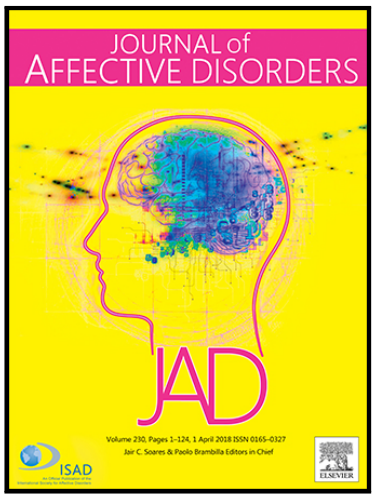
Screening for depression in primary care with Patient Health Questionnaire-9 (PHQ-9): A systematic review



Screening for depression in primary care with Patient Health Questionnaire-9 (PHQ-9): a systematic review

Luigi Costantini , Cesira Pasquarella , Anna Odone , Maria Eugenia Colucci , Alessandra Costanza , Gianluca Serafini , Andrea Aguglia , Martino Belvederi Murri , Vlasios Brakoulias , Mario Amore , S. Nassir Ghaemi , Andrea Amerio

PII: S0165-0327(20)32828-7
DOI: https://doi.org/10.1016/j.jad.2020.09.131
Reference: JAD 12535

To appear in: *Journal of Affective Disorders*

Received date: 12 March 2020
Revised date: 17 August 2020
Accepted date: 27 September 2020

Please cite this article as: Luigi Costantini , Cesira Pasquarella , Anna Odone , Maria Eugenia Colucci , Alessandra Costanza , Gianluca Serafini , Andrea Aguglia , Martino Belvederi Murri , Vlasios Brakoulias , Mario Amore , S. Nassir Ghaemi , Andrea Amerio , Screening for depression in primary care with Patient Health Questionnaire-9 (PHQ-9): a systematic review, *Journal of Affective Disorders* (2020), doi: https://doi.org/10.1016/j.jad.2020.09.131

Screening for depression in primary care with Patient Health Questionnaire-9 (PHQ-9): a systematic review

Running title: Screening for depression in primary care with PHQ-9.

Luigi Costantini[a1], Cesira Pasquarella[a], Anna Odone[b], Maria Eugenia Colucci[a], Alessandra Costanza[c,d], Gianluca Serafini[e,f], Andrea Aguglia[e,f], Martino Belvederi Murri[g], Vlasios Brakoulias[h], Mario Amore[e,f], S. Nassir Ghaemi[i,j], Andrea Amerio[e,f,i]

a. Department of Medicine and Surgery, University of Parma, Parma, Italy
b. Vita-Salute San Raffaele University, Milan, Italy
c. Department of Psychiatry, Faculty of Medicine, University of Geneva (UNIGE), Geneva, Switzerland
d. Department of Psychiatry, ASO Santi Antonio e Biagio e Cesare Arrigo Hospital, Alessandria, Italy
e. Department of Neuroscience, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health (DINOGMI), Section of Psychiatry, University of Genoa, Genoa, Italy
f. IRCCS Ospedale Policlinico San Martino, Genoa, Italy
g. Institute of Psychiatry, Department of Biomedical and Specialty Surgical Sciences, University of Ferrara, Ferrara, Italy

[1] Corresponding author at: Department of Medicine and Surgery, University of Parma, c/o Ospedale Maggiore, Via A. Gramsci 14, 43126 Parma, Italy. Phone: +39 0521 903831 Fax: +39 0521 903832. Email address: luigi.costantini1@studenti.unipr.it (L. Costantini)

h. School of Medicine, Western Sydney University, Blacktown Hospital, Sydney, NSW, Australia

i. Mood Disorders Program, Tufts Medical Center, Boston, MA, USA

j. Department of Psychiatry, Tufts University Medical School, Boston, MA, USA

HIGHLIGHTS

- Patient Health Questionnaire 9 (PHQ-9) has been widely validated for depression screening in primary care in high- and low-income countries.
- A two-stage screening is recommended for depression.
- A Mental Health Professional (MHP) should confirm the diagnosis by use of a semi-structured diagnostic interview.
- Systematic review according to PRISMA statement.

ABSTRACT Word count: 223

Background: Depression is a leading cause of disability. International guidelines recommend screening for depression and the Patient Health Questionnaire 9 (PHQ-9) has been identified as the most reliable screening tool. We reviewed the evidence for using it within the primary care setting.

Methods: We retrieved studies from MEDLINE, Embase, PsycINFO, CINAHL and the Cochrane Library that carried out primary care-based

depression screening using PHQ-9 in populations older than 12, from 1995 to 2018.

Results: Forty-two studies were included in the systematic review. Most of the studies were cross-sectional (N=40, 95%), conducted in high-income countries (N=27, 71%) and recruited adult populations (N=38, 90%). The accuracy of the PHQ-9 was evaluated in 31 (74%) studies with a two-stage screening system, with structured interview most often carried out by primary care and mental health professionals. Most of the studies employed a cut-off score of 10 (N=24, 57%, total range 5 – 15). The overall sensitivity of PHQ-9 ranged from 0.37 to 0.98, specificity from 0.42 to 0.99, positive predictive value from 0.09 to 0.92, and negative predictive value from 0.8 to 1.

Limitations: Lack of longitudinal studies, small sample size, and the heterogeneity of primary-care settings limited the generalizability of our results.

Conclusions: PHQ-9 has been widely validated and is recommended in a two-stage screening process. Longitudinal studies are necessary to provide evidence of long-term screening effectiveness.

Key words: depression, screening, primary care, PHQ-9.

Word count: 3257

INTRODUCTION

Depression represents a significant contributor to the global burden of disease and affects more than 300 million people in all communities across the world (World Health Organization, 2018). One in five people experiences a period of depression in their lives and it is the leading cause of disability worldwide. Burden of disease is a complex concept with different connotations, and covers the burden to the patient, caregiver, the health system, society and economy. Aside from the personal cost to sufferers and their families, the impact on the economy is vast, with the cost in Europe alone amounting to €92 billion a year, much of which is down to lost productivity (The Economist, 2014). Conversely, the recent economic crisis has overloaded the burden of mental disease and posed a further challenge to the prevention of psychiatric disorders (Odone et al., 2018)

International guidelines recommend screening for depression starting from primary care settings (Siu et al., 2016), while some concerns about possible harms of a massive screening have been raised (Thombs et al., 2012). A broad variety of depression screening tools have been proposed and validated. Nevertheless, there is urgent need of choosing one tool to reach a standardized and globally accepted approach (El-Den et al., 2018). Recently, the 9-item version of the Patient Health Questionnaire (PHQ-9) has been identified as the most reliable screening tool for depression (El-Den et al., 2018; Levis et al., 2019).

In recent years, the research around the diagnostic accuracy and psychometric properties of PHQ-9 has flourished within the international scientific community and several studies have been published (El-Den et al., 2018; Levis et al., 2017; Manea et al., 2017; Wu et al., 2019). This systematic review is the first to investigate how screening has been implemented in primary care settings using the PHQ-9.

Aim of the study

We systematically reviewed the literature to determine the clinical utility of the PHQ-9 as a screening tool for major depressive disorder within the primary care setting.

MATERIAL AND METHODS

This systematic review was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Liberati et al., 2009), as previously done (Amerio et al., 2016; Amerio et al., 2018).

