tested to ensure their level of reliability is comparable with the original version. This will be achieved through quantitative studies in primary care daily practice.

FPs in most of Europe are now able to use this tool in research studies within family practice and assess the severity of depression in their patients. The use of such a shared tool may have a great impact on the feasibility of doing research on depression in primary care in the future. We will be able to compare data between European countries more easily which will make it possible to undertake statistical reviews on the epidemiology and symptoms of depression throughout Europe. The use of the same instrument can support the conceptualization of the studied phenomenon across different studies, and the findings can then be compared. [21]

Limitations

To reduce the selection bias and to ensure the quality of the sample: the study was managed to ensure the involvement of FPs in the linguistic translation and that was a key point for us to pursue in this study. As described by many translators, when

discussing scientific translation work, it is essential that a “specialist” in the field (e.g. the field of primary care daily practice) take a last look at the translation. [20][51][52] He or she is the main arbiter of the quality of the final translation. [53] Thus, specific attention was paid to choosing FP researchers and certified bilingual translators with sufficient knowledge of health care terminology, to reduce selection bias.

The cultural control was rigorous. It involved a step-by-step analysis, in order to prevent confusion bias and linguistic problems. The intervention of a consensus group allowed several, gradual evaluations of each item, strengthening the accuracy of the validated translations and co-designing the end-result. This work is the result of multicultural collaborative research among European countries.

CONCLUSION

A translation of the HSCL-25 in which homogeneity is ensured, is now available for Spain and its culturally distinct regions of Galicia and Catalonia, as well as for France, Greece, Italy, Poland, Bulgaria, and Croatia. It is now ready to be tried out in actual and representative primary care populations in order to further validate its test-parameters.

List of abbreviations and definitions:

BT - Backward Translation
CE – Cultural Effect
DSM - Diagnostic and Statistical Manual of Mental Disorders
EGPRN - European General Practice Research Network
FPs - Family Practitioners
FT – Forward Translation
NPV - Negative Predictive Value
PPV - Positive Predictive Value
PRISMA - Preferred Reporting Items for Systematic reviews and Meta-Analyses
RAND - Research And Development
RAM - RAND Appropriateness Method

RAND/UCLA - Research and Development / University of California Los Angeles
Se - Sensitivity
Sp - Specificity
SRL - Systematic Review of literature

BIBLIOGRAPHY

1. King M, Nazareth I, Levy G, Walker C, Morris R, Weich S, Bellón-Saameño JA, Moreno B, Svab I, Rotar D, Rifel J, Maaroos H-I, Aluoja A, Kalda R, Neeleman J, Geerlings MI, Xavier M, de Almeida MC, Correa B, Torres-Gonzalez F: **Prevalence of common mental disorders in general practice attendees across Europe.** *Br J Psychiatry* 2008, **192**:362–7.

2. Ayuso-Mateos JL, Vázques-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, Gaite 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.

3. 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.

4. Torzsa P, Szeifert L, Dunai K, Kalabay L, Novák M: **[Diagnosis and therapy of depression in family practice].** *Orv Hetil* 2009, **150**:1684–93.

5. Jorm AF: **Mental health literacy. Public knowledge and beliefs about mental disorders.** *Br J Psychiatry* 2000, **177**:396–401.

6. Mitchell AJ, Vaze A, Rao S: **Clinical diagnosis of depression in primary care: a meta-analysis.** *Lance* 2009, **374**:609–619.

7. Ani C, Bazargan M, Hindman D, Bell D, Farooq MA, Akhanjee L, Yemofio F, Baker R, Rodriguez M: **Depression symptomatology and diagnosis: discordance between patients and physicians in primary care settings.** *BMC Fam Pract* 2008, **9**:1.

8. 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 MA, Codony M, Domingo-Salvany A, Ferrer M, Joo SS, Martínez-Alonso M, Matschinger H, et al.: **Psychotropic drug utilization in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project.** *Acta Psychiatr Scand Suppl* 2004, **109**:38–46.

9. Wells KB: **Caring for depression in primary care: defining and illustrating the policy context.** *J Clin Psychiatry* 1997, **58 Suppl 1**:24–7.

10. Schoenbaum M, Unützer J, McCaffrey D, Duan N, Sherbourne C, Wells KB: **The effects of primary care depression treatment on patients' clinical status and employment.** *Health Serv Res* 2002, **37**:1145–58.

11. Cuéllar I, Paniagua FA: **Handbook of Multicultural Mental Health : Assessment and Treatment of Diverse Populations.** Academic Press; 2000.

12. Marsella AJ, Yamada AM: **Culture and Mental Health. In Handbook of Multicultural Mental Health.** Elsevier; 2000:3–24.

13. Woltmann E, Grogan-Kaylor A, Perron B, Georges H, Kilbourne AM, Bauer MS: **Comparative Effectiveness of Collaborative Chronic Care Models for Mental Health Conditions Across Primary, Specialty, and Behavioral Health Care Settings: Systematic Review and Meta-Analysis**. *Am J Psychiatry* 2012, **169**:790–804.

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

15. Kirmayer LJ, Robbins JM, Dworkind M, Yaffe MJ: **Somatization and the recognition of depression and anxiety in primary care.** *Am J Psychiatry* 1993, **150**:734–741.

16. 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.

17. Bullinger M, Anderson R, Cella D, Aaronson N: **Developing and evaluating cross-cultural instruments from minimum requirements to optimal models**. *Qual Life Res* 1993, **2**:451–459.

18. Zhang J, Patel VL, Johnson TR, Shortliffe EH: **A cognitive taxonomy of medical errors**. *J Biomed Inform* 2004, **37**:193–204.

19. Mitchell AJ: **Clinical utility of screening for clinical depression and bipolar disorder.** *Curr Opin Psychiatry* 2012, **25**:24–31.

20. Beaton DE, Bombardier C, Guillemin F, Ferraz MB: **Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures**. *Spine (Phila Pa 1976)* 2000, **25**:3186–3191.

21. Brislin RW: **Comparative research methodology: Cross-cultural studies**. *Int J Psychol* 1976, **11**:215–229.

22. Jones PS, Lee JW, Phillips LR, Zhang XE, Jaceldo KB: **An adaptation of Brislin's translation model for cross-cultural research**. *Nurs Res* 2001, **50**:300–304.

23. Nabbe P, Le Reste JY, Guillou-Landreat M, Munoz Perez MA, Argyriadou S, Claveria A, Fernández 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* 2017, **39**:99–105.

24. 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.

25. Fröjdh K, Håkansson A, Karlsson I, Frojdh 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.

26. 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.

27. 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.

28. Sandanger I, Moum T, Ingebrigtsen G, Dalgard OS, S??rensen 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**:345–354.

29. Ventevogel P, De Vries G, Scholte WF, Shinwari NR, Faiz H, Nassery R, van den Brink W, Olff M: **Properties of the Hopkins symptom checklist-25 (HSCL-25) and the Self-Reporting Questionnaire (SRQ-20) as screening instruments used in primary care in Afghanistan**. *Soc Psychiatry Psychiatr Epidemiol* 2007, **42**:328–335.

30. Moum T: **Mode of administration and interviewer effects in self-reported symptoms of anxiety and depression**. *Soc Indic Res* 1998, **45**:279–318.

31. SPOONER D, PACHANA N: **Ecological validity in neuropsychological assessment: A case for greater consideration in research with neurologically intact populations**. *Arch Clin Neuropsychol* 2006, **21**:327–337.

32. Sousa VD, Rojjanasrirat W: **Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline.** *J Eval Clin Pract* 2011, **17**:268–74.

33. Maneesriwongul W, Dixon JK: **Instrument translation process: a methods review.** *J Adv Nurs* 2004, **48**:175–86.

34. Jones J, Hunter D: **Consensus methods for medical and health services research**. *Bmj Clin Res Ed* 1995, **311**:376–380.

35. HAS, Haute Autorité Santé: **Bases Méthodologiques Pour L'élaboration de Recommandations Professionnelles Par Consensus Formalisé.** 2006.

36. Linstone HA, Turoff M: **The Delphi Method : Techniques and Applications.** Addison-Wesley Pub. Co., Advanced Book Program; 1975.

37. Graham B, Regehr G, Wright JG: **Delphi as a method to establish consensus for diagnostic criteria.** *J Clin Epidemiol* 2003, **56**:1150–6.
38. 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.
39. De Villiers MR, De Villiers PJT, Kent AP: **The Delphi technique in health sciences education research.** *Med Teach* 2005, **27**:639–643.
40. Hasson F, Keeney S, McKenna H: **Research guidelines for the Delphi survey technique.** *J Adv Nurs* 2000, **32**:1008–1015.
41. 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.
42. Powell C: **The Delphi technique: myths and realities.** *J Adv Nurs* 2003, **41**:376–82.
43. Romm FJ, Hulka BS: **Developing criteria for quality of assessment: effect of the Delphi technique.** *Health Serv Res* 1979, **14**:309–312.
44. Jamieson S: **Likert scales: how to (ab)use them.** *Med Educ* 2004, **38**:1217–8.
45. Anadón M, Guillemette F, P: **La recherche qualitative est-elle nécessairement inductive ?** *Rech Qual* 2007, **Hors Série**:26–37.
46. Herdman M, Fox-Rushby J, Badia X: **"Equivalence" and the translation and adaptation of health-related quality of life questionnaires.** *Qual Life Res* 1997, **6**:237–47.
47. Yu DSF, Lee DTF, Woo J: **Issues and Challenges of Instrument Translation**. *West J Nurs Res* 2004, **26**:307–320.
48. Daugherty JC, Puente AE, Fasfous AF, Hidalgo-Ruzzante N, Pérez-Garcia M: **Diagnostic mistakes of culturally diverse individuals when using North American neuropsychological tests**. *Appl Neuropsychol* 2017, **24**:16–22.
49. Schnohr CW, Gobina I, Santos T, Mazur J, Alikasifuglu M, Välimaa R, Corell M, Hagquist C, Dalmasso P, Movseyan Y, Cavallo F, van Dorsselaer S, Torsheim T: **Semantics bias in cross-national comparative analyses: is it good or bad to have "fair" health?** *Health Qual Life Outcomes* 2016, **14**:70.
50. Hanrahan D, Sexton P, Hui K, Teitcher J, Sugarman J, London AJ, Barnes M, Purpura J, Klitzman R: **Linguistic and cultural challenges in communication and**

translation in ussponsored HIV Prevention research in emerging economies. *PLoS One* 2015, **10**:e0133394.

51. Skulmoski GJ, Hartman FT, Krahn J: **The Delphi Method for Graduate Research**. *J Inf Technol Educ* 2007, **6**:1.

52. Vesga O, Agudelo M, Salazar BE, Rodriguez C a, Zuluaga AF: **Generic vancomycin products fail in vivo despite being pharmaceutical equivalents of the innovator.** *Antimicrob Agents Chemother* 2010, **54**:3271–9.

53. Balliu C: **L'enseignement de la traduction médicale: pour une nouvelle pragmatique**. *Meta: Journal des Traducteurs* 1994, **39**:15–25.

CHAPTER 6

The french version of the hscl-25 scale: a cross-validation study set against the pse-9, in primary care daily practice.

Nabbe P
Le Reste JY
Guillou-Landreat M
Gatineau F
Le Floch B
Viala J
Montier T
Van Marwijk H
Van Royen P

ABSTRACT

Background

The Hopkins Symptom Checklist in 25 items (HSCL-25) helps to assess depression in Primary care. This self-administrated questionnaire is validated, reliable and ergonomic. A patient is considered 'depressive' if a score > 1.75 is obtained. We have translated it into French.

The aim of this study was to validate the test characteristics of the HSCL-25, in its French version (F-HSCL-25), by comparing the results with the Present State Examination-9 French version (F-PSE-9) results.

Method

Outpatients from three French General Practice settings (rural, semi-rural and urban) were recruited: approximately 20,000 outpatients among 17 GPs. Two groups were formed: F-HSCL-25 ≥1.75 and F-HSCL-25 <1.75. In order to obtain two balanced groups, a different method of randomization was chosen for each group. The F-PSE-9 was randomly administered to 1 in 2 patients in the F-HSCL-25 ≥1.75 group, and to 1 in 16 in the (much larger) F-HSCL-25 <1.75 group. The diagnostic performance was assessed and the test results obtained from both groups were compared with their F-PSE-9 results.

Results

Of the 1126 patients who completed the F-HCL-25, 886 joined the F-HSCL-25 <1.75 group and 240 the F-HSCL-25 ≥1.75 group. The overall prevalence of depression, using the F-HSCL-25, was 21% in these medical practices. The diagnostic performance of the F-HSCL-25 versus the external criteria (F-PSE-9) were as follows: Positive Predictive Value (PPV) 69.8%, Negative Predictive Value (NPV) 87%; Sensitivity 59.1%, and Specificity 91.4%.

Conclusion

The F-HSCL-25 is an appropriate diagnostic tool for depression in primary care in France due to its high specificity and high NPV. This pilot study will be extended throughout Europe, however, preliminary evidence suggests that the HSCL-25 is a suitable diagnostic tool for depression in primary care.

Keywords: Depression – Hopkins symptom Checklist 25 items – Validation studies – Psychometrics

INTRODUCTION

Major depression affects 4.4% of the world population [1-3]. Estimates of prevalence in the general population vary in Europe but are currently around 25% [4-6] . Furthermore, the prevalence is twice as high for women [7]. A prevalence increase of more than 18% was observed between 2005 and 2015 [8]. Within the French population, prevalence is estimated to be between 5% and 12% [9]. Currently, nearly 8 million French people have experienced, or will experience, depression during their lifetime [10]. Depression has a significant impact on emotional, social and occupational life and is a major risk factor for suicide [11].

The general practitioner (GP) diagnosis for major depression has a high specificity but a low sensitivity in routine care but, as GPs can also offer efficient follow-up, primary care is a good place to organize treatment [12,13]. This syndromic disorder is not easy to diagnose due to the wide variety of ways in which it may be presented [14]. In most European countries, GPs are the first, and often the only, physicians to take care of depressed patients but they generally have little time [15,16]. A fast, efficient and sensitive tool with a reasonable specificity and negative predictive value, would add value and save time, thereby improving performance management in primary care.

From the many diagnostic tools available for combined European research studies, the HSCL-25 has been selected, using a European consensus procedure, based on a systematic review of the literature. It combines high quality reliability, effectiveness and ergonomics with a conceptual connection to the DSM [17,18].
The HSCL-25 is a short-form diagnostic tool derived from HSCL-90 [19,20]. This is a comprehensive, systematized, semi-directed, clinical self-administered questionnaire [28][29].

The specificity is robust: between 0.78 to 0.88, the reliability (Alpha de Cronbach) is between 0.87 to 0.97 [21-24] . The HSCL-25 short length self-administered format is perfectly suited for use in busy primary care settings with many competing demands. It may represent a practical instrument to alert French GPs to potentially depressive or anxious symptomatology.

The score is based on 25 questions divided into two sub-sections related to the presence and intensity of symptoms of depression and anxiety experienced during the previous week. Patients select one of the four responses for each item on a 4-point Likert scale, ranging from 1 (strongly disagree) to 4 (completely agree). Completing the questionnaire takes between 5 and 10 minutes. The final score is calculated by dividing the sum of the scores of all the items by 25 (the final score ranges from 1.00 to 4.00). A diagnosis of Major Depression, defined as "a case requiring treatment," is generally above a threshold of 1.75 [25].

The HSCL-25 was translated into French using a well-established procedure in primary care, involving a forward/backward translation based on a Delphi procedure, combined with a cultural check to maintain linguistic and semantic reliability (appendix 1) [26,27]**.**

In 1993, Nettlebladt & al. evaluated the accuracy of the HSCL-25 as a primary care diagnostic questionnaire in Sweden [30]. They carried out a study in six Swedish primary healthcare centers in two districts, one rural and one semi-urban, to validate the HSCL-25 against the PSE-9 and establish a cut-off.
A cut-off of 1.55 indicated a patient at risk, but a cut-off of 1.75 specified that the patient needed treatment. A cut-off of 1.75 gave a sensitivity of 73%, a specificity of 76%, a Positive Predictive Value (PPV) of 58% and a Negative Predictive Value (NPV) of 86% [30].

The HSCL-25 is not currently used by French GPs, but is a potentially promising tool. The aim of this project, inspired by the Nettlebladt study, was to determine the external efficiency of the HSCL-25 French version (F-HSCL-25) in French general practice by comparing it with the Present State Examination-9 French version (F-

PSE-9), a widely accepted semi-structured clinical interview used extensively in psychiatry [29].

METHOD

Study design

A quantitative cross-validation study of the F-HSCL-25 in an adult French general practice population was carried out by the research team of the Soins primaires, Santé Publique, Registre des tumeurs de Bretagne Occidentale (EA 7479 SPURBO). It was a comparative, non-inferiority, multi-centered, survey. The study team constituted of two physician researchers, three GP trainees specifically trained in psychiatric assessment using the PSE-9 and using the CATEGO algorithms [29], a psychiatrist, a statistician, a GP research network of 20 GPs, a Data Manager and a Research Coordinator. The psychiatrist of Brest CHRU trained the GP trainees in psychiatric assessment and confirmed the validity of the clinical diagnoses. A multidisciplinary research network supported the study.
The inclusion period was 20 weeks. The duration of participation for each patient was 1 week. The study was conducted between June 2015 and February 2016.

Participants

The study was carried out in northern Finistère (Brittany, France) in three study centres (family practice offices affiliated to SPURBO). The population was a mix of patients from urban, semi-rural and rural environments. In the waiting room, before their primary care appointment, patients were given a leaflet explaining the study, an F-HSCL-25 scale and a consent form. Participants were recruited spontaneously to ensure the representativeness of the recruited population, after they had read the explanatory notice and completed the F-HSCL-25 (paper version).

Inclusion criteria

The patients needed to be adults (over 18 years). Patients had to give their written informed consent to participate. They completed the F-HSCL-25 self-assessment questionnaire and submitted it to the study team.

Exclusion criteria

To avoid possible cases of puerperal depression, which requires specific management, women with a reported pregnancy were not included in the study [31][32][33]. Also excluded were adults consulting for administrative purposes, patients known to be schizophrenic or having related disorders and patients requiring emergency care.

Sample size

Patients were placed in an HSCL+ group or an HSCL- group according to their scores : F-HSCL-25 score ≥1.75 (or HSCL+) and F-HSCL-25 score <1.75 (or HSCL-). To obtain two balanced groups for final analysis, one in two patients in the HSCL+ group were randomly administered an PSE-9 interview, and one in sixteen patients in the HSCL- group were administered an F-PSE-9. This process ensured the two groups were as comparable as possible.

The delay between interview and inclusion had to be between one week and one month in order to prevent bias in the results of the PSE-9 interview. This was particularly important where an F-HSCL-25 score of ≥1.75 initiated treatment by the GP.

These ratios assume a prevalence of depression between 5% and 12% which gives reasonable precision in estimating diagnostic performance [9]. At least 45 patients were needed per group to ensure a power of 80% in order to detect a difference of at least 50% in the number of people with a PSE-9+ result in the HSCL+ group, compared with 20% with a PSE-9+ result in the HSCL- group.
This required the recruitment of 810 patients. To compensate for those lost to follow-up, the research team decided to include 1100 patients.

The randomization was achieved independently, via computer software, excluding any human intervention in the selection.

Ethics

The entire study obtained the ethical agreement of the PPC (Protection of Persons Committee). Patients had to give their written, ethical consent to participate. (ID RCB: n°2014-A01790-47; reference CPP: CPP Ouest VI 872; N° Clinical Trial.gov: NCT02414711).
All patients with a score of ≥ 1.75 were informed by the investigating physician, that they could be depressed, in order to initiate the necessary care with their GPs, according to ethical principles and the ethical consent form.

Statistical analysis

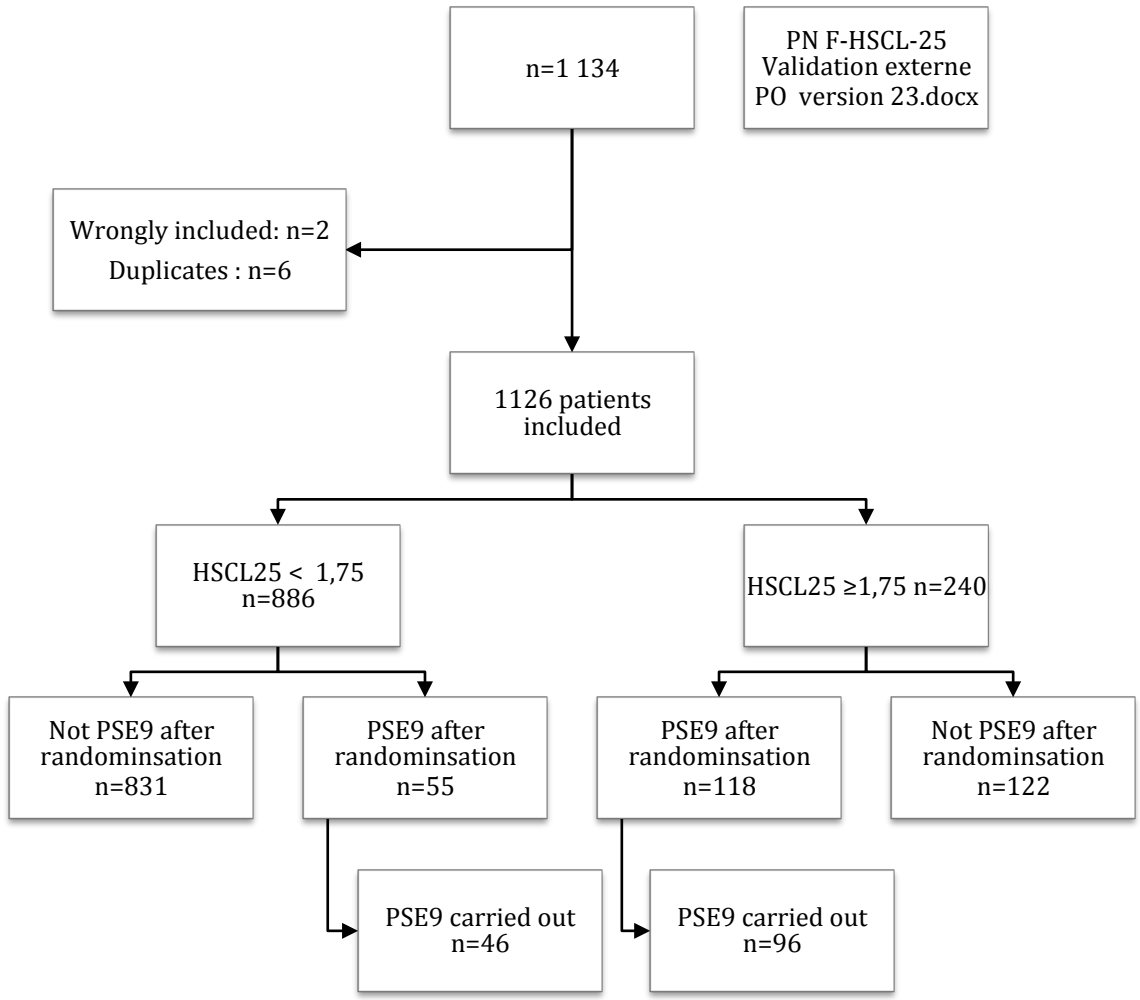
The data was analysed by the Data Management Unit of the Brest University Hospital (Brest CHRU), and the statistical analyses were carried out using SAS software version 9.4 and R version 3.2.0. The tests were carried out with an alpha risk of 5 %.
Descriptive Analysis: Quantitative variables are expressed as means, standard deviations, 25, 50 and 75 quantiles, minimum and maximum values. Qualitative variables are expressed as ratios and percentages.
Comparative Analysis: Univariate comparisons were carried out using relevant standard tests (Student's, Wilcoxon's, chi-squared and Fisher's tests).
External HSCL-25 validation: PPV and NPV were directly calculated, according to formulas based on a contingency table, but this was not possible for sensitivity and specificity. Due to a different artificial sampling step for the PSE-9 positive/negative patients groups, prevalence was not respected. The corrected proportions for the contingency table were calculated, taking into account the number of positive/negative patients and the number of included patients. The whole calculation is in appendix 2. For each parameter, 95% confidence intervals were computed by bootstrap using R library boot.

RESULTS

Clinical and demographic features

The Flow diagram (Fig. 1) shows the number of included patients who had filled in the HSCL-25, whether they were randomised to the PSE-9 group or not, and also shows those who took the PSE-9.

Fig 1. Flow diagram



1134 patients were selected: 2 patients were wrongly included (a pregnant patient and a patient with related disorders) and 6 were duplicates.

1126 patients filled in the HSCL-25 questionnaire. The two groups were created.

HSCL- group:

- 886 patients were randomized according to a ratio of 1/16.
- 831 did not take the PSE-9 test, the study ended for these patients

HSCL+ group:

- 240 patients were randomized according to a ratio of 1/2.
- 122 did not take the PSE-9 test, the study ended for these patients.

Prevalence pitfall

A prevalence established by the F-HSCL-25 of 21.3% was identified among patients consulting their GPs. At the beginning, the sample size was calculated according to prevalence between 5% and 12%. This led to some imbalance in the number of PSE-9 assessments being carried out in the HSCL+ and HSCL- groups.

The study included 1126 French outpatients consulting their GP. Patients were aged between 18 and 94 years. The median age was 59 years and the gender ratio (F/M) was 1.49, Table 1.

Table 1. Patients' characteristics

Variable	**Overall Population** (N=1126)	**Group F-HSCL-25 <1.75** (N=886)	**Group F-HSCL-25 ≥1.75** (N=240)	**inter-group comparisons**
Age				t(408.53)=3.66
Mean +/- SD	55.62 +/- 18.4	56.61 +/- 18.6	51.98 +/- 17.0	P<0.001
Median (q1-Q3)	59 (42 – 70)	61(42-72)	53(38 - 66)	
min-max	18-94	18-94	19-91	
Gender				Chi(1)=25.24
Male	452 (40.14%)	390 (44.02%)	62 (25.83%)	P<0.001
Female	674 (59.86%)	496 (55.98%)	178 (74.17%)	

*inter-group comparisons obtained by Student t test for quantitative variables and Chi² test for qualitative variables

Contingency

55 patients in the HSCL- group had to take the PSE-9. 9 were lost to follow-up; 118 patients in the HSCL+ group had to take the PSE-9. 22 were lost to follow-up. Contingency data are expressed in Table 2, Table 3 and Appendix 2.

Table 2. Contingency table HSCL-25/PSE-9, before prevalence correction

		PSE-9		TOTAL
		« Positive »	**« Negative »**	
HSCL-25	**« Positive »**	67	29	96
	« Negative »	6	40	46
TOTAL		73	69	142

Table 3. Estimated contingency table HSCL-25/PSE-9, after prevalence correction

		PSE-9		TOTAL
		« Positive »	**« Negative »**	
HSCL-25	**« Positive »**	21.12 (15%)	9.14 (6%)	30.26
	« Negative »	14.57 (10%)	97.16 (68%)	111.73
TOTAL		35.69	106.3	142

Outcomes

According to a prevalence of 21.3% (including prevalence corrections) and a cut-off of 1.75, accuracy data gave the following efficiency features, Table 4:

Table 4. Efficiency features

	Value	IC95% *
PPV	69.79	[60.61 – 78.98]
NPV	86.96	[77.22 – 96.69]
Sensitivity	59.17	[43.59 – 80.85]
Specificity	91.40	[88.49 – 94.06]

*Obtained by bootstrap

DISCUSSION

Main Findings

F-HSCL-25 adequately assessed major depression. It demonstrated a capacity to recognise a major depressive episode with a PPV greater than 60%. The specificity of 91% indicated efficiency in identifying significant depression in primary care settings. It is a useful first-line ergonomic diagnostic tool with a low number of false positive patients. The GPs' high depression diagnosis specificity, combined with this tool's efficiency in excluding non-depressive patients with a low margin of error, may serve to identify patients with depressive symptoms much more rapidly.

General discussion

Compared to the study by Nettlebladt, this study resulted in a lower sensitivity (59% versus 76%), it had a higher specificity (91% versus 73%). The prevalence of conspicuous psychiatric morbidity was lower (21% versus 33%). Previous studies showed similar results in terms of sensitivity and specificity [30,34].

A cut-off point of 1.75 was established for case definition in the original English version. According to Nettlebladt & al., choosing a lower cut-off point (1.55) tended to raise the sensitivity (89%), but also gave higher false positives (43%), making it less accurate. Screening capacity is improved at the expense of diagnostic capacity. Due to the average sensitivity rate and the high specificity in the French study, the HSCL-

25, with a cut-off point of 1.75, is valuable in diagnosing patients who require a specific treatment for depression.

The use of a different randomization for each group: a ratio of 1/2 for HSCL+ group, a ratio of 1/16 for HSCL- group, could explain the differences in terms of prevalence, sensitivity and specificity compared with Nettelbladt's study. Nevertheless, the difference in randomization ratios allowed us to balance the number of F-PSE-9 patients in our groups more closely.

A more recent Swedish study by Lundin & al. also examined the concordance between the HSCL-25 scale score and the DSM-IV depression and anxiety disorders using a well-known semi-structured psychiatric interview (SCAN) as a criterion standard [35]. It differs from the previously mentioned studies due to its large sample (8613 patients recruited) based on a general population although not a medical outpatients' population. It found that both the depression and anxiety scales of HSCL-25 performed well in detecting their respective DSM-IV disorders. A combined (global) scale also performed efficiently. Nettlebladt's diagnostic performance, with the cut-off >1.75, showed a higher sensitivity (67.1%), a lower specificity (78.4%), a much weaker PPV (29.8%) but a better NPV (94.6%) than this survey. Our results are comparable with the survey by Lundin and are better than the survey by Nettelbladt.

These results merit comparison with the external validity data of other tools for use in primary care. HSCL-25 like the HADS, is built along two axes: anxiety and depression. HADS has been tested in primary care. It has a higher sensitivity and specificity compared to HSCL-25 (between 0.84 and 0.96) [36]. The ergonomics of this tool seemed more complex to the researchers who preferred the HSCL-25 [18].
The PHQ-9 has a sensitivity between 0.77 and 0.88 and a specificity between 0.88 and 0.94 [37][38]. It is built on the PRIME-MD, not the DSM.
The tools are numerous; researchers will make their choices according to their objectives. Systematic reviews or Meta analyses would then be very useful [39,40].

Strengths

The strength of this study and its relevance for GPs lies in the fact it is specifically set in primary care.
Several types of data quality procedures were followed which increased the reliability of the results, including the appointment of a designated DRCI data manager at the Brest CHRU. Furthermore, the expertise of the stakeholders in the team was balanced to make data collection secure. A stratified randomization was used to ensure both satisfactory statistical power and affordable logistics.
Women accounted for 60% of the sample. The mean age was 59 years. These sample features were comparable to other studies in primary care settings (51 years). The sample characteristics are close to European population-based norms which make it feasible to generalize from these results [4]**.**

Selection bias

A prevalence of 21.3% was identified among patients consulting their GPs. At the beginning of the study, the sample size was calculated according to a prevalence of 5% to 12% in the general population. This study focused on a population which consulted the GP [41]. This prevalence was close to that in Hesbacher's study, but lower than those in Nettelbladt's and Golberg's studies [8,30,34] .

Overestimation of the prevalence is possible due to the internal structure of the HSCL-25. This may occur when anxiety and depression are considered separately; however, it is consistent when anxiety and depression are combined [42,43]. In research, the high NPV and specificity, which enable us to eliminate the false positives, also limit this bias. Therefore, physicians should take this into account in their clinical work. To increase the sensitivity, the HSCL-25 could be combined with a screening tool such as the PHQ-2 [44]. With Brittany currently having the highest rate of suicide in France, it is possible that the depression rate in this region may be higher than in France as a whole [45].

This difference has been taken into account in the statistical analysis. The number of subjects was reassessed during the study because of the unexpected distribution of

the patients in the two groups. The number of subjects necessary to guarantee the statistical power of the study did not depend on this prevalence but on the minimum number of patients placed in each subgroup. This imbalance does not influence the statistical power of the global study. There were 31 (17.9%) lost to follow-up out of the 173 subjects chosen to take the PSE-9 assessment. Other patients replaced them in accordance with the original randomization method. The protocol had entirely anticipated this bias by allowing for 20% to be lost to follow-up.

Information bias

The electronic observation book (eCRF) guaranteed the anonymity of the subjects, allocating them a number and keeping only the first two letters of the surname and first name and the date of birth. The eCRF allowed monitoring and enabled traceability of the study. A research assistant checked the validity and consistency of the information between the paper questionnaires and the eCRF. All collected data were compiled into a numeric database. At the end of the study, all information was checked one last time and the database was frozen before statistical work to prevent any information bias.

Confusion bias

All responses collected during the PSE-9 interviews were retrospectively analysed under the psychiatrist's supervision to avoid misinterpretations and to limit any confusion bias.

Implications

The F-HSCL-25 performs well in detecting symptoms of depression in French primary care and similarly, with its high sensitivity, provides suitable estimates for clinical research purposes. Its possible use by healthcare professionals with basic diagnostic skills in mental health could be an advantage in multidisciplinary research. As this study was carried out among unselected adult patients, further investigations could examine the performance of the HSCL-25 in its French version. This could

include specific samples in primary care, for example, in student populations or in elderly patients, as has already been carried out in Norway and in Sweden respectively [43,46].

CONCLUSION

The F-HSCL-25 demonstrated a capacity to detect symptoms of a major depressive episode. This useful first-line ergonomic diagnostic tool, combined with the GPs' high depression diagnosis specificity, may serve to identify patients with depressive symptoms much more rapidly.

The validation of this reliable and efficient tool throughout Europe, in its translated version, with the same study design, could be of significant epidemiological importance and facilitate the development of more collaborative research within Europe on the subject of depression.

List of abbreviations and definitions

Brest CHRU: Centre Hospitalier Régional Universitaire de Brest

CIC: Centre d'Investigation Clinique

CPP: Comité de Protection des Personnes

DSM IV / V: Diagnostic and Statistical Manual of Mental Disorders 4^{th} / 5^{th} Edition

DUMG: Département Universitaire de Médecine Générale

DRCI: Délégation à la Recherche Clinique et à l'Innovation

eCRF: electronic case report

F-HSCL-25: French version HSCL-25

GPs: General Practitioners

HSCL-25: Hopkins Symptom Checklist - 25 items

PHQ-2: Patient Health Questionnaire 2 items

PSE-9: Present State Examination in its 9^{th} version

Se: Sensitivity

Sp: Specificity

SPURBO = EA 7479 SPURBO: Soins primaires, Santé Publique, Registre des Tumeurs de Bretagne Occidentale

NPV: Negative Predictive Value

PPV: Positive Predictive Value

BIBLIOGRAPHY

1. World Health Organization: **WORLD HEALTH STATISTICS - MONITORING HEALTH FOR THE SDGs**. *World Heal Organ* 2016:1.121.

2. Filipovic-Pierucci A, Samson S, Fagot J-P, Fagot-Campagna A: **Estimating the prevalence of depression associated with healthcare use in France using administrative databases**. *BMC Psychiatry* 2017, **17**:1.

3. 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.

4. King M, Nazareth I, Levy G, Walker C, Morris R, Weich S, Bellon-Saameno JA, Moreno B, Svab I, Rotar D, Rifel J, Maaroos H-I, Aluoja A, Kalda R, Neeleman J, Geerlings MI, Xavier M, de Almeida MC, Correa B, Torres-Gonzalez F: **Prevalence of common mental disorders in general practice attendees across Europe**. *Br J Psychiatry* 2008, **192**:362–367.

5. **La dépression en Europe : faits et chiffres**. 2018.

6. Health Organization Regional Office for Europe W: **The European Mental Health Action Plan 2013-2020.** 2015.

7. Ayuso-Mateos JL, Vázques-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, Gaite 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.

8. Vos T, Allen C, Arora M, Barber RM, Brown A, Carter A, Casey DC, Charlson FJ, Chen AZ, Coggeshall M, Cornaby L, Dandona L, Dicker DJ, Dilegge T, Erskine HE, Ferrari AJ, Fitzmaurice C, Fleming T, Forouzanfar MH, Fullman N, Goldberg EM, Graetz N, Haagsma JA, Hay SI, Johnson CO, Kassebaum NJ, Kawashima T, Kemmer L, Khalil IA, Kyu HH, et al.: **Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–**

2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lance* 2016, **388**:1545–1602.

9. Beck F, Guignard R: **La dépression en France (2005-2010) : prévalence, recours au soin et sentiment d'information de la population**. *Sante Homme* 2012, **421**:43–45.

10. Beck F, Guilbert P, Gautier A: **Baromotre Santé 2005: Attitudes et Comportements de Santé.** *Editions I*, 2007.

11. Robert M., Léon C. DRE: **Comportements suicidaires en France métropolitaine : résultats du baromètre santé 2014.** 2016:1–8.

12. Hérique A, Kahn J-P: **Réalités et recommandations dans la prescription et l'observance des antidépresseurs en médecine générale : évaluation des pratiques dans le traitement de la dépression en Lorraine et Champagne-Ardenne**. *Encephale* 2009, **35**:73–79.

13. OMS: **CIM-10 à usage PMSI**. 2015:888.

14. Jorm AF: **Mental health literacy. Public knowledge and beliefs about mental disorders.** *Br J Psychiatry* 2000, **177**:396–401.

15. IFOP: **Les Français et le système de santé**. *La Vague* 2013, **FD/ AB N°1**.

16. Kovess-Masféty V, Saragoussi D, Sevilla-Dedieu C, Gilbert F, Suchocka A, Arveiller N, Gasquet I, Younes N, Hardy-Bayle M-C: **What makes people decide who to turn to when faced with a mental health problem? Results from a French survey.** *BMC Public Health* 2007, **7**:188.

17. Nabbe P, Le Reste JY, Guillou-Landreat M, Munoz Perez MA, Argyriadou S, Claveria A, Fernández 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* 2017, **39**:99–105.

18. Nabbe P, Le Reste JY, Guillou-Landreat M, Beck-Robert E, Assenova R, Lazic D, Czachowski S, Stojanovi-Špehar S, Hasanagic M, Lingner H, Clavería A, Fernandez San Martin MI, Sowinska A, Argyriadou S, Lygidakis C, Le Floch B, Doerr C, Montier T, Van Marwijk H, Van Royen P: **One consensual depression diagnosis tool to serve many countries: A challenge! A RAND/UCLA methodology**. *BMC Res Notes* 2018, **11**.

19. Derogatis LR, Lipman RS, Rickels K, Uhlenhuth EH, Covi L: **The hopkins symptoms checklist (HSCL): A Self-report Sympotoms Inventory**. *Behav Sci* 1974, **19**:1–11.

20. Derogatis LR, Unger R, Derogatis LR, Unger R: **Symptom Checklist-90-Revised**. In *The Corsini Encyclopedia of Psychology*. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2010.

21. Lipman RS, Covi L, Shapiro AK: **The Hopkins Symptom Checklist (HSCL)--factors derived from the HSCL-90.** *J Affect Disord* 1979, **1**:9–24.

22. 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.

23. 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.

24. 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.

25. Mattisson C, Bogren M, Horstmann V: **Correspondence between clinical diagnoses of depressive and anxiety disorders and diagnostic screening via the Hopkins Symptom Check List-25 in the Lundby Study**. *Nordic Journal of Psychiatry* 2012:1–10.

26. Brislin RW: **Back-Translation for Cross-Cultural Research**. *J Cross Cult Psychol* 1970:2.

27. Bartram D, Muniz J: **International Test Commission Guidelines.** 2016.

28. Lesage AD, Cyr M, Toupin J: **Reliable use of the Present State Examination by psychiatric nurses for clinical studies of psychotic and nonpsychotic patients.** *Acta Psychiatr Scand* 1991, **83**:121–124.

29. Wing JK, Cooper JE, Sartorius N: **Measurement and Classification of Psychiatric Symptoms : An Instruction Manual for the Pse and Catego Program.** *Cambridge Univ Press* 2011.

30. 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.

31. Grazioli R, Terry DJ: **The role of cognitive vulnerability and stress in the prediction of postpartum depressive symptomatology**. *Br J Clin Psychol* 2000, **39**:329–347.

32. Uher R, Payne JL, Pavlova B, Perlis RH: **MAJOR DEPRESSIVE DISORDER IN DSM-5: IMPLICATIONS FOR CLINICAL PRACTICE AND RESEARCH OF CHANGES FROM DSM-IV**. *Depress Anxiety* 2014, **31**:459–471.

33. Stowe ZN, Hostetter AL, Newport D: **The onset of postpartum depression: Implications for clinical screening in obstetrical and primary care**. *Am J Obstet Gynecol* 2005, **192**:522–526.

34. Hesbacher PT, Rickels K, Morris RJ, Newman H, Rosenfeld H: **Psychiatric illness in family practice.** *J Clin Psychiatry* 1980, **41**:6–10.

35. Lundin A, Hallgren M, Forsell Y: **The validity of the symptom checklist depression and anxiety subscales: A general population study in Sweden**. *J Affect Disord* 2015, **183**:247–252.

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. Kroenke K, Spitzer RL, Williams JBW: **The PHQ-9**. *J Gen Intern Med* 2001, **16**:606–613.

38. 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.

39. 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.

40. Mitchell AJ, Vaze A, Rao S: **Clinical diagnosis of depression in primary care: a meta-analysis.** *Lance* 2009, **374**:609–619.

41. White KL, Williams TF, Greenberg BG: **The ecology of medical care**. *N Engl J Med* 1961, **265**:885–92.

42. Al-Turkait FA, Ohaeri JU, El-Abbasi AHM, Naguy A: **Relationship between symptoms of anxiety and depression in a sample of arab college students**

using the Hopkins Symptom Checklist 25. *Psychopathology* 2011, **44**:230–241.
43. Skogen J, Øverland S, Orf S, Aarø L: **The factor structure of the Hopkins Symptoms Checklist (HSCL-25) in a student population : A cautionary tale .** :25.
44. Arroll B, Goodyear-Smith F, Crengle S, Gunn J, Kerse N, Fishman T, Falloon K, Hatcher S: **Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population**. *Ann Fam Med* 2010, **8**:348–353.
45. Kopp-Bigault C, Walter M, Thevenot A: **The social representations of suicide in France: An inter-regional study in Alsace and Brittany**. *Int J Soc Psychiatry* 2016, **62**:737–748.
46. Fröjdh K, Håkansson A, Karlsson I, Frojdh 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.

Appendix 1: HSCL-25 Original version / HSCL-25 French version

ITEMS N°	**HSCL-25 ORIGINAL VERSION**	F-HSCL-25
	Choose the best answer for how you felt over the past week	Veuillez choisir la réponse qui décrit le mieux comment globalement vous vous sentiez toute la semaine dernière
1	Being scared for no reason	Vous avez peur sans raison
2	Feeling fearful	Vous vous sentez effrayé
3	Faintness	Vous avez une sensation d'étourdissement
4	Nervousness	Vous vous sentez nerveux
5	Heart racing	Vous avez l'impression que votre cœur bat anormalement vite
6	Trembling	Vous avez la sensation de trembler
7	Feeling tense	Vous vous sentez tendu
8	Headache	Vous avez des maux de tête

9	Feeling panic	Vous vous sentez paniqué
10	Feeling restless	Vous vous sentez agité
11	Feeling low in energy	Vous manquez d'énergie
12	Blaming oneself	Vous ressentez une sensation de culpabilité
13	Crying easily	Vous pleurez facilement
14	Losing sexual interest	Vous ressentez un désintérêt pour la vie sexuelle
15	Feeling lonely	Vous avez une sensation de solitude
16	Feeling hopeless	Vous vous sentez désespéré
17	Feeling blue	Vous avez le cafard
18	Thinking of ending one's life	Vous avez pensé à mettre fin à votre vie
19	Feeling trapped	Vous vous sentez pris au piège
20	Worrying too much	Vous vous inquiétez trop
21	Feeling no interest	Plus rien ne vous intéresse
22	Feeling that everything is an effort	Tout est un effort pour vous
23	Worthless feeling	Vous avez le sentiment d'être bon à rien
24	Poor appetite	Vous avez perdu l'appétit
25	Sleep disturbance	Votre sommeil est perturbé

Appendix 2: Calculation of the F-HSCL-25 predictive values

Table 2. Contingency table HSCL-25/PSE-9, before prevalence correction

		PSE-9		**TOTAL**
		« Positive »	**« Negative »**	
HSCL-25	**« Positive »**	67 (69.79%)	29 (30.21%)	96
	« Negative »	6 (13.04%)	40 (86.96%)	46
TOTAL		73	69	142

We could calculate PPV and NPV directly from the contingency table, according to the following formulas:

PPV = TP / (TP + FP) = 67 / (67 + 29) = 0.70

NPV = TN / (TN + FN) = 40 / (40 + 6) = 0.87

However, the sampling step was artificial. It was determined by the protocol to improve the feasibility of the study, as 1/16 (HSCL-) and 1/2 (HSCL +) patient. The prevalence is not respected.