Information sources and search strategy

Studies were identified searching the electronic databases MEDLINE, Embase, PyscInfo, CINAHL, and the Cochrane Library. We combined free text terms and MeSH heading as following: ((primary[tiab] AND (care*[tiab] OR healthcare*[tiab] OR health[tiab])) OR ((general[tiab] OR

family[tiab]) AND (practitioner*[tiab] OR physician*[tiab] OR medic*[tiab])) OR GP[tiab] OR "Physicians, Primary Care"[Mesh] OR "Primary Health Care"[Mesh]) AND (Screening[tiab] OR (screening[tiab] AND (tool*[tiab] OR test*[tiab] OR instrument*[tiab] OR scale*[tiab] OR intervention*[tiab])) OR “secondary prevention”[tiab] OR "Mass Screening"[Mesh]) AND (“Patient Health Questionnaire”[tiab] OR PHQ*[tiab] OR "Patient Health Questionnaire"[Mesh]) AND (Depress*[tiab] OR ((unipolar[tiab] OR major[tiab]) AND (depress*[tiab] OR (“mood disorder*”[tiab]))) OR "Depression/prevention and control"[Mesh] OR “Depression”[Mesh] OR "Depressive Disorder"[Mesh]). The strategy was first developed in MEDLINE and then adapted for use in the other databases (Appendix). Studies in English, published from January 1st, 1995 to October 31st, 2018 were included. In addition, further studies were retrieved from reference listing of relevant articles and consultation with experts in the field.

Study population and study designs

We considered studies recruiting participants from primary care settings that focused on PHQ-9 screening of major depressive disorder (American Psychiatric Association, 2013; World Health Organization, 2004) in primary care settings. Studies conducted using other screening tools were excluded. Studies examining populations of both sexes older than 12 years of age were included.

Studies that focused on specific populations or that were carried out in specialized settings (e.g. hospital inpatient specialties) were excluded.

Studies that compared the PHQ-9 with a diagnostic tool based on DSM or ICD were included as well as studies that performed a screening without diagnostic comparison.

Both observational and experimental studies were included. Grey literature was considered. Secondary literature reports and book chapters were excluded. Studies included in former relevant systematic reviews and meta-analyses were individually evaluated. Studies not applying correct epidemiological methodology were excluded.

Outcomes

Primary outcomes were PHQ-9 sensitivity and specificity for the presence of major depressive disorder according to DSM or ICD criteria. Literature on the PHQ-9 suggests to adopt a cut-off score of 10 in a 2-stage screening, that is consistent with moderate severity of depression symptoms (Kroenke et al., 2010).

However, we also included studies using other cut-off values that yielded sensitivity above specificity, keeping the latter equal or above 75%. These are suggested as the optimal characteristics in order to use PHQ-9 by the authors of the questionnaire (Lowe et al., 2004; Spitzer et al., 1999).

Study selection and data extraction

Identified studies were independently reviewed for eligibility by two authors (LC, AA) in a two-step process: a first screening was performed based on titles and abstracts, then full texts were retrieved for a second screening. Disagreement was resolved by consensus.

Data were extracted by two reviewers (LC, AA) with the supervision of another author (AO) using an ad-hoc developed data extraction spreadsheet.

Data items

Information was extracted from each included study on: 1) study design, time and country of intervention, sample size and possible subsets; 2) demographic characteristics of the sample, such as age, sex, ethnicity, educational level, income, employment status, and health insurance coverage; 3) setting, language, and method of administration of PHQ-9, screening stages, positive and negative aspects highlighted in the reports; 4) reference diagnostic interview, cut-off scores considered, sensitivity, specificity, positive and negative predictive values.

Quality assessment

The same authors who performed data extraction (LC, AA) independently assessed the quality of selected studies using the checklist developed by Downs and Black both for randomized and non-randomized studies (Downs and Black, 1998). Disagreements by reviewers were resolved by consensus. Table 1 shows the quality assessment total score assigned to each study.

RESULTS

Study selection

One thousand fourteen potential studies were identified from the selected databases and after cross-checking references of relevant articles. Six hundred seventy-one studies were retrieved after duplicate removal. Studies were screened and selected as described in Figure 1. The search identified 42 studies that were included in the systematic review.

Study characteristics

Characteristics of included studies are reported in Table 1. Forty (95%) studies were cross-sectional (Ahmad et al., 2016; Azah et al., 2005; Ballou et al., 2016; Becker et al., 2002; Bhatta et al., 2018; Carey et al., 2014; Chen et al., 2016; Chen et al., 2010; Chen et al., 2013; Chen et al., 2006; Cheng and Cheng, 2007; Chowdhury et al., 2004; Fogarty et al. 2008; Ganguly et al., 2013; Gelaye et al., 2013; Gilbody et al., 2007; Harriss et al., 2018; Hong, 2018; Husain et al., 2007; Inagaki et al., 2013; Indu et al., 2018; Karekla et al., 2012; Kohrt et al., 2016; Kroenke et al., 2001; Kujawska-Danecka et al., 2016; Liu et al., 2011; Lotrakul et al., 2008; Lowe et al., 2004; Muñoz-Navarro et al., 2017; Muramatsu et al., 2007; Pilowsky et al., 2006; Rancans et al., 2018; Richardson et al., 2010; Sherina et al., 2012; Spitzer et al., 1999; Sung et al., 2013; Vrublevska et al., 2018; Wulsin et al., 2002; Yeung et al., 2008; Zuithoff et al., 2010), one was prospective cohort study (Aalsma et al., 2018) and one included

prospective, focus-group, and cross-sectional designs (Hanlon et al., 2015). The study sample sizes ranged from 93 to 3417 patients, with a total sample size of 35464 patients.

Studies were conducted between 1997 and 2017. Four studies did not report the time of implementation and were assumed to be carried out two years before their publication dates (Ballou et al., 2016; Chowdhury et al., 2004; Ganguly et al., 2013; Indu et al., 2018).

Demographics

Thirty-eight (90%) studies were carried out on adults, four (10%) on adolescents. With regard to the former subset, twenty-seven (71%) studies were carried out in high income countries (Ahmad et al., 2016; Ballou et al., 2016; Becker et al., 2002; Carey et al., 2014; Chen et al., 2016; Chen et al., 2006; Cheng and Cheng, 2007; Fogarty et al., 2008; Gilbody et al., 2007; Hanlon et al., 2015; Harriss et al., 2018; Hong, 2018; Inagaki et al., 2013; Karekla et al., 2012; Kroenke et al., 2001; Kujawska-Danecka et al., 2016; Liu et al., 2011; Lowe et al., 2004; Muñoz-Navarro et al., 2017; Muramatsu et al., 2007; Pilowsky et al., 2006; Rancans et al., 2018; Richardson et al., 2010; Spitzer et al., 1999; Sung et al., 2013; Vrublevska et al., 2018; Yeung et al., 2008; Zuithoff et al., 2010), as defined by the World Bank (World Bank, 2019). Eighteen (66%) of those 27 studies were conducted in the USA . Three (75%) studies conducted in adolescents were carried out in high income countries. Two of those were conducted in the USA. The comprehensive rate of females across the studies ranged between 64% and 74% in adults, between 46% and 58% in adolescents.