We could not apply the contingency table directly, according to the formulas for Se and Sp

Corrective formulas to obtain Se and Sp

The probability of the test being positive or negative from the contingency table should be calculated as follows:

The number of positive tests (HSCL ≥ 1.75) divided by the number of patients included: P (HSCL +) = (HSCL +) / N

The number of negative tests (HSCL <1.75) divided by the number of patients included: P (HSCL-) = (HSCL-) / N

N = 1126

P(HSCL+) = (HSCL+) / N = 240 / 1126 = 0.21

P(HSCL-) = (HSCL-) / N = 886 / 1126 = 0.79

Now we are able to calculate the corrected proportions for the contingency table:

Proportion of True Positive = PPV * P (HSCL +) = 0.70*0.21 = 0.15

Proportion of True Negative = NPV * P (HSCL-) = 0.87*0.79 = 0.68

Proportion of False positive = (1-PPV) * P (HSCL +) = (1-0.7)*0.21 = 0.06

Proportion of False Negative = (1-NPV) * P (HSCL-) (1-0.87)*0.79 = 0.10

Table 3. Estimated contingency table HSCL-25/PSE-9, after prevalence correction

		PSE-9		**TOTAL**
		« Positive »	**« Negative »**	
HSCL-25	**« Positive »**	21.12 (15%)	9.14 (6%)	30.26
	« Negative »	14.57 (10%)	97.16 (68%)	111.73
TOTAL		35.69	106.3	142

The corrected number on the contingency table can then be calculated by multiplying by the number of patients who have passed the PSE (142 outpatients).

Then directly apply the calculation formulas:
Se = TP / (TP + FN) = 21.12 / (21.12+35.69) = 0.59
Sp = TN / (TN + FP) = 97.16 / (97.16 + 9.14) = 0.91

The calculation of the NPV and the PPV from the initial or modified contingency table were, of course, identical.

This could be expressed concisely and applied rapidly by using the following corrective formulas directly:

*Se = PPV * P(HSCL+) / [P(HSCL+) * PPV] + [P(HSCL-) * (1-NPV)]*
*Sp = NPV * P(HCSL-) / [P(HSCL+) * PPV] + [P(HSCL-) * (1-NPV)*

Se= Sensitivity; Sp= Specificity; P: Prevalence; PPV = Positive Predictive Value; NPV = Negative Predictive Value; P(HSCL+) = Patient HSCL+ frequency; P(HSCL−) = Patient HSCL− frequency

CHAPTER 7

General discussion

In this chapter, the findings of the thesis are summarized, a critical comparison with existing literature is provided, strong and limitations are displayed and a general perspective is designed. Implications for practice, medical education and future research are discussed.

The goal of this research was to find the most interesting diagnosis tool for depression, adapted to daily practice, suitable for collaborative research work in Europe, throughout languages and cultures.

The specific objectives were:
To find diagnostic tools validated against a face-to-face psychiatrist examination, using DSM major depression criteria as Gold Standard.
to consensually select a tool, according to their qualities of effectiveness, reliability and ergonomics combined; suitable for research in daily General Practice.
To translate it into as many languages as participants in the study, ensuring the linguistic and semantic stability of the transfer
To validate the translated forms of this tool, starting in France, to propose a standardized validation protocol to the different countries of the study.

As a first step, the main results of each step are listed, then a comparison to the existing literature is undertaken, followed by an explanation of the strengths and weaknesses of the study, to conclude with teaching implications and openings for research in future.

What were the results of the systematic literature review?

The design of RSL has been designed to achieve a very "sharp" result. At the end of the PRISMA process, [1] 4 studies were selected [2][3][4][5]. They were worth the following 7 tools: GDS-30 and CESD-R, PSC-51 and HADS, GDS-5 and GDS-15, HSCL-25.

All the psychometric data of effectiveness were extracted from articles: Sensitivity, specificity, positive and negative predictive values. To allow a comparison of the effectiveness, the Youden index (Se + Sp-1) was calculated [6][7]. (see table below)

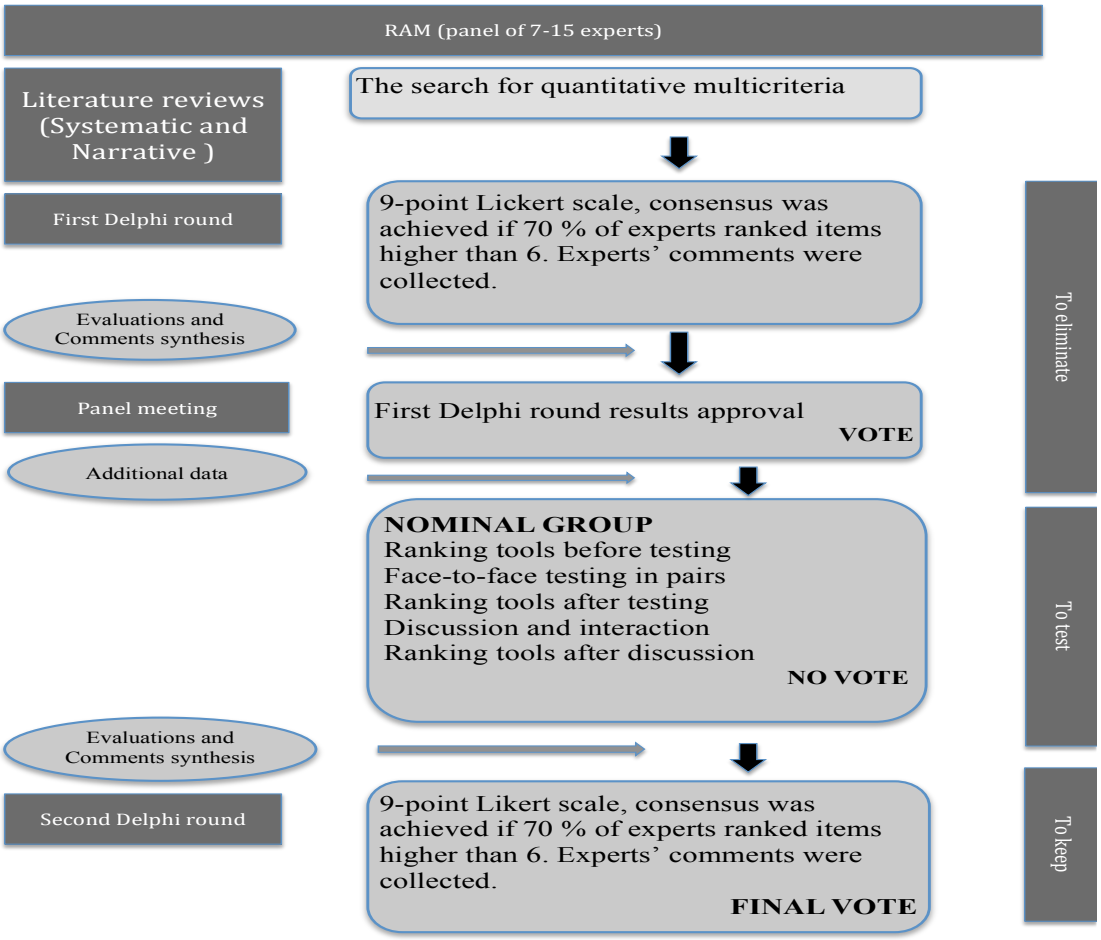
TITLE	First two Authors	Publication Year	Tool	Tool used in face to face interview	Interviewer using DSM criteria	Individuals	Se	Sp	PPV	NPV	YI	Mean Age
Usefulness of two instruments in assessing depression among elderly Mexicans in population studies and for primary care.	Sánchez-García S, Juárez-Cedillo 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	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	

physical symptoms in primary care.												
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.	Fröjdh K, Hakansson NA & al.	2004	HSCL-25	MADRS based on DSM-IV	Yes	74	0.94	0.94			0.88	78.5

What was the RAND/UCLA contribution?

The design of the RAND / UCLA (see figure below) allowed the comparison of the tools on quantitative and qualitative criteria [8]. Effectiveness, reliability and ergonomics were taken into account in order to find the tool with the best combination of the three. The effectiveness data came from the SRL. A narrative review was completed to extract reliability data (Cronbach's Alpha) [9]. Ergonomics data has been collected from the literature.

The HSCL-25 has emerged as the most convenient tool (see table below) [10][11].



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

What was the contribution of translations?

Using a qualitative, ecologically integrated procedure in daily practice, nine consensual translations of the HSCL-25 in three language families (Greek, Slavic and Romance) were obtained.

The methodology based on the Brislin model [12], ensured the ecological validity in family practice of the entire process, as well as maintaining the linguistic and semantic stability of language transfer [13][14]. The produced versions of the HSCL-25 were homogeneous to the original version. (see translations tables below)

ITEM	HSCL-25 ORIGINAL VERSION	GREECE	POLAND	BULGARIA	CROATIA
A	Choose the best answer for how you felt over the past week	Επιλέξτε την καλύτερη απάντηση για το πώς αισθανθήκατε την τελευταία εβδομάδα	Wybierz najlepszą odpowiedź	Изберете отговора, който най-добре описва как сте се чувствали през изминалата седмица	Izaberite jedan odgovor koji najbolje opisuje kako ste se osjećali tijekom prošlog tjedna:
1	Being scared for no reason	Είμαι τρομαγμένος/η χωρίς αιτία	Bać się bez powodu	Чувство за уплаха без причина	Bili ste bezrazložno uplašeni
2	Feeling fearful	Αισθάνομαι φοβισμένος /η	Poczucie strachu	Чувство за страх	Bojali ste se
3	Faintness	Αίσθημα λιποθυμιάς	Omdlenia	Отпадналост	Bili ste slabi
4	Nervousness	Νευρικότητα	Nerwowość	Нервност	Bili ste nervozni
5	Heart racing	Ταχυπαλμία	Kołatanie serca	Сърцебиене	Ubrzano vam je lupalo srce
6	Trembling	Τρεμούλα	Drżenia	Треперене	Drhtali ste
7	Feeling tense	Αισθάνομαι υπερένταση	Poczucie napięcia	Чувство за напрежение	Bili ste napeti
8	Headache	Πονοκέφαλος	Bóle głowy	Главоболие	Boljela vas glava
9	Feeling panic	Αισθάνομαι πανικό	Uczucie paniki	Чувство за паника	Bili ste u panici
10	Feeling restless	Αισθάνομαι ταραχή	Uczucie niepokoju	Чувство на безпокойство	Bili ste uznemireni
11	Feeling low in energy	Αισθάνομαι ότι δεν έχω ενέργεια	Poczucie braku energii	Усещане за понижена енергия	Niste imali dovoljno energije
12	Blaming oneself	Κατηγορώ τον εαυτό μου	Obwinianie samego siebie	Самообвинение	Okrivljavali ste se
13	Crying easily	Εύκολο κλάμα	Płaczliwość	Плачливост	Bili ste plačljivi
14	Losing sexual interest	Απώλεια σεξουαλικού ενδιαφέροντος	Utrata zainteresowań sferą seksualną	Загубата на сексуален интерес	Niste bili zainteresirani za spolni odnos
15	Feeling lonely	Αισθάνομαι μοναξιά	Poczucie osamotnienia	Чувство за самотност	Bili ste usamljem
16	Feeling hopeless	Αισθάνομαι απελπισμένος/η	Poczucie beznadziejności	Чувство за безнадежност	Osjećali ste sebeznadno
17	Feeling blue	Νοιώθω πεσμένος/η	Poczucie przygnębienia	Чувстам се нещастен	Bili ste sjetni
18	Thinking of ending one's life	Σκέφτομαι να δώσω τέλος στη ζωή	Myśli samobójcze	Мисли за самоубийство	Razmišljali ste da si oduzmete život
19	Feeling trapped	Αισθάνομαι παγιδευμένος /η	Poczucie uwięzienia	Чувстам се като в капан	Osjećali ste sekao da ste u klopci
20	Worrying too much	Ανησυχώ υπερβολικά	Zamartwianie się	Притеснявам се твърде много	Bili ste previše zabrinuti
21	Feeling no	Αισθάνομαι ότι τίποτε	Poczucie braku	Чувство за загуба	Bez interesa za bilo što

	interest	δεν είναι ενδιαφέρον	zainteresowań	на интерест	
22	Feeling that everything is an effort	Αισθάνομαι ότι για το καθε τί χρειάζεται να κάνω προσπάθεια	Poczucie, że wszystko jest ciężarem	Чувство, че всичко изисква усилие	Sve vam je bilo naporno
23	**feelings of Worthlessness**	Αισθάνομαι ότι δεν αξίζω τίποτε	Poczucie bezwartościowości	Чувство за безполезност	Osjećali ste se bezvrijedno
24	Poor appetite	Μείωση της όρεξης	Słaby apetyt	Лош апетит	Imali ste slab apetit
25	Sleep disturbance	Διαταραχές ύπνου	Zaburzenia snu	Нарушения на съня	Imali ste problema sa spavanjem

ITEM	HSCL-25 ORIGINAL VERSION	CASTILE	CATALONIA	GALICIA	ITALY	FRANCE
A	Choose the best answer for how you felt over the past week	Elija la respuesta que mejor describa cómo se ha sentido durante la semana pasada	Triï la millor resposta per indicar com s'ha sentit en la darrera setmana	Escolla a resposta que mellor describa como se sentiu durante a semana pasada	Scegliere la risposta più adatta su come ti sei sentito/a nell'ultima settimana	Veuillez choisir la réponse qui décrit le mieux comment globalement vous vous sentiez toute la semaine dernière
1	Being scared for no reason	Se asusta sin motivo	Estar espantat/espantada sense motiu aparent	Asústase sen motivo	Avere paura senza motivo	Vous avez peur sans raison
2	Feeling fearful	Siente miedo	Sentir por	Ten medo	Sentirsi impauriti	Vous vous sentez effrayé
3	Faintness	Debilidad	Debilitat	Debilidade	Sensazione di mancamento	Vous avez une sensation d'étourdissement
4	Nervousness	Nerviosismo	Nerviosisme	Nerviosismo	Esseri nervosi	Vous vous sentez nerveux
5	Heart racing	Palpitaciones	Cor accelerat	Palpitacións	Sentire il cuore battere veloce	Vous avez l'impression que votre cœur bat anormalement vite
6	Trembling	Tiembla	Tremola	Ten tremores	Tremore	Vous avez la sensation de trembler
7	Feeling tense	Se siente tenso/a	Sentir-se tens/a	Séntese tenso/a	Sensazione di tensione	Vous vous sentez tendu
8	Headache	Dolor de cabeza	Mal de cap	Dor de cabeza	Avere mal di testa	Vous avez des maux de tête
9	Feeling panic	Siente pánico	Sensació de pànic	Sente pánico	Sensazione di panico	Vous vous sentez paniqué
10	Feeling restless	Siente inquietud	Sensació d'inquietud	Séntese inquedo/a	Sensazione di irrequietezza	Vous vous sentez agité
11	Feeling low in energy	Siente que le falta energía	Sensació de manca d'energia	Sente que lle falta enerxía	Sentirsi senza energia	Vous manquez d'énergie
12	Blaming oneself	Se culpa a sí mismo/a	Culpar-se un/a mateix/a	Cúlpase a si mesmo/a	Avere sensi di colpa	Vous ressentez une sensation de culpabilité
13	Crying easily	Llora con facilidad	Plora fàcilment	Chora con facilidade	Piangere facilmente	Vous pleurez facilement
14	Losing sexual interest	Pierde el interés sexual	Pèrdua de l'interès sexual	Perda do interese sexual	Perdere l'interesse sessuale	Vous ressentez un désintérêt pour la vie sexuelle
15	Feeling lonely	Se siente solo/a	Sentir-se sol/a	Séntese só/soa	Sentirsi soli	Vous avez une sensation de solitude

16	Feeling hopeless	Se siente sin esperanza	Sentiment de desesperança	Séntese sen esperanza	Sentirsi senza speranza	Vous vous sentez désespéré
17	Feeling blue	Se siente triste	Sentir-se trist/a	Séntese triste	Sentirsi tristi	Vous avez le cafard
18	Thinking of ending one's life	Piensa en acabar con su vida	Pensa en treure's la vida	Pensa en acabar coa súa vida	Avere pensieri di togliersi la vita	Vous avez pensé à mettre fin à votre vie
19	Feeling trapped	Se siente atrapado/a	Sentir-se atrapat/atrapada	Séntese atrapado/a	Sentirsi intrappolati	Vous vous sentez pris au piège
20	Worrying too much	Se preocupa en exceso	Preocupar-se en excés	Preocúpase en exceso	Preoccuparsi troppo	Vous vous inquiétez trop
21	Feeling no interest	No siente interés por nada	Sentiment de manca d'interès	Non sente interese por nada	Non avere alcun interesse	Plus rien ne vous intéresse
22	Feeling that everything is an effort	Siente que todo le cuesta un esfuerzo	Sentir que tot és un esforç	Sente que todo lle supón un esforzo	Sentire che tutto è uno sforzo	Tout est un effort pour vous
23	**feelings of Worthlessness**	Se siente inútil	Sentir-se inútil	Séntese inútil	Sentirsi inutili	Vous avez le sentiment d'être bon à rien
24	Poor appetite	poco apetito	Pèrdua de la gana	Poco apetito	Avere poco appetito	Vous avez perdu l'appétit
25	Sleep disturbance	Problemas para dormir	Alteració de la son	Alteracións do sono	Disturbi del sonno	Votre sommeil est perturbé

SCALE INSTRUCTIONS ORIGINAL VERSION	GREECE	POLAND	BULGARIA	CROATIA
The HSCL-25 score is based on pencil-and-paper self-report of 25 questions about the presence and intensity of anxiety and depression symptoms over the last week.	Η βαθμολογία του HSCL-25 βασίζεται σε γραπτό ερωτηματολόγιο αυτοαξιολόγησης 25 ερωτήσεων σχετικά με την παρουσία και την ένταση των συμπτωμάτων άγχους και κατάθλιψης κατά την τελευταία εβδομάδα. Οι συμμετέχοντες	Ocena testu HSCL-25 oparta jest na kwestionariuszu 25 pytań, w którym zakreśla się na papierze obecność i nasilenie objawów lęku i depresji w ciągu ostatniego tygodnia.	Резултатът от HSCL-25 се основава на самостоятелно попълнен инструмент на хартиен носител, включващ 25 въпроса за наличието и интензивността на симптоми на тревожност и депресия през последната седмица.	HSCL-25 skor sastoji se od 25 pitanja koja se rješavaju jednostavno olovkom i papirom, a temelji se na samoprocjeni prisutnosti i intenzitetu ansksioznih i depresivnih simptoma tijekom prošlog tjedna.
Participants answer to one of four categories for each item on a four-point scale ranging from 1 to 4	απαντούν σε μία από τις τέσσερις κατηγορίες για κάθε ερώτημα σε μια κλίμακα εύρους τεσσάρων βαθμών με τιμές από 1 μέχρι 4.	Badani odpowiadają na jedno z czterech możliwych kategorii na skali mierzącej wartości od 1 do 4.	Участниците избират една от категориите за всяка позиция по скала от четири точки от 1.00 до 4.00.	Ispitanici odgovaraju jednom od četiri kategorija za svako pitanje na skali od 1-4.
1.“Not at all”	Καθόλου	Wcale	Съвсем не	Nimalo
2.“A little”	Λίγο	Trochę	Незначително	Malo
3.”Quite a bit”	Αρκετά	Znacznie	Съвсем малко	Dosta
4.“Extremely”	Πάρα πολύ	Bardzo mocno	Извънредно	Jako

SCALE INSTRUCTIONS ORIGINAL VERSION	CASTILE	CATALONIA	GALICIA	ITALY	FRANCE
The HSCL-25 score is based on pencil-and-paper self-report of 25 questions about the presence and intensity of anxiety and depression symptoms over the last week.	La puntuación HSCL-25 se basa en un cuestionario auto cumplimentado con lápiz y papel, de 25 preguntas sobre la presencia y la intensidad de ansiedad y síntomas depresivos en la última semana.	L'escala HSCL-25 es basa en un qüestionari auto administrat de 25 preguntes, sobre la presència i la intensitat de símptomes d'ansietat i depressió en la darrera setmana.	A puntuación HSCL-25 baséase nun cuestionario cumprimentado con lapis e papel, de 25 preguntas sobre a presenza e a intensidade de ansiedade e síntomas depresivos na última semana.	Il punteggio dell'HSCL-25 si basa sulla compilazione di un questionario di autovalutazione in cartaceo ("carta/penna") di 25 domande sulla presenza e intensità di sintomi di ansia e depressione nel corso dell'ultima settimana.	La HSCL-25 est un auto-questionnaire en 25 questions relatives à la présence et à l'intensité des symptômes d'anxiété et de dépression durant toute la semaine dernière.
Participants answer to one of four categories for each item on a four-point scale ranging from 1 to 4	Los/ las participantes responden una de cuatro categorías para cada ítem, en una escala de cuatro puntos que van desde 1 a 4.	Els/les participants responen a una de les quatre categories per a cada ítem en una escala de quatre punts que va de l'1 al 4.	Os participantes responden unha de catro categorías para cada ítem, nunha escala de catro puntos que van desde 1 a 4.	I partecipanti rispondono a una delle quattro categorie per ciascun sintomo su una scala di punteggio che va da 1 a 4.	Les participants cotent chaque proposition, sur une échelle en quatre points, cotée de 1 à 4.
1."Not at all"	En absoluto	Gens	En absoluto	Per niente	Pas du tout d'accord
2."A little"	Un poco	Una mica	Un pouco	Poco	Un peu d'accord
3."Quite a bit"	Bastante	Bastant	Bastante	Abbastanza	Plutôt d'accord
4."Extremely"	Mucho	Molt	Moito	Moltissimo	Complètement d'accord

SCALE INSTRUCTIONS ORIGINAL VERSION	GREECE	POLAND	BULGARIA	CROATIA
The HSCL-25 score is calculated by dividing the total score (sum score of items) by the number of items answered (ranging between 1.00 and 4.00). It is often used as the measure of distress.	Η βαθμολογία του HSCL-25 υπολογίζεται διαιρώντας τη συνολική βαθμολογία (αθροιστική βαθμολογία των ερωτημάτων), διά του αριθμού των ερωτημάτων που απαντήθηκαν (κυμαινόμενο μεταξύ του 1,00 έως 4,00). Συχνά χρησιμοποιείται για τη μέτρηση της δυσφορίας.	Wynik testu HSCL-25 jest obliczany poprzez podzielenie całkowitej liczby punktów (suma punktów z każdej pozycji testu) przez liczbę pozycji na które udzielono odpowiedzi (w skali od 1 do 4). Często służy on do pomiaru dystresu.	HSCL-25 резултатът се изчислява, като се раздели общият брой точки (сбор точки по критерий) на броя на отговорените критерии (вариращи между 1,00 и 4,00). Той често се използва като мярка за страдание.	Skor HSCL-25 se izračunava dijeljenjem ukupnog zbroja (zbroj skora pojedinih pitanja) s brojem odgovorenih pitanje (raspon od 1,00 do 4,00). Obično se koristi za mjerenje distresa.
The patient is considered as a "*probable psychiatric case"* if the mean rating on the HSCL-25 is ≥1.55.	Ο ασθενής θεωρείται σαν "πιθανό ψυχιατρικό περιστατικό" εάν η μέση βαθμολογία του HSCL-25 είναι >=1,55	Pacjenta uważamy za "prawdopodobny przypadek psychiatryczny" jeśli średnia ocena w teście HSCL-25 jest >/ (większa lub równa) 1,55.	Пациентът се приема като "вероятно психиатричен случай", ако средната оценка по HSCL-25 е ³ 1,55.	Pacijent se smatra « vjerojatno psihijatrijskim slučajem » ako je srednja vrijednost na HSCL-25 ≥ 1,55.
A cut-off value of ≥1.75 is generally used for diagnosis of major depression defined as "*a case, in need of treatment"*. This cut-off point is recommended as a valid predictor of mental disorder as assessed independently by clinical interview, somewhat depending on diagnosis and gender.	Το όριο του >= 1,75 γενικώς χρησιμοποιείται για τη διάγνωση της μείζονος κατάθλιψης που ορίζεται ως "περίπτωση που χρήζει θεραπείας". Αυτό το όριο συνίσταται σαν ένας έγκυρος προγνωστικός δείκτης ψυχικής διαταραχής, όπως εκτιμάται ανεξάρτητα από την κλινική εικόνα, η οποία εξαρτάται κάπως από τη διάγνωση και το φύλο.	Wartość graniczną>/ (większą lub równą) 1,75 ogólnie przyjmuje się w diagnozowaniu ciężkiej depresji, definiowanej jako „przypadek wymagający leczenia." Wartość ta jest zalecana jako istotny czynnik w przewidywaniu obecności choroby psychicznej, wymagającej jednak niezależnego wywiadu klinicznego i w pewnym sensie zależy od rozpoznania i płci.	Гранична стойност от ³ 1,75 обикновено се използва за диагностициране на тежка депресия и определя случая като "случай, нуждаещ се от лечение". Тази гранична стойност, получена независимо от клиничното интервю и зависеща до определена степен от диагнозата и пола, се препоръчва като валиден предиктор за психично разстройство.	Razdjelna točka (cut-off) ≥1,75 se koristi za dijagnozu velikog depresivnog poremećaja i to kao „slučaj koji zahtjeva liječenje". Razdjelna točka se preporuča kao validni prediktor mentalnog poremećaja podjednako kao i sama procjena neovisnim kliničkim intervjuom, dijelom ovisan o dijagnozi i spolu.
The administration time of HSCL 25 is 5 to 10 minutes.	Ο χρόνος χορήγησης του HSCL 25 είναι 5 έως 10 λεπτά.	Czas na wykonanie testu HSCL 25 wynosi od 5 do 10 minut.	Времето за провеждане HSCL-25 е от 5 до 10 минути.	Vrijeme za ispunjavanje HSCL-25 je 5-10 minuta.

SCALE INSTRUCTIONS ORIGINAL VERSION	CASTILE	CATALONIA	GALICIA	ITALY	FRANCE
The HSCL-25 score is calculated by dividing the total score (sum score of items) by the number of items answered (ranging between 1.00 and 4.00). It is often used as the measure of distress.	La puntuación del HSCL-25 se calcula dividiendo la puntuación total (sumando la puntuación de todos las preguntas) entre el número de respuestas (varía entre 1,00 y 4,00). Se usa habitualmente para medir el malestar psicológico.	La puntuació total del HSCL-25 es calcula dividint la suma de la puntuació dels diferents ítems pel número d'ítems contestats. El resultat total oscil·la entre 1,00 i 4,00. Aquesta escala sovint s'utilitza com a mesura del malestar psicològic.	A puntuación do HSCL-25 calcúlase dividindo a puntuación total (a suma de todas as preguntas) entre o número de respostas (cuxa puntuación oscila entre 1,00 e 4,00). Úsase de forma habitual para medir o nivel del malestar psicológico.	Il punteggio dell' HSCL-25 si calcola dividendo il punteggio totale (somma dei punteggi degli elementi) con il numero di elementi risposti (che variano da 1,00 a 4,00). Spesso si usa come misura di ansietà	Le score du HSCL- 25 se calcule en divisant la somme des cotations des propositions par le nombre de réponses reçues. Le résultat final est compris entre 1,00 à 4,00. Il est couramment utilisé pour mesurer la souffrance psychologi-que.
The patient is considered as a "*probable psychiatric case*" if the mean rating on the HSCL-25 is ≥1.55.	El/la paciente se considera un "probable caso psiquiátrico" si el valor medio del HSCL-25 es ≥1,55.	El/la pacient és considerat/considerada com a " probable cas psiquiàtric " si la qualificació mitjana del HSCL-25 és ≥ 1,55.	Considérase que o/a paciente é un "caso psiquiátrico probable" se o valor medio do HSCL-25 é ≥ 1,55.	Il paziente è considerato come un "probabile caso psichiatrico" se il punteggio medio dell'HSCL-25 è ≥1,55.	Le patient est considéré comme « probablement atteint d'un trouble psychiatrique » si le score moyen du HSCL-25 est supérieur ou égal à 1,55.
A cut-off value of ≥1.75 is generally used for diagnosis of major depression defined as "*a case, in need of treatment*". This cut-off point is recommended as a valid predictor of mental disorder as assessed independently by clinical	Por lo general se usa un valor de corte de ≥1,75 para el diagnóstico de depresión mayor, definida como "un caso que necesita tratamiento". Este valor de corte se considera un predictor válido de un trastorno mental, evaluado de forma independiente mediante	Generalment s'utilitza un punt de tall ≥1,75 per al diagnòstic de la depressió major i es defineix com " cas que precisa de tractament". Es recomana aquest punt de tall com un predictor vàlid de trastorn mental com ho seria l'avaluació independent per entrevista clínica, depenent en part del diagnòstic i del gènere.	Polo xeral, úsase un valor de corte ≥ 1,75 para diagnosticar a depresión maior, definida como "un caso que precisa tratamento". Este valor de corte recoméndase como un predictor válido dun trastorno mental, avaliado independentemente por medio de entrevistas clínicas, aínda que depende	Un cut-off che sia >=1,75 è normalmente usato per la diagnosi di depressione maggiore definita come "un caso che necessita di trattamento". Questo cut-off è raccomandato come un valido predittore di disordine mentale come valutato in modo	Un score supérieur ou égal à 1,75 diagnostique généralement une dépression caractérisée et définit « un patient nécessitant un traitement ». Ce seuil est considéré comme un score prédictif validé des troubles mentaux. Il a été évalué de manière

interview, somewhat depending on diagnosis and gender.	entrevista clínica, aunque depende en parte del diagnóstico y el género.		en parte do diagnóstico e do xénero.	indipendente da un colloquio clinico, dipendente in qualche modo dalla diagnosi e dal genere	indépendante par des études cliniques. Il varie peu quelles que soient les situations diagnostiques et le sexe.
The administration time of HSCL 25 is 5 to 10 minutes.	El tiempo de administración del HSCL-25 es de 5 a 10. minutos.	El temps d'administració del HSCL 25 és de 5 a 10 minuts.	O tempo de realización do HSCL-25 é de 5 a 10 minutos.	Il tempo di somministrazione dell'HSCL-25 è da 5 a 10 minuti.	Remplir le questionnaire HSCL-25 prend entre 5 et 10 minutes.

What the validation study gave?

A French form of HSCL-25 was produced: F-HSCL-25. It was then necessary to test the psychometric qualities of this French questionnaire.

The F-HSCL-25 allowed the recognition of a major depressive episode with a VPP greater than 60%. Its specificity of 91% indicates its effectiveness in the identification of the depression in general practice. It is a helpful first-line diagnostic tool with an insufficient number of false positive patients.

The diagnostic specificity of depression is high among general practitioners. [15] Combined with the effectiveness of this tool in excluding non-depressive patients with a low margin of error, effective synergy could be achieved.

Comparison to the existing literature

Categorical diagnostic tools are not widely used in general practice. Psychiatrists report the difficulty of having to combine the validity, utility and status of the disease into one tool. This would prevent clinicians from using it [16]. The international medical community doubts about the validity of the DSM. And it is restrained to use as a reference [17]. The GPs refuted the use of preformed tools in their practice [18]. Tools and scales from the DSM are used in research and not in daily practice.

Tools extracted from literature

GDS-30 was developed in 1982 to diagnose and quantify depression in elderly patients [19]. It was designed with 30 questions using a binary response mode. It is focused on the symptoms of the previous week. It is used for research purposes [20][21]. The GDS-5 and the GDS-15 are short versions of the GDS-30 for better ergonomics [22][23].

CESD-R was developed in 1977 to diagnose and quantify depression [24]. It was designed with 20 questions. The answers are on a 4-point Likert scale. It focuses on the symptoms of the previous week. It is also widely used in research [25].

PSC-51 is a list of physical symptoms in 51 elements. It is little used [3].

HADS was developed in 1983 to diagnose and quantify hospital depression [26]. This is a 14-question tool, using a 4-point Likert scale for answers. It is focused on the symptoms of the previous week. HADS has been widely and extensively used for clinical and research purposes [27]. It is translated into several languages [28]. Validated for use in primary care, it seems complex and rather "calibrated" for research purposes rather than daily practice [27][29][30].

HSCL-25 was developed in 1974 to diagnose and quantify depression [11]. This is a tool in 25 questions. She uses for the answers a scale of Likert in 4 points. It is focused on the symptoms of the previous week. It is widely used and well anchored

in primary care. It is specifically used with refugee and suffering populations [31][32][33][34][35][36][37].

Compare and select a tool: a challenge

A comparison of tools based on effectiveness criteria was not satisfactory enough, given the wide disparity in populations and sampling. Another way of comparison had to be found. RAM or RAND / UCLA was selected for its above-mentioned qualities in the method, which allowed comparison and a selection process.

This RAM was based on an SRL, which increased the quality level compared to the original design based on an unsystematic review [8]. The ergonomic factor was an important criterion in maintaining a relationship between patients and GPs. Through this process, the researchers demonstrated that ergonomics were decisive in choosing a tool suited to future research [38].

At the end of the first Delphi round, the PSC-51, the CES-DR and the GDS-30 were not kept for a too low efficiency. The GDS was also rejected in its short versions in 5 and 15 questions for too low reliability.

HSCL 25 and HADS have passed the first stage to be discussed.
In the final vote, the HSCL-25 was selected. Its ergonomic qualities were predominant, accordingly to the evolution of the votes during the test phase. In the end, the HSCL-25 best combined effectiveness, reliability and ergonomics for the diagnosis of depression in European primary care practice in a research orientation.

Its robust efficacy and reliability scores, [39][40][41] and its ergonomic qualities allowed for multicenter collaborative research across Europe anchored in primary care. It also allowed cross-disciplinary research between psychiatrists and MG.

But it is a self-questionnaire, which must be easily understood by the general population and outpatients without error of meaning. The group had to remain

vigilant about the translation process in several European languages. And its application in practice had to be demonstrated for every national translation.

Language-to-language transfer, a subtle exercise

Using a three-step qualitative procedure, ecologically embedded in primary care, nine consensus translations of HSCL-25 were obtained.

The translation by Delphi procedure has guaranteed the linguistic transfer from the original version to the target language.

To guarantee the semantic transfer, a blind back translation back made it possible an analysis of the two English versions by a linguist meticulously found the translation inconsistencies. Once detected, they were submitted to each national investigator, discussed in expert group, possibly rediscovered in each country to find the most appropriate formulation. This methodical and transcultural validation ensured the homogeneous transfer from one language to another as well as its reliability [13][14][12].

The Greek translation remained the most stable, followed by Bulgarian.
Question 17 "Feeling blue" was the most difficult to translate in most languages, followed by question 3 "Faintness" and question 5 "Heart racing".
Some scales needed a final adaptation in terms of tense (French, Croatian) and others in terms of gender (Greek, Italian and Hispanic).
The difficulty in transferring these semantic concepts from each of these questions shows the importance of the process of cultural adaptation in a language-to-language transfer [42].

This first step of transferring language to language / culture-to-culture, was not enough. The external and internal validity of each version had to be tested to ensure that the level of effectiveness and reliability psychometrics characteristics were comparable to the original version. Quantitative studies in daily practice were needed.

F-HSCL-25: its validation study

The validation study was anchored in family surgeries, with outpatients.
Its design has made it possible to extract end-to-end secure data. The relationship between the psychiatrist and the general practitioners was maintained throughout the study.
Women accounted for 60% of the sample. The average age was 59 years old. These characteristics of the sample were comparable to other studies in primary care settings (51 years old). The characteristics of the sample are close to European standards based on population. These examples of characteristics make it possible to generalize the results [43].
F-HSCL-25 correctly assessed major depression. She has demonstrated an ability to recognize a major depressive episode with a PPV greater than 60%. The 91% specificity indicated effectiveness in identifying significant depression in daily practice. It is a useful first-line ergonomic diagnostic tool with a low number of false positive patients. The GP high specificity depression diagnosis combined with the effectiveness of this tool to exclude non-depressive patients with a low margin of error, could become an effective synergy. Therefore, this could quickly improve the diagnosis.

Strengths and limits

This collaborative work followed a well-defined and rigorous methodology. The large-scale research team consisted of GP researchers from several countries and cultures. However, not all European countries were represented. Nevertheless, the members of this study cover a wide linguistic range: Romance, Greek, Germanic and Slavic languages were represented.

Regarding the Literature Review

Selection bias was always possible, but it was limited by the use of a multilingual team, blindly working in pairs of researchers, at all stages of the eligibility and inclusion process as well as by the extent of the search equation.

The information bias was possible but limited by the rigor of the research. A complete collection of all abstracts and full text articles has been compiled. No documents have been omitted. Relevant results, such as VPP and VPN, were not always present and were calculated.
The choice of the database is questionable, but it has directed research towards primary care.

Confusion bias was limited by using a group consensus procedure to establish the final list at each stage of the PRISMA process (identification, selection, eligibility and inclusion).

The research team made successive choices throughout the process to be as specific as possible and to maintain the ability to communicate with other health professionals [44]. Choosing and deliberately maintaining the face-to-face psychiatric examination as referral based on the DSM and choosing Youden's index, was an effective and robust way to compare the effectiveness of the tools [45].

It later became apparent that these intentional choices led to the elimination of some popular tools that had certain methodological limitations that prevented them from being validated according to our research method. The tools extracted by the literature review were not the tools most commonly used in practice. It was intentional; the goal was to select a research tool.

For example, the 4DSQ is validated against a mathematical model based on the population and not with respect to clinical comparison criteria [46][47].

The inclusion or not of the PHQ-9 was heavily discussed. PHQ-9 is an important tool [48]. It is extracted from the PRIME-MD (Primary Care Evaluation of Mental Disorders) or PHQ (Patient Health Questionnaire). [49] It has a high level of validation. Primary care is his field of action. In the Kroenke’s study, a validation procedure was performed in two subpopulations (a general population consulting in primary care centres and a population consulting gynaecology-obstetrics unit), against a psychiatric interview, using the SCID (Structured Clinical Interview for

DSM) as guideline and the PRIME-MD questions modified to match on the DSM in form and stratification [48]. The interview was held by phone. Nevertheless the scientific committee of the survey specified in the inclusion criteria that included questionnaire have to be face-to-face psychiatric interview to avoid any bias. Ina ddition the reference tool was the PRIME-MD, even though the DSM criteria were integrated it is not a real DSM based questionnaire like the PSE 9. To avoid confusion, the consensus was not to include this tool even if it is a well-known tool.

Regarding the RAND / UCLA

The quality of the expert panel was important for the overall quality level.
The panel complied with the requirements of variability in culture, language and practice. It was of sufficient size, 7 to 15 experts are recommended for a RAND / UCLA, between 10 and 11 experts participated in the study [8].

The deadlines for the Delphi rounds were short. Each judgment was made blindin order [50] to reduce information bias, each expert received a copy of all bibliographic sources of data provided.

The reliability data was based on Cronbach's alpha values. Cohen's kappa was not found for these tools. These values were extracted using a rapid literature review. A systematic review would have had more power, but it would have increased the duration of the study to the detriment of the feasibility [51].

The tools found in the literature have not been anonymized. The judgment of each expert could possibly take into account his own knowledge. Nevertheless, the opportunity to discuss at the intermediate nominal group helped to control this possible confusion bias.

Regarding the translation process

To reduce selection bias and ensure sample quality: the study was managed to ensure the participation of GPs in language translation and this was a key point to

pursue in this study. As many translators have described, when it comes to scientific translation work, it is essential that a "specialist" in the field (in this case the field of general practice that is common to us) looks at the translation [52][53]. It is the main arbiter of the quality of the final translation [54]. Thus, particular attention was paid to the choice of GPs researchers and certified bilingual translators with sufficient knowledge of medical terminology to reduce selection bias.

The step-by-step analysis of cultural control has helped to avoid confusion bias and linguistic problems related to the transfer. The intervention of a consensus group allowed for several progressive evaluations of each element, reinforcing the accuracy of the validated translations and jointly designing the final result. This work is the result of a multicultural collaborative research between European countries.

Regarding the validation study

Several types of data quality procedures were applied: a data manager designated by the Brest CHRU DRCI (Direction de la Recherche Clinique) was appointed to control the quality of the protocol and the progress of the study. This data manager also strictly controlled the data processing.
The expertise of each stakeholder of the team has made it possible to secure data collection at each stage.
We used stratified randomization to ensure both a satisfactory statistical power and an affordable logistics. This randomization was performed independently, via computer software, excluding any human intervention in the selection.

After the freezing of data, the data were entrusted to statisticians from the Brest Clinical Investigation Center (CIC). GP trainees specially trained to use the PSE-9 by a psychiatrist from Brest CHRU, performed the semi-structured validated psychiatric evaluation. The medically qualified trainees undertook recruitment and interviews by PSE-9. The psychiatrist then confirmed the validity of clinical diagnoses.

A multidisciplinary research network supported the study. A GP research network has enabled the recruitment of outpatients. The population was multi-centered,

drawn from rural, semi-rural and urban areas. Recruitment among outpatients in waiting rooms, attending consultations, ensured its representativeness. This recruitment was carried out over a short period, and in the same way, in three-study centre.

In 1993, Nettlebladt used the same method to evaluate the accuracy of HSCL-25 as a primary care diagnostic questionnaire in Sweden [35]. They conducted a study in six Swedish primary health care centre in two districts, one rural and the other semi-urban, to validate HSCL-25 against PSE-9 and establish a threshold. Although our study resulted in a lower sensitivity (59% vs. 76%), we had a higher specificity (91% vs. 73%). The prevalence of psychiatric morbidity was lower (21% versus 33%). Previous studies have shown similar results in terms of sensitivity and specificity [55]. Our sample of patients was a little larger (1146 versus 727), which could perhaps explain the observed differences.
According to Nettlebladt, the choice of a threshold at 1.55, tended to increase sensitivity (89%), but also gave higher false positives (43%), making it less accurate. Screening capacity was improved to the detriment of diagnostic capacity.

In the English version, the threshold of 1.55 is a warning criterion, the threshold of 1.75 is established as defining the patient requiring treatment for anxio-depressive syndrome. For the validation study, the threshold of 1.75 was considered. It gives a medium sensitivity but a high specificity. The level of 1.75 is effective for diagnosing patients requiring specific treatment of depression.