Twenty-two (52%) studies reported additional relevant demographic information, such as educational level (N=18, 43%), ethnic or linguistic composition (N=12, 29%), occupational status (N=8, 19%), health insurance (N=4, 11%), and residence (N=1, 3%) (Ahmad et al., 2016; Aalsma et al., 2018; Becker et al., 2002; Bhatta et al., 2018; Carey et al., 2014; Chen et al., 2016; Chen et al., 2013; Chen et al., 2006; Fogarty et al., 2008; Gelaye et al., 2013; Hanlon et al., 2015; Hong, 2018; Indu et al., 2018; Kohrt et al., 2016; Kroenke et al., 2001; Lotrakul et al., 2008; Muñoz-Navarro et al., 2017; Pilowsky et al., 2006; Rancans et al., 2018; Spitzer et al., 1999; Sung et al., 2013; Vrublevska et al., 2018). According to available data, on a subset of 15852 patients (45% out of the comprehensive sample size), the rate of individuals with educational level higher than Primary Education (UNESCO, 2011) was 71% (N=11247). Data on health insurance coverage were available for a subset of 6603 (20%) patients with a public health insurance coverage rate accounted for 57% (N=3780).

PHQ-9 screening process

Table 2 shows the characteristics of screening process drawn from the included studies, divided by age group to highlight the differences in depression screening among adolescents and adults (Siu et al., 2016).

The majority of the studies were carried out in community-based primary care practices (N=28, 67%); other settings were hospital-based primary care outpatient clinics (N=4, 10%), rural clinics (N=3, 7%), school-based programs (N=3, 7%), community-based prevention programs (N=2, 5%), a private-insurance healthcare facility (N=1, 2%), and a community pharmacy (N=1, 2%).

PHQ-9 was self-reported by patients in 34 studies (81%) and administered as an interview in the remaining eight studies (19%). PHQ-9 was validated in 20 languages besides English.

We retrieved information about implementation stages for 40 (95%) studies. Two studies included an ultra-brief screening scale before PHQ-9 was administered (Aalsma et al., 2018; Chen et al., 2006). The PHQ-9 was administered by General Practitioners (GPs), nurses, or medical students in ten studies (24%) (Becker et al., 2002; Bhatta et al., 2018; Chen et al., 2010; Chen et al., 2013; Chen et al., 2006; Cheng and Cheng, 2007; Gelaye et al., 2013; Spitzer et al., 1999; Sung et al., 2013; Wulsin et al., 2002). Most of the studies (N=31, 74%) adopted a two-stage screening system, in which a clinical interview confirmed or refused the preliminary PHQ-9 assessment. A Mental Health Professional (MHP), who was blind to PHQ-9 results, performed the diagnostic interview in 18 (43%) studies (Azah, 2005; Becker et al., 2002; Chen et al., 2010; Cheng and Cheng, 2007; Chowdhury et al., 2004; Gelaye et al., 2013; Hanlon et al., 2015; Hong, 2018; Indu et al., 2018; Kohrt et al., 2016; Kroenke et al., 2001; Muñoz-Navarro et al., 2017; Pilowsky et al., 2006; Rancans et al., 208; Spitzer et al., 1999; Vrublevska et al., 2018; Yeung et al., 2008; Zuithoff et al., 2010). Some studies developed a protocol for immediate referral of emergent cases such as suicidal ideation (N=2, 5%) (Ballou et al., 2016; Chen et al., 2010), implemented a formal staff training before carrying out the survey (N=5, 12%) (Bhatta et al., 2018; Chen et al., 2010; Chen et al., 2006; Cheng and Cheng, 2007; Chowdhury et al., 2004), and analyzed the staff compliance throughout the screening process (N=1, 2%) (Bhatta et al., 2018).

PHQ-9 Operating Characteristics

Table 3 shows the accuracy data of the PHQ-9 as evaluated in 31 (74%) studies that used different diagnostic interviews on 13459 participants. Fully structured and semi-structured interviews were considered separately. The main standardized diagnostic rating scales used were the Mini-International Neuropsychiatric Interview (MINI) (4004 patients, 30%), the Composite International Diagnostic Interview (CIDI) (2623 patients, 19%), the Structured Clinical Interview for DSM-IV (SCID) (2853 patients, 21%), and the Structured Clinical Assessment in Neuropsychiatry (SCAN) (2529 patients, 19%).

Overall, the cut-off scores ranged from a minimum of 5 to a maximum of 15 points, sensitivity from 0.37 to 0.98, specificity from 0.42 to 0.99, positive predictive value from 0.09 to 0.92, and negative predictive value from 0.8 to 1. A 10-point cut-off was applied in many of the studies (N=24, 57%). Considering 20 studies applying a 10-point cut-off and performing either a fully structured or semi-structured interview, sensitivity was 0.85 or higher in 9 studies (45%) (Bhatta et al., 2018; Chen et al., 2010; Chen et al., 2006; Cheng and Cheng, 2007; Chowdhury et al., 2004; Muñoz-Navarro et al., 2017) and specificity was 0.75 or more in 16 studies (80%) (Azah, 2005; Becker et al., 2002; Chen et al., 2016; Chen et al., 2010; Chen et al., 2013; Gilbody et al., 2007; Inagaki et al., 2013; Kroenke et al., 2001; Liu et al., 2011; Lotrakul et al., 2008; Muramatsu et al., 2007; Rancans et al., 2018; Spitzer et al., 1999; Vrublevska et al., 2018; Wulsin et al., 2002; Zuithoff et al., 2010). Sensitivity was higher than 0.9 in three studies that performed either SCID or CIDI (Gilbody et al., 2007; Kohrt et al., 2016; Muñoz-Navarro et al., 2017).

DISCUSSION

The PHQ-9 has been widely used in different primary care settings for the screening of depression. Most of the included studies were cross-sectional (N=40, 95%), conducted in high income countries (N=27, 71%) in adult population (N=38, 90%). PHQ-9 accuracy was evaluated in 31 (74%) studies with a two-stage screening system carried out by primary care and mental health professionals with either fully structured or semi-structured interviews.

Based on the results of our systematic review some observations can be made.

Cut-off score

The cut-off score approach proved to be more useful than the algorithm approach (He et al., 2019). In the last 20 years many of the researchers have used a cut-off score of 10 or higher, which is also the most represented among the reviewed studies. According to previous reviews, that was consistent with a severity measure of depressive symptoms evaluated with the same questionnaire (Kroenke et al., 2010). A meta-analysis defined acceptable cut-off points between 8 and 11 (Manea et al., 2012). Besides, an individual-participant data meta-analysis demonstrated that a retrospective selection of optimal cut-off led to the paradox of an increasing in sensitivity when the cut-off severity increased (Levis et al., 2017). The operating characteristics were maximized at a 10-point cut-off (Levis et al., 2019). Our review suggests that more variability across studies is to be expected in sensitivity than specificity.

Some authors emphasized the impact of age, context, educational level and social stigma on the choice of the cut-off in each specific setting (Chen et al., 2016; Gelaye et al., 2013). Authors of previous reviews analyzed this issue and recommended that researchers report the operating characteristics for the whole range of possible cut-off scores (Levis et al., 2017; Manea et al., 2012).

Diagnostic interviews

Given the complexity of the spectrum of depressive disorders (Amerio et al., 2014), the use of a structured interview based on DSM (Diagnostic and Statistical Manual of Mental Disorders) (American Psychiatric Association, 2013) criteria is recommended as a reliable way to validate a screening questionnaire. Nevertheless, a recent individual participant data meta-analysis showed that fully structured interviews tended to identify more cases of mild depression, whereas semi-structured interviews were more sensitive to severe cases (Levis et al., 2018).

Differences have also been reported among fully structured interviews: MINI, developed as a rapid diagnostic tool, tended to diagnose depression two times more than CIDI that provided a deeper diagnosis of depression (Levis et al., 2018). Similar issues have been reported in the selected studies (Zuithoff et al., 2010).