The use of a different randomization for each group: a ratio of 1/2 for the HSCL + group, a ratio of 1/16 for the HSCL group, could also explain our differences in terms of prevalence, sensitivity and specificity with the Nettelbladt's study . However, the difference in the randomization reports allowed us to balance the number of PSE-9 patients in our groups as closely as possible.
A more recent Swedish study examined the concordance between HSCL-25 and DSM-IV criteria for anxiety disorders and depression, using a semi-structured psychiatric interview, the SCAN (Schedules for Clinical Assessment in Neuropsychiatry) as a standard criterion [55]. It differs from the studies mentioned

above because of its large sample (8613 recruited patients) in the general population, and not focused on outpatients consulting in medical surgeries. Excellent agreement was found between HSCL-25 and DSM. In terms of sensitivity and specificity, their results are close to ours and confirm that the HSCL-25, with a threshold at 1.75, is definitely a diagnostic tool rather than screening.

Teaching implications

In medical education in general practice, trainees are often faced with the question of how to make a diagnosis of depression [56]. Many feel that they have difficulty detecting depression and therefore do not know who to treat and "when to hand over". Although this study is primarily research-based, the use of categorical tools can be of great help to these young physicians. They will be able to evaluate their practice with these tools and establish strong professional methods for the diagnosis of depression. As always, a tool is just a point of entry for diagnosis and for talking to the patient about labeling his symptoms. Students must learn to introduce a tool into the consultation; how to stimulate patients to use a tool; how to interpret, discuss and record the results, and then how to follow their patients with this help.

Students are looking for diagnostic tools to help them with their clinical approach. But most of the existing tools are in the Anglo-American language. Translation remains the most crucial step in the adoption of a well-developed instrument by another nation using a different language. Translation errors may distort the original intent of this instrument and compromise the validity and reliability of the resulting instrument [57]. Semantic problems affect comparability in international studies, since the same word is interpreted differently across countries and cultures. [58][59] Moreover, some terms and concepts may not exist in other languages, or may have additional connotations that retro translations do not always reveal. The challenges arise, not only because of the content of literal translation word for word, but also because of the linguistic form of language, such as tone and syntax [60].

Research Implications

The studies collected by the systematic review of the literature involved adult patients. Only 1 study out of 4 has a wide age range, between 20 and 80 years old. The other three studies included a population over 60 or 65 years old. 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, since in Europe, the working population aged 50 to 64 represents 1/3 of the active population aged between 20 to 64 years old [61]. In future studies, when there is discrimination in a population, there are other factors to consider, outside of age, for example, coping ability, is not only related to age [62]. Future research should ensure that these tools have a place in the treatment of adult patients, regardless of their age.

In the perspective of collaborative studies on depression in primary care, GPs show a good level of specificity in the diagnosis of depression according to DSM criteria but the choice of tools to share between GPs and psychiatrists will be a challenge [15]. The choice of a common tool could be based on statistical criteria but the choice could also be influenced by clinical criteria of utility [16]. Further research, using a standardized methodology, will be needed to select the best possible tool, in terms of reliability, effectiveness and ergonomics, to undertake collaborative studies at European level between general practitioners and psychiatrists, in the different fields of mental affections [63].

Using only effectiveness data is misleading when comparing tools. Therefore, tool selection should be based on several criteria, both quantitative and qualitative. The use of comparative multi-criteria models, including effectiveness and reliability, such as the COSMIN statement or qualitative selection procedure incorporating quantitative data such as RAND / UCLA are essential resources upstream of the collaborative research procedures, for the choice appropriate tools [64].
The time spent on the outpatient in general practice is a short time. As a result the tools must have a design that adapts to that time. In future research in general medicine, ergonomics is one of the major criteria of choice.

CONCLUSION

GPs in most European countries have now the possibility to use HSCL-25 in research studies in general practice and to assess the severity of depression in their patients.

The use of such a shared tool can have a great impact on the feasibility of doing research on depression in general practice in the future. We will be able to more easily compare data across European countries, which will enable us to undertake statistical reviews on the epidemiology and symptoms of depression across Europe. The use of the same instrument can support the conceptualization of the phenomenon studied through different studies, and the results would be comparable between populations but also between specialists of general practice and psychiatry.

Nevertheless this very specific tool has a limited PPV and should be used carefully in practice. Its combination with a screening tool could be of interest and the research team will follow on this new path of research to see if a combined test is feasible and efficient.

Bibliography

1. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, Clarke M, Devereaux PJ, Kleijnen J, Moher D: **The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration**. *PLoS Med* 2009, **6**:28.

2. 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.

3. 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.

4. 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.

5. Fröjdh K, Håkansson A, Karlsson I, Frojdh 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.

6. Youden WJ: **Index for rating diagnostic tests.** *Cancer* 1950, **3**:32–35.

7. Fluss R, Faraggi D, Reiser B: **Estimation of the Youden Index and its associated cutoff point.** *Biometrical J Biometrische Zeitschrift* 2005, **47**:458–472.

8. 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.

9. Tavakol M, Dennick R: **Making sense of Cronbach's alpha.** *Int J Med Educ* 2011, **2**:53–55.

10. 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.

11. Lipman RS, Covi ' L, Shapiro AK: **THE HOPKINS SYMPTOM CHECKLIST (HSCL) Factors Derived from the HSCL-90**. *J Affect Disord* 1979, **1**:9–24.

12. Brislin RW: **Back-Translation for Cross-Cultural Research**. *J Cross Cult Psychol* 1970:2.

13. Bullinger M, Anderson R, Cella D, Aaronson N: **Developing and evaluating cross-cultural instruments from minimum requirements to optimal models**. *Qual Life Res* 1993, **2**:451–459.

14. Herdman M, Fox-Rushby J, Badia X: **"Equivalence" and the translation and adaptation of health-related quality of life questionnaires.** *Qual Life Res* 1997, **6**:237–47.

15. Mitchell AJ, Vaze A, Rao S: **Clinical diagnosis of depression in primary care: a meta-analysis.** *Lance* 2009, **374**:609–619.

16. Kendell R, Jablensky A: **Distinguishing between the validity and utility of psychiatric diagnoses.** *Am J Psychiatry* 2003, **160**:4–12.

17. Cosgrove L, Krimsky S, Vijayaraghavan M, Schneider L: **Financial ties between DSM-IV panel members and the pharmaceutical industry.** *Psychother Psychosom* 2006, **75**:154–160.

18. Thomas-MacLean R, Stoppard J, Miedema BB, Tatemichi S: **Diagnosing depression: there is no blood test.** *Can Fam Physician* 2005, **51**:1102–1103.

19. 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.

20. 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.

21. 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.

22. 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.
23. Weeks SK, McGann PE, Michaels TK, Penninx BWJH: **Comparing various short-form Geriatric Depression Scales leads to the GDS-5/15.** *J Nurs Scholarsh an Off Publ Sigma Theta Tau Int Honor Soc Nurs Sigma Theta Tau* 2003, **35**:133–137.
24. Radloff LS: **The CES-D Scale: A Self-Report Depression Scale for Research in the General Population**. *Appl Psychol Meas* 1977, **1**:385–401.
25. 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.
26. Zigmond AS, Snaith RP: **The hospital anxiety and depression scale (HADS).** *Acta Psychiatr Scand* 2007, **67**:0–4.
27. 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.
28. 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.
29. Andrews B, Hejdenberg J, Wilding J: **Student anxiety and depression: comparison of questionnaire and interview assessments.** *J Affect Disord* 2006, **95**:29–34.
30. 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.
31. Ekblad S, Roth G: **Diagnosing posttraumatic stress disorder in multicultural patients in a Stockholm psychiatric clinic.** *J Nerv Ment Dis* 1997, **185**:102–107.
32. 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.
33. Oruc L, Kapetanovic A, Pojskic N, Miley K, Forstbauer S, Mollica RF, Henderson DC: **Screening for PTSD and depression in Bosnia and Herzegovina: validating**

the Harvard Trauma Questionnaire and the Hopkins Symptom Checklist. *International Journal of Culture and Mental Health* 2008:105–116.

34. 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.

35. 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.

36. 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.

37. Munk-Jørgensen P, Fink P, Brevik JI, Dalgard OS, Engberg M, Hansson L, Holm M, 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.

38. Hignett S, Carayon P, Buckle P, Catchpole K: **State of science: human factors and ergonomics in healthcare.** *Ergonomics* 2013, **56**:1491–503.

39. Sandanger I, Moum T, Ingebrigtsen G, Dalgard OS, S??rensen 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**:345–354.

40. Strand BH, Dalgard ODDS, 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).** *Psychiatry Interpers Biol Process* 2003, **57**:113–118.

41. 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.

42. Brislin RW: **Comparative research methodology: Cross-cultural studies**. *Int J Psychol* 1976, **11**:215–229.

43. King M, Nazareth I, Levy G, Walker C, Morris R, Weich S, Bellon-Saameno JA, Moreno B, Svab I, Rotar D, Rifel J, Maaroos H-I, Aluoja A, Kalda R, Neeleman J, Geerlings MI, Xavier M, de Almeida MC, Correa B, Torres-Gonzalez F: **Prevalence of common mental disorders in general practice attendees across Europe**. *Br J Psychiatry* 2008, **192**:362–367.

44. Zhang J, Patel VL, Johnson TR, Shortliffe EH: **A cognitive taxonomy of medical errors**. *J Biomed Inform* 2004, **37**:193–204.

45. Perkins NJ, Schisterman EF: **The Youden Index and the optimal cut-point corrected for measurement error.** *Biometrical J Biometrische Zeitschrift* 2005, **47**:428–441.

46. Terluin B, Van Marwijk HWJ, Adèr HJ, de Vet HCW, Penninx BWJH, Hermens MLM, Van Boeijen CA, Van Balkom AJLM, van der Klink JJL, 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.

47. Chambe J, Le Reste J-Y, Maisonneuve H, Sanselme A-E, 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**.

48. Kroenke K, Spitzer RL, Williams JB: **The PHQ-9: validity of a brief depression severity measure.** *J Gen Intern Med* 2001, **16**:606–613.

49. 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.

50. Elmer F, Seifert I, Kreibich H, Thieken AH: **Delphi method**. *Innovation* 2010, **30**:93–113.

51. Ganann R, Ciliska D, Thomas H: **Expediting systematic reviews: methods and implications of rapid reviews**. *Implement Sci* 2010, **5**:56.

52. Beaton DE, Bombardier C, Guillemin F, Ferraz MB: **Guidelines for the process of cross-cultural adaptation of self-report measures.** *Spine (Phila Pa 1976)* 2000, **25**:3186–91.

53. Skulmoski GJ, Hartman FT, Krahn J: **The Delphi Method for Graduate Research**. *J Inf Technol Educ* 2007, **6**:1.

54. Balliu C: **L'enseignement de la traduction médicale: pour une nouvelle pragmatique**. *Meta J des traducteurs* 1994, **39**:15–25.
55. Lundin A, Hallgren M, Forsell Y: **The validity of the symptom checklist depression and anxiety subscales: A general population study in Sweden**. *J Affect Disord* 2015, **183**:247–252.
56. GASK L, GOLDBERG D, LESSER AL, MILLAR T: **Improving the psychiatric skills of the general practice trainee: an evaluation of a group training course**. *Med Educ* 1988, **22**:132–138.
57. Yu DSF, Lee DTF, Woo J: **Issues and Challenges of Instrument Translation**. *West J Nurs Res* 2004, **26**:307–320.
58. Schnohr CW, Gobina I, Santos T, Mazur J, Alikasifuglu M, Välimaa R, Corell M, Hagquist C, Dalmasso P, Movseyan Y, Cavallo F, van Dorsselaer S, Torsheim T: **Semantics bias in cross-national comparative analyses: is it good or bad to have "fair" health?** *Health Qual Life Outcomes* 2016, **14**:70.
59. Daugherty JC, Puente AE, Fasfous AF, Hidalgo-Ruzzante N, Pérez-Garcia M: **Diagnostic mistakes of culturally diverse individuals when using North American neuropsychological tests**. *Appl Neuropsychol* 2017, **24**:16–22.
60. Hanrahan D, Sexton P, Hui K, Teitcher J, Sugarman J, London AJ, Barnes M, Purpura J, Klitzman R: **Linguistic and cultural challenges in communication and translation in ussponsored HIV Prevention research in emerging economies**. *PLoS One* 2015, **10**:e0133394.
61. Levasseur S: **Vieillissement de la population active**. *Rev l'OFCE* 2015, **6**:339–370.
62. 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.
63. 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.

64. 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.

SUMMARY

One consensual depression diagnosis tool to serve many countries: a challenge!

Abstract

Introduction: Depression is a common reason for consultation in general practice. Its variability makes its diagnosis difficult. An effective, reliable and ergonomic diagnostic tool would be an aid to research in general practice. The aim of this study was to find a consensual tool between general practitioners (GPs) and psychiatrists in several European countries.

Methods: A systematic literature review was undertaken to find validated tools in general practice against the psychiatrist. A consensus according to a RAM (RAND/UCLA Appropriateness Method) has selected one. It has been translated according to a procedure guaranteeing the stability and the ecology in general practice. A validation protocol has been produced to ensure the retention of psychometric qualities. The French external validation study was carried out.

Results: Seven tools were extracted: CESD-R, GDS 5-15-30 items, PSC-51, HADS, HSCL-25. Psychometric effectiveness data (Se, Sp, VPP, VPN) were collected. The HSCL-25 has been selected for its high combined qualities of effectiveness, reliability and ergonomics. It has been translated into 9 languages relating to 3 linguistic groups: Greek, Romance and Slavic languages. The French Validation Study has proven that the French form of HSCL-25 (F-HSCL-25) has high diagnostic performance (Se 59.4%, Sp 91.4%, VPP 69.8%, and VPN 86.9%) adapted to research in general practice.

Implication: HSCL-25 is a valid and effective tool for diagnosing depression in primary care. They could increase the diagnostic performance of GPs and foster collaborative research.

Summary

Background

Major depression affects 4.4% of the world's population. Prevalence estimates vary in Europe but are around 10% for people attending general practice, and the prevalence is twice as high for women. An increase of more than 18% was observed between 2005 and 2015.
Depression is a disease comprising contextual distress, anxiety and somatoform disorders. This disorder is not easy to diagnose, however, due to the wide variety of ways in which it may be presented. Patients themselves experience difficulties to express their suffering and display their own form of illness expression. Based in this inter-individual variability, the difficulties to diagnose and assess the severity of depression may overestimate or underestimate the distress level of their patients by clinicians. Those difficulties may lead to inappropriate care and cause public health issues.

It became clear that a single validated tool was needed by European General Practitioner (GP) researchers in order to allow multi-centred collaborative research, in daily practice, throughout Europe, interesting clinical, epidemiology and statistician comparison, and interesting both psychiatrists and Gps.
An international GP team, under the auspices of EGPRN (European General Practice Research Network) promoted a survey.

The aim was to select a consensually diagnostic tool for depression, validated against face-to-face psychiatric examination, according to the DSM-criteria. European GPs, could use this tool for research purposes in daily practice, according the best effectiveness, reliability and ergonomics combined.

Each step of the thesis used a specific method.
The first phase was an informal group consensus on the research protocol.
The Second was a systematic literature review (SRL) to extract validated tools on depression diagnosis.
The third was a consensus procedure, a RAND/UCLA, to select a single tool according efficiency criterion, reliability and ergonomics.
The fourth was a translation of the selected tool, in the language of each participating country. A forward/backward translation followed by a cultural check was used to maintain linguistic and semantic stability.
The fifth was the validation of the tool, at least in a country, in daily practice, with adults' outpatients.

Method

An international GP team conducted a systematic literature review, according PRISMA guideline, using the following databases: Pubmed, Cochrane and Embase, from 2000/01/01 to 2015/10/01. The SLR extracted tools were validated against the DSM. The Youden index was used as an effectiveness comparison criterion.

A RAM allowed a multi-criterion comparison based on effectiveness, reliability and ergonomics. The RAM combined the qualities of the Delphi process and the nominal group. GP researchers from different European countries were selected according following features: Good knowledge of English language, Academics, born in the participating countries, practising GP and EGPRN member. Reliability data (Cronbach's alpha) and ergonomics features were extracted from literature using a narrative review. Criterion to compare were: Youden index and Cronbach's alpha. Ergonomics were tested face-to-face.

The selected tool was translated, using a forward-backward translation. Two translators (an academic and an FP researcher) were recruited for the forward translation (FT). A panel of English-speaking FPs was set up in each country. A minimum size of 15 experts was requested. The panel of experts finalized the FT

using a Delphi procedure. Then, a different translator, who did not know the original version, undertook an English backward translation. Linguists compared the two English versions. Differences were listed to analyze the cultural impact of translation according to a multicultural consensus group.

To validate the test characteristics of the translated version, a validation study inserted in primary care daily practice, concerned outpatients was finalised. The tool should be compare to the Present State Examination-9 French version (F-PSE-9), using a psychiatric interview. A peculiar sample randomisation design should be used to allow feasibility in daily practice.

Results

Researchers identified 770 abstracts in three databases. After the removal of duplicates (n= 224) 546 abstracts were analysed. Fifty of the validity studies were eligible and finally 4 studies were included. In these 4 studies, the following tools were found: GDS-5, GDS-15, GDS-30, CESD-R, HADS, PSC-51 and HSCL-25. Figures on sensitivity, specificity, positive predictive value, and negative predictive value were collected. The Youden index was calculated.

Among these seven 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.

Each panel was composed with a maximum variation of researchers, teachers and practitioners. One to two Delphi rounds by country were sufficient. To ensure the original meaning, all versions were subjected to a cultural check. Translations into 9 languages were completed into Greek, Polish, Bulgarian, Croatian, Catalan, Galician, Spanish, Italian and French have been finalized without altering the meaning.

Outpatients from French General Practice settings (rural, semi-rural and urban) were recruited.
Two cut-offs characterize the HSCL-25: 1.55 pointed a risky patient, 1.75 pointed an anxio-depressive patient. Related to the French version (F-HSCL-25), two groups were formed: F-HSCL-25 ≥1.75 and F-HSCL-25 <1.75.
The randomization had taken place in both groups using a different ratio to pass F-PSE-9, given to 1 in 2 patients in the F-HSCL-25 ≥1.75 group, and to 1 in 16 in the (much larger) F-HSCL-25 <1.75 group. We assessed diagnostic performance comparing test results obtained in both groups with their F-PSE-9 results.

In total, 1126 patients filled in the F-HCL-25, of whom 886 had a negative result and 240 tested positive. The overall prevalence of depression, using the F-HSCL-25 (yes/no), was 21% in these physicians' surgeries. The diagnostic performance of the F-HSCL-25 versus the external criterion (F-PSE-9) was as follows: Positive Predictive Value (PPV) 69.8%, Negative Predictive Value (NPV) 87%; Sensitivity 59.1%, and Specificity 91.4%.

Discussion

Using efficiency data alone to compare tools could be misleading. Additional reliability, reproducibility and ergonomic data will be essential for making comparisons, using a method allowing a multi-criterion comparison possible. A multicultural consensus on one diagnostic tool for depression was obtained for the Hopkins Symptom Checklist in 25 items (HSCL-25).

The Hopkins Symptom Checklist-25 (HSCL-25) is a robust self-administrated questionnaire depression diagnosis validated tool. Effective, reliable and ergonomic, it can be used in daily practice. It helps to assess depression in primary care. A patient is considered 'depressive' if a score > 1.75 is obtained.

To allow GPs to use it, it had to be translated into various European languages. The entire translation process had to ensure homogeneity. It was translated into French.

The French validation study demonstrated that F-HSCL-25 was an appropriate diagnostic aid for depression in primary care, due to its high specificity and high NPV.
This pilot study will be extended throughout Europe, but preliminary evidence suggests that the HSCL-25 was a good transnational tool in primary care.

This tool could provide the opportunity to select homogeneous populations for European collaborative research in daily practice.

Key Words: Depression, Systematic review of literature, Diagnostic tool, Primary Care Research, RAND/ UCLA Appropriateness Method, Multicultural Consensus, Delphi Procedure, Depression Diagnosis Tool**,** Hopkins symptom Checklist 25 items, Validation studies, Psychometrics

RESUME FRANÇAIS

A la recherche d'un outil diagnostique de la dépression en médecine générale, simple, stable et efficace, pour favoriser les recherches collaboratives en France et en Europe.

Abstract

Introduction: La dépression est un motif fréquent de consultation en médecine générale. Sa variabilité rend son diagnostique difficile. Un outil diagnostique efficace, stable et ergonomique serait une aide en recherche en médecine générale. L'objectif de cette étude était de trouver un outil consensuel entre médecins généralistes (MG) et psychiatres sur plusieurs pays européens.

Méthodes: Une revue systématique de littérature a été entreprise pour trouver les outils validés en médecine générale contre le psychiatre. Un consensus selon une RAM en a sélectionné un. Il été traduit selon une procédure garantissant la stabilité du transfert et l'écologie en médecine générale. Un protocole de validation a été produit pour s'assurer de la conservation des qualités psychométriques. L'étude de validation externe française a été réalisée.

Résultats: Sept outils ont été extraits : CESD-R, GDS 5-15-30 questions, PSC-51, HADS, HSCL-25. Les données psychométriques d'efficacité (Se, Sp, VPP, VPN) ont été colligés. La HSCL-25 a été sélectionnée pour ses hautes qualités combinées d'efficacité, de stabilité et d'ergonomie. Elle a été traduite en 9 langues relatives à 3 groupe linguistiques : le grecque, les langues romanes et slaves. L'étude de validation française a prouvé que la forme française de la HSCL-25 (F-HSCL-25) a de hautes performances diagnostiques (Se 59,4%, Sp 91,4%, VPP 69,8%, VPN 86,9%) adaptées à la recherche en médecine générale.

Implication: la HSCL-25 est un outil valide et efficace pour le diagnostic de la dépression en soins primaires. Ils pourraient augmenter les performances

diagnostiques des MG et favoriser des recherches collaboratives.

INTRODUCTION

Depression et soins primaires

La dépression concerne 4.4 % de la population mondiale [1][2][3]. En Europe, sa prévalence est estimée à 10% [4]. En France, sa prévalence varie entre 5 et 12 % avec une nette prédominance pour les femmes [5].

Ce syndrome comprend des signes de détresse, d'anxiété et de multiples expressions somatiques [6][7]. Sa grande variabilité d'expression entre les patients et l'appétence variable à son diagnostique par les médecins rendent son diagnostique difficile [8], à l'origine de sur et sous diagnostiques et d'estimations erronées des niveaux de détresse [9][10]. La prise en charge médicale est alors inadaptée avec des conséquences en santé publique [11].

C'est la seconde maladie chronique prise en charge en médecine générale. Le médecin généraliste (MG) souvent en première ligne diagnostique et thérapeutique, est seul avec un temps contraint [9][12][13][14]. Les MG ne sont pas à l'aise avec les critères de définition de la dépression [15][16][17]. Pourtant, ils sont à la meilleure place pour organiser le suivi des soins au long court [18][19]. Les MG ont une sensibilité diagnostique basse (37,9%) mais une spécificité diagnostique élevée (89,7%) comme les autres spécialistes [20][21]. Pour abaisser ces barrières, [22] un outil diagnostique ergonomique à forte valeurs prédictives pourrait les aider et améliorer leur efficacité.

Apport d'un réseau de recherche européen

Des recherches collaboratives européennes seraient un atout pour améliorer la prise en charge des patients. Mais les différences de modèle de santé et d'objectifs sont autant d'obstacles. L'EGPRN (European General Practice Research Network) a développé un réseau de recherche pour les lever.
Concernant la dépression, l'EGPRN s'est donné pour objectif de trouver un outil diagnostique favorisant des recherches collaboratives en médecine générale en

Europe, malgré les différences de langue et de culture et de données de santé publique [4][23]. Cet outil devait avoir une taxonomie commune entre les MG et les psychiatres [24].

Constitution de l'équipe Européenne

En Octobre 2010, la problématique de recherche a été proposée à la communauté. Le recrutement de 10 pays s'est déroulé sur 2 ans pour se terminer en mai 2012, avec pour objectif d'identifier les outils et d'en sélectionner un consensuellement. Il devait être acceptable pour casser les barrières des MG à utiliser les outils académiques et permettre des recherches collaboratives. Il pourrait être éventuellement proposé aux MG pour améliorer leur performance diagnostique.

Les chercheurs devaient être académique (chercheurs universitaires ou membres d'un réseau de recherche), membres de l'EGPRN et indépendants de l'industrie pharmaceutique. Ils ont constitué un groupe multi culturel le plus large possible.
Ils devaient avoir des compétences linguistiques pour permettre : une revue de la littérature et un processus de traduction.
Ils devaient avoir des compétences académiques communautaires : basées sur les ressources universitaires et les réseaux de recherche et de pratique dans chaque pays.

Le management d'un groupe multi culturel pouvait réserver des difficultés de compréhension : liées à l'usage exclusif de l'anglais et des différentes cultures. L'autre difficulté a été de maintenir la cohésion du groupe. A chaque étape, l'agrément a été demandé aux membres. Ils ont participés activement à l'élaboration de chaque protocole. Une aide active et présentielle a été apportée ainsi qu'une aide à la publication. Les rangs de publication ont été établis avec les membres. Selon ces conditions, l'étude s'est déroulée sur 6 ans avec tous les membres.

Chausses trappes et objectifs, taxonomie des outils diagnostiques et question de recherche

Pour diagnostiquer la dépression, outre l'approche clinique, il existait de nombreux outils et algorithmes utilisables [20].

Le DSM (Diagnostic and Statistical Manual) est un outil catégoriel américain. Il est largement utilisé dans le monde [25][26] comme ses algorithmes et interviews structurés fabriqués à partir de lui et validés contre lui : le CIDI (Composite International Composite Interview), le SCID (Structured Clinical Interview for DSM disorders), le MINI (Mini International Neuropsychiatric Interview) [27][28][29]. Le DSM et ses interviews sont difficiles à utiliser en pratique, d'autres outils au design plus adapté comme le GDS (Geriatric Depression Scale) [30] ou la HSCL-25 (Hopkins Symptom Checklist en 25 questions) par exemples [31], ont été créés et validés directement contre le DSM ou contre ses dérivées.
Le HDRS ou HAM-D (Hamilton Depression rating Scale) est un outil longitudinal anglais. Largement utilisé dans le monde [32][33][34][35], li a donné naissance au BDI (Beck Depression Inventory) [36], et au MADRS (Montgomery-Asberg Depression Scale) [37][38].
Le PRIME-MD (Primary Care Evaluation of Mental Disorders) est une interview clinique structurée indépendante [39]. Cet outil anglais a évolué en une forme simple, le PHQ-9 (Patient Health Questionnaire) [40][41].
Cette liste n'est pas exhaustive.

Objectif et question de recherche

Un protocole de recherche a été écrit avec l'objectif de sélectionner l'outil le plus fiable et ergonomique. Il est développé dans la partie méthode qui suit. Il a été soumis au comité d'éthique de l'université de Bretagne Occidentale et à l'Agrément du CPP (Comité de Protection des Personnes) de l'Hôpital universitaire de Brest (ID RCB: n°2014-A01790-47; Référence CPP: CPP Ouest VI 872; N° enregistrement Clinical Trial.gov: NCT02414711).

La recherche devait porter sur une population adulte (plus de 18 ans). Conformément aux demandes du comité d'éthique, et parce qu'il y a un débat scientifique pour découvrir s'il y a une différence significative entre la forme

périnatale et d'autres formes de dépression, la dépression périnatale n'a pas été retenue [42][43][44][45].

La question de recherche a été la suivante : Quel outil pourrait être accepté et utilisé par les Médecins Généralistes, pour diagnostiquer la depression au sein des cabinets médicaux, à la fois efficace, stable, ergonomique et applicable dans les différents pays européens participants à l'étude ?

METHODE

DESIGN GENERAL

Premièrement, un consensus de groupe informel a permis de choisir la référence et bâtir le protocole. La seconde étape a été une revue systématique de la littérature, pour extraire les outils diagnostiques validés. La troisième a été une procédure de consensus, une RAM (Research and Development Appropriatness method) ou RAND/UCLA (Research and Development / university of California Los Angeles). La quatrième, une procédure de traduction aller/ retour suivi d'un contrôle culturel, pour maintenir la stabilité linguistique et sémantique. La cinquième, une étude de validation multi centrique, en pratique courante de médecine générale, chez des patients adultes.

Revue systematique de litterature (RSL)

Dans chaque pays, chaque chercheur a contacté un autre chercheur pour travailler en binôme, pour le processus d'exclusion des résumés et d'inclusion des articles.

Le premier objectif a été de choisir la référence pour extraire les outils qui lui étaient relatifs. Puis les chercheurs ont construit l'équation de recherche ; ils ont sélectionnées les critères d'inclusion et d'exclusion ; ils ont fait le choix des bases de données indexées.

Le choix de la référence était crucial, il allait donner une tendance forte à l'ensemble des résultats de l'étude. Pour être sur de ne pas faire d'erreur, les chercheurs ont au départ, à l'aide d'une revue de la littérature, recensé tous les outils validés utilisables

en médecine générale. Ceci a donné une taxonomie des outils. Cette revue a révélé que le DSM était largement utilisé comme référence internationale. C'est une référence diagnostique et catégorielle en accord avec les objectifs de l'étude. Le DSM est devenu le critère majeur d'éligibilité pour la RSL.

L'autre référence majeure a été le psychiatre. L'équipe a considéré que l'examen direct, face à face avec un psychiatre était la référence la plus appropriée.

Ces deux références se sont combinées pour donner le gold standard de l'étude : l'examen psychiatrique direct par le psychiatre, en face à face, utilisant les critères de la dépression majeure du DSM.

Les autres critères d'éligibilité devaient s'acquitter des prérequis de l'étude.
Le champ de la recherche était la médecine générale, en lien avec la psychiatrie.
Les adultes étaient la cible. L'enfance, l'adolescence, la grossesse et le post-partum étaient exclus.
Seuls les études de validation avec des données psychométriques d'efficacité seraient conservées. Les études de corrélation seraient exclues.
Les articles devraient être écrits dans au moins une des langues des experts ou en anglais.

La RSL a été conduite selon le guide PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) et chaque étape a été respectée [46][47][48].

Identification

Les bases suivantes ont été explorées : PubMed, Embase and Cochrane.

L'équation de recherche dans les bases suivante a été utilisée (format 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]) et adaptée à chaque base de donnée.

Un binôme de chercheurs internationaux français a entrepris la recherche documentaire en travaillant à l'aveugle et en regroupant les documents à la fin du processus d'identification. Ils ont compilé la liste des résumés qui répondaient aux critères. Cette liste a été répartie entre les équipes nationales, une fois les doublons supprimés.
Puis, chaque équipe nationale a entrepris les procédures d'exclusion des résumés avec un binôme national travaillant à l'aveugle. De plus, le binôme français travaillant à l'aveugle, a complété le même processus sur tous les résumés. Les deux équipes de binôme ont ensuite comparé leurs résultats pour parvenir à un consensus basé sur les critères qualitatifs. Tous les résumés admissibles ont été évalués pour identification.

Screening (criblage)

Critères d'inclusion:
Limité aux 15 dernières années.
Adultes et / ou patients âgés.
Anglais, Grec, Espagnol, Italien, Français, Allemand, Polonais.
Critères d'exclusion:
Pas de format IMRaD (Introduction, Méthodes, Résultats et Discussion) [49].
La dépression n'était pas le sujet principal.
Aucun outil de diagnostic identifié.
L'étude portait sur les enfants ou la grossesse ou la dépression du post-partum.
L'étude n'était pas dans un contexte de soins primaires.
Les outils ont été identifiés sans données de validité.

Eligibilité

La même méthode consensuelle à 2 binômes a été utilisée.

Les articles ont été exclus selon les critères suivants:
Le diagnostic de dépression n'était pas le sujet principal de l'étude.
Les données d'efficacité (sensibilité, spécificité, valeur prédictive positive, valeur prédictive négative) étaient absentes ou importées d'une autre étude.
La fiabilité était la seule donnée de validité mentionnée dans l'article.
La langue utilisée dans l'étude n'était pas l'anglais, le grec, l'espagnol, l'italien, le français, l'allemand ou le polonais.
Les chercheurs n'étaient pas des MG.
L'outil n'a été validé que par rapport à un autre outil de diagnostic sans examen psychiatrique en face à face, avec l'usage des critères de la dépression majeure du DSM.
L'outil était seulement un outil de dépistage.

À cette étape, les articles restants inclus ont été analysés par une équipe de deux chercheurs. Les outils ont été identifiés, Les propriétés psychométriques ont été collectées [50].

Procedure de consensus: la RAND / UCLA ou RAM

Critères à comparer

Les propriétés psychométriques (sensibilité, spécificité, valeurs prédictives positives et négatives) ne variaient pas suffisamment pour permettre une comparaison statistique et les populations étudiées étaient différentes.
Une revue narrative a été entreprise pour extraire les données de fiabilité (alpha de Cronbach, kappa de Cohen).
L'ergonomie était importante, mais la comparaison de cet aspect des outils était complexe en raison de sa grande variabilité.
Un consensus, prenant en compte des critères quantitatifs et qualitatifs, basé sur un panel d'experts européens, était la seule alternative pour assurer la comparaison [51].

Procédure de consensus

La méthode de pertinence RAM a été sélectionnée. Elle est approuvée par des instituts majeurs, tels que le NICE (Institut National pour la Santé et l'Excellence Clinique) au Royaume-Uni ou la HAS (Haute Autorité de Santé) en France. [52][53]

La RAM provoque une évaluation répétée pour classer la pertinence, l'objectivité et l'homogénéité d'une procédure (géopolitique, industrielle, médico-chirurgicale, etc.). Elle permet un choix consensuel dans la comparaison de processus complexes, selon un "processus Delphi modifié en deux tours distincts" fait d'une seule ronde avec un groupe nominal intercalé. L'effet leader est supprimé, les points de vue de chacun sont garantis [51][54][55]. Elle produit des critères d'adéquation, des indicateurs visibles de qualité, des critères de validité prédictive. [56]

Panel d'experts

Le panel d'experts a été constitué par les membres chercheurs MG de l'équipe. Il répondait aux critères de sélection raisonnée, avec la variation maximale possible de langues et de cultures. [54][57]

Première étape

La procédure Delphi a éliminé les outils les moins efficaces et les moins fiables. Les commentaires ne prenaient en compte que les données psychométriques et non ergonomiques.
Chaque expert devait évaluer l'efficacité et la fiabilité de chaque outil sur une échelle de Likert en 9 points [58] :
Cet outil est-il efficace pour le diagnostic de la dépression en soins primaires?
Cet outil est-il fiable pour le diagnostic de la dépression en soins primaires?
Le consensus était défini, si au moins 70% des experts scoraient à 7 ou plus l'assertion [53][59].

Deuxième étape

La réunion du panel devait confirmer les résultats et permettre un débat sans vote, aboutissant à une présentation des outils sélectionnés.

Les experts ont été invités à discuter des résultats de la première étape. Si plus de 70% des experts étaient d'accord, le premier Delphi était considéré réussi.

Puis le groupe nominal a eu lieu pour évaluer les déclarations suivantes:
"Cet outil est facile à utiliser en médecine générale".
"Cet outil pourrait facilement être introduit lors d'une consultation".
"Cet outil pourrait être compris par les patients".
"J'aime cet outil".
"Les patients pourraient être surpris par cet outil".

Les experts devaient évaluer chaque outil, avant et après les avoir testés face à face en binôme. L'objectif était d'évaluer si le test modifiait leur jugement. Ensuite, l'ergonomie a été débattue en groupe. La réunion s'est terminée avec une nouvelle évaluation finale. L'objectif était d'évaluer l'évolution du jugement de chaque expert.

La réunion entière a été entièrement audio et vidéo enregistrée pour le contrôle de qualité final. Aucun consensus final n'était requis à la fin de la réunion [51].

Troisième étape

L'objectif était le vote final. Chaque expert a reçu la transcription des discussions de manière indépendante. La dernière question était: «Quel est l'outil le plus approprié pour diagnostiquer la dépression chez les patients adultes, en médecine générale, en Europe, en termes d'efficacité, de fiabilité et d'ergonomie ?» Les experts ont été invités à voter sur chaque outil et à commenter leurs réponses.

Les traductions, une procédure aller-retour complétée par un contrôle culturel

L'objectif était la traduction sans perdre la stabilité linguistique et sémantique, et de rester dans le contexte de la médecine générale [60][61]. Une étude normalisée en

trois étapes a été menée parmi les pays participants, incluant: une traduction aller, une traduction retour et une vérification culturelle [62][63][64].

Les investigateurs nationaux (NI) ont mis en place dans leur pays deux équipes de traduction qui devaient travailler en aveugle, pour la traduction aller et retour. Les traducteurs devaient bien connaître la terminologie médicale. Chaque équipe de chaque pays, pour la traduction aller était composée d'un NI et d'un traducteur officiel. Pour les équipes de traduction retour, elles se composaient d'un ou deux MG et d'un traducteur officiel différent [60].

La traduction aller a été réalisé selon une procédure Delphi classique [52][59][65][66][67][68][69][70][71]. Les NI ont recruté un panel d'experts MG dans leur propre pays. Ils ont anonymisé les réponses et attribué un numéro d'identification par expert [53], au moins 15 participants devaient être présents à la fin de la dernière ronde.
Les experts MG devaient être originaire du pays et parler sa langue, être anglophone [61]. Plus de la moitié devait participer à des activités d'enseignement et / ou de recherche. La variation maximale a été évalué sur leur sexe, leur type de pratique, leur nombre d'années de pratique et leur nombre de publications [72].

Les biais de traduction linguistique liés aux aspects culturels de chaque pays étaient possibles; cela nécessitait un contrôle culturel et un processus d'adaptation pour assurer la stabilité sémantique. Une fois la traduction aller terminée, la traduction retour a été réalisée pour: (I) s'assurer que les problèmes de traduction linguistique ont été identifiés, (II) collecter les problèmes de traduction indépendants de la traduction linguistique [63][64][73][74][75].

Un MG chercheur et un linguiste PhD en langue anglaise ont analysé et comparé toutes les traductions retour anglaises avec la version originale. Ils ont statué s'il y avait une différence significative entre les deux formes. Leur rapport a été soumis en groupe de consensus constitué de l'ensemble des NI. La tâche était de clarifier la nature de l'anomalie dans trois domaines: (I) un problème de traduction retour, (II) un problème de traduction aller, (III) un effet culturel s'il n'y avait pas de problèmes

linguistiques, mais une adaptation sémantique nécessaire pour la compréhension des patients dans leur langue "de tous les jours".

À la fin du processus, une version définitive linguistiquement et sémantiquement stable de l'original, était produite pour chaque pays. Elle englobait la structure et l'ordre des questions, question par question, ainsi que la méthode d'utilisation.

L'étude de validation française

Il fallait s'assurer que les qualités psychométriques étaient conservées lors du transfert linguistique et culturel. Lié à la conception globale de l'étude, Il était nécessaire que les médecins généralistes et les psychiatres chercheurs travaillent ensemble pour conduire l'étude.

Une étude de validation croisée quantitative a été réalisée par l'équipe de recherche des Soins primaires, Santé Publique, Registre des tumeurs de Bretagne Occidentale (EA 7479 SPURBO). C'était une enquête comparative, de non infériorité, multicentrique. L'équipe d'étude était constituée de deux médecins chercheurs, de trois internes en médecine générale formés à l'évaluation psychiatrique au moyen d'un entretien clinique structurée, d'un psychiatre, d'un statisticien, de 20 MG, d'un gestionnaire de données et d'un coordonnateur de recherche.

La population était composée de patients issus d'environnements urbains, semi-ruraux et ruraux du nord Finistère (Bretagne, France). Dans la salle d'attente, les patients ont reçu un dépliant expliquant l'étude, un questionnaire et un formulaire de consentement. Les participants ont été spontanément recrutés.

Critère d'intégration

Les patients devaient être adultes (plus de 18 ans). Ils devaient donner leur consentement éclairé écrit pour participer. Ils ont rempli le questionnaire et l'ont soumis à l'équipe d'étude.

Critère d'exclusion

Les femmes ayant eu une grossesse rapportée n'ont pas été incluses [42][43][76]. Les patients consultant pour un certificat médical, souffrant de troubles du comportement et nécessitant des soins d'urgence ont également été exclus.
La taille de l'échantillon a été calculée en fonction de la prévalence de la dépression dans la population générale. Un entretien clinique structuré a été utilisé comme guide d'entretien.

Deux groupes de patients, dépressifs et non dépressifs, devaient être comparés. Pour des raisons logistiques, le pas d'échantillonnage aléatoire a été diffèrent pour chaque groupe. La période d'inclusion était de 20 semaines, la durée de participation pour chaque patient de 1 semaine. Pour palier au risque des perdus de vue, en tenant compte de la prévalence, inclure 1100 patients était nécessaire.

L'analyse finale des données a été effectuée après le gel de la base de données, une fois tous les entretiens vérifiés lors d'une réunion finale.

La VPP et la VPN ont été calculées en fonction du tableau de contingence. Les valeurs de sensibilité et de spécificité ne pouvant être obtenues directement, les pas d'échantillonnage aléatoire des deux sous groupes étant différents, une équation correctrice a été nécessaire:

$$\boldsymbol{Se} = \frac{\boldsymbol{PPV} * \boldsymbol{P}(+)}{\boldsymbol{P}(+) * \boldsymbol{PPV} + \boldsymbol{P}(-) * (\boldsymbol{1} - \boldsymbol{NPV})}$$

$$\boldsymbol{Sp} = \frac{\boldsymbol{NPV} * \boldsymbol{P}(-)}{\boldsymbol{1} - [\boldsymbol{P}(+) * \boldsymbol{PPV} + \boldsymbol{P}(-) * (\boldsymbol{1} - \boldsymbol{NPV})]}$$

P: Prévalence; VPP = valeur prédictive positive; NPV = valeur prédictive négative
P (+) = Fréquence de patients positif; $\mathrm{P}\,(-)$ = fréquence de patients négatifs
Les intervalles de confiance ont ensuite été obtenus par une méthode Bootstrap. (Percentile Bootstrap).

RESULTATS

Revue systématique de littérature

770 résumés ont été identifiés, 546 résumés ont été analysés après la suppression de 224 doublons; 50 études étaient éligibles, 4 études ont été incluses.
Sept outils ont été sélectionnés: les **GDS-5, 15 et 30** (Echelle de Depression Gériatrique en 5, 15 et 30 questions), la **HSCL-25** (Hopkins Symptoms Checklist en 25 questions), la HADS (Hospital Anxiety Depression Scale), le **PSC-51** (Physical symptom checklist en 51 questions) et le **CES-DR** (Center for Epidemiologic Studies Depression Scale-Revised).

L'échantillonnage initial complet, les données psychométriques d'efficacité (Se, Sp, VPP, VPN) ont été colligées ou calculées. Le calcul de l'index de Youden (sensibilité + spécificité -1) permettait de comparer l'efficacité des outils.

Procédure de consensus : RAND/UCLA

Un Panel de 11 experts MG, anglophones, représentant 8 pays européens, composé de 9 femmes et 2 hommes a participé. 9 pratiquaient en zones urbaines de plus de 5 000 habitants, 2 travaillaient dans des zones urbaines de 2 000 à 5 000 habitants.

Première ronde Delphi

Le HADS et le HSCL-25 ont été sélectionnés, les autres outils ont été éliminés pour efficacité ou fiabilité non suffisante.

Groupe Nominal

Une fois les confirmée résultats de la 1ère ronde. Le test d'ergonomie en groupe nominal a eu lieu. Si la HADS était initialement privilégié, le jugement des experts a

évolué en faveur de la HSCL-25 après test en face-à-face. Tous les commentaires ont été recueillis et joints au document qui a été envoyé aux experts pour la 3ème phase.

Deuxième Ronde Delphi

Les experts ont été invités à voter: "Quel est l'outil le plus approprié pour diagnostiquer la dépression chez les patients adultes en médecine générale, en Europe, en termes d'efficacité, de stabilité et de facilité d'utilisation?". La HSCL-25 a été élue.