Role and training of primary care operators

Few studies clearly reported the role of GPs and other primary care professionals throughout the screening process. Primary care operators should be trained in explaining the meaning of the score to the patients, in order to reduce possible harms from misinterpretation.

GPs should make use of the screening tool to detect and deepen the patient's experience of illness.

Studies that included a staff training or a compliance analysis suggested that such procedures can help addressing organizational factors that might affect screening efficiency and effectiveness.

Integration between mental health and primary care services

Many of the selected studies included a structured diagnostic interview performed by an MHP. That is a screening model that should be recommended in order to increase homogeneity and reliability of the reference test. Therefore, a real integration between Mental Health and Primary Care Services is essential to ensure a prompt patient-centered care. Structured diagnostic interviews should be timely performed, in order to lower the emotional consequences of positive screening results and lead patients to early treatments.

Future directions

New technologies could speed up the screening process, as reported in previous studies (Aalsma et al., 2018; Harriss et al., 2018). A digital implementation could be the best and simple way to administer PHQ-9, increasing the study sample size.

Few studies acted as spin-off examples of PHQ-9 application and adaptability to different settings including community pharmacies (Ballou et al., 2016), school-based programs (Bhatta et al., 2018) and community health campaigns (Harriss et al., 2018; Kujawska-Danecka et al.,

2016). These new settings might be taken in account in the near future and they may be the only access to health care for a substantial part of population. New settings coupled with the use of a valid screening tool provide a valuable opportunity to perform widespread screening for depressive disorders.

LIMITATIONS

The main strengths of this study are the extent of the review, the total sample size and systematic approach that was used to review the literature. A meta-analysis was not possible as the data lacked homogeneity. Also, results may be sensible to the methodological shortcomings of the primary studies. Inclusion of participants with current psychiatric diagnoses and comorbidities or currently taking psychotropic medication could have overestimated the clinical utility of the screening tool (Rice et al., 2016). Lack of longitudinal studies and small sample sizes tended to reduce the power of studies largely affecting their quality. Thirty-two (86%) studies were assigned a score equal or lower than 24/31 on the Downs and Black quality scale (Downs and Black, 1998). The studies adopted different approaches for the PHQ-9 administration and sequential study stages. Moreover, ultra-brief pre-screening yielded high risk of recall bias in two studies.

The heterogeneity of primary care services also limits the generalizability of the results. Primary care services are different across countries and this is more evident when comparing countries with high, middle, and low-income economies.

CONCLUSIONS

The PHQ-9 has been tested extensively for depression screening. It was widely validated as a screening tool in primary care services in such different countries and its psychometric reliability is established by now. Recently, a shorter 8-item equivalent has been validated (Wu et al., 2019). Our systematic review suggests that a two-stage screening carried out by primary care and mental health professionals is recommended. Longitudinal studies are necessary to provide evidence of long-term screening effectiveness.

SUPPLEMENTARY MATERIAL

The complete search strategy is available in the supplementary material of this article.

CONTRIBUTORS

Authors LC, CP, MEC, AC, MBM and AA designed the study and wrote the protocol. Studies were identified and independently reviewed for eligibility by two authors (LC, AA) in a two-step-based process. Data were extracted by two authors (LC, AA) and supervised by a third author (AO) using an ad-hoc developed data extraction spreadsheet. Authors LC, GS, MA, VB, AA, and SNG wrote the first draft of the manuscript.

AUTHOR AGREEMENT

Our manuscript has been approved by all authors.

ACKNOWLEDGMENTS

None.

CONFLICTS OF INTEREST

Dr. Costantini, Prof. Pasquarella, Prof. Odone, Prof. Colucci, Dr. Costanza, Prof. Serafini, Dr. Aguglia, Dr. Belvederi Murri, Prof. Brakoulias, Prof. Amore, and Dr. Amerio report no conflicts of interest. Prof. Ghaemi is employed by Novartis Institutes for Biomedical Research and holds equity in Novartis.

FUNDING SOURCE DECLARATION

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

REFERENCES

Aalsma, M.C., Zerr, A.M., Etter, D.J., Ouyang, F., Gilbert, A.L., Williams, R.L., Hall, J.A., Downs, S.M., 2018. Physician Intervention to Positive Depression Screens Among Adolescents in Primary Care. Journal of Adolescent Health 62, 212-218.

Ahmad, F., Shakya, Y., Ginsburg, L., Lou, W., Ng, P.T., Rashid, M., Ferrari, M., Ledwos, C., McKenzie, K., 2016. Burden of common mental disorders in a community health centre sample. Canadian Family Physician 62, e758-e766.

American Psychiatric Association, 2013. Diagnostic and Statistical Manual of Mental Disorders: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. American Psychiatric Association, Arlington, VA.

Amerio, A., Odone, A., Marchesi, C., Ghaemi, S.N., 2014. Is depression one thing or many? British Journal of Psychiatry 204, 488-488.

Amerio, A., Ossola, P., Scagnelli, F., Odone, A., Allinovi, M., Cavalli, A., Iacopelli, J., Tonna, M., Marchesi, C., Ghaemi, S.N., 2018. Safety and efficacy of lithium in children and adolescents: A systematic review in bipolar illness. European Psychiatry 54, 85-97.

Amerio, A., Stubbs, B., Odone, A., Tonna, M., Marchesi, C., Nassir Ghaemi, S., 2016. Bipolar I and II Disorders; A Systematic Review and Meta-Analysis on Differences in Comorbid Obsessive-Compulsive Disorder. Iran J Psychiatry Behav Sci 10, e3604-e3604.

Azah, M., Shah, M., Juwita, S., Bahri, I., Rushidi, W., Jamil, Y., 2005. Validation of the Malay version brief patient health Questionnaire (PHQ-9) among adult attending family medicine clinics. International Medical Journal 12, 259-263.

Ballou, J., Roark, A., Chapman, A., Huie, C., Marciniak, M., 2016. Implementation of depression screening in an independent community pharmacy. Journal of the American Pharmacists Association 56, e77.

Becker, S., Al Zaid, K., Al Faris, E., 2002. Screening for somatization and depression in Saudi Arabia: A validation study of the PHQ in primary care. International Journal of Psychiatry in Medicine 32, 271-283.

Bhatta, S., Champion, J.D., Young, C., Loika, E., 2018. Outcomes of Depression Screening Among Adolescents Accessing School-based Pediatric Primary Care Clinic Services. Journal of Pediatric Nursing 38, 8-14.

Carey, M., Jones, K.A., Yoong, S.L., D'Este, C., Boyes, A.W., Paul, C., Inder, K.J., Sanson-Fisher, R., 2014. Comparison of a single self-assessment item with the PHQ-9 for detecting depression in general practice. Family Practice 31.

Chen, I.P., Liu, S.I., Huang, H.C., Sun, F.J., Huang, C.R., Sung, M.R., Huang, Y.P., 2016. Validation of the Patient Health Questionnaire for Depression Screening Among the Elderly Patients in Taiwan. International Journal of Gerontology 10, 193-197.

Chen, S., Chiu, H., Xu, B., Ma, Y., Jin, T., Wu, M., Conwell, Y., 2010. Reliability and validity of the PHQ-9 for screening late-life depression in Chinese primary care. International journal of geriatric psychiatry 25, 1127-1133.

Chen, S., Fang, Y., Chiu, H., Fan, H., Jin, T., Conwell, Y., 2013. Validation of the nine-item Patient Health Questionnaire to screen for major depression in a Chinese primary care population. Asia-Pacific Psychiatry 5, 61-68.

Chen, T.M., Huang, F.Y., Chang, C., Chung, H., 2006. Using the PHQ-9 for depression screening and treatment monitoring for Chinese Americans in primary care. Psychiatric Services 57, 976-981.

Cheng, C.M., Cheng, M., 2007. To validate the Chinese version of the 2Q and PHQ-9 questionnaires in Hong Kong Chinese patients. Hong Kong Practitioner 29, 381-390.

Chowdhury, A.N., Ghosh, S., Sanyal, D., 2004. Bengali adaptation of Brief Patient Health Questionnaire for screening depression at primary care. Journal of the Indian Medical Association 102, 544-547.

Downs, S.H., Black, N., 1998. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. Journal of epidemiology and community health 52, 377-384.

El-Den, S., Chen, T.F., Gan, Y.L., Wong, E., O'Reilly, C.L., 2018. The psychometric properties of depression screening tools in primary healthcare settings: A systematic review. Journal of affective disorders 225, 503-522.

Fogarty, C.T., Sharma, S., Chetty, V.K., Culpepper, L., 2008. Mental health conditions are associated with increased health care utilization among urban family medicine patients. Journal of the American Board of Family Medicine 21, 398-407.

Ganguly, S., Samanta, M., Roy, P., Chatterjee, S., Kaplan, D.W., Basu, B., 2013. Patient health questionnaire-9 as an effective tool for screening of depression among indian adolescents. Journal of Adolescent Health 52, 546-551.

Gelaye, B., Williams, M.A., Lemma, S., Deyessa, N., Bahretibeb, Y., Shibre, T., Wondimagegn, D., Lemenhe, A., Fann, J.R., Vander Stoep, A., Andrew Zhou, X.H., 2013. Validity of the patient health questionnaire-9 for depression screening and diagnosis in East Africa. Psychiatry Research 210, 653-661.

Gilbody, S., Richards, D., Barkham, M., 2007. Diagnosing depression in primary care using self-completed instruments: UK validation of PHQ-9 and CORE-OM. British Journal of General Practice 57, 650-652.

Hanlon, C., Medhin, G., Selamu, M., Breuer, E., Worku, B., Hailemariam, M., Lund, C., Prince, M., Fekadu, A., 2015. Validity of brief screening questionnaires to detect depression in primary care in Ethiopia. Journal of affective disorders 186, 32-39.

Harriss, L.R., Kyle, M., Connolly, K., Murgha, E., Bulmer, M., Miller, D., Munn, P., Neal, P., Pearson, K., Walsh, M., Campbell, S., Berger, M., McDermott, R., McDonald, M., 2018. Screening for depression in young Indigenous people: building on a unique community initiative. Australian Journal of Primary Health 24, 343-349.

He, C., Levis, B., Riehm, K.E., Saadat, N., Levis, A.W., Azar, M., Rice, D.B., Krishnan, A., Wu, Y., Sun, Y., Imran, M., Boruff, J., Cuijpers, P., Gilbody, S., Ioannidis, J.P.A., Kloda, L.A., McMillan, D., Patten, S.B., Shrier, I., Ziegelstein, R.C., Akena, D.H., Arroll, B., Ayalon, L., Baradaran, H.R., Baron, M., Beraldi, A., Bombardier, C.H., Butterworth, P., Carter, G., Chagas, M.H.N., Chan, J.C.N., Cholera, R., Clover, K., Conwell, Y., de Man-van Ginkel, J.M., Fann, J.R., Fischer, F.H., Fung, D., Gelaye, B., Goodyear-Smith, F., Greeno, C.G., Hall, B.J., Harrison, P.A., Harter, M., Hegerl, U., Hides, L., Hobfoll, S.E., Hudson, M., Hyphantis, T.N., Inagaki, M., Ismail, K., Jette, N., Khamseh, M.E., Kiely, K.M., Kwan, Y., Lamers, F., Liu, S.I., Lotrakul, M., Loureiro, S.R., Lowe, B., Marsh, L., McGuire, A., Mohd-Sidik, S., Munhoz, T.N., Muramatsu, K., Osorio, F.L., Patel, V., Pence, B.W., Persoons, P., Picardi, A., Reuter, K., Rooney, A.G., da Silva Dos Santos, I.S., Shaaban, J., Sidebottom, A., Simning, A., Stafford, L., Sung, S., Tan, P.L.L., Turner, A., van Weert, H., White, J., Whooley, M.A., Winkley, K., Yamada, M., Thombs, B.D., Benedetti, A., 2019. The Accuracy of the Patient Health Questionnaire-9 Algorithm for Screening to Detect Major Depression: An Individual Participant Data Meta-Analysis. Psychotherapy and psychosomatics, 1-13.

Hong, S., Heng, Sheng, 2018. Use of patient health questionnaires (phq-9, phq-2 & phq-1) for depression screening in singapore primary care. The Singapore Family Physician 44, 6.

Husain, N., Waheed, W., Tomenson, B., Creed, F., 2007. The validation of personal health questionnaire amongst people of Pakistani family origin living in the United Kingdom. Journal of affective disorders 97, 261-264.

Inagaki, M., Ohtsuki, T., Yonemoto, N., Kawashima, Y., Saitoh, A., Oikawa, Y., Kurosawa, M., Muramatsu, K., Furukawa, T.A., Yamada, M., 2013. Validity of the Patient Health Questionnaire (PHQ)-9 and PHQ-2 in general internal medicine primary care at a Japanese rural hospital: A cross-sectional study. General Hospital Psychiatry 35, 592-597.

Indu, P.S., Anilkumar, T.V., Vijayakumar, K., Kumar, K.A., Sarma, P.S., Remadevi, S., Andrade, C., 2018. Reliability and validity of PHQ-9 when administered by health workers for depression screening among women in primary care. Asian journal of psychiatry 37, 10-14.

Karekla, M., Pilipenko, N., Feldman, J., 2012. Patient health questionnaire: Greek language validation and subscale factor structure. Comprehensive Psychiatry 53, 1217-1226.

Kohrt, B.A., Luitel, N.P., Acharya, P., Jordans, M.J., 2016. Detection of depression in low resource settings: validation of the Patient Health Questionnaire (PHQ-9) and cultural concepts of distress in Nepal. BMC Psychiatry 16, 58.

Kroenke, K., Spitzer, R.L., Williams, J.B., Lowe, B., 2010. The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. Gen Hosp Psychiatry 32, 345-359.

Kroenke, K., Spitzer, R.L., Williams, J.B.W., 2001. The PHQ-9: Validity of a brief depression severity measure. Journal of General Internal Medicine 16, 606-613.

Kujawska-Danecka, H., Nowicka-Sauer, K., Hajduk, A., Wierzba, K., Krzemiński, W., Zdrojewski, Z., 2016. The prevalence of depression symptoms and other mental disorders among patients aged 65 years and older – screening in the rural community. Family Medicine and Primary Care Review 18, 274-277.

Levis, B., Benedetti, A., Levis, A.W., Ioannidis, J.P.A., Shrier, I., Cuijpers, P., Gilbody, S., Kloda, L.A., McMillan, D., Patten, S.B., Steele, R.J., Ziegelstein, R.C., Bombardier, C.H., de Lima Osorio, F., Fann, J.R., Gjerdingen, D., Lamers, F., Lotrakul, M., Loureiro, S.R., Lowe, B., Shaaban, J., Stafford, L., van Weert, H., Whooley, M.A., Williams, L.S., Wittkampf, K.A., Yeung, A.S., Thombs, B.D., 2017. Selective Cutoff Reporting in Studies

of Diagnostic Test Accuracy: A Comparison of Conventional and Individual-Patient-Data Meta-Analyses of the Patient Health Questionnaire-9 Depression Screening Tool. American journal of epidemiology 185, 954-964.