Procédure de traduction aller/retour et contrôle culturel

Les experts ont participé à la procédure de traduction en tant qu'investigateurs nationaux (NI). Les caractéristiques du panel sont les mêmes.

Traduction aller

De 14 à 31 MG anglophones experts ont été recrutés par pays pour la procédure Delphi. Ils constituaient un panel européen de 215 MG (111 hommes, 104 femmes), avec en moyenne 16,4 années d'expérience. 20 travaillaient dans une ville de moins de 2000 habitants, 36 dans une ville de 2000 à 5000 habitants, 159 dans une ville plus de 5000 habitants. Il y a eu une seule ronde Delphi en Pologne, Bulgarie, en Allemagne et Espagne, deux rondes dans les autres pays. 273 sur 320 propositions ont été acceptées en un tour.

Traduction retour et vérification culturelle

Toutes les traductions retours ont été linguistiquement comparées à l'original, les différences colligées et soumises aux NI en groupe de consensus.

Pour la langue grecque
L'adaptation était principalement basée sur une adaptation au genre.

Pour les langues slaves, plusieurs questions ont nécessité une adaptation culturelle : Pologne: 6 ; Bulgarie: 1; Croatie: 8.
Par exemple, dans la plupart des cas, les problèmes étaient d'ordre conceptuel : « Feeling restless » a été traduit par « Anxiété » dans les trois langues, car il n'y avait pas de mots équivalents pour exprimer cette idée de « se sentir agité ».

Pour les langues romanes, plusieurs questions ont nécessité une adaptation culturelle : Italie: 1 ; France: 1 ; Castille: 5 ; Catalogne: 3 ; Galice: 4.
Par exemple en Français, l'imparfait était le temps issue de la traduction aller, usuellement considéré comme d'un usage plus ancien et « upper class », tous les temps ont été modifié pour le présent dans la version finale.
Par exemple, pour les langues hispaniques et l'italien, la traduction a dû être adaptée au genre ; en castillan pour être compris par le patient *« Feeling worthless »* est devenu *« Feeling useless »,* soit se sentir sans valeur et devenu se sentir sans utilité.

Pour toutes les langues
La question 17 « Feeling Blue », venant de la culture afro-américaine, et qui serait une contraction de *« having the blue devils »* et *« having the blows/hits to the soul »* a induit un effet culturel dans 6 des 9 langues. Une restitution mot à mot alors impossible a imposé une adaptation culturelle.

À la fin de l'analyse culturelle, le groupe de consensus a finalement conclu qu'il n'y avait plus de changement de sens, et la traduction a été finalisée dans les 9 langues.

L'étude de validation française

1134 patients sélectionnés: 1126 ont rempli le F-HSCL-25. 2 ont été inclus à tord et 6 étaient des doublons. Les deux groupes ont été créés.
Groupe F-HSCL-25 <1.75 : 886 patients ont été randomisés selon un ratio de 1/16 ; 831 n'ont pas passé le PSE-9 :

Groupe F-HSCL-25 ≥1.75 ; 240 patients ont été randomisés selon un ratio de 1/2 ; 122 n'ont pas passé le test PSE-9.
Les patients avaient entre 18 et 94 ans. L'âge médian était de 59 ans.

Une prévalence de 21,3% a été relevée au sein des cabinets. La taille de l'échantillon a été calculée selon la prévalence de la dépression en population générale entre 5% et 12%. Cela a conduit à un excès dans le nombre d'évaluations par PSE-9.

Contingence

55 patients du groupe F-HSCL-25 <1.75 ont passé le PSE-9 ; 9 ont été perdus de vue;
118 patients du groupe F-HSCL-25 ≥1.75 ont passé le PSE-9 ; 22 ont été perdus de vue.
Pour une prévalence de 21,3%, pour un F-HSCL-25 avec un seuil de 1,75, il a été constaté une sensibilité de 59% (intervalle de confiance de 95%), une spécificité de 91% (IC 95%), une VPP de 70% (95 % CI) et une VPN de 87% (IC 95%).

DISCUSSION GENERALE

Au final de la RSL [46], 4 études ont été retenues [77][78][79][80]. Elles validaient les 7 outils suivants : GDS-30 et CESD-R, PSC-51 et HADS, GDS-5 et GDS-15, HSCL-25. L'ensemble des données psychométriques d'efficacité (Sensibilité, spécificité, valeurs prédictives positives et négatives) et d'échantillonnage [81][82] n'était pas suffisante pour permettre une comparaison

La comparaison multicritère des outils sur l'efficacité, la stabilité et l'ergonomie a été obtenue selon une RAM [51]. Une rapide revue a extrait les données de stabilité [83]. Les données d'ergonomie ont été colligées de la littérature. La HSCL-25 est apparu comme l'outil le plus intéressant. [84][85][86] .

Selon une procédure qualitative, neuf traductions consensuelles dans trois familles de langue ont été obtenues. La méthodologie [87] a assuré la validité écologique en médecine générale et le maintient des stabilités linguistique et sémantique [73][75] .

La F-HSCL-25 (forme française de la HSCL-25) a reconnu un épisode dépressif majeur avec une VPP supérieure à 60%. Sa spécificité de 91% indiquait son efficacité significative en cabinet de médecine générale avec un faible nombre de patients faux positifs.

La spécificité diagnostique de la dépression est élevée chez les médecins généralistes [20] . Combinée à l'efficacité de cet outil une synergie efficace pourrait être obtenue.

Comparaison à la littérature existante

Les psychiatres affirment la difficulté d'avoir à combiner la validité, l'utilité et le statut de la maladie dans un seul outil, ce qui empêcherait les cliniciens de les utiliser [88]. La communauté médicale internationale doute de la validité du DSM. Elle est réticente à l'utiliser comme référence [89]. Les MG rebutent à utiliser des outils pré formatés dans leur pratique [10]. Pour toutes ces raisons, les outils et échelles ont principalement été développés pour la recherche plutôt que la pratique courante.

La GDS-30 a été développé en 1982 pour diagnostiquer et quantifier la dépression chez les patients âgés [30]. Conçue en 30 questions avec un mode de réponse binaire, elle est centrée sur les symptômes de la semaine précédente. Elle est largement utilisée à des fins de recherche [90][91]. Les GDS-5 et GDS-15 sont ses versions courtes conçues pour une meilleure ergonomie [92][93].

Le CESD-R a été développé en 1977 pour diagnostiquer et quantifier la dépression [94]. Conçu en 20 questions, les réponses sont sur une échelle de Likert en 4 points. Il est centré sur les symptômes de la semaine précédente. Il est largement utilisé en recherche [95].

La PSC-51 est une liste de symptômes physiques en 51 éléments. Elle est peu utilisée [78].

La HADS a été développé en 1983 pour diagnostiquer et quantifier la dépression en milieu hospitalier [96]. Conçue en 14 questions, les réponses sont sur une échelle de Likert en 4 points. Elle est centrée sur les symptômes de la semaine précédente. Elle a été largement utilisée à des fins cliniques et de recherche [96][97]. Traduite en plusieurs langues [98], elle est validée pour une utilisation en soins primaires, elle semble complexe et plutôt « calibrée » pour la recherche [97][99][100].

La HSCL-25 a été développé en 1974 pour diagnostiquer et quantifier la dépression. [85]. Conçue en 25 questions, les réponses sont sur une échelle de Likert en 4 points. Elle est centrée sur les symptômes de la semaine précédente. Elle est largement utilisée dans les soins primaires. Elle est spécifiquement utilisée auprès des populations réfugiées et en souffrance [101][102][103][104][105][106][107].

Une comparaison sur les seuls critères d'efficacité n'était pas satisfaisante, compte tenu de la grande disparité de populations et d'échantillonnage. La RAM a permis de contourner cet obstacle. Basée sur une RSL, elle a augmenté son niveau de qualité par rapport au design original se fondant sur une revue non systématique [50]. Les chercheurs ont démontré par ce processus que l'ergonomie était déterminante dans le choix d'un outil adapté à la recherche [108]. Si la HSCL 25 et la HADS ont passé la première étape pour être discutées. Dans le vote final, c'est la HSCL-25 qui l'a emporté pour ses qualités ergonomiques.

La HSCL-25 est un instrument de test psychologique multidimensionnel pour l'évaluation des symptômes psychologiques et de détresse [84][85][86]. Ses scores robustes d'efficacité et de fiabilité [109][110][111] et ses qualités ergonomiques, pourraient permettre une recherche collaborative dans toute l'Europe, ancrée dans les soins primaires et des recherches transversales entre psychiatres et MG.

Le transfert langue à langue est un exercice subtil. La traduction grecque est restée la plus stable, suivie par le bulgare. La question 17 « Feeling blue » a été la plus

difficile à traduire. Certaines échelles ont eut besoin d'une ultime adaptation en termes de temps, d'autres en termes de genre. La difficulté à transférer les concepts sémantiques montre l'importance du processus d'adaptation culturelle dans un transfert langue à langue [112].

L'étude de validation française s'est ancrée dans les cabinets de médecine générale. La relation entre le psychiatre et les médecins généralistes a été maintenue tout au long de l'étude. Les caractéristiques de l'échantillon sont comparable aux autre études en soins primaires, européennes, [113].

Forces et limites

Tous les pays européens n'étaient pas représentés. Néanmoins les membres de cette étude couvrent une large gamme linguistique: les langues romanes, grecques, germaniques et slaves étaient représentées.

Concernant la Revue de littérature

Le biais de sélection a été limité par l'utilisation d'une équipe multilingue, travaillant à l'aveugle par paires de binômes de chercheurs, à toutes les étapes du processus de sélection et d'inclusion et par l'étendue de l'équation de recherche.

Pour limiter le biais d'information, une collection complète de tous les résumés et de tous les articles en texte intégral a été rassemblée. Aucun document n'a été omis. Le choix de la base de données est discutable, mais il a orienté la recherche vers les soins primaires.

Le biais de confusion a été limité en utilisant une procédure consensuelle de groupe pour établir la liste finale à chaque étape du processus PRISMA.

L'équipe de recherche a fait des choix successifs, pour être aussi précis que possible et maintenir la capacité de communiquer avec d'autres professionnels de

santé [24]. Choisir et maintenir délibérément l'entretien direct psychiatrique en face à face comme référence, basé sur le DSM et choisir l'index de Youden, était un moyen efficace et robuste de comparer l'efficacité des outils [114].

L'inclusion ou non du PHQ-9 a longuement été débattu. C'est un outil important de soins primaires [115]. Issu du PRIME-MD (ou PHQ : Patient Health Questionnaire) [41], il a un niveau de validation élevé. On observe dans l'étude de Kroenke, une procédure de validation dans deux sous population (une population générale consultant en centres de soins primaires et une population consultant en unité de gynécologie obstétrique), contre un entretien psychiatrique, utilisant comme guide d'entretien le SCID et le format de question du PRIME-MD modifié pour correspondre au DSM [115]. L'entretien était téléphonique, il était spécifié dans les critères d'inclusion que l'entretien psychiatrique devait être réalisé en face à face pour éviter tout biais. La référence était le PRIME-MD, même si les critères du DSM étaient intégrés. Pour éviter un biais de confusion, le consensus du comité scientifique a été de ne pas intégrer cette étude.

Concernant la RAND/UCLA

Le panel s'est conformé aux exigences de variabilité dans la culture, la langue et la pratique. Il a été de taille suffisante, la RAM recommande 7 à 15 experts, entre 10 et 11 ont participés à l'étude [51].

Les délais pour les rondes Delphi ont été courts. Chaque jugement a été effectué à l'aveugle [116]. Pour réduire le biais d'information, chaque expert a reçu un exemplaire de toutes les sources bibliographiques des données fournies.
Les données de fiabilité, extraites selon une revue rapide de littérature, étaient basées sur l'alpha de Cronbach, aucun Kappa de Cohen n'a été trouvé. Une revue systématique aurait eu plus de puissance, mais aurait augmenté la durée de l'étude au détriment de la faisabilité [117].

Concernant le processus de traduction

Pour réduire le biais de sélection et assurer la qualité de l'échantillon,: le design a assuré la participation des MG lors de la traduction linguistique aller. Comme l'ont décrit de nombreux traducteurs lorsqu'il est question de traduction scientifique, il est essentiel qu'un «spécialiste» du domaine se penche en tant qu'arbitre sur la traduction [74][54][118].

L'analyse étape par étape du contrôle culturel, basé sur la comparaison des formes anglaises originale et retour, a permis d'éviter les biais de confusion et les problèmes linguistiques liés au transfert. L'intervention d'un groupe de consensus a permis plusieurs évaluations progressives de chaque élément, renforçant l'exactitude des traductions validées et concevant en commun le résultat final. Ce travail est le résultat d'une recherche collaborative multiculturelle entre pays européens.

Concernant l'étude de validation

Plusieurs types de procédures de qualité des données ont été appliqués: un gestionnaire de données a contrôlé la qualité du protocole et le déroulement de l'étude et le traitement des données. L'expertise de chaque partie prenante de l'équipe a permis de sécuriser la collecte de données à chaque étape.
Nous avons utilisé la randomisation stratifiée pour assurer à la fois une puissance statistique satisfaisante et une logistique abordable. Cette randomisation a été réalisée indépendamment, via un logiciel informatique, excluant toute intervention humaine.

Des internes de médecine générale (IMG) spécialement formés pour utiliser le PSE-9 par un psychiatre agréé du CHRU de Brest, ont réalisé l'évaluation psychiatrique validée semi-structurée. Les IMG ont entrepris le recrutement et les entrevues. Un

psychiatre a ensuite confirmé la validité des diagnostics cliniques. Après gel des données, l'analyse statistique a été confiée au Centre d'Investigation Clinique de Brest.

Un réseau de recherche multidisciplinaire a soutenu l'étude. Un réseau de recherche en MG a permis le recrutement. La population était multicentrique et diversifiée. Le recrutement parmi les patients dans les salles d'attente assurait sa représentativité. Il a été effectué sur une courte période, de la même manière dans les trois centres d'étude.

En 1993, Nettlebladt a utilisé la même méthode pour évaluer l'exactitude du HSCL-25 en tant que questionnaire diagnostic en soins primaires, dans six centres de soins dans deux districts, l'un rural et l'autre semi-urbain, contre le PSE-9 et établir un seuil. Notre étude a abouti à une sensibilité plus faible (59% contre 76%), mais une spécificité plus élevée (91% contre 73%), pour un seuil à 1,75. La prévalence de la morbidité psychiatrique était plus faible (21% contre 33%). Des études antérieures ont montré des résultats similaires en termes de sensibilité et de spécificité [106][119]. Notre échantillon de patients était un peu plus grand (1146 contre 727), ce qui pourrait peut-être expliquer les différences observées.

Selon Nettlebladt, un seuil à 1,55, avait tendance à augmenter la sensibilité (89%), mais également donnait plus de faux positifs (43%). La capacité de dépistage était améliorée au détriment de la diagnostic. Dans la version originale, le seuil de 1.55 est un critère d'alerte, le seuil de 1,75 défini un patient requérant un traitement pour syndrome anxio-dépressif. Pour l'étude de validation, c'est le seuil de 1,75 qui a été considéré.

L'utilisation d'un pas de randomisation différent pour chaque groupe pourrait aussi expliquer nos différences avec l'étude de Nettelbladt. Cependant la différence de pas de randomisation nous a permis d'équilibrer le nombre de patients PSE-9 au plus près.

Une étude suédoise plus récente a examiné la concordance entre la HSCL-25 et les critères des troubles anxio-dépressifs du DSM [120]. Son échantillonnage est grand (8613 personnes). Mais il est en population générale et non dans une population médicale consultant en cabinet. Une excellente concordance a été constatée entre HSCL-25 et DSM. Leurs résultats proches des nôtres, confirment la HSCL-25 avec un seuil à 1,75, comme définitivement un outil de diagnostique plus que de dépistage.

Implications pédagogiques

En éducation médicale, les IMG sont souvent confrontés à la question de savoir comment faire un diagnostic de dépression. Beaucoup ont le sentiment qu'ils ont de la difficulté à détecter la dépression et, par conséquent, ils ne savent pas qui traiter et « quand ils doivent passer la main ». Cette étude est principalement axée sur la recherche, toutefois l'utilisation d'outils catégoriels peut être d'une grande aide pour ces IMG. Ils seront en mesure d'évaluer leur pratique avec ces outils et d'établir des méthodes professionnelles solides pour le diagnostic de la dépression. Comme toujours, un outil n'est qu'un point d'entrée pour le diagnostic et pour la conversation avec le patient. Les élèves doivent apprendre à introduire un outil dans la consultation; comment stimuler les patients à utiliser un outil; comment interpréter, discuter et enregistrer les résultats, et par la suite, comment suivre leurs patients avec cette aide.

La plupart des outils existant sont en langue anglo-américaine. La traduction reste l'étape la plus cruciale dans l'adoption d'un instrument dans une autre langue. Des erreurs de traduction peuvent fausser l'intention originale de l’instrument et compromettre sa validité et sa fiabilité [121]. Les problèmes sémantiques affectent la comparabilité dans les études internationales [122][123]. Certains termes et concepts peuvent ne pas exister dans d'autres langues, ou peuvent avoir des connotations supplémentaires que les rétro-traductions ne révèlent pas toujours. Les défis surgissent, non seulement à cause du contenu de la traduction littérale mot à mot, mais aussi à cause de la forme linguistique du langage comme le ton et la syntaxe [124].

Implications dans la recherche

Les études recueillies par la RSL concernaient des adultes. Peut-on considérer l'âge comme seul facteur discriminant, alors qu'en Europe, la population active de 50 à 64 ans représente 1/3 de la population active âgée de 20 à 64 ans [125] ? Dans les études futures, d'autres facteurs seront à prendre en compte comme par exemple la capacité à faire face [126].

Dans la perspective d'études collaboratives, psychiatres et MG devront travailler ensemble. La spécificité diagnostique des MG concernant dépression est élevée [20]. Le choix d'outils communs, basé sur des critères statistiques ne sera pas suffisant, les critères d'utilité et d'ergonomie seront essentiels [88]. D'autres recherches, impliquant psychiatres et MG avec une méthodologie standardisée, seront nécessaires pour choisir les meilleurs outils possibles, pour des études collaboratives à l'échelle européenne, dans les différents champs des affections mentales [127].
Par conséquent, l'utilisation de modèles comparatifs multicritères incluant efficacité et fiabilité, comme celui de l'énoncé COSMIN ou de procédure de sélection qualitative intégrant des données quantitatives comme la RAM sont des ressources indispensables en amont des procédures de recherche collaborative [128].
Conclusion

Les MG de nombreux pays d'Europe sont désormais en mesure d'utiliser la HSCL-25 dans des recherches en médecine générale afin de diagnostiquer la dépression chez leurs patients.

L'utilisation d'un tel outil partagé rend faisable la recherche sur la dépression dans les soins primaires en Europe. Ceci permettra d'entreprendre des revues statistiques comparatives épidémiologiques et séméiologiques et soutiendra la conceptualisation des phénomènes étudiés.

Néanmoins, cet outil très spécifique a une valeur prédictive positive limitée. Il doit être utilisé avec prudence en pratique courante. Sa combinaison avec un outil de dépistage pourrait être intéressante. L'équipe de recherche va maintenant s'atteler à suivre ce nouveau chemin, afin de trouver la meilleure combinaison ergonomique et efficace de cet outil avec un outil de dépistage.

BILBLIOGRAPHIE

1. World Health Organization: **WORLD HEALTH STATISTICS - MONITORING HEALTH FOR THE SDGs**. *World Heal Organ* 2016:1.121.

2. Filipovic-Pierucci A, Samson S, Fagot J-P, Fagot-Campagna A: **Estimating the prevalence of depression associated with healthcare use in France using administrative databases**. *BMC Psychiatry* 2017, **17**:1.

3. 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.

4. Ayuso-Mateos JL, Vázques-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, Gaite 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.

5. Beck F, Guignard R: **La dépression en France (2005-2010) : prévalence, recours au soin et sentiment d'information de la population**. *Sante Homme* 2012, **421**:43–45.

6. Dezetter A, Briffault X, Alonso J, Angermeyer MC, Bruffaerts R, de Girolamo G, De Graaf R, Haro JM, König HH, Kovess-Masfety V: **Factors Associated With Use of Psychiatrists and Nonpsychiatrist Providers by ESEMeD Respondents in Six European Countries**. *Psychiatr Serv* 2011, **62**:143–151.

7. **OMS | La dépression**. *WHO* 2016.

8. Jorm AF: **Mental health literacy. Public knowledge and beliefs about mental disorders.** *Br J Psychiatry* 2000, **177**:396–401.

9. Ani C, Bazargan M, Hindman D, Bell D, Farooq MA, Akhanjee L, Yemofio F, Baker R, Rodriguez M: **Depression symptomatology and diagnosis: discordance between patients and physicians in primary care settings.** *BMC Fam Pract* 2008, **9**:1.

10. Thomas-MacLean R, Stoppard J, Miedema BB, Tatemichi S: **Diagnosing depression: there is no blood test.** *Can Fam Physician* 2005, **51**:1102–1103.

11. Nutting PA, Rost K, Dickinson M, Werner JJ, Dickinson P, Smith JL, Gallovic B: **Barriers to Initiating Depression Treatment in Primary Care Practice**. *J Gen Intern Med* 2002, **17**:103–111.

12. Torzsa P, Szeifert L, Dunai K, Kalabay L, Novák M: **[Diagnosis and therapy of depression in family practice].** *Orv Hetil* 2009, **150**:1684–93.

13. IFOP: **Les Français et le système de santé**. *La Vague* 2013, **FD/ AB N°1**.

14. Kovess-Masféty V, Saragoussi D, Sevilla-Dedieu C, Gilbert F, Suchocka A, Arveiller N, Gasquet I, Younes N, Hardy-Bayle M-C: **What makes people decide who to turn to when faced with a mental health problem? Results from a French survey.** *BMC Public Health* 2007, **7**:188.

15. Alonso J, Codony M, Kovess V, Angermeyer MC, Katz SJ, Haro JM, De Girolamo G, De Graaf R, Demyttenaere K, Vilagut G, Almansa J, Lépine JP, Brugha TS: **Population level of unmet need for mental healthcare in Europe.** *Br J Psychiatry* 2007, **190**:299–306.

16. Demyttenaere K, Bruffaerts R, Posada-Villa J, Gasquet I, Kovess V, Lepine JP, Angermeyer MC, Bernert S, de Girolamo G, Morosini P, Polidori G, Kikkawa T, Kawakami N, Ono Y, Takeshima T, Uda H, Karam EG, Fayyad JA, Karam AN, Mneimneh ZN, Medina-Mora ME, Borges G, Lara C, de Graaf R, Ormel J, Gureje O, Shen Y, Huang Y, Zhang M, Alonso J, et al.: **Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys.** *JAMA* 2004, **291**:2581–90.

17. 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.

18. Hérique A, Kahn J-P: **Réalités et recommandations dans la prescription et l'observance des antidépresseurs en médecine générale : évaluation des**

pratiques dans le traitement de la dépression en Lorraine et Champagne-Ardenne. *Encephale* 2009, **35**:73–79.
19. OMS: **CIM-10 à usage PMSI**. 2015:888.
20. Mitchell AJ, Vaze A, Rao S: **Clinical diagnosis of depression in primary care: a meta-analysis.** *Lance* 2009, **374**:609–619.
21. Aben I, Verhey F, Beusmans G, Lodder J: **Recognition and treatment of post-stroke depression in general practice**. *Huisarts Wet* 2003, **46**.
22. Coventry PA, Hays R, Dickens C, Bundy C, Garrett C, Cherrington A, Chew-Graham C: **Talking about depression: a qualitative study of barriers to managing depression in people with long term conditions in primary care**. *BMC Fam Pract* 2011, **12**:10.
23. Lehti A, Hammarström A, Mattsson B: **Recognition of depression in people of different cultures: a qualitative study**. *BMC Fam Pract* 2009, **10**:53.
24. Zhang J, Patel VL, Johnson TR, Shortliffe EH: **A cognitive taxonomy of medical errors**. *J Biomed Inform* 2004, **37**:193–204.
25. **Apa: *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV*.** *Volume 4th*. American Psychiatric Association; 1994(VI).
26. American Psychiatric Association: **Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5)**. *Diagnostic Stat Man Ment Disord 4th Ed TR* 2013:280.
27. Kessler RC, Andrews G, Mroczek D, Ustun B, Wittchen H-U: **The World Health Organization Composite International Diagnostic Interview short-form (CIDI-SF)**. *Int J Methods Psychiatr Res* 1998, **7**:171–185.
28. First MB, Spitzer RL, Gibbon M, Williams JBW: **Structured Clinical Interview for DSM-IV Axis I Disorders, Clinician Version (SCID-CV).** American Psychiatric Press; 1997.
29. Sheehan D V, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, Hergueta T, Baker R, Dunbar GC: **The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10.** *J Clin Psychiatry* 1998, **59 Suppl 2**:22-33;quiz 34-57.

30. 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.

31. 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.

32. Hamilton M: **Rating depressive patients.** *J Clin Psychiatry* 1980, **41**(12 Pt 2):21–4.

33. Hedlund JLUVBW: **The Hamilton rating Scale for Depression: a comprehensive review**. *J Oper Psychatrie* 1979, **10**:149–165.

34. Bagby RM, Ryder AG, Schuller DR, Marshall MB: **The Hamilton Depression Rating Scale: Has the Gold Standard Become a Lead Weight?** *Am J Psychiatry* 2004, **161**:2163–2177.

35. Trajković G, Starčević V, Latas M, Leštarević M, Ille T, Bukumirić Z, Marinković J: **Reliability of the Hamilton Rating Scale for Depression: A meta-analysis over a period of 49years**. *Psychiatry Res* 2011, **189**:1–9.

36. Beck AT, Steer RA, Carbin MG: **Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation**. *Clin Psychol Rev* 1988, **8**:77–100.

37. Montgomery SA, Asberg M: **A new depression scale designed to be sensitive to change**. *Br J Psychiatry* 1979, **134**:382–389.

38. Williams JBW, Kobak KA: **Development and reliability of a structured interview guide for the Montgomery-Åsberg Depression Rating Scale (SIGMA)**. *Br J Psychiatry* 2008, **192**:52–58.

39. Spitzer RL, Williams JB, Kroenke K, Hornyak R, McMurray J: **Validity and utility of the PRIME-MD patient health questionnaire in assessment of 3000 obstetric-gynecologic patients: the PRIME-MD Patient Health Questionnaire Obstetrics-Gynecology Study.** *Am J Obstet Gynecol* 2000, **183**:759–69.

40. Kroenke K, Spitzer RL, Williams JB: **The PHQ-9: validity of a brief depression severity measure.** *J Gen Intern Med* 2001, **16**:606–613.

41. 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.

42. Grazioli R, Terry DJ: **The role of cognitive vulnerability and stress in the prediction of postpartum depressive symptomatology**. *Br J Clin Psychol* 2000, **39**:329–347.
43. Stowe ZN, Hostetter AL, Newport D: **The onset of postpartum depression: Implications for clinical screening in obstetrical and primary care**. *Am J Obstet Gynecol* 2005, **192**:522–526.
44. 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.
45. Whiffen VE: **The comparison of postpartum with non-postpartum depression: a rose by any other name.** *Journal of psychiatry & neuroscience : JPN* 1991:160–165.
46. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, Clarke M, Devereaux PJ, Kleijnen J, Moher D: **The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration**. *PLoS Med* 2009, **6**:28.
47. Counsell C: **Formulating questions and locating primary studies for inclusion in systematic reviews.** *Ann Intern Med* 1997, **127**:380–387.
48. 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.
49. 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.
50. Nabbe P, Le Reste JY, Guillou-Landreat M, Munoz Perez MA, Argyriadou S, Claveria A, Fernández 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* 2017, **39**:99–105.
51. 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.

52. HAS, Haute Autorité Santé: **Bases Méthodologiques Pour l'élaboration de Recommandations Professionnelles Par Consensus Formalisé.** 2006.

53. 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.

54. Skulmoski GJ, Hartman FT, Krahn J: **The Delphi Method for Graduate Research**. *J Inf Technol Educ* 2007, **6**:1.

55. Letrilliart L, Vanmeerbeek M: **À la recherche du consensus : quelle méthode utiliser ?** *exercer* 2011, **99**:170–177.

56. McGory ML, Shekelle PG, Ko CY: **Development of quality indicators for patients undergoing colorectal cancer surgery.** *J Natl Cancer Inst* 2006, **98**:1623–1633.

57. Noblit G, Hare R: **Meta-Ethnography: Synthesizing Qualitative Studies.** Sage; 1988, **11**.

58. Hassan TB, Barnett DB: **Delphi type methodology to develop consensus on the future design of EMS systems in the United Kingdom.** *Emerg Med J* 2002, **19**:155–9.

59. Jamieson S: **Likert scales: how to (ab)use them.** *Med Educ* 2004, **38**:1217–8.

60. Jones PS, Lee JW, Phillips LR, Zhang XE, Jaceldo KB: **An adaptation of Brislin's translation model for cross-cultural research**. *Nurs Res* 2001, **50**:300–304.

61. SPOONER D, PACHANA N: **Ecological validity in neuropsychological assessment: A case for greater consideration in research with neurologically intact populations**. *Arch Clin Neuropsychol* 2006, **21**:327–337.

62. Cuéllar I, Paniagua FA: **Handbook of Multicultural Mental Health : Assessment and Treatment of Diverse Populations.** Academic Press; 2000.

63. Sousa VD, Rojjanasrirat W: **Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline.** *J Eval Clin Pract* 2011, **17**:268–74.

64. Maneesriwongul W, Dixon JK: **Instrument translation process: a methods review.** *J Adv Nurs* 2004, **48**:175–86.

65. Jones J, Hunter D: **Consensus methods for medical and health services research**. *Bmj Clin Res Ed* 1995, **311**:376–380.

66. Linstone HA, Turoff M: **The Delphi Method : Techniques and Applications.** Addison-Wesley Pub. Co., Advanced Book Program; 1975.
67. Graham B, Regehr G, Wright JG: **Delphi as a method to establish consensus for diagnostic criteria.** *J Clin Epidemiol* 2003, **56**:1150–6.
68. De Villiers MR, De Villiers PJT, Kent AP: **The Delphi technique in health sciences education research.** *Med Teach* 2005, **27**:639–643.
69. Hasson F, Keeney S, McKenna H: **Research guidelines for the Delphi survey technique.** *J Adv Nurs* 2000, **32**:1008–1015.
70. Powell C: **The Delphi technique: myths and realities.** *J Adv Nurs* 2003, **41**:376–82.
71. Romm FJ, Hulka BS: **Developing criteria for quality of assessment: effect of the Delphi technique.** *Health Serv Res* 1979, **14**:309–312.
72. Anadón M, Guillemette F: **La recherche qualitative est-elle nécessairement inductive?** *Rech Qual* 2006.
73. Bullinger M, Anderson R, Cella D, Aaronson N: **Developing and evaluating cross-cultural instruments from minimum requirements to optimal models**. *Qual Life Res* 1993, **2**:451–459.
74. Beaton DE, Bombardier C, Guillemin F, Ferraz MB: **Guidelines for the process of cross-cultural adaptation of self-report measures.** *Spine (Phila Pa 1976)* 2000, **25**:3186–91.
75. Herdman M, Fox-Rushby J, Badia X: **"Equivalence" and the translation and adaptation of health-related quality of life questionnaires.** *Qual Life Res* 1997, **6**:237–47.
76. Uher R, Payne JL, Pavlova B, Perlis RH: **MAJOR DEPRESSIVE DISORDER IN DSM-5: IMPLICATIONS FOR CLINICAL PRACTICE AND RESEARCH OF CHANGES FROM DSM-IV**. *Depress Anxiety* 2014, **31**:459–471.
77. 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.
78. 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.

79. 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.

80. Fröjdh K, Håkansson A, Karlsson I, Frojdh 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.

81. Youden WJ: **Index for rating diagnostic tests.** *Cancer* 1950, **3**:32–35.

82. Fluss R, Faraggi D, Reiser B: **Estimation of the Youden Index and its associated cutoff point.** *Biometrical J Biometrische Zeitschrift* 2005, **47**:458–472.

83. Tavakol M, Dennick R: **Making sense of Cronbach's alpha.** *Int J Med Educ* 2011, **2**:53–55.

84. Derogatis LR, Unger R, Derogatis LR, Unger R: **Symptom Checklist-90-Revised**. In *The Corsini Encyclopedia of Psychology*. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2010.

85. 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.

86. Lipman RS, Covi ' L, Shapiro AK: **THE HOPKINS SYMPTOM CHECKLIST (HSCL) Factors Derived from the HSCL-90**. *J Affect Disord* 1979, **1**:9–24.

87. Brislin RW: **Back-Translation for Cross-Cultural Research**. *J Cross Cult Psychol* 1970:2.

88. Kendell R, Jablensky A: **Distinguishing between the validity and utility of psychiatric diagnoses.** *Am J Psychiatry* 2003, **160**:4–12.

89. Cosgrove L, Krimsky S, Vijayaraghavan M, Schneider L: **Financial ties between DSM-IV panel members and the pharmaceutical industry.** *Psychother Psychosom* 2006, **75**:154–160.

90. 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.

91. 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.
92. 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.
93. Weeks SK, McGann PE, Michaels TK, Penninx BWJH: **Comparing various short-form Geriatric Depression Scales leads to the GDS-5/15.** *J Nurs Scholarsh an Off Publ Sigma Theta Tau Int Honor Soc Nurs Sigma Theta Tau* 2003, **35**:133–137.
94. Radloff LS: **The CES-D Scale: A Self-Report Depression Scale for Research in the General Population**. *Appl Psychol Meas* 1977, **1**:385–401.
95. 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.
96. Zigmond AS, Snaith RP: **The hospital anxiety and depression scale (HADS).** *Acta Psychiatr Scand* 2007, **67**:0–4.
97. 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.
98. 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.
99. Andrews B, Hejdenberg J, Wilding J: **Student anxiety and depression: comparison of questionnaire and interview assessments.** *J Affect Disord* 2006, **95**:29–34.
100. 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.
101. Ekblad S, Roth G: **Diagnosing posttraumatic stress disorder in multicultural patients in a Stockholm psychiatric clinic.** *J Nerv Ment Dis* 1997, **185**:102–107.

102. 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.

103. Oruc L, Kapetanovic A, Pojskic N, Miley K, Forstbauer S, Mollica RF, Henderson DC: **Screening for PTSD and depression in Bosnia and Herzegovina: validating the Harvard Trauma Questionnaire and the Hopkins Symptom Checklist**. *International Journal of Culture and Mental Health* 2008:105–116.

104. 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.

105. 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.

106. 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.

107. Munk-Jørgensen P, Fink P, Brevik JI, Dalgard OS, Engberg M, Hansson L, Holm M, 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.

108. Hignett S, Carayon P, Buckle P, Catchpole K: **State of science: human factors and ergonomics in healthcare.** *Ergonomics* 2013, **56**:1491–503.

109. Sandanger I, Moum T, Ingebrigtsen G, Dalgard OS, S??rensen 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**:345–354.

110. Strand BH, Dalgard ODDS, 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).** *Psychiatry Interpers Biol Process* 2003, **57**:113–118.

111. 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.

112. Brislin RW: **Comparative research methodology: Cross-cultural studies**. *Int J Psychol* 1976, **11**:215–229.

113. King M, Nazareth I, Levy G, Walker C, Morris R, Weich S, Bellon-Saameno JA, Moreno B, Svab I, Rotar D, Rifel J, Maaroos H-I, Aluoja A, Kalda R, Neeleman J, Geerlings MI, Xavier M, de Almeida MC, Correa B, Torres-Gonzalez F: **Prevalence of common mental disorders in general practice attendees across Europe**. *Br J Psychiatry* 2008, **192**:362–367.

114. Perkins NJ, Schisterman EF: **The Youden Index and the optimal cut-point corrected for measurement error.** *Biometrical J Biometrische Zeitschrift* 2005, **47**:428–441.

115. Kroenke K, Spitzer RL, Williams JBW: **The PHQ-9**. *J Gen Intern Med* 2001, **16**:606–613.

116. Elmer F, Seifert I, Kreibich H, Thieken AH: **Delphi method**. *Innovation* 2010, **30**:93–113.

117. Ganann R, Ciliska D, Thomas H: **Expediting systematic reviews: methods and implications of rapid reviews**. *Implement Sci* 2010, **5**:56.

118. Balliu C: **L'enseignement de la traduction médicale: pour une nouvelle pragmatique**. *Meta J des traducteurs* 1994, **39**:15–25.

119. Hesbacher PT, Rickels K, Morris RJ, Newman H, Rosenfeld H: **Psychiatric illness in family practice.** *J Clin Psychiatry* 1980, **41**:6–10.

120. Lundin A, Hallgren M, Forsell Y: **The validity of the symptom checklist depression and anxiety subscales: A general population study in Sweden**. *J Affect Disord* 2015, **183**:247–252.

121. Yu DSF, Lee DTF, Woo J: **Issues and Challenges of Instrument Translation**. *West J Nurs Res* 2004, **26**:307–320.

122. Schnohr CW, Gobina I, Santos T, Mazur J, Alikasifuglu M, Välimaa R, Corell M, Hagquist C, Dalmasso P, Movseyan Y, Cavallo F, van Dorsselaer S, Torsheim T: **Semantics bias in cross-national comparative analyses: is it good or bad to have "fair" health?** *Health Qual Life Outcomes* 2016, **14**:70.

123. Daugherty JC, Puente AE, Fasfous AF, Hidalgo-Ruzzante N, Pérez-Garcia M: **Diagnostic mistakes of culturally diverse individuals when using North American neuropsychological tests**. *Appl Neuropsychol* 2017, **24**:16–22.

124. Hanrahan D, Sexton P, Hui K, Teitcher J, Sugarman J, London AJ, Barnes M, Purpura J, Klitzman R: **Linguistic and cultural challenges in communication and translation in ussponsored HIV Prevention research in emerging economies**. *PLoS One* 2015, **10**:e0133394.

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

126. 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**. *J Am Med Dir Assoc* 2013, **14**.

127. 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.

128. 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.

REMERCIEMENTS, ACKNOWLEDGMENTS

Je voudrais exprimer ma plus profonde gratitude à Paul Van Royen pour m'avoir donné l'honneur de travailler avec lui.
Dans une vie, c'est une chance de rencontrer un homme comme lui. Il a toujours été présent et à l'écoute, malgré ses lourdes responsabilités et son temps précieux.
Je me souviens d'une soirée à Cracovie où nos échanges ont été riches et amicaux, malgré l'obstacle de l'usage de l'anglais.
Encore merci, car sans vous, je ne serais pas ici.

Je voudrais exprimer ma plus profonde gratitude à Harm Van Marwijk pour m'avoir donné l'honneur de travailler avec lui. Autant de patience et de bienveillance mérite mon plus profond respect.
Je me souviens de Dubrovnik, car c'est là, par Harm, que tout va débuter. Sans Harm, je ne serais pas non plus ici.

Avoir la chance de travailler avec Harm et Paul aura été un tournant majeur de la seconde partie de ma vie.

Je voudrais exprimer ma plus profonde gratitude à Tristan Montier. Tristan, c'est le chercheur qui un jour se dit qu'il faut quitter le laboratoire pour aller à la rencontre de l'autre. C'est grâce à des hommes comme lui, véritables ponts entre les savoirs, que le monde peut changer. Je me souviens de nos échanges et j'ai la chance de l'avoir rencontrer.

Je voudrais dire ici ma profonde pensée pour notre Doyen, Christian Berthou ; sans son soutien constant, tous ces travaux n'auraient pas vue le jour.

J'ai une profonde pensée pour mes deux amis, Jean Yves et Bernard. Dans les moments difficiles de ma vie, ils ont toujours été présents. Et encore aujourd'hui, ils vont m'accompagner dans un avenir serein.

Sans les personnes qui nous entourent, nous sommes peu de choses. Alors, je voudrais dire ici mon profond remerciement à l'ensemble des membres de notre département, de tous âges, de tout horizon: Marie Barais, Benoit Chiron, Pierre Barraine pour les plus anciens et Jeremy Derriennic, Delphine Le Goff, Michele Odorico, Sophie Lalande pour les plus jeunes. Et je n'oublie pas Florence Morvan, Florence Gatineau, Alex Gillman notre traductrice, Morgane Guillou-Landreat notre psychiatre préférée. Et tous nos internes, sans qui ce travail aurait été impossible.

J'ai une pensée particulière pour mon associée, Bénédicte Bodin, qui a dû supporter mes états d'âmes, tout au long de mon cursus de "thésard".

Je pense aussi à mes collègues du CNGE, avec lesquels j'ai toujours plaisir à échanger: Jean Pierre Lebeau, Denis Pouchain, Isabelle Aubain, Alain Mercier, Anne Marie Lher et la liste est bien trop longue à citer.

Mes remerciements profonds vont à l'équipe européenne, et à ceux qui sont devenus au fil du temps des amis: Djurdjica Lazic, Radost Assenova, Heidrun Lingner, Slawomir Czachowski, Ana Claveria, Miquel Angel Munoz.

Maintenant,

Je voudrais dire ici l'amour et le pardon à ma famille, pour lesquels l'épreuve d'une thèse de sciences n'est pas facile. Je pense à mes filles Marie et Lucie que j'embrasse, et qui m'ont toujours soutenu. Mais aussi à mon ex-épouse, à qui je n'ai pas rendu la vie facile.

Mais je pense aujourd'hui, profondément, à la femme qui me rend heureux, Elisabeth, qui sait toujours me supporter et me faire aller de l'avant. Sans elle, rien de ce qui se fait maintenant serait possible. Sans elle, rien de ce qui se fera plus tard sera possible.

Pour finir, je voudrais dire à mes parents tout l'amour que je leur porte. Sans ma maman, qui m'a toujours soutenue, je ne suis pas sur que j'en serais là !

I would like to express my deepest gratitude to Paul Van Royen for giving me the honor of working with him.
In a lifetime, it's a chance to meet a man like him. He has always been present and attentive, despite his heavy responsibilities and his precious time.
I remember an evening in Krakow where our exchanges were rich and friendly, despite the obstacle of the use of English.
Thanks again, because without you, I would not be here.

I would like to express my deepest gratitude to Harm Van Marwijk for giving me the honor of working with him. So much patience and kindness deserve my deepest respect.
I remember Dubrovnik, because it's there, by Harm, that everything will start. Without Harm, I would not be here either.

Having the chance to work with Harm and Paul has been a major turning point in the second half of my life.

I would like to express my deepest gratitude to Tristan Montier. Tristan, it is the researcher who one day says that it is necessary to leave the laboratory to go to meet the other one. It is thanks to men like him, true bridges between knowledge, that the world can change. I remember our exchanges and I'm lucky to have met him.

I would like to say here my deep thought for our Dean, Christian Berthou; without his constant support, all this work would not have seen the day.

I have a deep thought for my two friends, Jean Yves and Bernard. In the difficult moments of my life, they have always been present. And even today, they will accompany me in a serene future.

Without the people around us, we are not much. So, I would like to say here my deepest thanks to all the members of our department, of all ages, from all horizons: Marie Barais, Benoit Chiron, Pierre Barraine for the oldest and Jeremy Derriennic,

Delphine Le Goff, Michele Odorico, Sophie Lalande for the youngest. And I do not forget Florence Morvan, Florence Gatineau, Alex Gillman our translator, Morgane Guillou-Landreat our favorite psychiatrist, and all our trainees, without whom this work would have been impossible.

I have a particular thought for my partner, Bénédicte Bodin, who had to endure my moods, throughout my course of "PhD student".

I am also thinking of my colleagues from the CNGE, with whom I always enjoy to discuss: Jean Pierre Lebeau, Denis Pouchain, Isabelle Aubain, Alain Mercier, Anne Marie Lher and the list is too long to mention.

My deepest thanks go to the European team, and to those who have become over time, friends: Djurdjica Lazic, Radost Assenova, Heidrun Lingner, Slawomir Czachowski, Ana Claveria, Miquel Angel Munoz.Now,

I would like to say here the love and forgiveness to my family, for whom the test of a science thesis is not easy. I think of my daughters Marie and Lucie whom I embrace, and who have always supported me. But also to my ex-wife, to whom I did not make life easy.

But today I think, deeply of the woman who makes me happy, Elisabeth, who always knows how to support me and make me go forward. Without it, nothing that is done now would be possible. Without it, nothing that will be done later will be possible.

Finally, I would like to tell my parents all the love I have for them. Without my mom, who has always supported me, I'm not sure that I would be there!

CURRICULUM VITAE

Personal

Name Patrice Nabbe

Birthdate 03 février 1963

Nationality French

E-mail Patrice.nabbe@univ.brest.fr

Place of practice

Cabinet medical de la baie

Zone artisanale la gare

29890 PLOUNEOUR BRIGNOGAN PLAGES FRANCE

Tel 00 33 2 98 83 51 31

University

Université de Bretagne Occidentale

Unité de formation et de recherche en médecine

EA 7479 SPURBO

Département universitaire de médecine générale

22 av Camille Desmoulins

29200 Brest

E-mail : Patrice.nabbe@univ-brest.fr

Cell phone: 00 33 6 07 63 14 90

Education

1986

- Diplôme d'Etudes Universitaires Générales – Sciences de la Nature et de la vie Faculté des sciences de la vie et de la terre de Dijon– Université de Bourgogne

1997

- Diplôme d'état de Docteur en Médecine – Qualification en Médecine Générale – Faculté de médecine de Dijon – Bourgogne

2009

- Diplôme Inter Universitaire – Formation complémentaire pour les médecins généralistes en gynécologie/obstétrique – Université de Bretagne Occidentale

2010

- Qualitative research in primary Health Care september 2010: Antwerp university, (Bergium).
- International course on research in primary health care : UMC Amsterdam, (Nederlands), september 2010.

2011

- International summer course on research in primary health care , second year: Maastricht University (Nederlands), september 2011.

2012

- International Spring course on research in primary health care, second year: Marburg University (Germany), June 2012.
- International Summer course on primary Health care, Antwerp University (Belgium), October 2012 second year.

PhD student at university of Antwerp, Belgium and Université de Bretagne occidentale, France. Tutors Paul Van Royen, MD, PhD, Dean; Harm Van Marwijk, MD, PhD; Tristan Montier MD, PhD, HDR. Topic : the research objective of the thesis was to select a tool that could be used and accepted by GPs to diagnose depression in adults' outpatients. It had to be efficient, reliable and easy to use in daily practice. This tool should be applicable in the European countries.

Professional Practice

1997 Family Physician, group practice in Nevers (Nièvre, France)
2002 Family Physician, practice alone in Plouneour Brignogan Plages (Finistère, France)
2011 Family Physician, Group practice in Plouneour Brignogan Plages (Finistère, France)
2008 Member of the medical staff of Lesneven's Hospital
2011 In charge of the medical information department of Lesneven's Hospital
2011 Animator of the peer group of FPs of Lanmeur
2011 Member of the directory of Lesneven's Hospital

University Titles

2004 Master internship university, General practice, University of Medicine, Brest
2005 Tutor at university of medicine Brest.
2005 Clinical teacher, General practice, University of Medicine, Brest
2008 In charge of " Pays de Lesneven - Côte des légendes" 's university primary care clinic
2010 Coordinator assistant of study General Practice Specialist Diploma
2011 Lecturer at university of medicine Brest.
2014 Member of ERCR SPURBO (commute in EA 7479 SPURBO, 2018)
2016 Director of the département de médecine générale, University of medicine Brest

Scientific Work

Original articles in French or other national languages in national publications (9)

1. Nabbe P, Le Reste JY. Les Bretons chez les Dalmates à la découverte du concept de morbidité. Exercer 2010 ; 92 (supp.3) :70S-1S.
2. Barais M, Cadier S, Chiron B, Barraine P, Nabbe P, Le Reste JY. Ejaculation prématurée : stratégies utilisées pour aborder le sujet en médecine générale. exercer 2011;95:10-5.
3. Coppens M, Barraine P, Barais M, Nabbe P, Berkhout C, Stolper E, Le Reste JY. L'intuition en médecine générale : validation française du consensus néerlandais « gut feelings ». exercer 2011;95:16-20.
4. Frèche B, Le Grand-Penguilly J, Le Reste JY, Nabbe P, Barrais M, Le Floch B. Les débuts et les modalités d'exercice des étudiants de la faculté de Brest sont-ils influencés par le SASPAS ? exercer 2011;95:21-4.
5. Le Reste JY, Le Lez N, Le Floch B, Barraine P, Cadier S, Nabbe P. Infections broncho pulmonaires nosocomiales en EHPAD. *Le suivi des recommandations suffit pour le traitement.* exercer 2011;95:25-9.
6. Cadier S, Le Reste JY, Barraine P, Chiron B, Barais M, Nabbe P, Le Floch B, Gut-Gobert C. Création d'une liste hiérarchisée d'objectifs par la méthode du groupe nominal. Exercer 2011 ;97 :80-4.
7. Le Reste JY, Millet C, Chiron B, Barais M, Cadier S, Barraine P, Nabbe P, Le Floch B. Comment les médecins généralistes ressentent-ils la place des médicaments antidémentiels dans la prise en charge de leurs patients atteints de la maladie d'Alzheimer ? exercer 2012;100:4-10.
8. Marzo C, Lygidakis C, Rigoni S, Nabbe P, Lazic D, Assenova R, Doerr C, Lingner H, CZachowski S, Munoz M, Argyradiou S, Claveria A, Hasaganic M, Le Floch B, Deriennic J, Van Marjwijk H, Van Royen P, Le Reste JY. Definisione della multimorbidita in MG : una revisione sistematica. M.D. Medicinae Doctor. 2013 ;1:32-4.
9. Derriennic J, Le Reste JY, Nabbe P, Lazic D, Stampar A, Radost A, LYgidakis C, Sowinska A, Vince JF, Czachowsky S, Van Royen P, Liétard C. Biopsychosocial ou psychosocial : quel terme choisir dans la définition de la multimorbidité en médecine générale ? Exercer, 2015, 118 : 68,73.

Original articles in English (14, 2 first author)

1. Cadier S, Hummers-Pradier E, Barais M, Barraine P, Chiron B, Le Floch B, Nabbe P, Le Reste JY. Audit about Medical Decision: Data Transmission Concerning Patients with

Dementia Entering French Nursing Homes Does Not Confirm the Diagnosis. International Journal of Family Medicine 2010 ; Article ID 857581.

2. Le Reste JY, Nabbe P, Lygidakis C, Doerr C, Lingner H, Czachowski S, Munoz M, Argyriadou S, Claveria A, Calvez A, Barais M, Lietard C, Van Royen P, Van Marwijk H. A research group from the European General Practice Research Network (EGPRN) explores the concept of Multimorbidity for further research into Long Term Care. Journal of the American Medical Director Association 2012;14(2):132-133.
3. Le Reste, JY, Nabbe, P., Manceau, B., Lygidakis, C., Doerr, C., Lingner, H., Czachowski, S., Munoz M, Argyriadou S, Claveria A, Bower P, Barais M, 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;*14*(5), 319-325.
4. Le Reste JY, Coppens M, Barais M, Nabbe P, Le Floch B, Chiron B, Dinant GJ, Berkhout C, Stolper E, Barraine P. The transculturality of "gut feelings". results from a french delphi consenus survey. Eur J Gen Pract. 2013 apr 16.
5. Le Reste JY, Chiron B, Le Floch B, Nabbe P, Barrais M, Mansourati J, Cadier S, Barraine P, Lietard C. There are considerable drawbacks to oral anticoagulant for monitoring patients at home which should lead family physicians to discuss alternative or enhanced solutions: a cross-sectional study. BMC Cardiovasc Disord. 2013 Sep 11;13:71. doi: 10.1186/1471-2261-13-71.
6. Czachowski S, Le Reste JY, Sowinska A, Nabbe P, Lazic D, Lygidakis C, Lingner H, Assenova R, Argyradiou S, Doerr C, Claveria A, Deriennic J, Munoz M, Van Marjwijk H, Lietard C, Van Royen P. The EGPRN new concept of multimorbidity in General Practice. The Polish Study. Probl Med Rodz 2014; 3(43):77-80.
7. Marzo C, Lygidakis C, Nabbe P, Lazic D, Assenova R, Doerr C, Lingner H, Czachowski S, Munoz M, Argyradiou S, Claveria A, Hasaganic M, Le Floch B, Deriennic J, Van Marwijk H,Van Royen P, Le Reste, J. Y. Definizione della multimorbidità in MG : una revisione sistematica. *Medicinae Doctor*, 2014;1,32–34.
8. Kasuba Lazic D, Le Reste JY, Murgic L, Petricek G, Katic M, Ozvacic-Adzic Z, Cerovecki Nekic V, Nabbe P, Hasanagic M, Assenova R, Lygidakis H, Lingner H, Doerr C, Czachowski S, Sowinska A, Le Floch B, Munoz M, Argyriadou S, Van Marwijk H, Lietard C, Van Royen P. Say it in Croatian – Croatian translation of the EGPRN definition of Multimorbidity using a Delphi consensus technique. Collegium Antropologicum. 2014;38(3):1027-1032.
9. Le Reste JY, Nabbe P, Rivet C, Lygidakis C, Doerr C, Czachowski S, Lingner H, Argyriadou S, Lazic D, Assenova R, Hasaganic M, Munoz M, Thulesius H, Le Floch B, Derriennic J, Sowinska A, Van Marwijk H, Lietard C, Van Royen P. The European General Practice Research Network presents the translations of its comprehensive definition of Multimorbidity in Family Medicine in ten European languages. PLoS One. 2015 Jan 21;10(1):e0115796. doi: 10.1371/journal.pone.0115796. eCollection 2015.

10. Chambe J, Le Reste J-Y, Maisonneuve H, Sanselme A-E, 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(4):474-9
11. Le Reste JY, Nabbe P, Lingner H, Kasuba Lazic D, Assenova R, Munoz M, Sowinska A, Lygidakis C, Doerr C, Czachowski S, Argyriadou S, Valderas J, Le Floch B, Deriennic J, Jan T, Melot E, Barraine P, Odorico M, Lietard C, Van Royen P, Van Marwijk H. What research agenda could be generated from the European General Practice Research Network concept of Multimorbidity in Family Practice? (A nominal group survey). *BMC Family Practice* 2015;16:125.
12. Le Reste JY, Nabbe P, Lazic D, Assenova R, Lingner H, Czachowski S, Argyriadou S, Sowinska A, Lygidakis C, Doerr C, Claveria A, Le Floch B, Derriennic J, Odorico M, Van Marwijk H, Van Royen P. How do General Practitioners recognize the definition of Multimorbidity? A European qualitative study. Eur J Gen Prac. in print
13. Nabbe P, Le Reste JY, Guillou-Landreat M, Munoz Perez MA, Argyriadou S, Claveria A, Fernández 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. 2017 Jan;39:99-105.
14. P. Nabbe, J. Y. Le Reste, M. Guillou-Landreat, E. Beck-Robert, R. Assenova, D. Lazic, S. Czachowski, S. Stojanović-Špehar, M. Hasanagic, H. Lingner, A. Clavería, M. I. Fernandez San Martin, A. Sowinska, S. Argyriadou, C. Lygidakis, B. Le Floch, C. Doerr, T. Montier, H. Van Marwijk, P. Van Royen. One consensual depression diagnosis tool to serve many countries: a challenge! A RAND/UCLA methodology. BMC Res Notes. 2018; 11: 4. Published online 2018 Jan 3. doi: 10.1186/s13104-017-3111-x

Oral communications in national meetings (53)

1. Cadier S, Le Roux V, Nabbe P, Le Reste JY. Description de la population de l'hôpital local de Lanmeur étude démographique et de l'activité médicale. Poster 2ème congrès de la médecine générale Lyon, 12-14 juin 2008.
2. Bail P, Barraine P, Le Floch B, Cadier S, Le Reste JY, Nabbe P. Comment faire face aux contraintes du DES de MG en conservant le cadre pédagogique du tutorat ? Réflexions au terme de 10 ans de mise en place du tutorat à Brest. Communication orale au congrès du CNGE Angers Novembre 2008
3. Cadier S, Barraine P, Nabbe P, Gut-Gobert C, Leroyer C, le Reste JY. Etude SPIFP (Spirometry in Family Practice). Priorisation des objectifs par la Technique du groupe nominal. Communication orale 3ème congrès de la médecine générale Nice, juin 2009.
4. Keruzoré B, Cadier S, Barraine P, Nabbe P, Le Floch B, Nowak E, Le Reste JY. Déterminants non cliniques de prescription d'antibiotiques dans la rhinopharyngite de l'enfant

et concordance de ces déterminants avec les pratiques de la population des médecins généralistes Finistériens. Communication orale 3ème congrès de la médecine générale Nice, juin 2009.

5. Coppens M, Cadier S, Le Floch B, Nabbe P, Barraine P, Chiron B, Jaubert C, Barais M, Le Reste JY. Sixième sens en médecine générale. Appropriation des critères néerlandais par les enseignants associés de médecine générale. Communication orale au congrès du CNGE, Toulouse, Novembre 2009.
6. Nabbe P, Le Floch B, Barraine P, Créach P, Gasnier AL, Cadier S, Barais M, Chiron B, Le Reste JY. Visites sur site des services universitaires de médecine générale ambulatoires (SUMGA) de Bretagne occidentale en 2009. L'expérience Brestoise. Communication orale au congrès du CNGE, Toulouse, Novembre 2009.
7. Barraine P, Nabbe P, Le Floch B, Chiron B, Jaubert C, Barais M, Cadier S, Le Reste JY. Cursus recherche à Brest : une formation théorique au service de la recherche en soins primaires. Communication orale au congrès du CNGE, Toulouse, Novembre 2009.
8. Le Floch B, Barraine P, Nabbe P, Chiron B, Barais M, Cadier S, Le Reste JY. Stage d'externat utilisation de projets de recherche courts. Communication orale au congrès du CNGE, Toulouse, Novembre 2009.
9. Barais M, Chiron B, Cadier S, Barraine P, Nabbe P, Le Floch B, Le Reste JY. Ejaculation prématurée : stratégies pour aborder le sujet en médecine générale. Communication orale congrès de la Médecine Générale, Nice, France. Juin 2010.
10. Barais M, Breuilly-Leveau C, Chiron B, Cadier S, Barraine P, Nabbe P, Le Floch B, Le Reste JY. Contraception de l'adolescente : pourquoi les généralistes loupent le coche ? Communication orale au congrès de la Médecine Générale, Nice, France. Juin 2010.
11. Nabbe P, Le Reste JY, Cadier S, Barraine P, Le Floch B, Chiron B, Barais M. Gestion de l'oubli de pilule par les internes Brestois de médecine générale. Communication orale au congrès de la Médecine Générale, Nice, France. Juin 2010.
12. Le Floch B, Le Floch PY, Lemey C, Maguet J, Letty E, Gestin Y, Sarni N, Nabbe P, Le Reste JY. Alcolado. Communication orale au congrès du CNGE, Rouen, Novembre 2010.
13. Freche B, Penguilly J, Le Floch B, Barais M, Barraine P, Chiron B, Nabbe P, Cadier S, Le Reste JY. Les débuts et les modalités d'exercice des étudiants issus de la faculté de Brest sont-ils influencés par leur SASPAS ? Communication orale au congrès du CNGE, Rouen, Novembre 2010.
14. Chiron B, Le Reste JY, Barais M, Barraine P, Le Floch B, Nabbe P. Peut-on se fier aux INR prélevés à domicile ? Communication orale au congrès inter régional du grand ouest, Tours, janvier 2011.
15. Le Floch B, Zacharewicz B, Barba D, Nabbe P, Barraine P, Chiron B, Barais M, Cadier S, Le Reste JY. Déterminants de la prise d'opiacées chez les marins pêcheurs. Communication orale au 5ème congrès de la médecine générale, Nice, juin 2011.

16. Le Floch B, Pérès F, Barba D, Nabbe P, Barraine P, Chiron B, Barais M, Cadier S, Le Reste JY. Le déficit en vitamine D : Prévalence de l'insuffisance et de la carence en vitamine D. Communication orale au 5ème congrès de la médecine générale, Nice, juin 2011.
17. Le Reste JY, Martin JC, Chiron B, Nabbe P, Le Floch B, Collet Y, Senecail B. Echographie en Médecine Générale. Communication orale au 5ème congrès de la médecine générale, Nice, juin 2011.
18. P Nabbe, JY Le Reste, Chiron B, Le Prielec A, B Manceau, H Van Marwijk, P Van Royen. Etude FPDM (Dépression et Multimorbidité en médecine de famille), projet de revue systématique de la littérature : quels sont les outils utilisés pour diagnostiquer la dépression en médecine générale ? Communication orale au 5ème congrès de la médecine générale, Nice, juin 2011.
19. Le Reste JY, Nabbe P, Manceau B, Chiron B, Munoz MA, Claveria A, Fonsecca C, Argyradiou S, Lygidakis C, Czachowsky S, Mueller C, Hummers Pradier E, Van Marwijk H, Van Royen P. The FPDM (Family Practice Depression and Multimorbidity) Study: revue systématique de la littérature pour trouver les critères de définition de la multimorbidité. Communication orale au 5ème congrès de la médecine générale, Nice, juin 2011.
20. J.Y. Le Reste, P. Nabbe, H. Lingner, C. Lygidakis, B. Chiron, M.A. Munoz, S. Argyradiou, S. Czachowski , A. Calvez, C. Berkhout, H. Van Marwijk, P. Van Royen, C Lietard. The FPDM (Family Practice Depression and Multimorbidity) Study : définition exhaustive de la multimorbidité. Communication orale au congrès du CNGE, Bordeaux, Novembre 2011.
21. P. Nabbe, J.Y. Le Reste, H. Lingner, C. Lygidakis, B. Chiron, M.A. Munoz, S. Argyradiou, S. Czachowski , A. Calvez, H. Van Marwijk, P. Van Royen, C Lietard. Etude FPDM (Dépression et Multimorbidité en médecine de famille) : Revue systématique de la littérature : quels sont les outils utilisés pour diagnostiquer la dépression en Médecine Générale. Communication orale au congrès du CNGE, Bordeaux, Novembre 2011.
22. Le Floch B, L'Echelard S, Bovay F, Chiron B, Calvez A, Barais M, Nabbe P, Barraine P, Le Reste JY. Lietard C. Quels facteurs positifs déterminent la satisfaction chez les médecins généralistes installés ? Enquête qualitative par focus groupes. Communication orale au congrès du CNGE, Bordeaux, Novembre 2011.
23. Le Floch B, Granja V, Chiron B, Barais M, Calvez A, Nabbe P, Barraine P, Le Reste JY. Lietard C. Quelle vision positive des étudiants de second cycle ont-ils sur la médecine générale ? Communication orale au congrès du CNGE, Bordeaux, Novembre 2011.
24. J.Y. Le Reste, P. Nabbe, C. Lygidakis, C. Doerr, M.A. Munoz, S. Argyradiou, S. Czachowski , A. Calvez, C. Berkhout, H. Van Marwijk, P. Van Royen, C Lietard. The FPDM (Family Practice Depression and Multimorbidity) Study: définition académique de la multimorbidité. Communication orale au congrès inter régional du grand ouest, Tours, janvier 2012.
25. Robert E, Nabbe P, Le Prielec A, Czachowski S, Doer C, Argyriadou S, Chiron B, Claveria A, Fernandez San Martin MI, Lingnier H, Lygidakis C, Munoz Perez MA, Van Marjwick H, Le Reste JY, Van Royen P, Liétard C. Etude FPDM (Dépression et Multimorbidité en médecine de famille) : Revue systématique de la littérature : quels sont les outils validés utilisés pour

diagnostiquer et dépister la dépression en médecine générale ? Communication orale au congrès inter régional du grand ouest, Tours, janvier 2012.

26. Calvez A, Barais M, Le Reste JY, Le Floc'h B, Nabbe P, Chiron B, Barraine P. Place du sens de l'alarme de « gut feeling » chez les médecins généralistes de terrain en situation d'urgence. Communication orale au congrès inter régional du grand ouest, Tours, janvier 2012.
27. Le Floch B, Granja V, Chiron B, Barais M, Calvez A, Nabbe P, Barraine P, Le Reste JY. Lietard C. Quelle vision positive les étudiants de second cycle ont-ils sur la médecine générale ? Communication orale au congrès inter régional du grand ouest, Tours, février 2012.
28. Le Floch B, Helou J, Cadier S, Barais M, Barraine P, Chiron B, Calvez E, Nabbe P, Le Reste JY. Pratique des médecins généralistes du dépistage et du diagnostic de l'hypovitaminose D. Présentation d'un poster au congrès inter régional du grand ouest, Tours, février 2012.
29. Bovay F, Le Floch B, Le Reste JY, L'Echelard S, Liétard C, Nabbe P. Satisfaction chez les médecins généralistes installés. Enquête qualitative par entretiens semi-dirigés. Communication orale au congrès inter régional du grand ouest, Tours, février 2012.
30. Argyriadou S, Lygidakis C, Lygera A, Le Reste JY, Nabbe P, Manceau B, Chiron B, Munoz MA, Claveria A, Fonsecca C, Czachowsky S, Mueller C, Hummers Pradier E, Van Marwijk H, Van Royen P. Searching criteria for multimorbidity in General Practice, by using systematic review. Panellenic meeting of General Practitioners, Greece, 2012.
31. Nabbe P, Robert E, Le Prielec A, Czachowski S, Doer C, Argyriadou S, Chiron B, Claveria A, Fernandez San Martin MI, Lingnier H, Lygidakis C, Munoz Perez MA, Van Marjwick H, Le Reste JY, Van Royen P, Liétard C. Etude FPDM (Dépression et Multimorbidité en médecine de famille) : Revue systématique de la littérature : quels sont les outils validés utilisés pour diagnostiquer et dépister la dépression en médecine générale ? Communication orale au 6ème congrès de la médecine générale, Nice, juin 2012.
32. J.Y. Le Reste, P. Nabbe, C. Lygidakis, C. Doerr, M.A. Munoz, S. Argyradiou, S. Czachowski , A. Calvez, C. Berkhout, H. Van Marwijk, P. Van Royen, C Lietard. The FPDM (Family Practice Depression and Multimorbidity) Study: définition académique de la multimorbidité. Communication orale au 6ème congrès de la médecine générale, Nice, juin 2012.
33. J.Y. Le Reste, B. Bodin, P. Nabbe, A. Calvez, B. Le Floch, M. Barais, C. Lietard. Critères de définition de la multi morbidité en médecine générale : Etude qualitative par entretiens individuels semi-dirigés avec les maîtres de stage des universités brestois. Etude FPDM (Family Practice Depresion and Multimorbidity). Communication orale au 6ème congrès de la médecine générale, Nice, juin 2012.
34. K. Mignotte, J.Y. Le Reste, B. Bodin, P. Nabbe, A. Calvez, B. Le Floch, M. Critères de définition de la multi morbidité en médecine générale : Etude qualitative par entretiens individuels semi-dirigés avec les maîtres de stage des universités brestois assortis d'une analyse cognitivo discursive. Etude FPDM (Family Practice Depresion and Multimorbidity). Communication orale au 6ème congrès de la médecine générale, Nice, juin 2012.

35. Nabbe P, Le Reste JY, Robert E, Guine F, Calvez A, Czachowski S, Doerr C, Lygidakis C, Argyriadou D, San Martin Fernandez MI, Lingnier H, Munoz Perez MA, Claveria A, Van Marjwick H, Liétard C et Van Royen P. Étude FPDM (dépression et multimorbidité en médecine familiale): Design d'un projet de consensus européen pour s'accorder sur un outil unique et validé pour diagnostiquer la dépression en médecine générale, Communication orale au congres du CNGE, Lyon, Novembre 2012
36. Le Reste JY, Nabbe P, Bosseur K, Calvez A, Barais M, Le Floch B, Chiron B, Barraine P, Van Marjwick H, Van Royen P, Liétard C. FPDM: (Family Practice Depression and multimorbidity): Critères de définition de la multimorbidité. Communication orale au congrès du CNGE, Lyon, Novembre 2012.
37. Le Floch B, Montfort A, Chiron B, Barais M, Calvez A, Nabbe P, Barraine P, Le Reste JY. Lietard C. Satisfaction des internes de Médecine Générale Etude qualitative par entretiens semi-dirigés. Communication orale au congrès du CNGE, Lyon, Novembre 2012.
38. Le Reste JY, Nabbe P, Le Pabic A, Calvez A, Barais M, Barraine P, le Floch B, Van Marjwick H, Liétard C and Van Royen P. FPDM: (Family Practice Depression and multimorbidity): traduction aller retour de la définition de la multimorbidité en français. Communication orale au congrès du CNGE, Lyon, Novembre 2012.
39. Le Reste JY, Nabbe P, Bosseur K, Calvez A, Barais M, Le Floch B, Chiron B, Barraine P, Van Marjwick H, Van Royen P, Liétard C. FPDM: (Family Practice Depression and multimorbidity): Multimorbidité, comment est compris le terme condition par les médecins généralistes ? Communication orale au congrès inter régional du grand ouest, Tours, février 2013.
40. Nabbe P, Le Reste JY, Guiné F, Beck-Robert Emilie, Mercier A, Calvez A, Freche B, Huas C, Laporte C, Pouchain D, Huas D, Vaillant H, Auger-Aubin I, Cadwallader JS, Baumann-Coblenz L, Martinez L, Frappé P, Lebeau JP, Liétard C et Van Royen P. Étude FPDM (dépression et multimorbidité en médecine familiale): Un outil diagnostique de la dépression, validé et unique ; Un consensus. Une RAND/UCLA. Communication orale au congrès inter régional du grand ouest, Tours, février 2013.
41. Le Reste JY, Nabbe P, Leroy Franck, Lever Delphine, Bosseur K, Gouez S, Le Mao Y, Calvez A, Barais M, Le Floch B, Chiron B, Barraine P, Van Marjwick H, Van Royen P, Liétard C. Etude FPDM (Family Practice Depression and multimorbidity) : Multimorbidité, quelle compréhension du terme « condition médicale » ont les médecins généralistes ? Communication orale au 7ème congrès de la médecine générale, Nice, juin 2013.
42. Le Reste JY, Nabbe P, Le Pabic A, Calvez A, Barais M, Barraine P, le Floch B, Van Marjwick H, Van Royen P, Lietard C. Etude FPDM (Family Practice Depression and multimorbidity): traduction aller retour de la définition de la multimorbidité en français. Communication orale au 7ème congrès de la médecine générale, Nice, juin 2013.
43. Le Reste JY, Nabbe P, Czachowski S, Doerr C, Lygidakis C, Lazic D, Hasaganic M, Assenova R, Sowinska A, Argyriadou S, Lingner H, Calvez A, Van Marjwick H, Liétard C and Van Royen P. Etude FPDM (Family Practice Depression and multimorbidity): traduction aller retour de la définition de la multimorbidité en Allemand, Bosniaque, Bulgare, Croate,

Français, Grec, Italien et Polonais. . Communication orale au congrès du CNGE, Clermont Ferrand, Novembre 2013.

44. Le Reste JY, Nabbe P, Lygidakis C, Doer C, Czachowski S, Lazic D, Argyriadou S, Lingner H, Hasaganic M, Assenova R, Sowinska A, Deriennic J, Le Floch B, Van Marjwick H et Liétard C et Van Royen P. Traduction aller retour et homogénéité de la définition EGPRN du concept de multimorbidité en Europe. Communication orale au 8ème congrès de la médecine générale, Paris, Avril 2014.
45. Melot E, Le Reste JY, Nabbe P, Lygidakis C, Doer C, Czachowski S, Lazic D, Argyriadou S, Lingner H, Hasaganic M, Assenova R, Sowinska A, Deriennic J, Le Floch B, Van Marjwick H et Liétard C et Van Royen P. Une définition homogène de la multimorbidité pour la recherche à travers l'Europe. Présentation au congrès du CNGE, Lille, Novembre 2014.
46. Le Reste JY, Deriennic J, Nabbe P, Doer C, Argyriadou S, Lingner H, Lygidakis C, Czachowski S, Lazic D, Hasaganic M, Assenova R, Sowinska A, Deriennic J, Melot E, Le Floch B, Van Marjwick H, Liétard C, Van Royen P. Une définition homogène de la multimorbidité pour la recherche à travers l'Europe. Communication orale à la journée inter régional de Tours de recherche en médecine générale, Tours, juin 2015
47. Melot E, Degironde C, Martin JC, Derrienic J , Chiron B, Barais M, Nabbe P, Lefloch B, Querellou S, Bressolette L, Senecail B, Collet M, Le Reste JY. Indication de l'échographie en médecine générale en 2012, revue systématique de la littérature. Communication orale à la journée inter régional de Tours de recherche en médecine générale, Tours, juin 2015.
48. Le Floch B, Le Reste JY, Drevillon S, Gicquel A, Haerlé W, Derriennic J, Chiron B, Barais M, Barraine P, Nabbe P, Lietard C. Satisfaction des médecins généralistes : une revue systématique de la littérature. Communication orale à la journée inter régional de Tours de recherche en médecine générale, Tours, juin 2015
49. Nabbe P, Lancelot P, Le Reste JY, Le Floch B, Chiron B, Barais M, Derriennic J, Melot E, Liétard C. Consensus d'experts sur une traduction en français d'une echelle d'autoévaluation de la dépression, la « Hopkins Symptom Checklist 25 items » via une procédure Delphi et une traduction aller retour en anglais. Communication orale à la journée inter régional de Tours de recherche en médecine générale, Tours, juin 2015.
50. Nabbe P, Sowinska A, Melot E, Claveria A, Czachowski S, Doer C, Asenova R, Stojanovic-Spehar S, Hasanagic M, Lazic D, Lingnier H, Lygidakis C, Argyriadou S, Fernandez San Martin MI, Munoz Perez MA, Le Floch B, Chiron B, Derriennic J, Lancelot P, Van Marwijk H, Van Royen P, Liétard C, Le Reste JY. Consensus d'experts sur une traduction en français d'une échelle d'auto-évaluation anglaise de la dépression, la « Hopkins Symptom Checklist en 25 questions », via une traduction aller retour avec contrôle culturel, selon une procédure Delphi avec contrôle d'expert. CNGE, Dijon, Novembre 2015
51. Nabbe P, Le Reste JY, Odorico M, Gatineau F, Le Floch B, Le Goff D, Derriennic J, Melot E, Lalande S, Le Graet D, Guilcher E, Corvez H, Montier T, Van Marwijk, Van Royen P. FPDM (Family Practice Depression and Multimorbidity): Validation en Médecine Générale de la

version Française de la Hopkins Symptoms Check List-25 items (HSCL-25). CNGE, Grenoble, Novembre 2016

52. Nabbe P, Le Reste JY, Odorico M, Gatineau F, Le Floch B, Le Goff D, Derriennic J, Melot E, Lalande S, Le Graet D, Guilcher E, Corvez H, Montier T, Van Marwijk, Van Royen P. HSCL-25 Versus PSE-9 ; French Validation study. CMGF Mars 2017
53. Nabbe P, Le Reste JY, Guillou-Landreat M, Le Floch B, Montier T, Van Marwijk H, Van Royen P. What is the internal validation and dimensionality in the translation of HSCL-25 in French, in the diagnosis of depression in primary care? CNGE Montpellier Nov 2017

Oral communications in international meetings (39)

1. Cadier S, Cornec B, Nabbe N, Varnoux M, Bail P, Leroyer C, Le Reste JY, Oger E. Patients treated by anticoagulants wish more education. Communication orale au Congrès EGPRN (European General Practitioner Research Network) Budapest, Hongrie, Octobre 2008.
2. Chiron B, Claux F, Bensassi M, Cadier S, Nabbe P, Barais M, Barraine P, Le Floch B, Mansouratti J, Le Reste JY. INR out patient heads or tails? Congrès EGPRN (European General Practitioner Research Network) Plovdiv, Bulgarie, Mai 2010.
3. Nabbe P, Barraine P, Barrais M, Chiron B, Le Floch B, Le Reste JY. Missed pill, management by GP's trainees. Congrès EGPRN (European General Practitioner Research Network) Plovdiv, Bulgarie, Mai 2010.
4. Barais M, Cadier S, Barraine P, Chiron B, Le Floch B, Nabbe P, Le Reste JY. Premature ejaculation: strategies to tackle the topic in family practice. Congrès EGPRN (European General Practitioner Research Network) Plovdiv, Bulgarie, Mai 2010.
5. Le Floch B, Zacharewicz B, Barba D, Nabbe P, Barraine P, Chiron B, Barais M, Cadier S, Le Reste JY. Determinants of opiate intake among fishermen. Congrès EGPRN (European General Practitioner Research Network) Zurich, Suisse, Octobre 2010.
6. Barraine P, Calvez A, Barais M, Chiron B, Le Floch B, Nabbe P, Le Reste JY. Serious pathologies are always addressed to emergencies with serious criteria, aren't they ? Congrès EGPRN (European General Practitioner Research Network) Zurich, Suisse, Octobre 2010.

7. Barais M, Breuilly-Leveau C, Cadier S, Barraine P, Chiron B, Le Floch B, Nabbe P, Le ResteJY. Teenagers' contraception : how do French GPs miss it? Congrès EGPRN (European General Practitioner Research Network) Zurich, Suisse, Octobre 2010.
8. Lapprand H, Le Reste JY, Le Floch B, Nabbe P, Barraine P, Barais M, Chiron B, Cadier S Evaluation of GPs motyivation to learn motivational interviewing for obesity treatment. Congrès EGPRN (European General Practitioner Research Network) Zurich, Suisse, Octobre 2010.
9. Nabbe P, Le Reste JY, Van Marwijk H, Chiron B, Barais M, Cadier S, Le Floch B, Barraine P. The FPDM (family practice depression and Multimorbidity) Study: Project for multimorbidity and depression after 50. Back draft. Congrès EGPRN (European General Practitioner Research Network) Zurich, Suisse, Octobre 2010.
10. Le Reste JY, Nabbe P, Manceau B, Chiron B, Munoz MA, Claveria A, Fonsecca C, Argyradiou S, Ligydakis H, Czachowsky S, Mueller C, Hummers Pradier E, Van Marwijk H, Van Royen P. The FPDM (family practice depression and Multimorbidity) Study: Project for systematic review of literature to find criteria for multimorbidity definition. Congrès EGPRN (European General Practitioner Research Network) Nice, France, Mai 2011.
11. Nabbe P, Le Reste JY, Le Prielec A, Chiron B, Munoz MA, Claveria A, Fonsecca C, Argyradiou S, Ligydakis H, Czachowsky S, Mueller C, Hummers Pradier E, Van Marwijk H, Van Royen P. The FPDM (family practice depression and Multimorbidity) Study: Project for systematic review of literature to find tools for depression diagnosis in primary care. Congrès EGPRN (European General Practitioner Research Network) Nice, France, Mai 2011.
12. Le Reste JY, Martin JC, Chiron B, Nabbe P, Le Floch B, Collet Y, Senecail B. Echography in general practice a systematic literature review. Congrès Wonca Europe Warsaw Septembre 2011.
13. Le Reste JY, Nabbe P, Manceau B, Chiron B, Munoz MA, Claveria A, Fonsecca C, Argyradiou S, Ligydakis H, Czachowsky S, Lingner H, Hummers Pradier E, Van Marwijk H, Van Royen P. The FPDM (family practice depression and Multimorbidity) Study: systematic review of literature to find criteria for multimorbidity definition former results. Congrès Wonca Europe Warsaw Septembre 2011.

14. Le Reste JY, Nabbe P, Manceau B, Chiron B, Munoz MA, Claveria A, Fonsecca C, Argyradiou S, Ligydakis H, Czachowsky S, Lingner H, Hummers Pradier E, Van Marwijk H, Van Royen P. The FPDM (family practice depression and Multimorbidity) Study: systematic review of literature to find criteria for multimorbidity definition, results of axial coding. Congrès EGPRN Krakow Octobre 2011.
15. Calvez A, Le Reste JY, Le Lez N, Chiron B, Le Floch B, Nabbe P, Cadier S.Broncho pulmonary infections: following guideline is far enough for french nursing homes". Congrès EGPRN Krakow Octobre 2011.
16. Argyriadou S, Lygidakis C, Lygera A, Le Reste JY, Nabbe P, Manceau B, Chiron B, Munoz MA, Claveria A, Fonsecca C, Czachowsky S, Mueller C, Hummers Pradier E, Van Marwijk H, Van Royen P. ΑΝΑΖΗΤΩΝΤΑΣ ΤΑ ΚΡΙΤΗΡΙΑ ΓΙΑ ΤΟΝ ΟΡΙΣΜΟ ΤΗΣ ΠΟΛΥΝΟΣΗΡΟΤΗΤΑΣ ΣΤΗ ΓΕΝΙΚΗ ΙΑΤΡΙΚΗ ΔΙΑ ΜΕΣΟΥ ΤΗΣ ΣΥΣΤΗΜΑΤΙΚΗΣ ΒΙΒΛΙΟΓΡΑΦΙΚΗΣ ΑΝΑΣΚΟΠΙΣΗΣ : Searching criteria for multimorbidity in General Practice, by using systematic review Panhellenic meeting of general practitioners - ELEGEIA. Kastro Kyllinos, April 2012
17. Nabbe P, Le Reste JY Le Prielec A, Robert E, Czachowski S, Doer C, Lygidakis C, Argyriadou S, Chiron B, San Martin Fernandez MI, Lingnier H, Munoz Perez MA, Claveria A, Van Marjwick H, Liétard C, Van Royen P. Study FPDM (Depression and multimorbidity in family medicine): Systematic literature review: what validated tools are used to screen or diagnose depression in general practice? Congrès EGPRN Ljubljana May 2012.
18. J.Y. Le Reste, P. Nabbe, C. Lygidakis, C. Doerr, M.A. Munoz, S. Argyradiou, S. Czachowski , A. Calvez, B. Chiron, H. Van Marwijk, C Lietard, P. Van Royen. The FPDM (Family Practice Depression and Multimorbidity) Study: A pragmatic definition of Multimorbidity from scientific medical literature. Congrès EGPRN Ljubljana May 2012.
19. Calvez A, Bodin B, Le Reste JY, Nabbe P, Chiron B, Barais M, Le Floch B, Lietard C Qualitative approach of Multimorbidity by semi-structured interviews with French GPs. Congrès EGPRN Ljubljana May 2012.
20. Le Reste JY, Nabbe P, Manceau B, Chiron B, Munoz MA, Fernandez J, Fonsecca C, Argyradiou S, Lygidakis H, Czachowsky S, Linger H, Mueller C, Doerr C, Hummers Pradier E, Van Marwijk H, Van Royen P, C Lietard. The FPDM (Family Practice

Depression and Multimorbidity) Study: Systematic review of literature to find criteria for multimorbidity definition. Congrès Wonca Europe Vienna July 2012.

21. Nabbe P, Le Reste JY Le Prielec A, Robert E, Czachowski S, Doer C, Lygidakis C, Argyriadou S, Chiron B, San Martin Fernandez MI, Lingnier H, Munoz Perez MA, Claveria A, Van Marjwick H, Liétard C, Van Royen P. Study FPDM (Depression and multimorbidity in family medicine): Systematic review of the literature: what validated tools are used for depression diagnosis and screening in general practice? Congrès Wonca Europe Vienna July 2012.
22. Calvez A, Le Reste JY, Balez R, Chiron B, Amouroux R, Barraine P, Barais M, Le Floch B, Nabbe P, Lietard C. An assessment communication tool to improve physician/patient therapeutic alliance training in medical education. Communication orale au congrès EGPRN Antwerpen, October 2012. Eur J Gen Pract, March 2013, Vol. 19, No. 1 , 29-51
23. Le Reste JY, Nabbe P, Lygidakis C, Doer C, Czachowski S, Lazic D, Argyriadou S, Lingner H, Hasaganic M, Assenova R, Sowinska A, Le Floch B, Van Marjwick H and Liétard C and Van Royen P. What is the translation of Multimorbidity definition in Bosnian, Bulgarian, Croatian, French, German, Greek, Italian and Polish? Communication orale au congrès EGPRN Kusadasi, May 2013.
24. Le Reste JY, Nabbe P, Lygidakis C, Doer C, Czachowski S, Lazic D, Argyriadou S, Lingner H, Hasaganic M, Assenova R, Sowinska A, Le Floch B, Van Marjwick H and Liétard C and Van Royen P. FPDM: For an international definition of Multimorbidity in General Practice what lies behind the term “condition” for French and Polish GPs? Communication orale au congrès EGPRN Kusadasi, May 2013.
25. Nabbe P, Le Reste JY, Robert E, Czachowski S, Doer C, Asenova R, Stojanovic-Spehar S, Hasanagic M, Lazic D, Lingnier H, Lygidakis C, Argyriadou S, Claveria A, Fernandez San Martin MI, Munoz Perez MA, Van Marjwick H and Van Royen P and Liétard C. FPDM (Family Practice Depression and Multimorbidity): A European Consensus on a diagnostic depression tool in primary care. Communication orale au congrès EGPRN Kusadasi, May 2013.
26. Le Reste JY, Nabbe P, Lygidakis C, Doer C, Czachowski S, Lazic D, Argyriadou S, Lingner H, Hasaganic M, Assenova R, Sowinska A, Deriennic J, Le Floch B, Van Marjwick H and Liétard C and Van Royen P. EGPRN’s Multimorbidity definition

translation and homogeneity into 8 European languages. Communication orale au congrès EGPRN Malta, October 2013.

27. Nabbe P, Le Reste JY, Robert E, Czachowski S, Doer C, Asenova R, Stojanovic-Spehar S, Hasanagic M, Lazic D, Lingnier H, Lygidakis C, Argyriadou S, Claveria A, Fernandez San Martin MI, Munoz Perez MA, Van Marjwick H and Van Royen P and Liétard C. Depression: the search for a diagnostic tool in Primary Care which will enable collaborative, interdisciplinary, Europe-wide research within General Practice. Communication orale au congrès EGPRN Malta, October 2013.
28. Le Reste JY, Nabbe P, Doer C, Argyriadou S, Lingner H, Lygidakis C, Czachowski S, Lazic D, Hasaganic M, Assenova R, Sowinska A, Deriennic J, Le Floch B, Van Marjwick H and Liétard C and Van Royen P. Bosnian, Bulgarian, Croatian, French, German, Greek, Italian and Polish GPs do recognize the EGPRN definition of Multimorbidity. Communication orale au congrès EGPRN Barcelona, May 2014.
29. Le Reste JY, Nabbe P, Lygidakis C, Doer C, Czachowski S, Lazic D, Argyriadou S, Lingner H, Deriennic J, Hasaganic M, Assenova R, Sowinska A, Le Floch B, Van Marjwick H and Liétard C and Van Royen P. EGPRN's Multimorbidity definition translation and meta-ethnographic analysis into 8 European languages. Communication orale au congrès WONCA Lisboa, July 2014.
30. Le Reste JY, Nabbe P, Doer C, Argyriadou S, Lingner H, Lygidakis C, Czachowski S, Lazic D, Hasaganic M, Assenova R, Sowinska A, Deriennic J, Melot E, Le Floch B, Van Marjwick H and Liétard C and Van Royen P. Bosnian, Bulgarian, Croatian, French, German, Greek, Italian and Polish General Practitionners add the core competencies of General Practice to the EGPRN definition of Multimorbidity. Communication orale au congrès EGPRN Heraklion Crete, October 2014.
31. Le Reste JY, Nabbe P, Doer C, Argyriadou S, Lingner H, Lygidakis C, Czachowski S, Lazic D, Hasaganic M, Assenova R, Sowinska A, Deriennic J, Melot E, Le Floch B, Van Marjwick H and Liétard C and Van Royen P. European General Practitioners recognize the EGPRN definition of Multimorbidity in clinical practice. Communication orale au congrès EGPRN Heraklion Crete, October 2014.
32. Nabbe P, Le Reste JY, Le Floch B, Czachowski S, Doer C, Asenova R, Stojanovic-Spehar S, Hasanagic M, Lazic D, Lingnier H, Lygidakis C, Argyriadou S, Claveria A, Fernandez San Martin MI, Munoz Perez MA, Derrienic J, Melot E., Barais M., Van Marwijk H, Van Royen P et Liétard C. FPDM (Family Practice Depression and

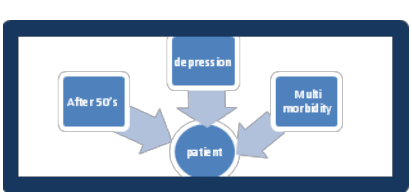
Multimorbidity): the Hopkins Symptoms Checklist-25 items (HSCL-25) translation in 10 European languages. Congrès EGPRN (European General Practitioner Research Network) Crête, October 2014.