Levis, B., Benedetti, A., Riehm, K.E., Saadat, N., Levis, A.W., Azar, M., Rice, D.B., Chiovitti, M.J., Sanchez, T.A., Cuijpers, P., Gilbody, S., Ioannidis, J.P.A., Kloda, L.A., McMillan, D., Patten, S.B., Shrier, I., Steele, R.J., Ziegelstein, R.C., Akena, D.H., Arroll, B., Ayalon, L., Baradaran, H.R., Baron, M., Beraldi, A., Bombardier, C.H., Butterworth, P., Carter, G., Chagas, M.H., Chan, J.C.N., Cholera, R., Chowdhary, N., Clover, K., Conwell, Y., de Man-van Ginkel, J.M., Delgadillo, J., Fann, J.R., Fischer, F.H., Fischler, B., Fung, D., Gelaye, B., Goodyear-Smith, F., Greeno, C.G., Hall, B.J., Hambridge, J., Harrison, P.A., Hegerl, U., Hides, L., Hobfoll, S.E., Hudson, M., Hyphantis, T., Inagaki, M., Ismail, K., Jette, N., Khamseh, M.E., Kiely, K.M., Lamers, F., Liu, S.I., Lotrakul, M., Loureiro, S.R., Lowe, B., Marsh, L., McGuire, A., Mohd Sidik, S., Munhoz, T.N., Muramatsu, K., Osorio, F.L., Patel, V., Pence, B.W., Persoons, P., Picardi, A., Rooney, A.G., Santos, I.S., Shaaban, J., Sidebottom, A., Simning, A., Stafford, L., Sung, S., Tan, P.L.L., Turner, A., van der Feltz-Cornelis, C.M., van Weert, H.C., Vohringer, P.A., White, J., Whooley, M.A., Winkley, K., Yamada, M., Zhang, Y., Thombs, B.D., 2018. Probability of major depression diagnostic classification using semi-structured versus fully structured diagnostic interviews. The British journal of psychiatry : the journal of mental science 212, 377-385.

Levis, B., Benedetti, A., Thombs, B.D., 2019. Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. BMJ (Clinical research ed.) 365, l1476.

Liberati, A., Altman, D.G., Tetzlaff, J., Mulrow, C., Gotzsche, P.C., Ioannidis, J.P., Clarke, M., Devereaux, P.J., Kleijnen, J., Moher, D., 2009. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS medicine 6, e1000100.

Liu, S.I., Yeh, Z.T., Huang, H.C., Sun, F.J., Tjung, J.J., Hwang, L.C., Shih, Y.H., Yeh, A.W.C., 2011. Validation of Patient Health Questionnaire for depression screening among primary care patients in Taiwan. Comprehensive Psychiatry 52, 96-101.

Lotrakul, M., Sumrithe, S., Saipanish, R., 2008. Reliability and validity of the Thai version of the PHQ-9. BMC Psychiatry 8.

Lowe, B., Spitzer, R.L., Grafe, K., Kroenke, K., Quenter, A., Zipfel, S., Buchholz, C., Witte, S., Herzog, W., 2004. Comparative validity of three screening questionnaires for DSM-IV depressive disorders and physicians' diagnoses. Journal of affective disorders 78, 131-140.

Manea, L., Gilbody, S., McMillan, D., 2012. Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis. Cmaj 184, E191-196.

Muñoz-Navarro, R., Cano-Vindel, A., Medrano, L.A., Schmitz, F., Ruiz-Rodríguez, P., Abellán-Maeso, C., Font-Payeras, M.A., Hermosilla-Pasamar, A.M., 2017. Utility of the PHQ-9 to identify major depressive disorder in adult patients in Spanish primary care centres. BMC Psychiatry 17.

Muramatsu, K., Miyaoka, H., Kamijima, K., Muramatsu, Y., Yoshida, M., Otsubo, T., Gejyo, F., 2007. The patient health questionnaire, Japanese version: validity according to the mini-international neuropsychiatric interview-plus. Psychol Rep 101, 952-960.

Odone, A., Landriscina, T., Amerio, A., Costa, G., 2018. The impact of the current economic crisis on mental health in Italy: evidence from two representative national surveys. Eur J Public Health 28, 490-495.

Pilowsky, D.J., Olfson, M., Gameroff, M.J., Wickramaratne, P., Blanco, C., Feder, A., Gross, R., Neria, Y., Weissman, M.M., 2006. Panic disorder and suicidal ideation in primary care. Depression and Anxiety 23, 11-16.

Rancans, E., Trapencieris, M., Ivanovs, R., Vrublevska, J., 2018. Validity of the PHQ-9 and PHQ-2 to screen for depression in nationwide primary care population in Latvia. Annals of General Psychiatry 17, N.PAG-N.PAG.

Rice, D.B., Thombs, B.D., 2016. Risk of Bias from inclusion of currently diagnosed or treated patients in studies of depression screening tool accuracy: A cross-sectional analysis of recently published primary studies and meta-analyses. PLoS ONE 11.

Richardson, L.P., McCauley, E., Grossman, D.C., McCarty, C.A., Richards, J., Russo, J.E., Rockhill, C., Katon, W., 2010. Evaluation of the patient health questionnaire-9 item for detecting major depression among adolescents. Pediatrics 126, 1117-1123.

Sherina, M.S., Arroll, B., Goodyear-Smith, F., 2012. Criterion validity of the PHQ-9 (Malay version) in a primary care clinic in Malaysia. The Medical journal of Malaysia 67, 309-315.

Siu, A.L., Bibbins-Domingo, K., Grossman, D.C., Baumann, L.C., Davidson, K.W., Ebell, M., García, F.A.R., Gillman, M., Herzstein, J., Kemper, A.R., Krist, A.H., Kurth, A.E., Owens, D.K., Phillips, W.R., Phipps, M.G., Pignone, M.P., 2016. Screening for depression in adults: US preventive services task force recommendation statement. JAMA - Journal of the American Medical Association 315, 380-387.

Spitzer, R.L., Kroenke, K., Williams, J.B., 1999. 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 282, 1737-1744.

Sung, S.C., Low, C.C.H., Fung, D.S.S., Chan, Y.H., 2013. Screening for major and minor depression in a multiethnic sample of Asian primary care patients: A comparison of the nine-item Patient Health Questionnaire (PHQ-9) and the 16-item Quick Inventory of Depressive Symptomatology - Self-Report (QIDS-SR16). Asia-Pacific Psychiatry 5, 249-258.

The Economist, 2014. The Global Crisis of Depression. The Low of 21st Century? Summary Report, The Economist Events, p. 14.

Thombs, B.D., Coyne, J.C., Cuijpers, P., De Jonge, P., Gilbody, S., Ioannidis, J.P.A., Johnson, B.T., Patten, S.B., Turner, E.H., Ziegelstein, R.C., 2012. Rethinking recommendations for screening for depression in primary care. CMAJ 184, 413-418.

UNESCO, 2011. International Standard Classification of Education (ISCED).

Vrublevska, J., Trapencieris, M., Rancans, E., 2018. Adaptation and validation of the Patient Health Questionnaire-9 to evaluate major depression in a primary care sample in Latvia. Nordic journal of psychiatry 72, 112-118.

World Bank, 2019. World Bank List of Economies, 07/2019 ed, Washington.

World Health Organization, 2004. ICD-10 : international statistical classification of diseases and related health problems : tenth revision.