33. Le Reste JY, Nabbe P, Doer C, Argyriadou S, Lingner H, Lygidakis C, Czachowski S, Lazic D, Hasaganic M, Assenova R, Sowinska A, Deriennic J, Melot E, Le Floch B, Van Marjwick H and Liétard C and Van Royen P. European General Practitioners recognize the EGPRN definition of Multimorbidity in clinical practice. Communication orale au congrès EGPRN Timisoara Romania , May 2015.
34. Melot E, Degironde C, Martin JC, Derriennic J , Chiron B, Barais M, Nabbe P, Lefloch B, Querellou S, Bressolette L, Senecail B, Collet M, Le Reste JY. Indications of ultrasonography in general practice, a systematic review. Présentation orale à l'EGPRN Timisoara Romania, May 2015.
35. Nabbe P, Le Reste JY, Le Floch B, Czachowski S, Doer C, Asenova R, Stojanovic-Spehar S, Hasanagic M, Lazic D, Lingnier H, Lygidakis C, Argyriadou S, Claveria A, Fernandez San Martin MI, Munoz Perez MA, Derrienic J, Melot E, Odorico M, Van Marwijk H, Van Royen P et Liétard C. FPDM (Family Practice Depression and Multimorbidity): the Hopkins Symptoms Checklist-25 items (HSCL-25), completed translation in 10 European languages. Communication orale au congrès EGPRN Edirne Turkey, October 2015.
36. Le Reste JY, Le Floch B, Nabbe P, Melot E, Derrienic J, Odorico M, Le Goff D, Letissier A, Claveria A, Lingner H, Czachowsky S, Sowinska A, Buczkowski K, Kasuba Lazic D, Hamulka D, Hoffman R, Petek D, Dexter D, Buono N, Thulesius A. A RAND UCLA procedure to select the best reliable tool to assess Therapeutic Alliance. (TATA STUDY). A pilot study. Communication orale au congrès EGPRN Edirne, Turkey, October 2015.
37. Nabbe P, Odorico M, Gatineau F, Le Floch B, Le Goff D, Derriennic J, Melot E, Morvan F, Lalande S, Nowak E, Le Graet D, Guilcher E, Corvez H, Barraine P, Barais M, Van Marwijk, Van Royen P, Le Reste JY. FPDM (Family Practice Depression and Multimorbidity): The French version of the Hopkins Symptoms Check List-25 items (HSCL-25), validation in general practice. Congres EGPRN, May, 2016
38. Nabbe P, Le Reste JY, Odorico M, Gatineau F, Le Floch B, Le Goff D, Derriennic J, Melot E, Lalande S, Le Graet D, Guilcher E, Corvez H, Montier T, Van Marwijk, Van Royen P. , FPDM (Family Practice Depression and Multimorbidity): The French version of the Hopkins Symptoms Check List-25 items (HSCL-25), validation in general practice. Leipzig, Octobre 2016

39. Nabbe P, Le Reste JY, Guillou-Landreat M, Le Floch B, Montier T, Van Marwijk H, Van Royen P. What is the internal validation and dimensionality in the translation of HSCL-25 in French, in the diagnosis of depression in primary care? EGPRN; Dublin Oct 2017

ANNEXE A

General Protocol



Project for multimorbidity and depression after 50.

The FPDM (Family Practice Depression and Multimorbidity Study): Depression line

What pragmatic and practical tool European GP's could use to diagnose depression in general practice, according to their validity, consistency and feasibility?

Introduction, Aim of FPDM:

To detect and diagnose depression in a risk population

Risk population: multimorbid patients over 50 years

Background:

Depression for people aged 55 or older is very frequent especially after a second factor of somatic co-morbidity. It is much more chronic than in people younger than 55, and more difficult to detect and acknowledge (2)(3)(4)(5) Gp's are the first port of call in most European countries, particularly for older subjects, but they seem to be less comfortable with actual tools for diagnosis and definition for those two diagnoses.(6)

Primary care patients seem also less comfortable with those diagnoses. As GPs try to provide personal, contextual and integrated care, this may explain their reluctance to use psychiatric labels such as depression especially in the context of a somatic illness, as they separate the patient from his or her mental state (7). However, it is not clear whether such reluctance is actually beneficial to patients (8). Not either if it is cost effective for the national health systems (9).

Multimorbidity is a new concept close to co-morbidity with a global vision in addition (10)(11). This concept is deeply in touch with the GP's core competencies as described by WONCA (12) and especially with the holistic modeling core competency. It could also be a help to detect frail patients (13) in primary care before decompensating. Therefore multiple definitions of multimorbidity are coexisting (14)(15) and the misunderstanding of the concept is obvious in literature (16)(17).

Treatment in primary care may consist first of preventive integrated care management for persons with depressive symptoms and multimorbidity without depression and second for actual treatment for depression, psychological or medical (i.e., antidepressants). Both are the subjects of further study.

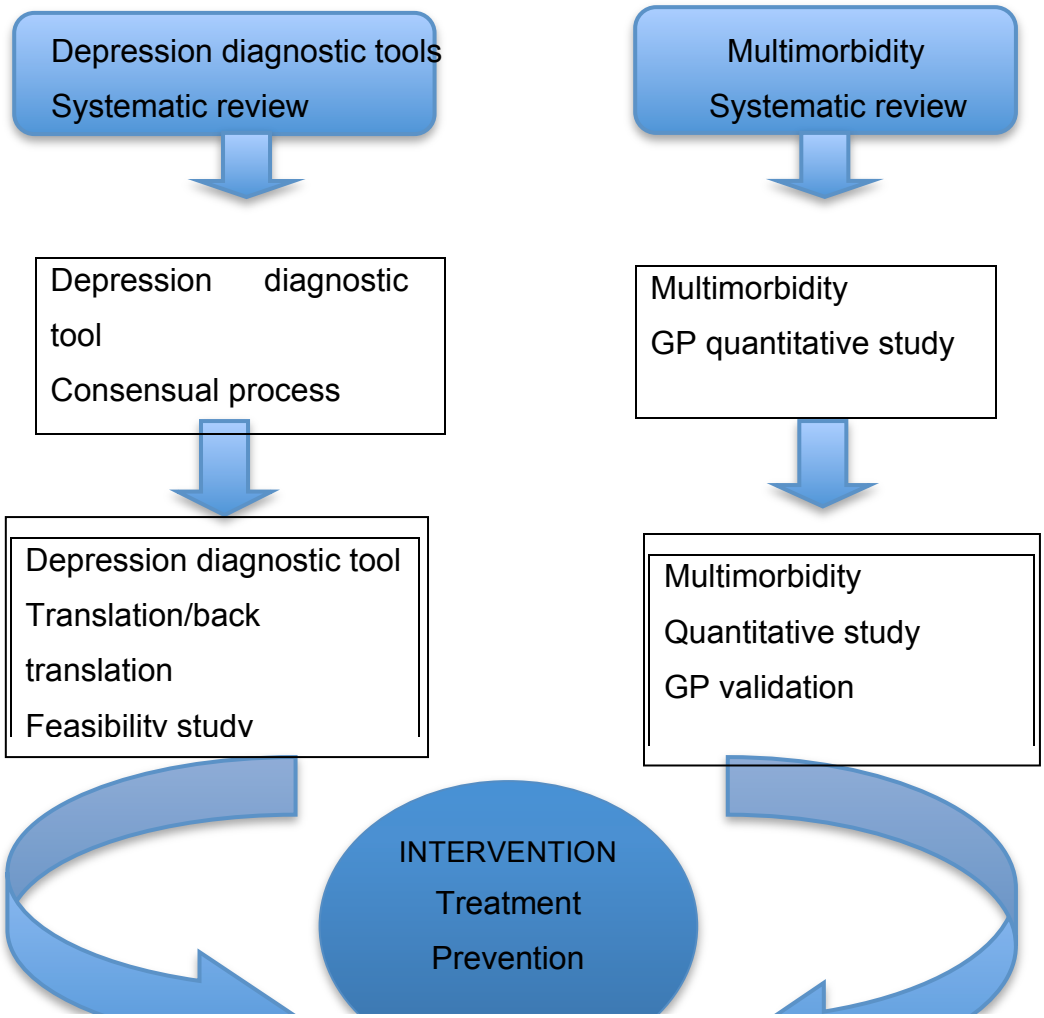
The European General Practitioner research network (EGPRN) is interested as a support and dissemination network in that study. He will support the study in each meeting by booking rooms and presentation devices (twice a year).

The primary care carer's community need a practical, stable and validated tool to diagnose depression translated into each European community language for further research and practice use and need a clear and academic definition about Multimorbidity.

Daily Board:

Jean Yves Le Reste, Harm Van Marwijk, Patrice Nabbe, Paul Van Royen, Claire Liétard

The global frame of FPDM



AIM OF THESIS, GLOBAL SUMMARY AND FRAMEWORK

Global Aim: To Find A Consensual Diagnostic Tool Of Depression, According Goal Of Fpdm

Research Question:

What Pragmatic And Practical Tool European Gp's Could Use To Diagnose Depression In General Practice, According To Their Validity, Consistency And Feasibility?

EGPRN members do this study across Europe with 11 EGPRN teams. Spain, Catalonia, Italia, Greece, Germany, Poland, Croatia, Bosnia, Belgium, Nederland and France are participating in this study.

BACKGROUND

Large and undeclared differences exist between incidence end prevalence rates of depression in general practice, probably related to conceptual differences and different objectives when diagnosing (3).

For FPDM, we need a clear knowledge about tools using to detect and to diagnose depression in adult patients (not only elderly patients because lots of tools can be used in unselected population) (4). Their tools should be validated, pragmatic, practical, reproducible and feasible and easy to use by European's Gps.

For FPDM, we have to choose a single diagnostic tool of depression, common to all members of FPDM, in order to reduce the selection bias or diagnostic imperfection.

A formal consensus procedure is necessary to find such tool. This research is based on the conceptualized process of consensus: RAND/UCLA. The RAND/UCLA method (RAM) was developed in the mid 1980's as an instrument to enable the measurement of the overuse and underuse of medical and surgical procedures. RAM is considered as a peculiar modified Delphi method. RAM includes an experts meeting with the advantages of the panel discussion over a method by independent reflection. (18)(19)(20).

RAM FRAMEWORK:

Review of literature

In order to increase the level of tools selection, a systematic review had been preferred to a simple review. It will be conduct to find the existing validated tools used to detect and to diagnose depression in adult patients. Then a tool will be selected according to its qualities of validity, reproducibility. Systematic review will met the quality criteria according the PRISMA checklist (21).

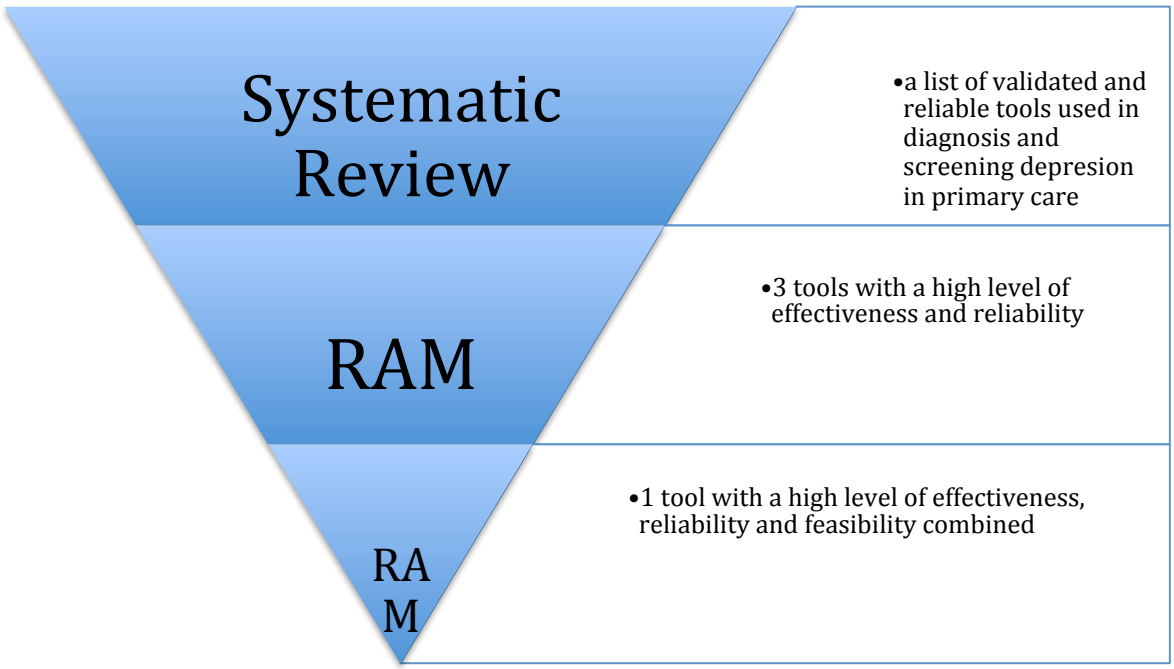
Rand Ucla Method (RAM): a Delphi procedure modified: select a single tool with the highest levels of efficacy, reliability and feasibility

This will be done using a Delphi procedure between European partners, in order to rank tools by level of efficacy and reliability. We will retain the first 3 tools with the highest level of validity and reliability for the experts.
During the panel meeting experts in pairs will test the three remaining tools. Then the comparative feasibility of these tools will be discussed in groups.
Then a Delphi procedure will be done in order to rank these 3 tools according to their feasibility.
At the end of this final round, one tool will remain and the RAM is over.

Panel meeting and Delphi procedure: to select a single tool according the better feasibility

Then the choosed tool will be translated into the languages of the European partners. The quality and objectivity of the translation will be validated with a forward backward translation methodology using a Delphi procedure for each participating European teams. Finally, the selected tool will be tested in practice in the field of real general practitioners practice.
By extension, this tool selected with a systematic review of literature, chosen by European primary care researchers and finally tested in practice, will be of interest for all GPs throughout Europe**.**

GLOBAL FRAME OF FPDM DEPRESSION LINE: RAND/UCLA PROCESS



AND AFTER RAM

Delphi Procedure To Translate Tool

Forward/backward translation in order to translate without loss of meaning the selected tool

Tool Testing In The Field

Test translated tool in each teams' language in order to define the real feasibility in practice.

Materials and design

What we need

We need a multinational team for each step (richness and triangulation about systematic review, robust validation for Delphi procedure):

Systematic review

Abstract's and article's selection

Data mining in articles

Delphi procedure and panel meeting

Criteria's selection in order to rank tools

Delphi procedure

Consensus on the translation of the tools in each language

What training for our PhD

Every student in PhD will have to follow the research courses in primary health care of Antwerp University and SICMA doctoral tools of Brest. They will have to complete the qualitative and quantitative courses of those universities. Our design is large enough to need qualitative and quantitative designs and will be an excellent training for the PhD students. At the end of the study all of them will be expert researchers in both qualitative and quantitative research. They will find opportunities to go on with this topic, as it is a very relevant subject for primary health care patients.

Design for each step

Meeting and check point at every EGPRN meeting as a pre conference workshop. Every team presents its work, troubles and questions during this session (one complete day, at least half day). Each national team presentation is in English, must be followed by a word document arguing the presentation. Presentation must be held in 15 minutes, questions for 15 minutes. Selection of trouble points for discussion and solution with all groups if needed.

FIRST STEP: REVIEWS: A SYSTEMATIC REVIEW

we have different questions:
What tools use GP's to diagnose depression in adult and elder people?
How and where those tools are used or in which situation are they used (Screening, diagnosis)?
What are the validity, reproducibility, and feasibility of each tool?
How to choose the most adapted tool easily to use in general practice in Europe?

The research question

To answer those questions the best research question seems to be:
What practical tool European Gp's could use to diagnose depression in general practice, according to their validity, consistency and feasibility?

Aim of the study

A methodical systematic review: proposition of method

A base of recruitment of abstract large: to keep all abstracts in IMRAD format, talking about tools used in diagnosis and detection depression, in adult and elderly population, in primary care
But a result of selection of abstract narrow: to keep only abstract with numerical data about validity of tools
We feel secure not to forget studies and to keep the right ones.

Quality criterias of the review
The systematic review should be consistent with quality standards PRISMA. At the end of the review, it must be possible to apply the checklist PRIMA to all results (21)

A list of Key words:
Depression definition or depression criteria or depression diagnosis or depressive disorders or depressive syndromes
Tools or scales or questionnaires or screening

Primary care or family practice or general, practice
Data bases used:
PubMed, Embase, Cochrane

Equation (pubmed format):
("Depression"[mh]) AND ("Physicians, Family"[mh] OR "Primary Health Care"[mh] OR "Family Practice"[mh]) AND ("Tool"[mh] OR "Scale"[mh] OR "questionnaire"[mh] OR "Criteria"[mh] OR "screening"[mh] OR "Diagnosis"[mh]) AND (English[lang] OR French[lang] OR Spanish[lang] OR German[lang] OR Portuguese[lang] OR Polish[lang] OR Italian[lang] OR Dutch[lang])
Equations for Embase and Cochrane are in the format suitable for each database.
Equation takes into account all the keywords.
The limits of the research are:
Adults and older people (we have taken as the limit age limit PubMed, age 19 +)
Language: are selected as mother languages used in the EGPRN teams.

Abstracts selections

The French team will do the databases research for abstract. They will send to each national team its national language abstract plus its part of the English abstracts.
Each national team will do inclusion or exclusion of its abstracts. The French team will include or exclude all abstracts (in order to give more reliability).
Inclusion will be achieved through a table of inclusion by all teams. It will contain both inclusion and exclusion, but also information on the abstracts

Information criteria's in abstract:
Title
Author
Scoring
Inclusion criteria's of abstract:
Abstract: yes
Language: English, Dutch, German, Polish, Greek, Italian, Spanish, French, Portuguese

Depression as major topic (is in research question, and criteria or definition is in results or discussion)
Tools for depression diagnosis or screening are named (or scales or questionnaires)
Primary care
IMRAD format
Numeric value of validity (sensitivity, specificity, negative, …), reproducibility or feasibility (the goal of this systematic review is to find valid tools in depression)

Exclusion criteria's of abstract:
Abstract: no
Language: another language of language of inclusion
Depression isn't a major topic
Not tools used or named for depression diagnosis or screening
Not in primary care
Not in IMRAD format
The field of the Study is about child, post partum blues or pregnancy (don't forget the global project, is about multimorbidity and depression after 50, it is logical to reject the studies about post partum blues, pregnancy or child)
And no numeric value of validity, reproducibility or feasibility
Validate list of abstract is send back to the French team, which will compare them with their own finding.
In case of discordance, Paul Van Royen and Harm Van Marjwick will judge the final inclusion.

Articles selection

The validated list of abstracts will be cuted for each national team (its mother language validated abstract plus its part of English abstracts). The French team will do this work. Each national team will receive his mother language articles plus English articles. Each team will have to include or exclude the articles after reading according to the inclusion exclusion criteria.
The French team will include or exclude all articles, using the same criteria's.

Inclusion criteria's of articles:

Language: English, Dutch, German, Polish, Greek, Italian, Spanish, French, Portuguese

Depression as major topic (is in research question, and criteria or definition is in results or discussion); There is a clearly focused research question, an appropriate methodology, recruitment is representative or not, drop out (if cohort study) are followed, if a questionnaire is used it is a validated one, results are generalizable.

Tools for depression diagnosis or screening are named

The field of study is in Primary care

Researchers are GP's

Article in IMRAD format

Numeric value of validity against a reference standard (sensitivity, specificity, …), reproducibility or feasibility are in articles (the goal of this meta analysis is to find valid tools in depression)

<u>Exclusion criteria's of articles:</u>

Language: another language of language of inclusion

Depression isn't a major topic

Not tools used or named for depression diagnosis or screening

Not in primary care

No face-to face psychiatric examination

Researchers are not GP's

Not in IMRAD format

The field of the study is about child, post partum blues or pregnancy (don't forget the global project, is about multimorbidity and depression after 50, is logical to reject the studies talk about post partum blues, pregnancy or child)

And no numeric value of validity, reproducibility or feasibility

Validate list of articles is send back to the French team, which will compare them with their own finding.

In case of discordance, Paul Van Royen and Harm Van Marjwick will judge the final inclusion.

Data mining in each article

The French team does a first table of data mining. Each national team gives proposition in order to increase the quality of the extraction. A final result: the table of data mining is the synthesis of all propositions of each national team.
All data directly related to the research question must be extracted:
Name of tool
Editorial data (review, author, impact factor)
Data about the sample of the study (size, population, situation of using, etc.)
Data about typology of the tool (origin, number of items,etc.)
Data about validity: reference test, Sensibility, Specificity
Data about efficacy: Youden index, Area under the curve
Data about reproducibility: cronbach's alpha
Data about feasibility (number of items, test time, etc.)

Method - instruction

Two local researchers in each team (national and French team) will do this inclusion/exclusion system separately.
Two local researchers in each team (national and French team) will do the data mining at the same time separately.
Patrice Nabbe will collate the two inclusion lists separately (the French complete list versus each national list).
They will agree at the end on their final inclusion and extraction list. In case of persistent disagreement they will send the final list plus the troubleshooting articles to Patrice Nabbe who give a consensual judgment.
They will agree at the end on their final inclusion list. In case of persistent disagreement they will send the final lists plus the troubleshooting articles to Paul Van Royen and Harm Van Marjwick who will judge the disagreement for final agreement.
At the end of this process we will have a new article lists that will be send to each national team according to the same partition (for each team all is mother language article plus a ninth of the English included articles).

Tools selection

All tools with test of validity compared to other tools have been extracted from the review of literature. But only the tools that are compared to a “gold standard” will be retained.
Are considered as “gold standard”, the tools traditionally used reference test, because of their seniority, their wide dissemination and acceptance as the gold standard by the international community.

expected results: identification of tools
As stated in the research question, the goal is to find valid instruments used in the diagnosis and screening for depression in primary care .
With this method, we hope to find and identify the validated tools used in that case, with an European agreement.

SECOND STEP: PROJECT WORK FOR A EUROPEAN CONSENSUS ON A VALIDATED TOOL AMONG THE TOOLS USED TO DIAGNOSE DEPRESSION IN FAMILY MEDICINE, BASED ON EFFECTIVENESS, RELIABILITY AND EASE TO USE. UTILIZATION OF A RAND/UCLA METHOD (RAM)

In the first part, a methodical systematic review of literature was completed. The result was the identification of validated tools versus “Gold Standard” .
All numerical values of validity and reliability were extracted from included articles. Values of efficacy were extracted or calculated. They are very different and done on different populations. Nevertheless they are comparable and could lead to a consensus.
Feasibility, which is the last important matter of quality, is far more complicated to ensure. We found many different ways to compare feasibility (number of items, test duration, time frame of question, method of inquiry and score range, etc.). But unlike efficacy and reliability those values are never the same and do not share a consensus.
A consensus methodology based on an expert panel looked as the only alternative to ensure comparison between those tools. They will first have to select the best tools on efficacy and reliability data (which are available with the result of the

systematic review). Then they will have to test and reflect about feasibility while comparing the tests.
The RAND/UCLA Appropriateness Method (RAM) is the most appropriate consensus method. It has been approved by NICE and HAS (Haute Autorité de Santé)(20)

Methodology

The RAND/UCLA Appropriateness Method (RAM) was developed in the mid 1980's as an instrument to enable the measurement of the overuse and underuse of medical and surgical procedures (22).The appropriateness criteria developed in early RAM studies were used as a tool to measure performance retrospectively (22). RAND/UCLA is considered as a peculiar modified Delphi method: it's a "two-round modified Delphi process" which includes an experts meeting between two rounds of independent ratings, based on the result of a literature review(22)(19).
Population:

The panel includes 7 to 15 voluntary EGPRN members from all over Europe. To be included an EGPRN member should speak English fluently, be registered as a researcher in a research team or having publications in a scientific journal with impact factor and be a GP.

First Round: Delphi round to assess efficacy and reliability

In the first round, each tool is rated using the data of efficacy and reliability from the systematic review. Each panelist receives by mail four sets of data:
Result of the literature review with data about validity (Sensibility, Specificity) and efficacy (Youden Index) for each tool.
Sample data (number, features, etc.)
The sources articles
Reliability data (Cronbach alpha)

Each set of data is shown in a synthetic view, in order to facilitate the rating work.

The experts do not receive the tools details in order not to be too subjective in that rating.

For each tool, experts will individually answer one question: does this tool seem effective and reliable enough to be appropriate for the diagnosis of depression in primary care? They will have to rate the efficacy and the reliability on a single global likert scale from 1 to 9 (1: highly not appropriate to 9: highly appropriate) using their own best clinical judgment with no interaction among experts (22). Panelists are encouraged to explain their opinion in quick notes for each tool.

The French team collects the individual results. All the tools are kept at the end of the first round.

Panel meeting

The aim of the second round is to disentangle feasibility on the tools. It takes place in a panel meeting during the EGPRN meeting in October 2012 (Antwerp, Belgium).

There is one moderator to handle the panel meeting and one observer to ensure the quality of the meeting. The panel moderator is a French member.

a/ select by panel meeting the three most effective and reliable tools

Experts receive an individualized document showing the distribution of all the experts' first round rating and comments, with their own specific rating.(19)

An ultimate discussion is held based on the individual results from the first round (22).The aim is to confirm the three most effective and reliable tools.

b/ rating by feasibility

The panelists receive the complete test forms of the three selected tools with details about feasibility of each tool (number of question, duration, etc.). They work in pairs (selected ~~in~~ a random way) and each one completes all the tests with his partner.

After testing, the panelists have to rate individually the selected tools. Each panelist has to answer one question for each test: does this test seem feasible in real GP's practice? They rate on a Likert scale from 1 (absolutely not feasible) to 9 (easily and completely feasible) with comments for each test.

Then, another discussion is held based on those individual results. The panelists are encouraged to discuss their rating for each tool in the light of their knowledge so

that? all the panelists rated (22). Then, the panelists have to re-rate individually the tools, regardless of whether their rating is unchanged from the beginning of the second round (22)

This discussion is audio-recorded and video taped.

No consensus is done at that time and the moderator has to promote the idea that no consensus has to be reached during the panel discussion

The results of the panel is a rating of each tool (median of its score) and the verbatim of the discussion(1)(22). Tools are classified into three levels of feasibility according to their medians:

Feasible: panel median of 7 to 9, without disagreement

Uncertain: panel median of 4 to 6 or 7 to 9 with disagreement

Not feasible: panel median of 1 to 3 without disagreement

Disagreement is a distribution of medians simultaneously with at least 30% of individual scores between 1 and 3 AND 30% of individual scores between 7 and 9 (19)(23)(24).

Third Round: Delphi round to Develop a consensus.

In this round, 15 days after the panel meeting, each panelist receives by mail the results of the meeting (i.e. the final rating of feasibility for each tool with the complete verbatim of the panel discussion) (1).

A classification of the three tools selected in the second round is proposed to each panelist.

The number 1 is the “most appropriate tool to diagnose depression in general practice”, the number 3 is the least appropriate” and the number 2 is for the intermediate tool. The results are collected for each expert.

The addition of scores for each tool aimed to determine the best (the lowest score is the best tool). This tool can be considered as the most effective, reliable, easy to use and practical tool of depression diagnosis in general practice.

If there is no clear consensus, another individual rating will be made.

This tool selected with a systematic review of literature, chosen by European primary care researchers and finally tested in practice, will be of interest for all GPs throughout Europe**.**

It will be submitted to the daily board of the study for final validation (Jean Yves Le Reste, Harm Van Marwijk, Patrice Nabbe, Paul Van Royen, Claire Lietard).

THIRD STEP: ADAPT EACH TOOL FOR EACH LANGUAGE

A Delphi procedure will be done, with forward/backward translation by each country.

The English version of the tool is proposed to each team by the French team. Delphi procedure will be done after the translation from English to native language to ensure its validity. Then the validated translated version in native language will be translated back to English and send to the French team to verify its homogeneity with the baseline English tool.

Translation should respect all the rules of forward/backward translation.

Research question (for each translation):

What is the translation of the diagnostic tool in our native language?

Research population: native expert GPs, English speakers, still in Gp practice and having teaching or research activities and not involved in the research.

Methodology: Forward backward translation using a Delphi consensus procedure. (25)(26)(27)(28)(29)(30)(31)

The forward translation will be done from English to native language by two translators (one medical and one official translator).

The Delphi consensus procedure will be held with 20 to 30 expert Gps. We will propose them the English definition and its translation into our native language. This proposition will be done using emails (each participant should be contacted separately to avoid contamination which is the basic methodology for Delphi procedure: so **no mailing list**). As many as needed Delphi round will be conducted to reach consensus.

Participants will rank translation from 1 (absolutely no agreement) to 9 (fully agreement). The participant should explain each rank under 7.

Consensus is defined as at least 70 % of the participants rating 7 or above the consensual definition.
With the consensual translation in native language two other native/English translators will do a backward translation from native language to English. It will be submitted to the daily board of the study for final validation (Jean Yves Le Reste, Harm Van Marwijk, Patrice Nabbe, Paul Van Royen, Claire Liétard).

FOURTH STEP AND FINAL STEP: TESTING IN THE FIELD

The chosen and translate tool will be tested in reality in practice.
Each team in the country will test the selected tool. The purpose will be to test the feasibility of tool in the field of consultation.
At this stage of the study, the procedure is not yet fully defined. It will be defined at a meeting of the FPDM working group.

TIME SCHEDULE
Review:
Agreement on abstract selection: May 2011
Agreement on article selection and data mining: November 2011
International redaction and submission: April 2012
Delphi procedure:
Tool ranking and ultimate choice of tool: October 2012
International redaction and submission: May 2013
Translation/back translation: November 2013
International redaction and submission: May 2014
Feasibility testing in practice of the selected tool: November 2014

PUBLICATION PLAN

Systematic review: one article proposed to BMJ
Delphi procedure, tool ranking: one article about methodology, one article about results
Delphi procedure, translation/back translation: one article per country

Feasibility testing of selected tool: one article per country

Publication rank:

Authorship credit is based only on substantial contribution to: conception and design, or analysis and interpretation of data drafting the article or revising it critically for important intellectual content and final approval of the version to be published (from BMJ criteria). Participation solely in the acquisition of funding or the collection of data does not justify authorship. The final decision rests with the daily board. The order of the authors depends further on the number of investigators and PhD student in every site.

Publication acceptations:

Each proposal for a publication with FPDM-depression data will be submitted to the daily board as an IMRAD abstract of 300 words. The daily board will check the proposal for overlap with other plans and potential combined or conflicting interests. These interests can concern the submitted publication plans or the use of data to which certain persons are explicitly involved. If there is no overlap or there are no conflicting interests, the proposal will, with a positive advise from the board, be accepted. When there is a possibility of overlap or when there are (possible) conflicting interests, the submitter of the proposal will be informed about this, with the request to adjust the plans.

The submission of a proposal implies concrete plans for a publication, etc. At which will be worked on a short time basis. When within half a year after submitting a proposal there is no provable activities in that direction, or when after a year no publication has been submitted to a journal, the subject can be released for other interested parties.

There is a maximum of two proposals that can be submitted as first author at the same time. Only after the presentation of the paper to a journal, a new proposal can be submitted.

Senior researchers who acquire extra funds have priority at the submission of publication plans on the theme of the extra fund.

Researchers who are not part of FPDM can, after consultation by the board taking advice with the most involved researchers, submit a proposal for data-analysis and publications. A senior researcher of FPDM will always be a member of this research group and will be co-author of the publications.

Authorship

The one, who “pulls” the article and has the most important role in writing it, is the first author. He/she is responsible for the contents of the article.
The first author determines in consultation with the board that the co-authors are and in which order. The board considers the investments and contributions of the FPDM members, like authorship of the original proposal and local coordination. When the first author is a PhD-student, this happens in consultation with the (co-) promoters. Possible conflicts will be put before the board.
(Co) promoter(s) who are primarily responsible for the supervision of PhD-students are (also) responsible for the integrity of the work as a whole, from inception to published article.
Other co-authors should have made an important contribution to the design of the study, data collection and/or writing of the paper. Consider the three conditions of the Requirements for Authorship sometimes an acknowledgement is more suitable then a co-authorship.
All authors get to inspect the article at least twice before it is forwarded to a magazine.
It is the responsibility of the first author that the guide-lines according authorship are followed, that the sample and the research-methodology are described correctly and that references to former relevant FPDM-publications are made. To anticipate carelessness in this, the board should see every manuscript before presented for publication.
The international FPDM team (named as “FPDM Team”) will always be in the author list of any publication as the last name of authors.

Overview of publications

An overview of all publications and publication plans, as well as publications in journals as chapters in books, are kept by the FPDM board.
Changes in publications and publication plans (as mentioned under a.), are to be communicated to the NESDA-secretary - preferably by e-mail.

The following information is needed:
Date of sending the manuscript and name of the journal, including possible changes
Changes in authors or title
Date of acceptance
At publication: full reference
Withdrawal (decision not to publish)

From the moment of presentation and after being inserted in the overview, all manuscripts are kept in an archive. As soon as a paper is send to a journal, the first author will send a (digital or paper) copy of this to the board. After publication a reprint of the final article will be send to the board.
Abstracts of congress papers should also be sending to the board.

Particular circumstances

If in publication plans FPDM data are used in combination with data collected by other research groups, the publication plan will be judged by both the FPDM board and the board of the other research group.

Everybody is free to publish is own national results (after board reviewing) but for the systematic reviews and for the final articles at each step all the national teams has to be named. For the reviews articles the one on depression should have Patrice Nabbe as first author (this is compulsory for the French team because he has a PhD on the topic.

Interested teams:

Miguel-Angel Munoz from Barcelona (Spain)

Ana Claveria From Vigo (Spain)

Jean Yves Le Reste and Patrice Nabbe from Brest

Stella Argyriadou from Greece

Harris Argyriadou from Italy

Harm Van Marwijk from Amsterdam

Eva Hummers Pradier from Germany

Slawomir Czachowsky from Poland

Djurdjica Lazic from Croatia

Melida Hasaganic from Bosnia

REFERENCE LIST

1. Bourrée F, Michel P, Salmi LR. Consensus methods: Review of original methods and their main alternatives used in public health. Revue d'Épidémiologie et de Santé Publique. 2008 Dec;56(6):e13–e21.
2. Licht-Strunk E, Windt DVD. The prognosis of depression in older patients in general practice and the community. A systematic review. Family Practice [Internet]. 2007 [cited 2011 Nov 6];24(December 2006):168–80. Available from: http://fampra.oxfordjournals.org/content/24/2/168.short
3. Licht-strunk E, Beekman ATF, Haan MD, Marwijk HWJV. The prognosis of undetected depression in older general practice patients . A one year follow-up study. Journal of Affective Disorders [Internet]. 2009;114(1-3):310–5. Available from: http://dx.doi.org/10.1016/j.jad.2008.06.006
4. Mitchell AJ, Vaze A, Rao S. Clinical diagnosis of depression in primary care: a meta-analysis. Lance [Internet]. 2009;374(9690):609–19. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19640579
5. 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 British Medical Journal [Internet]. 2009;338(7692):a3079. Available from: http://www.bmj.com/cgi/doi/10.1136/bmj.a3079
6. Lehti A, Hammarström A, Mattsson B. Recognition of depression in people of different cultures: a qualitative study. BMC Family Practice [Internet]. 2009;10:53. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2723088&tool=pmcentrez&rendertype=abstract
7. Wittkampf KA, Zwieten MV, Smits FT, Schene AH. Patients ' view on screening for depression in general practice. Canadian Journal Of Psychiatry. Revue Canadienne De Psychiatrie. 2008;(August):438–44.
8. Licht CMM, De Geus EJC, Seldenrijk A, Van Hout HPJ, Zitman FG, Van Dyck R, et al. Depression is associated with decreased blood pressure, but antidepressant use increases the risk for hypertension. Hypertension [Internet]. 2009;53(4):631–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19237679

9. Van't Veer-Tazelaar P, Smit F, Van Hout H, Van Oppen P, Van Der Horst H, Beekman A, et al. Cost-effectiveness of a stepped care intervention to prevent depression and anxiety in late life: randomised trial. [Internet]. 2010. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20357310

10. Starfield B. Global health, equity, and primary care. Journal of the American Board of Family Medicine JABFM [Internet]. 2007;20(6):511–3. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17954856

11. Boyd CM, Weiss CO, Halter J, Han KC, Ershler WB, Fried LP. Framework for evaluating disease severity measures in older adults with comorbidity. The journals of gerontology Series A Biological sciences and medical sciences [Internet]. 2007;62(3):286–95. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17389726

12. Europe W. THE EUROPEAN DEFINITION OF GENERAL PRACTICE / FAMILY MEDICINE. Europe [Internet]. 2002;54(2):149–56. Available from: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:The+European+definition+of+general+practice/family+medicine#0

13. Gobbens RJJ, Van Assen MALM, Luijkx KG, Wijnen-Sponselee MT, Schols JMGA. Determinants of frailty. Journal of the American Medical Directors Association [Internet]. 2010;11(5):356–64. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20511103

14. Van Den Akker M, Buntinx F, Metsemakers JF, Van Der Aa M, Knottnerus JA. Psychosocial patient characteristics and GP-registered chronic morbidity: a prospective study. Journal of Psychosomatic Research [Internet]. 2001;50(2):95–102. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11274666

15. Vandaele S. Morbidity, comorbidity et multimorbidity: polysémie et pléthore terminologique. PHARMATERM, Bulletin terminologique de l'industrie pharmaceutique. 2003;14(2):1–14.

16. Van Den Bussche H. Multimorbidité : état de la question. La revue de médecine générale. 2009;jan(259):31–2.

17. Fortin M, Lapointe L, Hudon C, Vanasse A. Multimorbidity is common to family practice: is it commonly researched? Canadian Family Physician [Internet]. 2005;51(2):245. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1472978&tool=pmcentrez&rendertype=abstract

18. Fitch K, Bernstein SJJ, Aguilar MDD, Burnand B, LaCalle JRR, Lazaro P, et al. The RAND / UCLA Appropriateness Method User ' s Manual [Internet]. Santa Monica: RAND; 2001. Available from: http://www.rand.org

19. Letrilliart L, Vanmeerbeek M. À la recherche du consensus : quelle méthode utiliser ? exercer. 2011;99(99):170–7.

20. Haute Autorité Santé. Bases méthodologiques pour l'élaboration de recommandations professionnelles par consensus formalisé [Internet]. 2006;Available from: http://www.has-sante.fr/portail/upload/docs/application/pdf/base_methodo_CFE.pdf

21. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. PLoS Medicine [Internet]. 2009;6(7):28. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2707010&tool=pmcentrez&rendertype=abstract

22. Fitch K, Bernstein SJ, Aguilar MD, Burnand B, LaCalle JR, Lazaro P, et al. The RAND/UCLA Appropriateness Method User's Manual. 2001.

23. McGory ML, Shekelle PG, Ko CY. Development of quality indicators for patients undergoing colorectal cancer surgery. Journal Of The National Cancer Institute. 2006;98(22):1623–33.

24. Shekelle PG, MacLean CH, Morton SC, Wenger NS. Assessing care of vulnerable elders: methods for developing quality indicators. Annals of Internal Medicine. 2001;135(8 Pt 2):647–52.

25. Jones J, Hunter D. Consensus methods for medical and health services research. Bmj Clinical Research Ed. [Internet]. 1995;311(7001):376–80. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2550437&tool=pmcentrez&rendertype=abstract

26. Nadeau L, Rousseau C, Séguin Y, Moreau N. [Preliminary qualitative evaluation of a shared-care mental health programme with youths in Montréal: facing institutional and cultural uncertainty]. Santé mentale au Québec. 2009 Jan;34(1):127–42.

27. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. Journal of Advanced Nursing [Internet]. 2000;32(4):1008–15. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11095242

28. Hastie R, Kameda T. The robust beauty of majority rules in group decisions. Psychological Review [Internet]. 2005;112(2):494–508. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15783295
29. Skulmoski GJ, Hartman FT, Krahn J. The Delphi Method for Graduate Research. Journal of Information Technology Education. 2007;6:2–21.
30. Novakowski N, Wellar B. Using the Delphi technique in normative planning research: methodological design considerations. Environment and Planning - Part A. 2008;40(6):1485–500.
31. De Villiers MR, De Villiers PJT, Kent AP. The Delphi technique in health sciences education research. Medical Teacher. 2005;27(7):639–43.

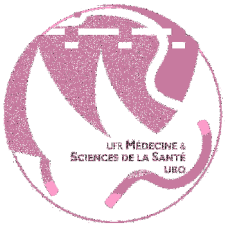
ANNEXE B

Project Work For A European Consensus On A Validated Tool Among The Tools Used To Diagnose Depression In Family Medicine, Based On Effectiveness, Reliability And Ease To Use.

FPDM Study: Project for multimorbidity and depression after 50. FPDM (Family Practice Depression and Multimorbidity Study), Depression Branch of FPDM

Object: FPDM Study, Project work for a European Consensus on a validated tool among the tools used to diagnose depression in family medicine, based on effectiveness, reliability and ease to use.

Patrice Nabbe – Jean Yves Le Reste



Département Universitaire de Médecine Générale

22, avenue Camille Desmoulins CS 93837 – 29238 – Brest CEDEX 3

Tél : 02 98 01 65 52 – fax : 02 98 01 64 74

Name and adress date, Brest

Madam, Sir,

The FPDM (Family Practice Depression and Multimorbidity) group is actually working on the validation of a tool used to diagnose depression in general practice in Europe, in patients over 50 years old, with at least two factors of multimorbidity. In this perspective, we appeal to you as an expert recognized for his competence and experience of general practice.

The FPDM group has already carried out a methodical systematic review of literature in order to collect the validated tools. These results will enable to sustain the research to obtain a consensus based on an expert panel.

Hereby letter, we thank you to participate in the second part of this project. The enclosed documents, ie the study's proposal and the results of the literature review, are aimed to give you informations about the research's methodology and the necessary data to answer individually to the first part of a RAND/UCLA Appropriateness Method (RAM).
We are asking you to fill out the rating forms and returns them within X DAYS, that is no later than XXXXXXXX.

The second round of the RAM will take place during the EGPRN meeting in October 2012 (Antwerp, Belgium). We assure you that all the collected informations will be confidential and will be exclusively used to the FPDM project.

Fifteen days after the Congress, you will receive the results of the meeting to answer to the third round. If no clear consensus is obtained after this round, you will be invited to participate in another individual rating.

We thank you for your interest in this project.
Yours faithfully,

signatures

BACKGROUND

Depression for people aged 55 or older is very frequent especially after a second factor of somatic co-morbidity. It is much more chronic than in people younger than 55, and more difficult to detect and acknowledge (1)(2)(3)(4). Gp's are the first port of call in most European countries, particularly for older subjects, but they seem to be less comfortable with actual tools for diagnosis and definition for those two diagnoses (5).

Depression occurs frequently but it may be difficult to detect and acknowledge in general medical settings. A major problem is that there is no objective test for the diagnosis. Incidence and prevalence rates of depression differ in general practice across Europe, probably related to conceptual differences and different objectives when diagnosing. There is also a large overlap between depression and contextual distress, anxiety and somatoform disorders in primary care. General practitioners (GPs) thus experience problems when detecting and diagnosing depression in their patients, which may lead to over prescription of antidepressant drugs. They are the first port of call in most European countries, but they seem to be less comfortable with actual tools for diagnosis and definition.

We need better knowledge about tools used by GPs or a primary care system in the field of depression in adult patients. We have the following research questions:
a) What tools do GPs actually use to diagnose depression in adult and elderly people?
b) What are the validity, reliability and effectiveness of each tool?
c) How to compare these tools?

These questions are the necessary steps to answer the final research question: "What diagnostic tool for depression, validated, reliable and easy to use, the European GPs could consensually use in general practice?"

STUDY DESIGN AND METHODOLOGY:

In the first part of the FPDM study, a methodical systematic review of literature with ten national teams of the EGPRN was completed. The result was the identification of these tools validated versus “a gold standard”.

Numerical values of validity, in terms of consistency and reliability were extracted from articles, in studies with different populations. This difference on populations does not make it possible to have a consensus based only on a strict comparison of quantitative numerical datas.

Feasibility, which is the last important matter of quality, is far more complicated to ensure. We found many different ways to compare feasibility (number of items, test duration, time frame of question, method of inquiry and score range, etc.). But unlike efficacy and reliability those values are never the same and do not share a consensus.

A Consensus based on an European expert panel looked as the only alternative to ensure comparison between those tools (6). They will first have to select the best tools on efficacy and reliability data (which are available with the result of the systematic review). Then they will have to test and reflect about feasibility while comparing the tests.

The RAND/UCLA Appropriateness Method (RAM) seems the more appropriate consensus method. It has been approved by National Healthcare Organizations (NICE and HAS in France) (7)(8).

METHODOLOGY

The RAND/UCLA Appropriateness Method (RAM) was developed in the mid 1980’s as an instrument to enable the measurement of the overuse and underuse of medical and surgical procedures (6).The appropriateness criteria developed in early RAM studies were used as a tool to measure performance retrospectively (6). RAND/UCLA is considered as a peculiar modified Delphi method: it’s a “two-round modified Delphi process” which includes an experts meeting between two rounds of independent ratings, based on the result of a literature review made in a first part(6)(9). It brings the advantages of a peer’s reflection and each panelist has equal weight in determining the final rating.

The RAM is only one of several methods that have been developed to identify the collective opinion of experts (6). In the RAM, repeated quantitative assessment is used by all the expert to encourage relevance, objectivity and homogeneity when ranking the proposal (8). The RAM has been shown to produce appropriateness criteria and quality indicators that have face, construct and predictive validity (10).
The Delphi rounds, thank to the anonymity of the responses, permits to avoid the influence of a leader opinion, and the panel meeting gives the panelists the opportunity to discuss their rating and judgments face to face evenly (6)(11).

ACTORS

THE STEERING GROUP

The objective of the steering group is:
To write the scientific project and recruit the scoring group
To ensure that the trading group has all the data for the evaluation procedure
To formulate questions for the Delphi round
To ensure that deadlines are met, possibly by raising the group scoring members

The steering group is made up of French members.

THE SCORING GROUP

The panel includes 9 to 15 voluntary EGPRN members from all over Europe(6). To be included an EGPRN member should speak English fluently, be registered as a researcher in a research team or having publications in a scientific journal with impact factor and be a GP (12).

A 9-member Rand panels is large enough to permit diversity of representation and a maximum of 15 members ensure that all a chance to participate the discussion(6).

Whatever the scoring round, members of the scoring group have to complete the questionnaires in full (7). To each question, scoring member had to circle a number

from 1 to 9. Responses between two numbers or circling two numbers are forbidden(7). All questions have to be rated(7).
In case of missing values, a member of the steering group will re-contact the member of the scoring group(6)(7).

THE MODERATOR

The moderator of the panel meeting is characterized by its neutrality and ensure a balance between speaking each member of the scoring group (7).

Delphi Round and panel meeting

FIRST ROUND: DELPHI ROUND TO ASSESS EFFICACY AND RELIABILITY

In the first round, each tool is rated using the data of efficacy and reliability from the systematic review. Each panelist receives by mail four sets of data:
Result of the literature review with data about validity (Sensibility, Specificity) and efficacy (Youden Index) for each tool.
Sample data (number, features, etc.)
The 4 sources articles
Reliability data (Cronbach alpha)

Each set of data is shown in a synthetic view, in order to facilitate the rating work.
The experts do not receive details of the tools not to be subjective in their rating.
For each tool, experts will individually answer three questions: *"Does this tool seem effective enough to be appropriate for the diagnosis of depression in primary care*?", "D*oes this tool seem reliable enough to be appropriate for the diagnosis of depression in primary care*?", 3"D*oes this tool seem effective and reliable enough to be appropriate for the diagnosis of depression in primary care*? "
They will have to rate the efficacy and the reliability on a single global–Likert scale from 1 to 9 (1: highly not appropriate to 9: highly appropriate) using their own best clinical judgment with no interaction among experts (6). Panelists are encouraged to explain their opinion in quick notes for each tool.

The French team collects the individual results.
The first result rating must be returned before 4 weeks by mail, and each panelist can contact by mail a member of the steering group if he needs help(6).
All the tools are kept at the end of the first round.

SECOND ROUND: PANEL MEETING

The aim of the second round is to disentangle feasibility on the tools. It takes place in a panel meeting during the EGPRN meeting in October 2012 (Antwerp, Belgium).
There is one moderator to handle the panel meeting and one observer to ensure the quality of the meeting. The panel's moderator is a French member.

1/ selects by panel meeting the three most effective and reliable tools

Experts receive an individualized document showing the distribution of all the experts' first round rating and comments, with their own specific rating (9).
An ultimate discussion is held based on the individual results from the first round (6).The aim is to confirm the three most effective and reliable tools.

2/ rating by feasibility

The panelists receive the complete test forms of the three selected tools with details about feasibility of each tool (number of question, duration, etc.). They work in pairs (selected in a random way) and each one completes all the tests with his partner.
After testing, the panelists have to rate individually the selected tools. Each panelist has to answer one question for each test: *does this test seem feasible in real GP's practice?* They rate on a Likert scale from 1 (absolutely not feasible) to 9 (easily and completely feasible) with comments for each test.
Then, another discussion is held based on those individual results. The panelists are encouraged to discuss their rating for each tool in the light of their knowledge so that all the panelists rated (6). Then, the panelists have to re-rate individually the tools, regardless of whether their rating is unchanged from the beginning of the second round (6).

This discussion is audio-recorded and video taped.

No consensus is done at that time and the moderator has to promote the idea that no consensus has to be reached during the panel discussion.

The results of the panel is a rating of each tool (median of its score) and the verbatim of the discussion(8)(6). Tools are classified into three levels of feasibility according to their median:
Feasible: panel median of 7 to 9, without disagreement
Uncertain: panel median of 4 to 6 or 7 to 9 with disagreement
Not feasible: panel median of 1 to 3 without disagreement

Disagreement is a distribution of medians simultaneously with at least 30% of individual scores between 1 and 3 and 30% of individual scores between 7 and 9 (9)(10)(13).

THIRD ROUND: DELPHI ROUND TO DEVELOP A CONSENSUS

In this round, 15 days after the panel meeting, each panelist receives by mail the results of the meeting (i.e. the final rating of feasibility for each tool with the complete verbatim of the panel discussion) (8).
A classification of the three tools selected in the second round is proposed to each panelist.
The number 1 is the “most appropriate tool to diagnose depression in general practice”, the number 3 is the least appropriate” and the number 2 is for the intermediate tool. The results are collected for each expert.
The addition of scores for each tool aimed to determine the best (the lowest score is the best tool). This tool can be considered as the most effective, reliable, easy to use and practical tool of depression diagnosis in general practice.
If there is no clear consensus, another individual rating will be made.

BIBLIOGRAPHY

1. Licht-Strunk E, Windt DVD. The prognosis of depression in older patients in general practice and the community. A systematic review. Family Practice. 2007;24(December 2006):168–80.
2. Licht-strunk E, Beekman ATF, Haan MD, Marwijk HWJV. The prognosis of undetected depression in older general practice patients . A one year follow-up study. Journal of Affective Disorders. 2009;114(1-3):310–5.
3. Mitchell AJ, Vaze A, Rao S. Clinical diagnosis of depression in primary care: a meta-analysis. Lance [Internet]. 2009;374(9690):609–19. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19640579
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 British Medical Journal. 2009;338(7692):a3079.
5. Lehti A, Hammarström A, Mattsson B. Recognition of depression in people of different cultures: a qualitative study. BMC Family Practice. 2009;10:53.
6. Fitch K, Bernstein SJ, Aguilar MD, Burnand B, LaCalle JR, Lazaro P, et al. The RAND/UCLA Appropriateness Method User's Manual [Internet]. 2001. Available from: http://www.rand.org/content/dam/rand/pubs/monograph_reports/2011/MR1269.pdf
7. HAS. Bases méthodologiques pour l'élaboration de recommandations professionnelles par consensus formalisé. 2006.
8. Bourrée F, Michel P, Salmi LR. Consensus methods: Review of original methods and their main alternatives used in public health. Revue d'Épidémiologie et de Santé Publique [Internet]. 2008 Dec [cited 2012 Apr 25];56(6):e13–e21. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0398762008008559
9. Letrilliart L, Vanmeerbeek M. À la recherche du consensus: quelle méthode utiliser? exercer. 2011;22(99):170–7.
10. McGory ML, Shekelle PG, Ko CY. Development of quality indicators for patients undergoing colorectal cancer surgery. Journal Of The National Cancer Institute [Internet]. 2006;98(22):1623–33. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17105985

11. Frappé P. Initiation à la recherche [Internet]. Global Media Santé; 2008. Available from: http://www.amazon.fr/Initiation-à-recherche-Paul-Frappé/dp/2919616056

12. Haute Autorité Santé. Bases méthodologiques pour l'élaboration de recommandations professionnelles par consensus formalisé. 2006;

13. Shekelle PG, MacLean CH, Morton SC, Wenger NS. Assessing care of vulnerable elders: methods for developing quality indicators. Annals of Internal Medicine [Internet]. 2001;135(8 Pt 2):647–52. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11601947

DATA ABOUT EFFICACY: SENSIBILITY, SPECIFICITY EXTRACTED OF ARTICLES, YOUDEN INDEX CALCULATED

According to the methodology of the systematic review, the articles were included thanks to numerical datas of validity or efficacy, with or without associated reliability datas.

The only values of validity or efficacy that are found in all the articles are Sensibility and Specificity. The other datas are not the same between the articles and can't be used.

In the first part of the RAND/UCLA Appropriateness Method, you only have to rate the validated tools with the datas of validity used to calculate Efficacy (Youden index), and data of reliability (Cronbach alpha) from the articles and from additional sources.

We don't use Area Under the Curve to compare Efficacy because of lake of information in the selected articles to extract or to calculate this. Additionally, we can't use Cohen Kappa to compare Reliability because of lake of data in the selected articles and because of lack of means in additional sources.

EFFICACY: YOUDEN INDEX CALCULATED WITH DATA RETRIEVAL

The Youden Index is often used as a summary measure of the receiver operating curve (1). Its measure the effectiveness of a diagnostic marker (1).
Youden Index is the maximum difference between Sensitivity (the probability of correctly classifying diseased individuals) and 1-Specificity (the probability of incorrectly classifying health individuals) (2)

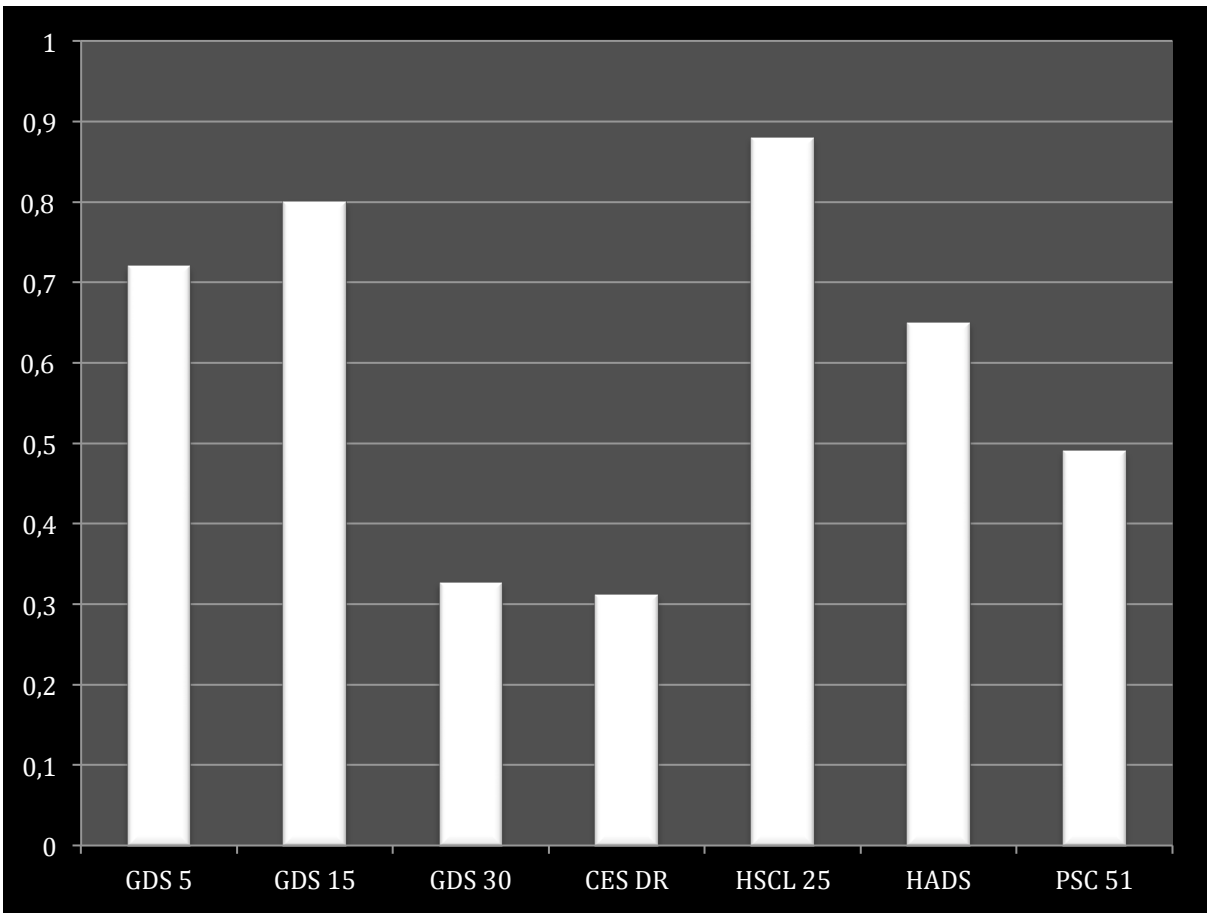
Youden Index= (Sensibility + Specificity) - 1

This index ranges between 0 and 1, with a value of 1 indicating perfect diagnostic effectiveness, and 0 indicating an ineffective test.

All Youden index were calculated by the steering group and not extracted from the selected articles.

Tools	**Cutpoint**	**Initial Size**	**Final Size**	**Sensibility**	**Specificity**	**Youden Index**
GDS 5	≥ 2	350	301	0,86	0,86	0,72
GDS 15	≥ 5			0,82	0,98	0,8
GDS 30	> 9	534	206	0,538	0,789	0,327
CES-DR	≥ 16			0,82	0,492	0,312
HSCL 25	> 1,75	475	74	0,94	0,94	0,88
HADS	≥ 15	1046	473	0,85	0,8	0,65
PSC 51	≥ 5			0,9	0,59	0,49

Tool	**Initial size**	**Final size**	**Aged of included patients**	**Mean Age**	**Gender**
GDS 5	350	301	>64 years,	74,3	57,8% women
GDS 15					
GDS 30,	534	206	≥ 60 years,	71,2+/- 6,8	65,5% women
CES-DR					
HSCL 25	475	74	≥ 65 years	78,5 (5,5)	61,7% women
HADS	1046	473	25 to 80 years	43,6 to 53,9	67,5% Women
PSC-51					



RATING

“Does this tool seem effective enough to be appropriate for the diagnosis of depression in primary care?”

You have to range each tool with a note from 1 to 9 in a discontinous Likert’s Scale rating:

1: Extremely Inappropriate

5: Uncertain

9: Extremely Appropriate

Tools	**Likert’s Scale and Comments**
GDS 5	*Appropriateness:*

	1 2 3 4 5 6 7 8 9
	Comment:
GDS 15	*Appropriateness:* 1 2 3 4 5 6 7 8 9
	Comment:
GDS 30	*Appropriateness:* 1 2 3 4 5 6 7 8 9
	Comment:
CES-	*Appropriateness:*

DR	1 2 3 4 5 6 7 8 9
	Comment:
HSCL 25	*Appropriateness:* 1 2 3 4 5 6 7 8 9
	Comment:
HADS	*Appropriateness:* 1 2 3 4 5 6 7 8 9
	Comment:

PSC 51	*Appropriateness:* 1 2 3 4 5 6 7 8 9
	Comment:

BIBLIOGRAPHY

1. Ortega Orcos R, Salinero Fort MA, Kazemzadeh Khajoui A, Vidal Aparicio S, Dios de Valle R. Validación de la versión española de 5 y 15 ítems de la Escala de Depresión Geriátrica en personas mayores en Atención Primaria. Revista Clinica Espanola. 2007;207(11):559–62.

2. Sánchez-garcía S, Juárez-cedillo T, García-gonzález JJ, Espinel-bermúdez C, J Gallo J, A Wagner F, et al. Usefulness of two instruments in assessing depression among elderly Mexicans in population studies and for primary care. Salud publica de mexico. 2008;50(6):447–56.

3. Frojdh K, Hakansson A, Karlsson I. The Hopkins Symptom Checklist-25 is a sensitive case-finder of clinically important depressive states in elderly people in primary care. International Journal of Geriatric Psychiatry. 2004;19:386–90.

4. W.M. de Waal M, A. Arnold I, Spinhoven P, A H Eekhof J, J J Assendelft W, M. van Hemert A. The role of comorbidity in the detection of psychiatric disorders with checklists for mental and physical symptoms in primary care. Social Psychiatry and Psychiatric Epidemiology. 2009;44:78–85.

DATA ABOUT RELIABILITY: CRONBACH ALPHA FROM THE ARTICLES AND FROM ADDITIONAL SOURCES

CRONBACH ALPHA

Cronbach alpha (α) is a test reliability technique that requires only a single test administration to provide a unique estimate of the reliability for a given test (3). It's a useful coefficient for assessing internal consistency reliability (4).
When items are used to form a scale, they need to have internal consistency: the items should have to measure the same thing, so they should be correlated with one another (3)(5).

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum_{i=1}^{k}\sigma_{Y_i}^2}{\sigma_X^2}\right)$$

k : Number of items

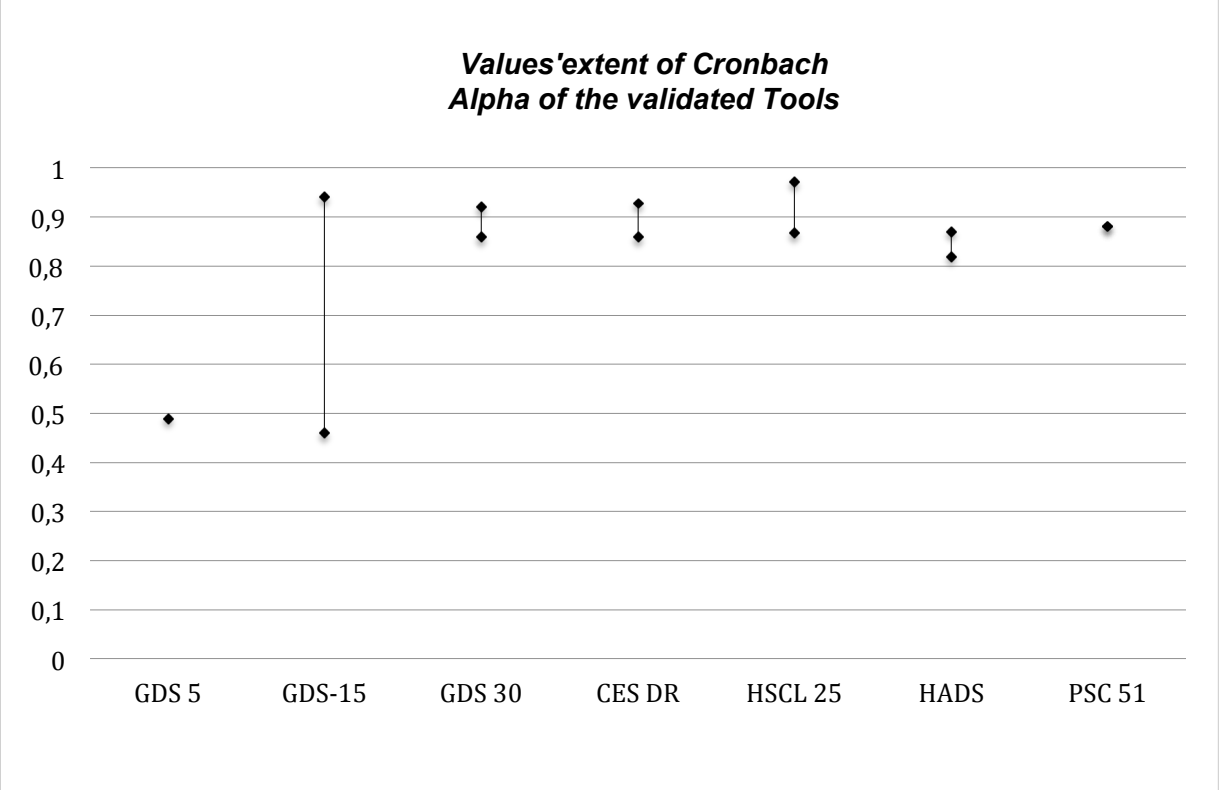
$\sigma_{Y_i}^2$: Variances of i^{th} individual item

σ_X^2 : Variance of the total score formed by

Based upon the formula, the size of alpha is determinate by number of items, and by the range of all the possible values of each item (variance), that means is determinate by the sample size. We choose to keep only additional sources with sample size higher than 100 peoples, to increase the strength of alpha values.

If the items making up the score are all identical and so perfectly correlated, alpha=1 (3).If the items are all independent, alpha=0 (3).
For comparing group, alpha values of 0,7 to 0,8 are regarded as satisfactory, but for the clinical application, much higher values of alpha are needed (the minimum alpha is 0,9)(3). It should be noted than an alpha of 0,8 is probably a reasonable goal (5).
Once validated tools clearly identified, further research was conducted in literature in order to complete reliability data of the validated tools:

-In Google Scholar® and Mendeley® with this equation: Cronbach's alpha (individually) + name of tool.
-In Pubmed®, research equation: « reliability » and « depression » or « depressive disorder » and « scale ».



Cronbach Alpha: data from additional sources

NB: cut-off point : 100 patients (studies with sample < 100 patients were excluded)

Scales	Samples n=	Cronbach Alpha	References
GDS 5	126	0.49	1, 2
GDS 15	960	0.75	3
	816	0.72	4
	586	0.76	5
	153 + 459	0.77	6
	194	0.80	7
	2032	0.46	8
	1034	0.82	1, 9
	187	0.77	1, 10
	4253	0.8	1, 11
	126	0.8	1, 2
	168+103	0.94	1, 12
	121	0.88	1, 13
	407	0.84	1, 14
	333	0.79	15
GDS 30	534	0.87	16
	40 + 461	0.89	1, 17
	187	0.86	1, 10
	126	0.87	1, 2
	100 + 95	0.87	1, 18
	200	0.92	1, 19
	407	0.90	1, 14
	333	0.88	15
CES-DR	534	0.86	16
	245	0.888	20
	Sample 1: 6971 Sample 2: 243	0.923 0.928	21
HSCL-25	6886	0.93	22

	180	0.8676	23
	159	0.97	24
HADS	302	0.87	25
	747	0.82-0.83	26
PSC 51	473	0.88	27

RATING

“Does this tool seem reliable enough to be appropriate for the diagnosis of depression in primary care?”

You have to range each tool with a note from 1 to 9 in a discontinous Likert’s Scale rating :

- 1: Extremely Inappropriate
- 5: Uncertain
- 9: Extremely Appropriate

Tools	**Likert’s Scale and Comments**
GDS 5	*Appropriateness:* 1 2 3 4 5 6 7 8 9
	Comment:
GDS 15	*Appropriateness:* 1 2 3 4 5 6 7 8 9
	Comment:

GDS 30	*Appropriateness:* 1 2 3 4 5 6 7 8 9
	Comment:
CES-DR	*Appropriateness:* 1 2 3 4 5 6 7 8 9
	Comment:
HSCL 25	*Appropriateness:* 1 2 3 4 5 6 7 8 9

	Comment:
HADS	*Appropriateness:* 1 2 3 4 5 6 7 8 9
	Comment:
PSC 51	*Appropriateness:* 1 2 3 4 5 6 7 8 9
	Comment:

BIBLIOGRAPHY

Vrantsidis F, Haralambous B, Lin X, Runci S, Rayner V, Dow B, et al. The Assessment of Older People with dementia and depression of Culturally and Linguistically Diverse Backgrounds : A review of current practice and the development of guidelines for Victorian Aged Care Assessment Services Literature Review July 2011. 2011.

Chattat R, Ellena L, Cucinotta D, Savorani G, Mucciarelli G. A study on the validity of different short versions of the Geriatric Depression Scale. Arch Gerontol Geriatr.2001;suppl 7,81-86

Friedman B, J.Heisel M, L.Delavan R. Psychometric Properties of the 15-Item Geriatric Depression Scale in Functionally Impaired, Cognitively intact, Community-Dwelling Elderly Primary Care Patients. Journal of the American Geriatrics Society. 2005;53:1570–6

Weeks SK, McGann PE, Michaels TK et al. Comparing various short-form geriatric depression scales leads to the GDS-5/15. J Nurs Scholarsh 2003;35:133–137.

Van Marwijk HWJ, Wallace P, DeBock GH et al. Evaluation of the feasibility, reliability and diagnostic value of shortened versions of the geriatric depression scale. Br J Gen Pract 1995;45:195–199.

Jang Y, Small BJ, Haley WE. Cross-cultural comparability of the Geriatric Depression Scale: Comparisons between older Koreans and older Americans. Aging Ment Health 2001;5:31–37.

D'Ath P, Katona P, Mullan E et al. Screening, detection and management of depression in elderly primary care attenders. I. The acceptability and performance of the 15 Item Geriatric Depression Scale (GDS15) and the development of short versions. Fam Pract 1994;11:260–266.

Incalzi RA, Cesari M, Pedone C et al. Construct validity of the 15-item Geriatric Depression Scale in older medical inpatients. J Geriatr Psychiatry Neurol 2003;16:23–28.

Kam W B, Chiu H F K. Assessing Psychological Well-being of the Old-Old: A Comparative Study of the GDS-15 and GHQ-12. Clinical Gerontologist 1998;19(1):65-75

Liu CY, Shu Yu CHL, Yang YY. Correlations between scores on Chinese versions of long and short forms of the Geriatric Depression Scale among elderly Chinese. Psychological Reports 1998;82:211-14

Nyunt MSZ, Gones C, Niti M, Ng TP. Criterion-based validity and reliability of the Geriatric Depression Screening Scale (GDS-15) in a larger validation sample of community-living Asian older adults. Aging and Mental Health 2009;13(3):376-82

Fountoulakis KN, Tsolaki M, Iacovdies A, Yesavage J, O'Hara R, Kazis A, et al. The validation of the short form of the Geriatric Depression Scale (GDS) in Greece. Aging Clin Exp Res 1999;11:367-72

Chaaya M, Sibai AM, El Rouelheb Z, Chemaitelly H, Chahine LM, Al-Amon H, et al. Validation of the Arabic version of the short Geriatric Depression Scale (GDS-15). International Psychogeriatrics 2008;20(3):571-81

Muy AC, Kang SY, Chen LM, Domanski MD. Reliability of the Geriatric Depression Scale for Use Among Elderly Asian Immigrants in the USA. International Psychogeriatrics 2003;15(3):253-71

Jongenelis K, Pot AM, Eisses AMH, Gerritsen DL, Derksen M. Diagnostic accuracy of the original 30-Item and shortened versions of the Geriatric Depression Scale in nursing home patients. International Journal of Geriatric Psychiatry. 2005;20:1067–74.

Sánchez-garcía S, Juárez-cedillo T, García-gonzález JJ, Espinel-bermúdez C, J Gallo J, A Wagner F, et al. Usefulness of two instruments in assessing depression among elderly Mexicans in population studies and for primary care. Salud publica de mexico. 2008;50(6):447–56.

Chan ACM. Clinical validation of the Geriatric Depression Scale (GDS) Chinese Version. Journal of Aging and Health 1996;8:238-53

Segulin N, Deponte A. The evaluation of depression in the elderly: A modification of the geriatric depression scale (GDS). Archives of Gerontology and Geriatrics 2007;44:105-12

Wrobel NH, Farrag MF. A Preliminary Report on the Validation of the Geriatric Depression Scale in Arabic. Clinical Gerontologist 2006;9(4):33-46

Kobayashi-Gutiérrez A, Martinez-Bonilla G, Bernard-Medina AG, Troyo-Sanroman R, Gonzales-Dias V, Castro-Contreras E, et al. Depression and its correlation with in patients pain in the rheumatology service of a Mexican teaching hospital. Rheumatology International. 2009;29:1169–75.

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 Research. 2011;186:128–32.

Strand BH, Dalgard ODDS, Tambs K. 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). Psychiatry: Interpersonal and Biological Processes. 2003;5(1).

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. International Journal. 2008;1(2).

Mouanoutoua VL, Brown LG. Hopkins Symptom Cheklist-25 Hmong version: A screening instrument for psychological distress. Journal of personality assessement. 1995;64(2):376–83.

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(1):6.

Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. Journal of Psychosomatic Research. 2002;52(2):69–77.

W.M. de Waal M, A. Arnold I, Spinhoven P, A H Eekhof J, J J Assendelft W, M. van Hemert A. The role of comorbidity in the detection of psychiatric disorders with checklists for mental and physical symptoms in primary care. Social Psychiatry and Psychiatric Epidemiology. 2009;44:78–85.

DATA ABOUT EFFICACY AND RELIABILITY COMBINED

Using data about efficacy and reliability combined, thank you to answer now to the following question:

"Does this tool seem effective and reliable enough to be appropriate for the diagnosis of depression in primary care?"

Tools	**Likert's Scale and Comments**
GDS 5	*Appropriateness:* 1 2 3 4 5 6 7 8 9
	Comment:
GDS 15	*Appropriateness:* 1 2 3 4 5 6 7 8 9
	Comment:

GDS 30	*Appropriateness:* 1 2 3 4 5 6 7 8 9
	Comment:
CES-DR	*Appropriateness:* 1 2 3 4 5 6 7 8 9
	Comment:
HSCL	*Appropriateness:* 1 2 3 4 5 6 7 8 9
	Comment:

HADS	*Appropriateness:* 1 2 3 4 5 6 7 8 9
	Comment:
PSC 51	*Appropriateness:* 1 2 3 4 5 6 7 8 9
	Comment:

TIME SCHEDULE

Each Delphi round shall not exceed 6 weeks.
The first Delphi round will take place between August 1 and September 15.
The panel meeting will take place at the congress on the EGPRN in Antwerp in October.
The second Delphi round will take place between November 15 and December 31.
The final results will be announced January 30

PUBLICATION PLAN

Publication rank:

Authorship credit is based only on substantial contribution to: conception and design, or analysis and interpretation of data drafting the article or revising it critically for important intellectual content and final approval of the version to be published (from BMJ criteria). Participation solely in the acquisition of funding or the collection of data does not justify authorship. The final decision rests with the daily board. The order of the authors depends further on the number of investigators and PhD student in every site.

Publication acceptations:

Each proposal for a publication with FPDM-depression data will be submitted to the daily board as an IMRAD abstract of 300 words. The daily board will check the proposal for overlap with other plans and potential combined or conflicting interests. These interests can concern the submitted publication plans or the use of data to which certain persons are explicitly involved. If there is no overlap or there are no conflicting interests, the proposal will, with a positive advise from the board, be accepted. When there is a possibility of overlap or when there are (possible) conflicting interests, the submitter of the proposal will be informed about this, with the request to adjust the plans.

The submission of a proposal implies concrete plans for a publication, etc. At which will be worked on a short time basis. When within half a year after submitting a proposal there is no provable activities in that direction, or when after a year no publication has been submitted to a journal, the subject can be released for other interested parties.

There is a maximum of two proposals that can be submitted as first author at the same time. Only after the presentation of the paper to a journal, a new proposal can be submitted.

Senior researchers who acquire extra funds have priority at the submission of publication plans on the theme of the extra fund.

Researchers who are not part of FPDM can, after consultation by the board taking advice with the most involved researchers, submit a proposal for data-analysis and publications. A senior researcher of FPDM will always be a member of this research group and will be co-author of the publications.

AUTHORSHIP

The one who “pulls” the article and has the most important role in writing it, is the first author. He/she is responsible for the contents of the article.
The first author determines in consultation with the board who the co-authors are and in which order. The board considers the investments and contributions of the FPDM members, like authorship of the original proposal and local coordination. When the first author is a PhD-student, this happens in consultation with the (co-) promoters. Possible conflicts will be put before the board.
(Co) promoter(s) who are primarily responsible for the supervision of PhD-students are (also) responsible for the integrity of the work as a whole, from inception to published article.
Other co-authors should have made an important contribution to the design of the study, data collection and/or writing of the paper. Consider the three conditions of the Requirements for Authorship sometimes an acknowledgement is more suitable then a co-authorship.
All authors get to inspect the article at least twice before it is forwarded to a magazine.
It is the responsibility of the first author that the guide-lines according authorship are followed, that the sample and the research-methodology are described correctly and that references to former relevant FPDM-publications are made. To anticipate carelessness in this, the board should see every manuscript before presented for publication.
The international FPDM team (named as “FPDM Team”) will always be in the author list of any publication as the last name of authors..

Overview of publications

An overview of all publications and publication plans, as well as publications in journals or chapters in books, is kept by the FPDM board.
Changes in publications and publication plans (as mentioned under a.), are to be communicated to the NESDA-secretary - preferably by e-mail.

The following information is needed:
Date of sending the manuscript and name of the journal, including possible changes
Changes in authors or title
Date of acceptance
At publication: full reference
Withdrawal (decision not to publish)

From the moment of presentation and after being inserted in the overview, all manuscripts are kept in an archive. As soon as a paper is send to a journal, the first author will send a (digital or paper) copy of this to the board. After publication a reprint of the final article will be send to the board.
Abstracts of congress papers should also be sended to the board.

Particular circumstances

If in publication plans FPDM data are used in combination with data collected by other research groups, the publication plan will be judged by both the FPDM board and the board of the other research group.

Everybody is free to publish is own national results (after board reviewing) but for the systematic reviews and for the final articles at each step all the national teams has to be named. For the reviews articles the one on depression should have Patrice Nabbe as first author (this is compulsory for the French team because he has a PhD on the topic) and for the one on multimorbidity a should have Jean Yves le Reste as first author as he have a PhD too on this topic.

INTERESTED TEAMS

Miguel-Angel Munoz from Barcelona (spain) only for translation procedure
Ana Claveria From Vigo (Spain)
Jean Yves Le Reste and Patrice Nabbe from Brest
Stella Argyriadou from Greece
Harris Argyriadou from italy

Harm Van Marjwijk from Amsterdam

Eva Hummers Pradier, Christa Doerr, Lingner Heidrun from Germany

Slawomir Czachowsky from Poland

Djurdjica Lazic from Croatia

Melida Hasaganic from Bosnia

Radost Assenova from Bulgaria

RESULTS OF THE METHODICAL SYSTEMATIC REVIEW, AND ALL NUMERICAL DATAS EXTRACTED FROM THE SELECTED ARTICLES (complementary data)

"What tools are validated against face-to-face psychiatric examination using DSM-IV, to diagnose depression in general practice for adult patients?" (1)(2)(3)(4)

Tools	Title	Year	Journal	First Author
GDS 5 GDS 15	Validacion de la version Espanola de 5 y 15 items de la Escala de Depression Geriatrica en personas mayores de Atencion primaria	2007	Rev Clin Esp	R. Ortega Ortos
GDS 30 CES DR	Usefulness of two instruments in assessing depression among elderly Mexicans in population studies and for primary care	2008	Salud Publica Mex	S. Sanchez-Garcia
HSCL-25	The Hopkins Symptom Cheklist-25 is a sensitive case-finder of clinically important depressive states in elderly people in primary care	2004	Int J Geriatr Psychitry	K. Frojdh
HADS PSC 51	The role of comorbidity in the detection of psychiatric disorders with checklist for mental and physical symptoms in primary care	2009	Soc Psychiatry Psychiatr Epidemiol	M.W.M. de Waal

Bibiography:

1. Ortega Orcos R, Salinero Fort MA, Kazemzadeh Khajoui A, Vidal Aparicio S, Dios de Valle R. Validación de la versión española de 5 y 15 ítems de la Escala de Depresión Geriátrica en personas mayores en Atención Primaria. Revista Clinica Espanola. 2007;207(11):559–62.

2. Sánchez-garcía S, Juárez-cedillo T, García-gonzález JJ, Espinel-bermúdez C, J Gallo J, A Wagner F, et al. Usefulness of two instruments in assessing depression among elderly Mexicans in population studies and for primary care. Salud publica de mexico. 2008;50(6):447–56.

3. Frojdh K, Hakansson A, Karlsson I. The Hopkins Symptom Checklist-25 is a sensitive case-finder of clinically important depressive states in elderly people in primary care. International Journal of Geriatric Psychiatry. 2004;19:386–90.

4. W.M. de Waal M, A. Arnold I, Spinhoven P, A H Eekhof J, J J Assendelft W, M. van Hemert A. The role of comorbidity in the detection of psychiatric disorders with checklists for mental and physical symptoms in primary care. Social Psychiatry and Psychiatric Epidemiology. 2009;44:78–85.