World Health Organization, 2018. Fact Sheet on Depression.

Wu, Y., Levis, B., Riehm, K.E., Saadat, N., Levis, A.W., Azar, M., Rice, D.B., Boruff, J., Cuijpers, P., Gilbody, S., Ioannidis, J.P.A., Kloda, L.A., McMillan, D., Patten, S.B., Shrier, I., Ziegelstein, R.C., Akena, D.H., Arroll, B., Ayalon, L., Baradaran, H.R., Baron, M., Bombardier, C.H.,

Butterworth, P., Carter, G., Chagas, M.H., Chan, J.C.N., Cholera, R., Conwell, Y., de Man-van Ginkel, J.M., Fann, J.R., Fischer, F.H., Fung, D., Gelaye, B., Goodyear-Smith, F., Greeno, C.G., Hall, B.J., Harrison, P.A., Harter, M., Hegerl, U., Hides, L., Hobfoll, S.E., Hudson, M., Hyphantis, T., Inagaki, M.D., Jette, N., Khamseh, M.E., Kiely, K.M., Kwan, Y., Lamers, F., Liu, S.I., Lotrakul, M., Loureiro, S.R., Lowe, B., McGuire, A., Mohd-Sidik, S., Munhoz, T.N., Muramatsu, K., Osorio, F.L., Patel, V., Pence, B.W., Persoons, P., Picardi, A., Reuter, K., Rooney, A.G., Santos, I.S., Shaaban, J., Sidebottom, A., Simning, A., Stafford, M.D., Sung, S., Tan, P.L.L., Turner, A., van Weert, H.C., White, J., Whooley, M.A., Winkley, K., Yamada, M., Benedetti, A., Thombs, B.D., 2019. Equivalency of the diagnostic accuracy of the PHQ-8 and PHQ-9: a systematic review and individual participant data meta-analysis. Psychol Med, 1-13.

Wulsin, L., Somoza, E., Heck, J., 2002. The feasibility of using the Spanish PHQ-9 to screen for depression in primary care in Honduras. Primary Care Companion to the Journal of Clinical Psychiatry 4, 191-195.

Yeung, A., Fung, F., Yu, S.C., Vorono, S., Ly, M., Wu, S., Fava, M., 2008. Validation of the Patient Health Questionnaire-9 for depression screening among Chinese Americans. Comprehensive Psychiatry 49, 211-217.

Zuithoff, N.P., Vergouwe, Y., King, M., Nazareth, I., van Wezep, M.J., Moons, K.G., Geerlings, M.I., 2010. The Patient Health Questionnaire-9 for detection of major depressive disorder in primary care: consequences of current thresholds in a crosssectional study. BMC family practice 11, 98.

Table 1. Characteristics of included studies and populations

Reference	Country	Time	Sample Size	Age *(mean±SD or range)*	Sex (♀)	Other Demographic Data	Study Design	Quality Score

1.1. Studies carried out in high-income economies (GNI per capita ≥ $12,376[$])

Table 1. Characteristics of included studies and populations

Reference	Country	Time	Sample Size	Age *(mean±SD or range)*	Sex (♀)	Other Demographic Data	Study Design	Quality Score
Aalsma, M. C., 2018	USA	2014-2015	2038	14±2	53%	Main ethnicity/language group: Afro-American (60%) Public Insurance coverage: 53.2%	Prospective cohort study	18/31
Ahmad, F., 2016	Canada	2014	75	36.5±12.7	65%	Main ethnicity/language group: Latin America (32.0%) Unemployment: 49%	Cross-sectional	19/31
Ballou, J., 2016	USA		150	>17			Abstract of a prospective cross-sectional study	
Becker, S., 2002	Saudi Arabia	2000-2001	431	18-80	54%	Higher than Primary Education: 11.4%	Cross-sectional	16/31
Bhatta, S., 2018	USA	2017	144	14.8±13.4	58%	Main ethnicity/language group: Hispanic (93%) Public Insurance coverage: 69.1%	Cross-sectional	15/31
Carey, M., 2014	Australia	2010-2014	1004	52.4±18.3	61%	Higher than Primary Education: 70.3% Public Insurance coverage: 21.7%	Cross-sectional	20/31
Chen, T. M., 2006	USA	2003	3417	43.16±14.79	55%	Main ethnicity/language group: Chinese (98.6%) Public Insurance coverage: 69.6%	Cross-sectional	17/31

Table 1. Characteristics of included studies and populations

Reference	Country	Time	Sample Size	Age *(mean±SD or range)*	Sex (♀)	Other Demographic Data	Study Design	Quality Score
Chen, I. P., 2016	Taiwan	2009-2012	634	>18	59%	Higher than Primary Education: 37% Unemployment: 36%	Cross-sectional	26/31
Cheng, C. M., 2007	Hong Kong SAR, China	2004	357	18-90	59%		Multi-center cross-sectional	18/31
Fogarty, C. T., 2008	USA	2001-2002	367	18-44 (68.8%)	61%	Main ethnicity/language group: Afro-American (46.8%) Higher than Primary Education: 83.4%	Cross-sectional	17/31
Gilbody, S., 2007	UK	2006	96	42.5±13.6	77%		Cross-sectional	23/31
Harriss, L. R., 2018	Australia	2016	199	15-25	35%		Cross-sectional	14/31
Hong, C. L. C. 2018	Singapore	2011	400	21-65	65%	Main ethnicity/language group: Chinese (52%) Higher than Primary Education: 92.7% Unemployment 18.8%	Cross-sectional	14/31
Husain, N., 2007	UK	2006	218	16-64			Cross-sectional	14/31
Inagaki, M., 2013	Japan	2010	511	21-102	59%		Cross-sectional	23/31

Table 1. Characteristics of included studies and populations

Reference	Country	Time	Sample Size	Age *(mean±SD or range)*	Sex (♀)	Other Demographic Data	Study Design	Quality Score
Karekla, M., 2012	Cyprus	2010	520	21.67±4.94	74%		Cross-sectional	19/31
Kroenke, K., 2001	USA	1997-1998	3000	>18	66%	Main ethnicity/language group: Caucasian (79%) Higher than Primary Education: 87%	Cross-sectional	24/31
Kujawska-Danecka, H, 2016	Poland	2005-2007	93	>65	63%		Cross-sectional.	14/31
Liu, S. I., 2011	Taiwan	2007-2008	1532	>18	61%		Cross-sectional	25/31
Löwe, B., 2004	Germany	2000-2001	1619	41.7±13.8	67%		Cross-sectional	26/31
Muñoz-Navarro, R., 2017	Spain	2014	260	18-65	71%	Higher than Primary Education: 61.1% Unemployment: 43.4%	Cross-sectional	
Muramatsu, K., 2007	Japan	2001-2002	131	43.4±16.4	60%		Cross-sectional	17/31
Pilowsky, D. J., 2006	USA	1998-2003	2043	51.7±12.3	76%	Main ethnicity/language group: Hispanic (78.6%)	Cross-sectional	14/31