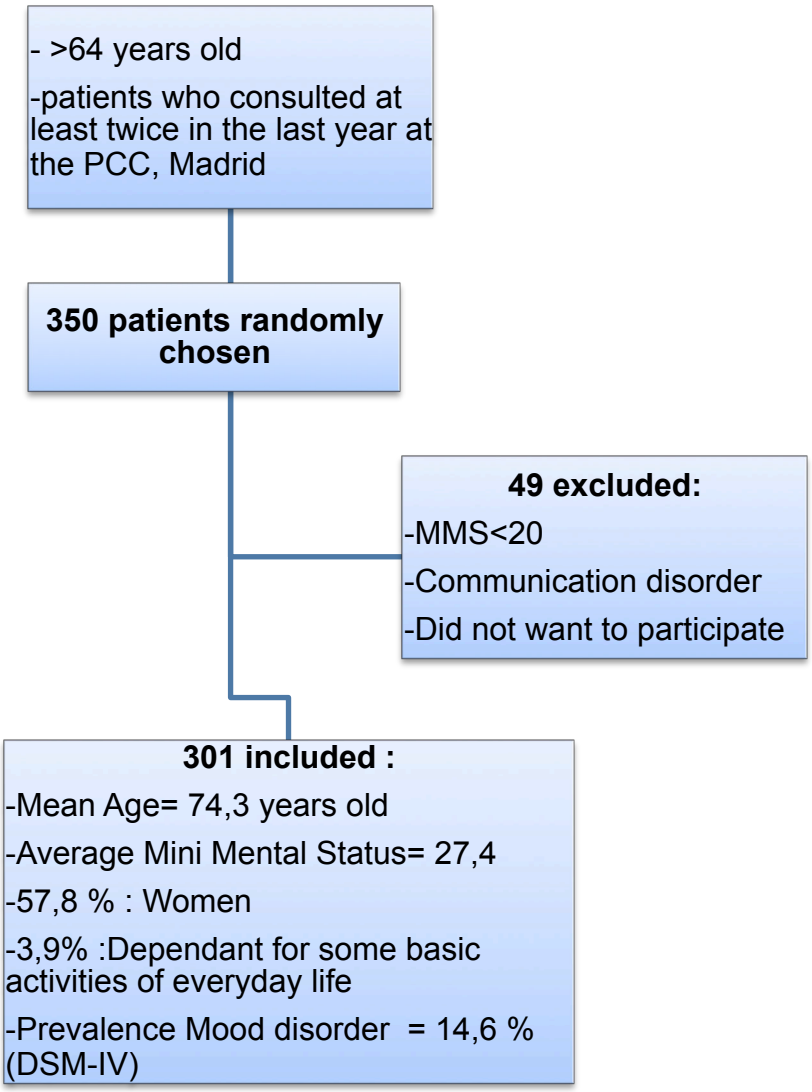
Tool	**Validity (versus DSM-IV)**							**Efficacy (Versus DSM-IV)**		**Reliability**		**Correlation**
	Se		*Sp*		*VPP*	*VPN*	*LR+*	*LR-*	*AUC*		*Cronbach Alpha*	*Cohen kappa*	
GDS 5	0,86		0,86		0,5	0,97	6	0,16	0,86			K=0,85 (GDS 5 or GDS 15? not information)	
GDS 15	0,82		0,98		0,86	0,96	35,03	0,19	0,90				
GDS 30											0,87		
	0,538		0,789		0,608	0,737							
CES-DR											0,86		
	0,82		0,492		0,496	0,818							
HSCL-25	0,94		0,94										
HADS	0,85	HADS and PSC 51: 0,75	0,8	HADS and PSC 51: 0,85					0,91	PSC 51 and HADS: 0,92			HADS versus PSC-51: 0,6
PSC 51	0,9		0,59						0,86		0,88		

DETAILS ABOUT SAMPLES IN THE SELECTED ARTICLES TO CALCULATE EFFICACY OF EACH VALIDATED TOOLS (Complementary data)

Tool	**Initial size**	**Final size**	**Aged of included patients**	**Mean Age**	**Gender**
GDS 5	350	301	>64 years,	74,3	57,8% women
GDS 15					
GDS 30,	534	206	≥ 60 years,	71,2+/- 6,8	65,5% women
CES-DR					
HSCL 25	475	74	≥ 65 years	78,5 (5,5)	61,7% women
HADS	1046	473	25 to 80 years	43,6 to 53,9	67,5% Women
PSC-51					

1/ GDS 5 and GDS 15, 5-item and 15-item Geriatric Depression Scale

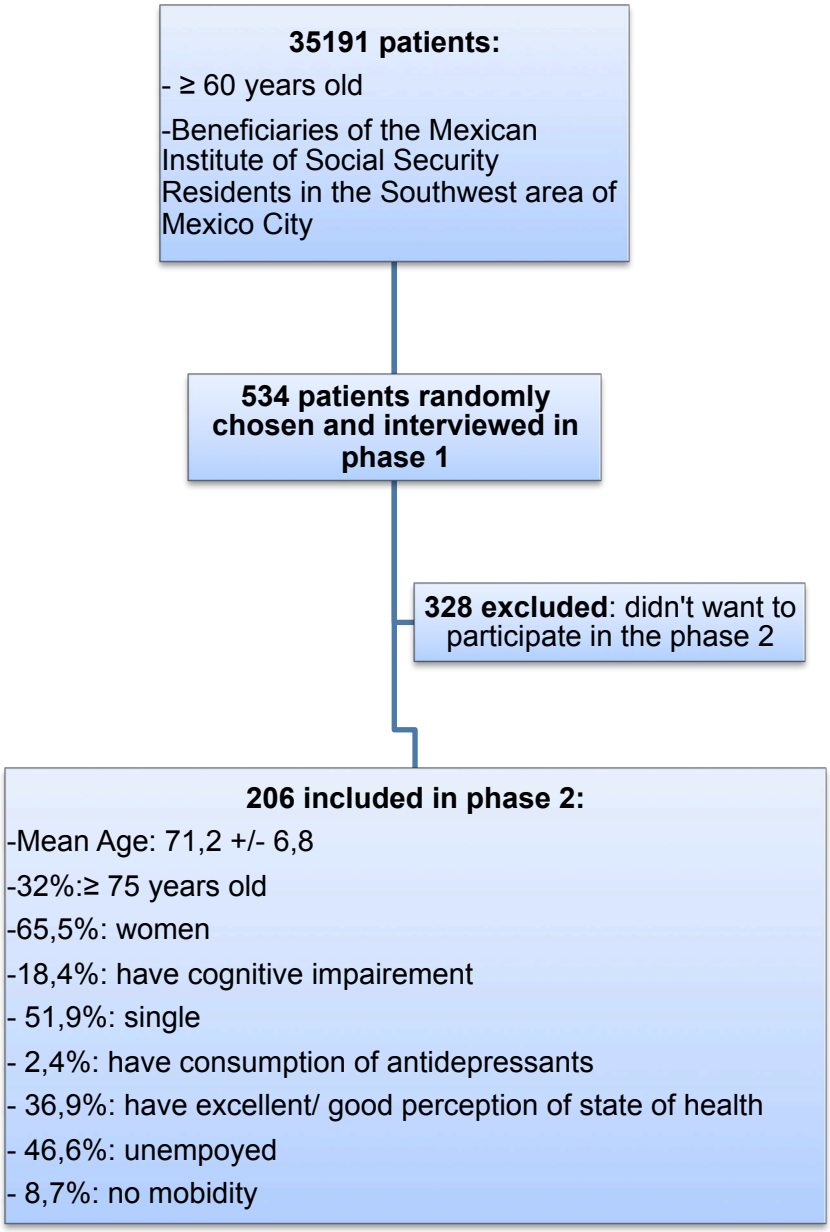
GDS 5 and GDS 15 were tested on the same sample



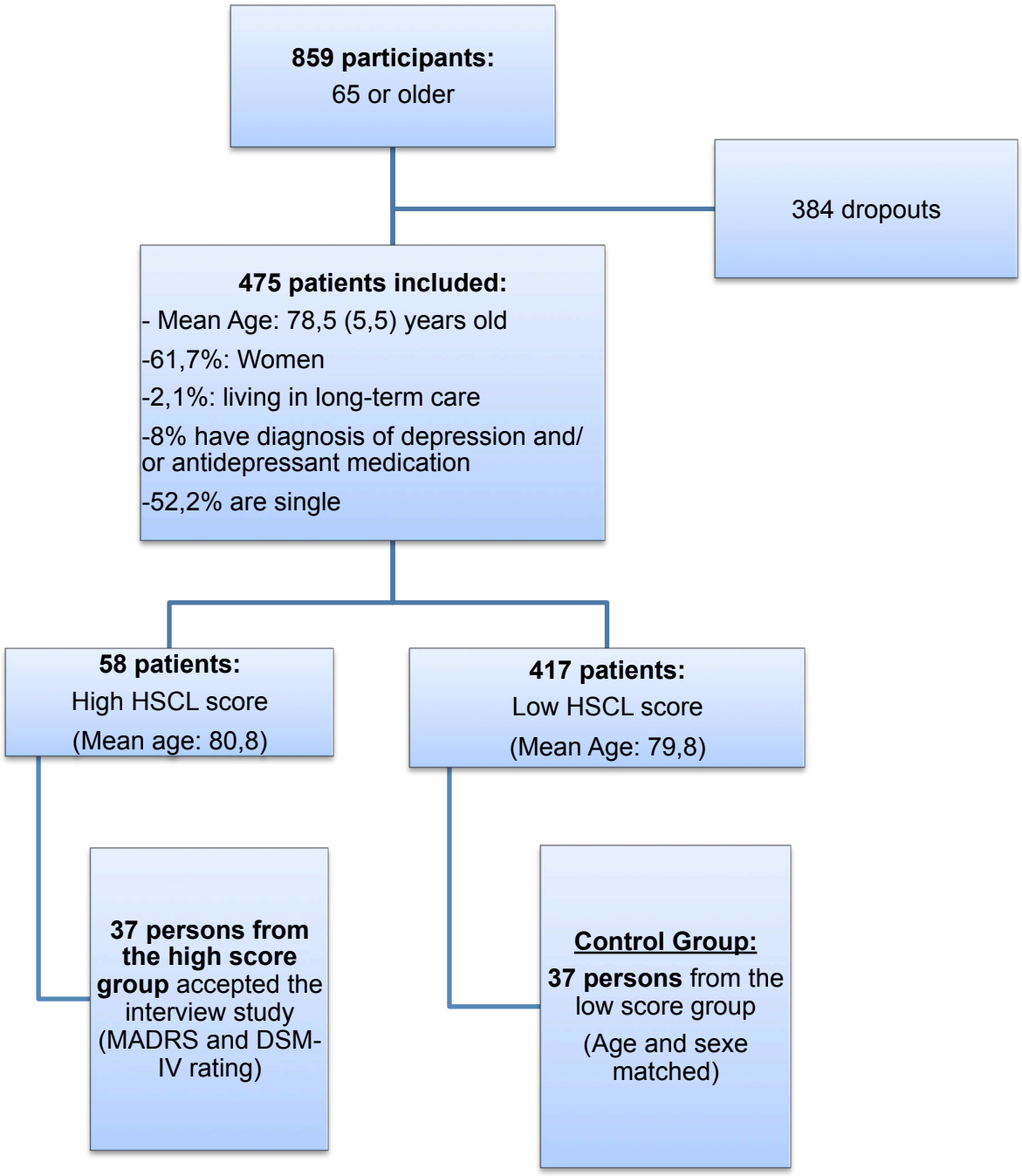
2/ GDS 30 and CES-DR, 30- Item Geriatric Depression Scale and Center for Epidemiologic Studies Depression Scale-Revised

GDS 5 and GDS 15 were tested on the same sample.

Cronbach alpha was calculated on the first phase with 534 individuals, and Validity was calculated on the second phase with 206 individuals.

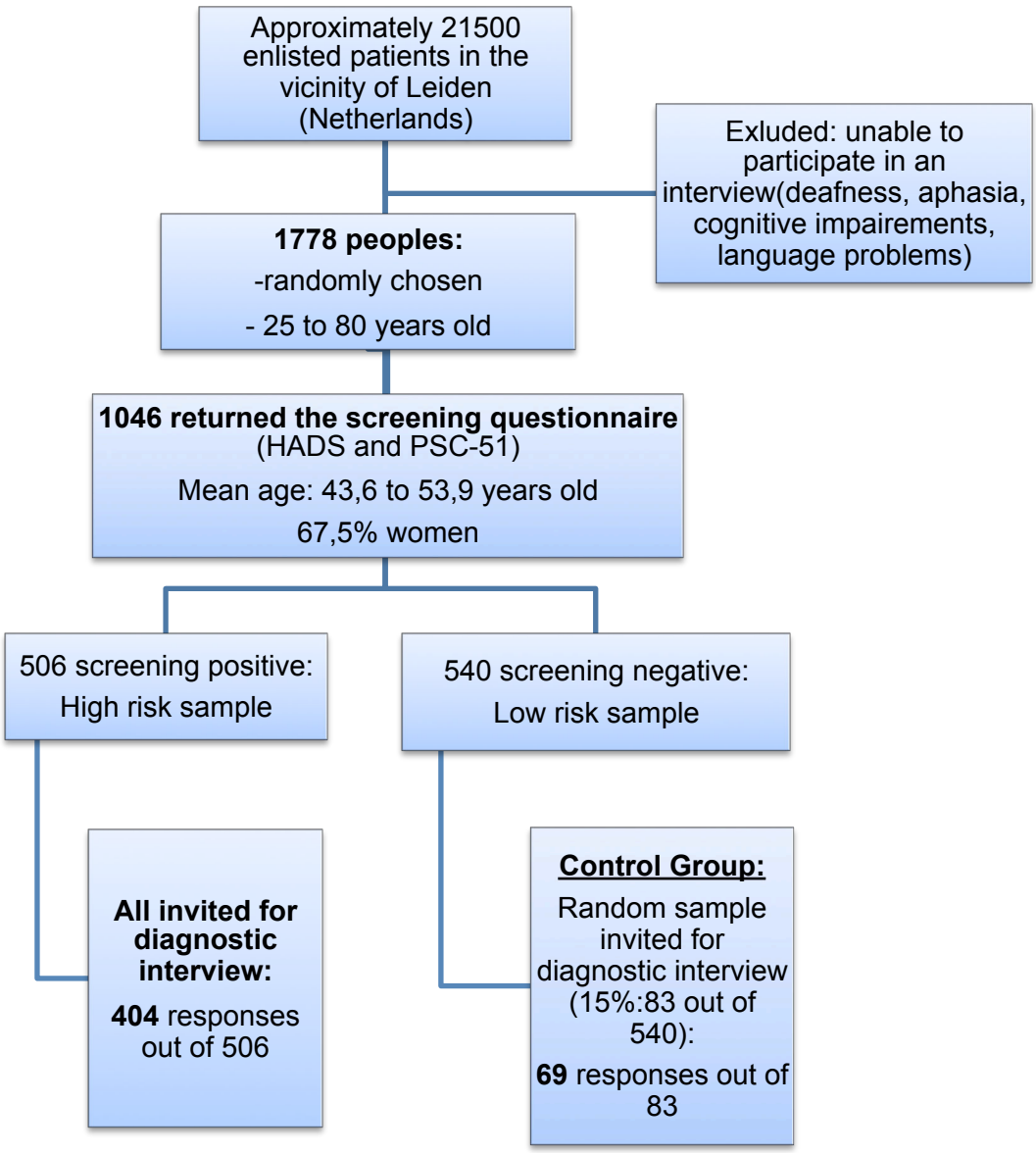


3/ HSCL, 25-item Hopkins Symptom Checklist:



4/ HADS and PSC 51, Hospital Anxiety and Depression Scale and 51-item Physical Symptom Checklist

HADS and PSC-51 were tested on the same sample.



Annexe 3: informed consent

Département Universitaire de Médecine Générale

22, avenue Camille Desmoulins CS 93837 – 29238 – Brest CEDEX 3

Tél : 02 98 01 65 52 – fax : 02 98 01 64 74

INFORMATION NOTICE

International Investigator Senior Coordinator

Name: Nabbe Patrice

Address: Département de médecine générale, Faculté de Médecine de Brest, 22, avenue Camille Desmoulins, 29238 Brest cedex 3

International Developer

Département Universitaire de Médecine Générale – 22 avenue Camille Desmoulins - 29238 Brest Cedex 3

National investigator senior coordinator:

Name:

Address:

National developer:

Dear colleagues

You are invited to participate in a survey by Nabbe Patrice (MD) (trainee in PhD) and Beck Robert Emilie (trainee in MD). The "Département universitaire de medecine générale de Brest (Université de Bretagne Occidentale)" is the national developer of that survey. He is responsible for it and assume its organization.

If you decide to participate you will be asked to sign a consent form. This signature will confirm that you did agree to participate.

Course of study

Rand UCLA (Delphi round and panel meeting). This Rand UCLA will be fully anonymised and it will be impossible for a study reader to identify you.

Potential risk of study

There are no risks associated with your participation in this study

Potential benefits of the study

There is no potential benefit to this study

Voluntary participation

Your participation to this study is entirely voluntary.

You are free to refuse to participate and to terminate your participation in the study at any time and without incurring any liability or any injury of this fact and without causing consequences.
In this case you must inform the investigator of your decision

In the event that you withdraw your consent, we will conduct a computer processing of your personal data unless written objection on your part.
During the study, your investigator will notify you, if new facts might affect your willingness to participate in the study.

Obtaining complementary informations

If desired, Patrice Nabbe (MD) or local national investigator (phone number), who can be reached at telephone number: 00 33 298 016 552 at any time can answer all your questions about the study.

At the end of the study, and at your request, your investigator will inform you of the overall results of this research.

Confidentiality and use of medical or personal data

As part of biomedical research in which the DUMG Brest, Patrice Nabbe (MD) and your national investigator offer to participate, a treatment of your personal data will be used to analyse the results of research in light of the objective of that study which was presented to you.
To this end, the data collected, including any survey and the data on your lifestyle will be forwarded to the promoter of the research where the data will be processed in this study.
Those data will be anonymized and their identification will be held with a code number.
Staff involved in the study is subject to professional secrecy.
These data may also, under conditions ensuring their confidentiality be transmitted to the national or European health authorities.
Under the provisions of Law you have the right to access and modify. You also have the right to object to the transmission of data covered by professional secrecy.

If you agree to participate in this study, thank you to complete and sign the consent form. You will keep a copy of it.

Consent Form

Promoter : Département Universitaire de Médecine Générale – 22 avenue Camille Desmoulins - 29238 Brest Cedex 3

Dr:……

Address: …….

Local investigator name

Address: ……..

University:

Asked me to participate in a Rand UCLA for FPDM Study (Depression line)

I had time to reflect on my involvement in this study. I am aware that my participation is completely voluntary and that the study will entail no additional cost to my charge.

I can, at any time, decide to leave the study without giving reasons for my decision and that it does without consequences.

I understood that the data collected during the research would be protected in accordance to confidentiality. They can only be accessed by persons subject to professional secrecy belonging to the team-investigating physician, mandated by the promoter.

I accept the computerized processing of personal data in accordance with the data protection act. I have been informed of my right to access and rectify data concerning me.

My consent does not absolve the responsibilities of the organizers of this research. I retain all my rights guaranteed by Law.

Done in two originals

at……, the dd/mm/yyyy

Name, first name of investigator: Name, first name of the interviewee:

Signature:

ANNEXE C

Published articles and ClinicalTrials.gov

European Psychiatry 39 (2017) 99–105

Contents lists available at ScienceDirect

European Psychiatry

journal homepage: http://www.europsy-journal.com

Review

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

P. Nabbe [a,*], J.Y. Le Reste [a,**], M. Guillou-Landreat [b], M.A. Munoz Perez [c], S. Argyriadou [d], A. Claveria [e], M.I. Fernández San Martín [c], S. Czachowski [f], H. Lingner [g], C. Lygidakis [h], A. Sowinska [i], B. Chiron [a], J. Derriennic [a], A. Le Prielec [a], B. Le Floch [a], T. Montier [j], H. Van Marwijk [k], P. Van Royen [l]

[a] Department of general practice, université de Bretagne Occidentale, ERCR SPURBO, Brest, France
[b] Department of addictology, université de Bretagne Occidentale, ERCR SPURBO, Brest, France
[c] IDIAP jordi GOL unitat de support a la recerca, Barcelona, Spain
[d] The Greek Association of General Practitioners (ELEGEIA), Thessaloniki, Greece
[e] Galician National Health Service, Universidad de Vigo, Rúa Oporta, 1, 36201 Vigo, Spain
[f] Department of Family Doctor, University Nicolaus Copernicus, Torun, Poland
[g] Centre for public health and healthcare, Hannover Medical School, Hannover, Germany
[h] Associazione Italiana Medici di Famiglia (AIMEF), Bologna, Italy
[i] Department of English, Nicolaus Copernicus University, Torun, Poland
[j] 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
[k] Primary care research center, Williamson building, Oxford Road, Manchester, United Kingdom
[l] Department of Primary and Interdisciplinary Care, Faculty of Medicine and Health Sciences, Universiteit Antwerpen, Antwerpen, Belgium

ARTICLE INFO

Article history:
Received 28 April 2016
Received in revised form 5 August 2016
Accepted 6 August 2016
Available online

Keywords:
Depression
Systematic review of literature
Diagnostic tool
Primary Care Research

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.

Abbreviations: 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, 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.

* Corresponding author. Tel.: +33 6 07 63 14 90.
** Co-corresponding author.
E-mail addresses: patrice.nabbe@univ-brest.fr (P. Nabbe), lereste@univ.brest.fr (J.Y. Le Reste).

http://dx.doi.org/10.1016/j.eurpsy.2016.08.004
0924-9338/

1. 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 worldwide (from 2.2 to 10.4%), probably due to conceptual differences and different objectives when diagnosing [3,5–8] and sociodemographic 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 focused on major depressive disorder according the DSM. Bipolar depressive disorders, are not covered by the fields in this research and have not been the concern of this research [16,17]. Minor depressive disorders or mood disorders have not been taken into account because the diagnosis is not clearly defined, particularly where older patients are concerned [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 [23]. Combining DSM with skilled professionals (psychiatrist, psychologists...) creates a robust standard or an external criterion with which to evaluate the tools for research purposes.

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

2. Method

2.1. Systematic review according to the PRISMA Guidelines [24,25]

2.1.1. Research group constitution

An international group of researchers in primary care, including a psychiatrist, from France, Spain, Portugal, Catalonia (Spain), 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.

2.1.2. Step 1: inclusion of articles

2.1.2.1. 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.

2.1.2.2. 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 [26];
- 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 the post-partum experience. However, in accordance with the demands of our ethical committee, and because there is a scientific debate to discover whether there is a significant difference between perinatal and other forms of depression, perinatal depression was not retained [27,28];

- the study was not in a primary care setting;
- the tools were identified without validity data.

2.1.2.3. 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.

2.1.3. Step 2: data extraction and selection of tools

A team of two researchers analysed 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 [29]. Researchers ensured that validity data was calculated on the findings of each individual study and not extracted from elsewhere.

3. Results

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

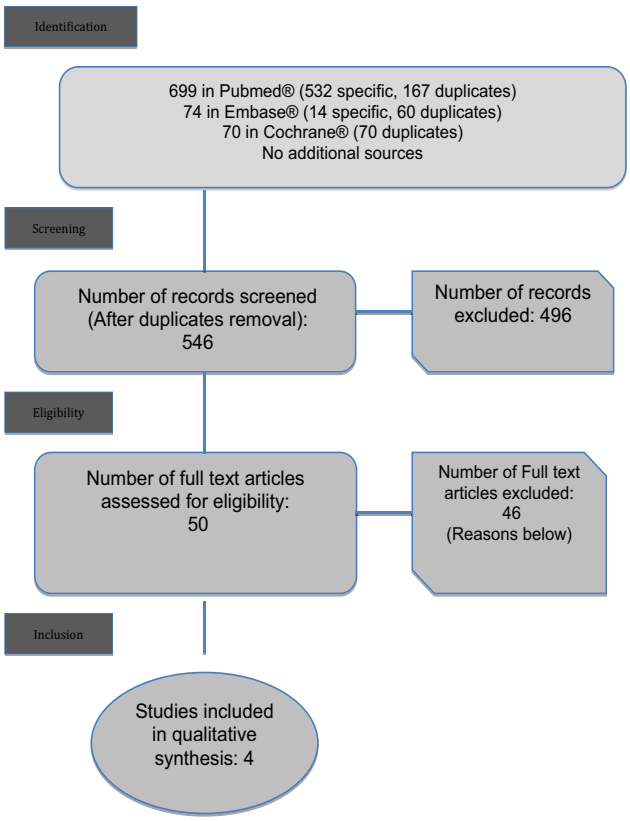


Fig. 1. Articles inclusion (related to PRISMA guideline).

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.

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).

3.1. 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.

3.2. 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.

3.3. 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.

3.4. Concerning the study of Fröjdh

This concerned a Swedish population over 65 years old; 37 individuals participated in a sample 58 individuals out of 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.

4. 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 [30], PSC-51 and HADS [31], the GDS-5 and GDS-15 [32], HSCL-25 [33].

4.1. Comparison with existing literature

Retaining a psychiatric examination based on the DSM was an effective means of comparing the efficiency of the tools [34]. 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 [35] is validated against a population-based mathematical model and not against clinical comparison criteria [36]. The PHQ-9, which is also very popular, is a follow-up tool [37], validated against the Hamilton Scale for follow-up and often used as a diagnostic tool [38]. 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 [39] 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 [40] 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 [41]. 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 [42] 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 [43,44]. The GDS-5 and GDS-15 are short versions of the GDS-30 [45,46] designed for better ergonomic use. The CESD-R was developed in 1977 [47] 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 [48] in research. The PSC-51 is a 51-item physical symptoms list. PSC-51 is little used [31]. The HADS was developed in 1983 to diagnose and quantify depression in hospital [49]. 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 [50] in research. The HSCL-25 was developed in 1974 to diagnose and quantify depression [51]. 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 [52,53].

4.2. 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

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	Mean Age
Usefulness of two instruments in assessing depression among elderly Mexicans in population studies and for primary care	Sánchez-García, Juárez-Cedillo & al. [30]	2008	CES-DR	Semi-structured tool based on the DMS-IV	Yes	206	0.82	0.49	0.5	0.88	0.31	71.2
The role of comorbidity in the detection of psychiatric disorders with checklists for mental and physical symptoms in primary care	De Waal, Arnold & al. [31]	2009	GDS HADS	SCAN 2.1 based on DSM-IV	Yes	473	0.54 0.65	0.79 0.79	0.61	0.74	0.33 0.44	48.8
Validation of 5 and 15 items Spanish version of the geriatric depression scale in elderly subjects in primary health care setting	Ortega-Orcos, Salinero Fort & al. [32]	2007	PSC-51 GDS-5	Clinical Diagnosis of Depression (using DSM-IV criteria)	Yes	301	0.9 0.86	0.59 0.87	0.51	0.97	0.49 0.72	74.3
The Hopkins Symptom Checklist-25 is a sensitive case-finder of clinically important depressive states in elderly people in primary care	Fröjdh, Hakansso & al. [33]	2004	GDS-15 HSCL-25	MADRS based on DSM-IV	Yes	74	0.82 0.94	0.98 0.94	0.86	0.97	0.79 0.88	78.5

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.

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).

4.3. 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.

4.4. Research implications

The studies collected by the systematic literature review involved adult patients. Only 1 study in 4 has a wide age range, between 20 and 80 years. The other 3 studies included a population over 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 [54]? In future studies, when discriminating within a population, there are other factors which should be taken into account, apart from age, for example, the ability to cope, which is not only age-related [55]. 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 [40]. 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 [56].

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 [46,57] 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 analysing 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.

5. 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.

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.

Financial disclosures

The study had a grant of 8000 Euros from the European General Practice Research Network (EGPRN).

Disclosure of interest

The authors declare that they have no competing interest.

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.

References

[1] Licht-Strunk E, Windt D, Van Der, Van Der Windt DAWM, Van Marwijk HWJ, De Haan M, et al. The prognosis of depression in older patients in general practice and the community. A systematic review. Fam Pract 2007;24:168–80.
[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;114:310–5.
[3] Mitchell AJ, Vaze A, Rao S. Clinical diagnosis of depression in primary care: a meta-analysis. Lance 2009;374:609–19.
[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, Vázques-Barquero JL, Dowrick C, Lehtinen V, Dalgard OS, Casey P, et al. Depressive disorders in Europe: Prevalence figures from the ODIN study. Br J Psychiatry 2001;179:308–16.
[6] Alonso J, Angermeyer MC, Bernert S, Bruffaerts R, Brugha TS, Bryson 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;109:21–7.
[7] Paykel ES, Brugha T, Fryers T. Size and burden of depressive disorders in Europe. Eur Neuropsychopharmacol 2005;15(4):411–23.
[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, et al. 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;438–44.
[13] Licht CMM, De Geus EJC, Seldenrijk A, Van Hout HPJ, Zitman FG, Van Dyck R, et al. Depression is associated with decreased blood pressure, but antidepressant use increases the risk for hypertension. Hypertension 2009;53:631–8.
[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;44:550–5.
[16] Smith DJ, Griffiths E, Kelly M, Hood K, Craddock N, Simpson S. 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;381:1663–71.
[18] Gallarda T, Lôo H. Dépression et personnes âgées. Encephale 2009;35(3):269–80.
[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.
[21] Mittal VA, Walker EF. Diagnostic and statistical manual of mental disorders. Psychiatry Res 2011;158–9.
[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 2000;50:284–7.
[23] Zimmerman M, Galione J. Psychiatrists' and nonpsychiatrist physicians' reported use of the DSM-IV criteria for major depressive disorder. J Clin Psychiatry 2010;71:235–8.
[24] Counsell C. Formulating questions and locating primary studies for inclusion in systematic reviews. Ann Intern Med 1997;127:380–7.
[25] 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;70:642–53.
[26] Sollaci LB, Pereira MG. The introduction, methods, results, and discussion (IMRAD) structure: a fifty-year survey. J Med Libr Assoc 2004;92:364–7.
[27] 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.

[28] Whiffen VE. The comparison of postpartum with non-postpartum depression: a rose by any other name. J Psychiatr Neurosci JPN 1991;160–5.
[29] Smits N. A note on Youden's J and its cost ratio. BMC Med Res Methodol 2010;10:89.
[30] Sánchez-García S, Juárez-Cedillo T, García-González JJ, Espinel-Bermúdez C, Gallo JJ, Wagner FA, et al. Usefulness of two instruments in assessing depression among elderly Mexicans in population studies and for primary care. Salud pública Méx 2008;50:447–56.
[31] 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.
[32] 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;207:559–62.
[33] Fröjdh K, Håkansson A, Karlsson I, Frojdh 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;19:386–90.
[34] Perkins NJ, Schisterman EF. The Youden Index and the optimal cut-point corrected for measurement error. Biometrical J Biometrische Zeitschrift 2005;47:428–41.
[35] Terluin B, Van Marwijk HWJ, Adèr HJ, de Vet HCW, Penninx BWJH, Hermens MLM, et al. 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.
[36] Chambe J, Le Reste J, Maisonneuve H, Sanselme A, Oho-Mpondo J, Nabbe P, et al. Evaluating the validity of the French version of the Four-Dimensional Symptom Questionnaire with differential item functioning analysis. Fam Pract 2015;32:474–9.
[37] Kroenke K, Spitzer RL, Williams JBW, Löwe B. The patient health questionnaire somatic, anxiety, and depressive symptom scales: a systematic review. Gen Hosp Psychiatry 2010;32:345–59.
[38] Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960;23:56–62.
[39] Gelaye B, Williams MA, Lemma S, Deyessa N, Bahretibeb Y, Shibre T, et al. Validity of the Patient Health Questionnaire-9 for depression screening and diagnosis in East Africa. Psychiatry Res 2013;210:653–61.
[40] Kendell R, Jablensky A. Distinguishing between the validity and utility of psychiatric diagnoses. Am J Psychiatry 2003;160:4–12.
[41] 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.
[42] Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, et al. Development and validation of a geriatric depression screening scale: a preliminary report. J Psychiatr Res 1983;17:37–49.
[43] 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.
[44] 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;19:465–71.
[45] 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;16:914–21.
[46] 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;35:133–7.
[47] Radloff LS. The CES-D Scale: A Self-Report Depression Scale for Research in the general population. Appl Psychol Meas 1977;1:385–401.
[48] 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;186:128–32.
[49] Zigmond AS, Snaith RP. Hospital Anxiety and Depression Scale (HADS). Ann Gen Psychiatry 1983;67:361–70.
[50] 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.
[51] 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.
[52] Ekblad S, Roth G. Diagnosing posttraumatic stress disorder in multicultural patients in a Stockholm psychiatric clinic. J Nerv Ment Dis 1997;185:102–7.
[53] Jones L. Exposure to political violence and psychological well-being in bosnian adolescents: a mixed method approach. Clin Child Psychol Psychiatr 2005;157–76.
[54] Levasseur S. Vieillissement de la population active. Rev OFCE 2015;6:339–70.
[55] Le Reste JY, Nabbe P, Manceau B, Lygidakis C, Doerr C, Lingner H, et al. 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. JAMA 2013;14(5):319–25.
[56] 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 2014;152:65–75.
[57] Mokkink LB, Terwee CB, Gibbons E, Stratford PW, Alonso J, Patrick DL, et al. 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.

Nabbe et al. BMC Res Notes (2018) 11:4
https://doi.org/10.1186/s13104-017-3111-x

BMC Research Notes

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 Reste[1], M. Guillou-Landreat[2], E. Beck-Robert[1], R. Assenova[3], D. Lazic[4], S. Czachowski[5], S. Stojanović-Špehar[6], M. Hasanagic[7], H. Lingner[8], A. Clavería[9], M. I. Fernandez San Martin[10], A. Sowinska[11], S. Argyriadou[12], C. Lygidakis[13], B. Le Floch[1], C. Doerr[14], T. Montier[15], H. Van Marwijk[16] and P. Van Royen[17]

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

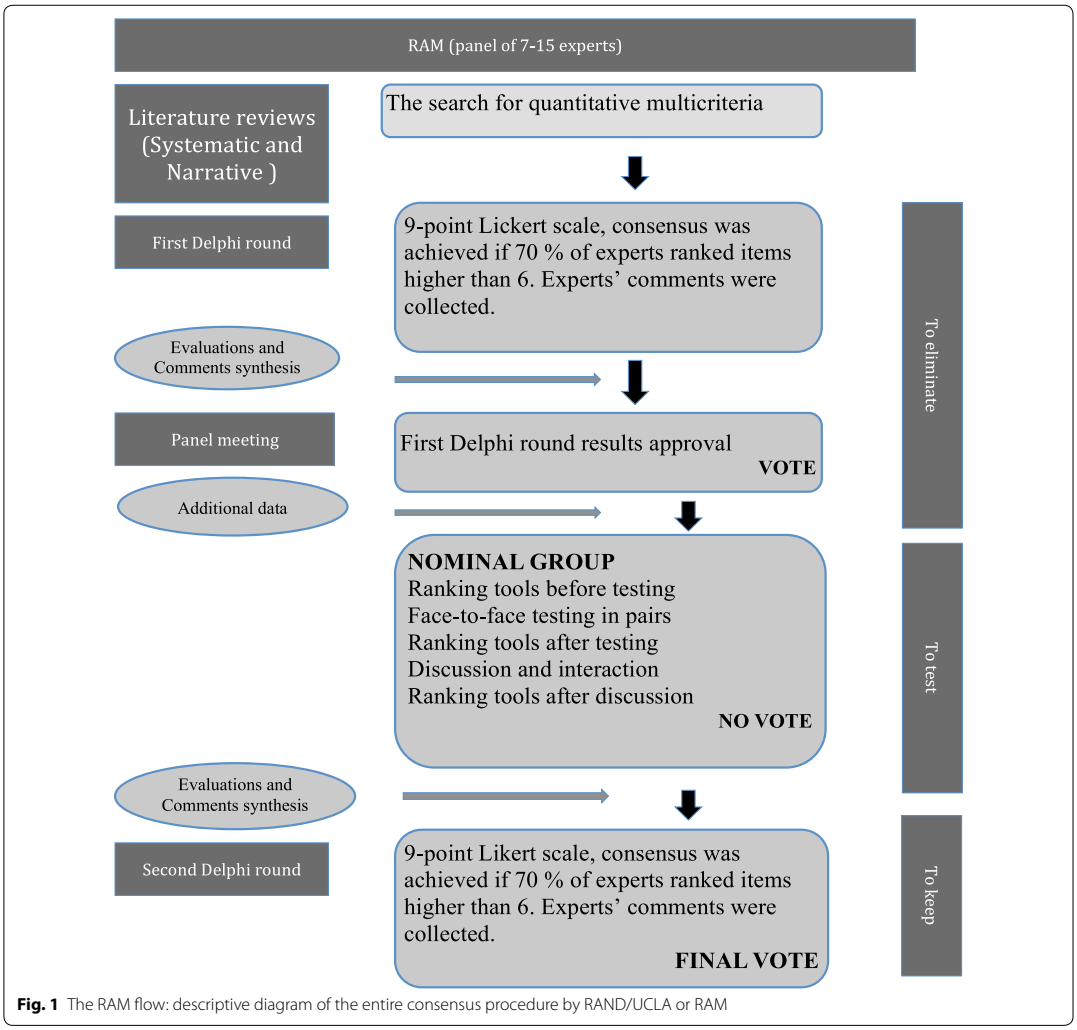


Fig. 1 The RAM flow: descriptive diagram of the entire consensus procedure by RAND/UCLA or RAM

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 transversal 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 Stampar", 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 (ELEGEIA), 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 proofreader 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. enregisterment 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

Received: 20 October 2017 Accepted: 21 December 2017
Published online: 03 January 2018

References

1. 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.
2. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med. 2006;3:2011–30.
3. World Health Organization. The World Health Report 2001: Mental Health: new understanding. New Hope: World Health Organization; 2001.
4. 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.
5. 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.
6. Zhang J, Patel VL, Johnson TR, Shortliffe EH. A cognitive taxonomy of medical errors. J Biomed Inform. 2004;37:193–204.
7. 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.
8. 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.
9. Nabbe P, Le Reste JY, Guillou-Landreat 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.
10. 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.
11. 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.
12. HAS, Haute Autorité Santé. Bases méthodologiques pour l'élaboration de recommandations professionnelles par consensus formalisé. Saint-Denis La Plaine: HAS; 2006.
13. 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.
14. Letrilliart L, Vanmeerbeek M. À la recherche du consensus : quelle méthode utiliser? Exercer. 2011;99:170–7.
15. McGory ML, Shekelle PG, Ko CY. Development of quality indicators for patients undergoing colorectal cancer surgery. J Natl Cancer Inst. 2006;98:1623–33.
16. Skulmoski GJ, Hartman FT, Krahn J. The delphi method for graduate research. J Inf Technol Educ. 2007;6:1.
17. 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.
18. 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.
19. Derogatis LR, Unger R, Derogatis LR, Unger R. Symptom checklist-90-revised. The corsini encyclopedia of psychology. Hoboken: Wiley, Inc.; 2010.
20. Lipman RS, Covi L, Shapiro AK. The Hopkins Symptom Checklist (HSCL)–factors derived from the HSCL-90. J Affect Disord. 1979;1:9–24.
21. 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.
22. 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.
23. 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;57:119–23.
24. Hignett S, Carayon P, Buckle P, Catchpole K. State of science: human factors and ergonomics in healthcare. Ergonomics. 2013;56:1491–503.
25. 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;1:105–16.
26. 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.
27. 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.
28. 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.
29. Munk-Jørgensen P, Fink P, Brevik JI, Dalgard OS, Engberg M, Hansson L, Holm M, 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.
30. Zigmond AS, Snaith RP. Hospital Anxiety and Depression Scale (HADS). Ann Gen Psychiatry. 1983;67:361–70.
31. 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.
32. 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.
33. Andrews B, Hejdenberg J, Wilding J. Student anxiety and depression: comparison of questionnaire and interview assessments. J Affect Disord. 2006;95:29–34.
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.
35. De Waal MWM, Arnold IA, Spinhoven P, Eekhof JAH, Assendelft WJJ, Van Hemert AM, De Waal MWM, Arnold ÆIA, Spinhoven ÆP, Eekhof ÆJAH, Van Hemert ÆAM. 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.
36. 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.
37. 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.

Validation of the French Translation of the Scale HSCL25 in the Diagnosis of Depression in Primary Care (HSCL25)

This study has been completed.

Sponsor:
University Hospital, Brest

Information provided by (Responsible Party):
University Hospital, Brest

ClinicalTrials.gov Identifier:
NCT02414711

First received: March 19, 2015
Last updated: January 11, 2017
Last verified: January 2017
History of Changes

Full Text View | Tabular View | No Study Results Posted | Disclaimer | How to Read a Study Record

Purpose

Depression has emotional consequences, social and important business. Reaching at least 7.5% of French aged 15 to 85 years. General Practitioners are the first professional consulted in terms of frequency of consultations. Indeed, in primary care, 10 to 25% of patients with depression that General Practitioners are often alone to support and follow. But General Practitioners are faced with diagnostic difficulties facing the diagnosis of depression, because if the patient does not engage itself, no tool is validated in France to help them in their diagnosis. However, standardized questionary exist that allow to overcome this difficulty. Questionary used in other european countries. Among them, the HSCL25 scale was selected by a systematic review of literature and a RAND UCLA type of consensus procedure. This HSCL25 scale is the most efficient for the diagnosis of depression in terms of efficiency tool, reproducibility and ergonomics. However, it remains necessary to confirm its validity in terms of the French translation and in terms of efficiency.

Condition	Intervention
Depression	Other: HSCL25 scale Other: PSE9 questionary

Study Type: Interventional
Study Design: Allocation: Randomized
Intervention Model: Single Group Assignment
Masking: Open Label

Official Title: Validation of the French Translation of the Scale HSCL25 in the Diagnosis of Depression in Primary Care

Further study details as provided by University Hospital, Brest:

Primary Outcome Measures:

- To validate the French translation of the HSCL25 scale [Time Frame: 1 month]
  The score between HSCL25 scale and PSE9 questionary will be compared
- To assess the HSCL25 scale diagnosis efficacy of depression in primary care ambulatory practice [Time Frame: 1 month]
  The score between HSCL25 scale and PSE9 questionary will be compared

Enrollment: 1128
Study Start Date: June 2015
Study Completion Date: October 2016
Primary Completion Date: February 2016 (Final data collection date for primary outcome measure)

Arms	Assigned Interventions
No depressed If the result of the HSCL25 scale is inferior than 1.75, then the patient is considered as no depressed. 50 of these patients have to have a psychological interview with the PSE9 questionary to validate the result of the HSCL25 scale.	Other: HSCL25 scale The patient must answer to 25 questions of the HSCL25 scale. In function of the results, the patients was diagnosed depressed (≥1.75) or not depressed (<1.75) Other: PSE9 questionary On the 900 patients recruited, 50 patients who was diagnosed depressed or no depressed have to have a psychological interview with the PSE9 questionary to validate the first diagnosis done by the HSCL25 scale

Depressed If the result of the HSCL25 scale is superior or equal to 1.75, then the patient is considered as depressed. 50 of these patients have to have a psychological interview with the PSE9 questionary to validate the result of the HSCL25 scale.	Other: HSCL25 scale The patient must answer to 25 questions of the HSCL25 scale. In function of the results, the patients was diagnosed depressed (≥1.75) or not depressed (<1.75) Other: PSE9 questionary On the 900 patients recruited, 50 patients who was diagnosed depressed or no depressed have to have a psychological interview with the PSE9 questionary to validate the first diagnosis done by the HSCL25 scale

Eligibility

Ages Eligible for Study: 18 Years and older (Adult, Senior)
Sexes Eligible for Study: All
Accepts Healthy Volunteers: No

Criteria

Inclusion Criteria:

- Adult (18 and over) consultant in primary care

Exclusion Criteria:

- Child or young person under 18
- Women with a declared pregnancy to prevent puerperal depression whose support is different.
- Adult consultant to obtain a medical certificate
- Psychotic patients and\or requiring immediate care

Contacts and Locations

Choosing to participate in a study is an important personal decision. Talk with your doctor and family members or friends about deciding to join a study. To learn more about this study, you or your doctor may contact the study research staff using the Contacts provided below. For general information, see Learn About Clinical Studies.

Please refer to this study by its ClinicalTrials.gov identifier: NCT02414711

Locations

France

Cabinet médical en zone urbaine
Brest, France, 29200

Cabinet médical en zone rurale - Pôle Santé Universitaire
Lanmeur, France, 29620

Cabinet médical en zone semi-rurale
Plounéour Trez, France, 29890

Sponsors and Collaborators

University Hospital, Brest

Investigators

Principal Investigator: Patrice NABBE, GP GP department

More Information

Responsible Party: University Hospital, Brest
ClinicalTrials.gov Identifier: NCT02414711 History of Changes
Other Study ID Numbers: HSCL25 Validation
Study First Received: March 19, 2015
Last Updated: January 11, 2017

Keywords provided by University Hospital, Brest:
Depression
Primary care
General Practitioners

Additional relevant MeSH terms:
Depression
Depressive Disorder
Behavioral Symptoms
Mood Disorders
Mental Disorders

ClinicalTrials.gov processed this record on March 03, 2017

Titre : Un outil diagnostique consensuel de la dépression pour servir de nombreux pays: un défi!

Mots clés : Depression, Outil diagnostique, médecine générale

Résumé : Introduction: La dépression est un motif fréquent de consultation en médecine générale. Sa variabilité rend son diagnostique difficile. Un outil diagnostique efficace, stable et ergonomique serait une aide en recherche en médecine générale. L'objectif de cette étude était de trouver un outil consensuel entre médecins généralistes (MG) et psychiatres sur plusieurs pays européens.

Méthodes: Une revue systématique de littérature été entreprise pour trouver les outils validés en médecine générale contre le psychiatre. Un consensus selon une RAM en a sélectionné un. Il été traduit selon une procédure garantissant la stabilité du transfert et l'écologie en médecine générale. Un protocole de validation a été produit pour s'assurer de la conservation des qualités psychométriques. L'étude de validation externe française a été realise.

Résultats: Sept outils ont été extraits : CESD-R, GDS 5-15-30 questions, PSC-51, HADS, HSCL-25. Les données psychométriques d'efficacité (Se, Sp, VPP, VPN) ont été colligés. La HSCL-25 a été sélectionnée pour ses hautes qualités combinées d'efficacité, de stabilité et d'ergonomie. Elle a été traduite en 9 langues relatives à 3 groupe linguistiques : le grecque, les langues romanes et slaves. L'étude de validation française a prouvé que la forme française de la HSCL-25 (F-HSCL-25) a de hautes performances diagnostiques (Se 59,4%, Sp 91,4%, VPP 69,8%, VPN 86,9%) adaptées à la recherche en médecine générale.

Implication: la HSCL-25 est un outil valide et efficace pour le diagnostic de la dépression en soins primaires. Ils pourraient augmenter les performances diagnostiques des MG et favoriser des recherches collaboratives.

Title : One consensual depression diagnosis tool to serve many countries: a challenge!

Keywords : Depression, Diagnostic tool, General Practice

Abstract : Depression is a common reason for consultation in general practice. Its variability makes its diagnosis difficult. An effective, reliable and ergonomic diagnostic tool would be an aid to research in general practice. The aim of this study was to find a consensual tool between general practitioners (GPs) and psychiatrists in several European countries.

Methods: A systematic literature review was undertaken to find validated tools in general practice against the psychiatrist. A consensus according to a RAM (RAND/UCLA Appropriateness Method) has selected one. It has been translated according to a procedure guaranteeing the stability and the ecology in general practice. A validation protocol has been produced to ensure the retention of psychometric qualities. The French external validation study was carried out.

Results: Seven tools were extracted: CESD-R, GDS 5-15-30 items, PSC-51, HADS, HSCL-25. Psychometric effectiveness data (Se, Sp, VPP, VPN) were collected. The HSCL-25 has been selected for its high combined qualities of effectiveness, reliability and ergonomics. It has been translated into 9 languages relating to 3 linguistic groups: Greek, Romance and Slavic languages. The French Validation Study has proven that the French form of HSCL-25 (F-HSCL-25) has high diagnostic performance (Se 59.4%, Sp 91.4%, VPP 69.8%, and VPN 86.9%) adapted to research in general practice.

Implication: HSCL-25 is a valid and effective tool for diagnosing depression in primary care. They could increase the diagnostic performance of GPs and foster collaborative research.