Table 1. Characteristics of included studies and populations

Reference	Country	Time	Sample Size	Age *(mean±SD or range)*	Sex (♀)	Other Demographic Data	Study Design	Quality Score
						Higher than Primary Education: 61.1% Unemployment: 16.6%		
Rancans, E., 2018	Latvia	2014-2017	1467	53.57±29.97	69%	Higher than Primary Education: 84.6%	Cross-sectional	24/31
Richardson, L.P., 2010	USA	2007-2008	442	13-17		Main ethnicity/language group: Caucasian (71%)	Cross-sectional	17/31
Spitzer, R. L., 1999	USA	1997-1998	3000	46±17.2	66%	Main ethnicity/language group: Caucasian (79%)	Cross-sectional	25/31
Sung, S. C., 2013	Singapore	2011	400	36±10.5	65%	Main ethnicity/language group: Chinese (52%) Higher than Primary Education: 96% Unemployment: 18.8%	Cross-sectional	21/31
Vrublevska, J., 2017	Latvia	2014	324	>18	66%	Main ethnicity/language group: Latvian 60% Higher than Primary Education: 85.2%	Cross-sectional	20/31
Yeung, A., 2008	USA	2004-2005	1940	50±17	79%		Cross-sectional	18/31
Zuithoff, N.P., 2010	Netherlands	2003-2004	1338	51±17	63%		Cross-sectional	22/31

Table 1. Characteristics of included studies and populations

Reference	Country	Time	Sample Size	Age *(mean±SD or range)*	Sex (♀)	Other Demographic Data	Study Design	Quality Score
1.2. Studies carried out in upper middle-income economies (GNI per capita: $3,996 - $12,375[$])								
Azah, N., 2005	Malaysia	2001-2003	265	18-79	62%		Cross-sectional	16/31
Chen, S., 2010	China	2008	364	>60	57%	Higher than Primary Education: 45.8%	Cross-sectional	18/31
Chen, S., 2013	China	2009-2010	2639	44.8±13.2	56%	Higher than Primary Education: 83.6% Unemployment: 37.1%	Multi-center, cross-sectional	20/31
Lotrakul, M., 2008	Thailand	2006-2007	924	45±14.3	74%	Higher than Primary Education: 37.7%	Cross-sectional	19/31
Sherina, M. S., 2012	Malaysia	2009-2010	895	18-81	100%		Cross-sectional	17/31
1.3. Studies carried out in lower middle-income economies (GNI per capita: $1,026 - $3,995[$])								

Table 1. Characteristics of included studies and populations

Reference	Country	Time	Sample Size	Age *(mean±SD or range)*	Sex (♀)	Other Demographic Data	Study Design	Quality Score
Chowdhury, A. N., 2004	India		200	37±11	>50%		Cross-sectional	15/31
Ganguly, S., 2013	India		233	14-18	46%		Cross-sectional	20/31
Indu, P. S., 2018	India		238	18-60	100%	Higher than Primary Education: 80.8%	Cross-sectional	17/31
Wulsin, L., 2002	Honduras	2001	199	32.30	100%			14/31
1.4. Studies carried out in low income economies (GNI per capita≤ $1,025[$])							Cross-sectional	
Gelaye, B., 2013	Ethiopia	2011	926		61%	Higher than Primary Education: 56.8%	Cross-sectional	15/31
Hanlon, C., 2015	Ethiopia	2013	306	32.27±16.34	62%	Residence: Urban (63.3%), Rural (36.7%) Higher than Primary Education: 56.9% Unemployment 3.9%	Prospective, Focus group, Cross-sectional	21/31
Kohrt, B.A., 2016	Nepal	2013	125	>18	50%	Higher than Primary Education: 39%		17/31

Table 1. Characteristics of included studies and populations

Reference	Country	Time	Sample Size	Age *(mean±SD or range)*	Sex (♀/♂)	Other Demographic Data	Study Design	Quality Score

[$] *World Bank Classification, 2018-2019*

Table 2. Characteristics of PHQ-9 screening in included studies.

Reference	Setting	Administration	Language	Study Stages	Highlights
2.1. Studies carried out in adolescent populations (age: 12-17)					
Aalsma, M. C., 2018	Electronic Medical Records (EMR) from pediatric primary care clinics	Electronic Self-report	English	1) Recruitment in waiting room 2) Pre-screening: PHQ-2 3) PHQ-9 filled out, only if positive PHQ-2 -Automatic computerized scoring 4) PCP prompts and automatic feedback on his indications	Implementing a depression screening algorithm within an existing Computer Decision Support System is feasible. Need to mechanisms to ensure adolescent self-report. Organizational factors must be studied.
Bhatta, S., 2018	Pediatric school-based primary care clinic	Self-report	English	1) Formal education training of clinic staff 2) PHQ-9 filled out in a private exam room 3) Weekly documentation of staff compliance 4) Post-implementation retrospective chart review 5) Screening protocol included a diagram of interventions.	Improved awareness of adolescents about depression and mental health status. Human and Organizational factors can affect the screening efficiency. Electronic

Table 2. Characteristics of PHQ-9 screening in included studies.

Reference	Setting	Administration	Language	Study Stages	Highlights
					implementation may be desirable. Episodic illness may have been confounding factors.
Ganguly, S., 2013	Four English medium schools	Self-report	English	1) PHQ-9 and other scales filled out 2) PCP results analysis and clinical interview	PHQ-9 may provide a measure of depression severity.
Richardson, L.P., 2010	Private insurance healthcare facilities	Phone interview	English	1) Sending of invitation letter 2) Screening phone interview with PHQ-2/PHQ-9 3) Diagnostic phone interview on a subset of patients	PHQ-9 does not investigate irritability, which is included in DSM-IV criteria for depression in youth.

2.2. Studies carried out in adult and elderly populations (age: ≥18)

Table 2. Characteristics of PHQ-9 screening in included studies.

Reference	Setting	Administration	Language	Study Stages	Highlights
Ahmad, F., 2016	Community health centres	Digital Self-report	English Spanish	1) PHQ-9 administered in waiting rooms 2) Scoring	High rates of probable depression justify a systematic assessment in primary care and readiness to case management. E-health-mediated assessments enhance the screening capacity of primary care clinics.
Azah, N., 2005	Family clinic	Self-report	Malay	1) PHQ-9 filled out in waiting room 2) PHQ review and scoring 3) MHP diagnostic interview of all positive cases and a subset of negative cases 4) Follow up of positive cases	Socio-cultural differences, education level and need of guidance in completing the questionnaire may affect the result. Classification of depression is different between CIDI (ICD-10) and PHQ-9 (DMS-IV).

Table 2. Characteristics of PHQ-9 screening in included studies.

Reference	Setting	Administration	Language	Study Stages	Highlights
Ballou, J., 2016	Independent, community pharmacy	Self-report (2/3) Interview (1/3)	English	1) PHQ-9 administration 2) Pharmacist's score interpretation and counselling 3) Positive cases referred to their primary care provider -Emergency protocol for urgent/emergent crises	PHQ-9 administration can be implemented in a community pharmacy workflow and increases access to care.
Becker, S., 2002	Primary care hospital-based outpatient clinic	Self-report	Arabic	1) PHQ-9 filled out 2) PCP visit 3) MHP diagnostic interview on a subset of patients	Prevalence of depressive disorder is similar in developing and developed countries.
Carey, M., 2014	12 general practices	Electronic self-report	English	1) PHQ-9 filled out at reception	
Chen, T. M., 2006	Community Health Centre	Self-report unless difficulty with reading	English Chinese	1) Staff training 2)Pre-screening: three-item questionnaire 3)PHQ-9 interview of positive pre-screening patients by nurses. 4) Primary care physician's diagnosis confirmation and treatment discussion.	PHQ may measure depression severity and monitor treatment progress.

Table 2. Characteristics of PHQ-9 screening in included studies.

Reference	Setting	Administration	Language	Study Stages	Highlights
Chen, S., 2010	Primary care clinics	Self-report	Chinese	1) Nurse-assisted PHQ-9 administration 2) MHP interview of eligible subjects -Emergency measures for severe depression and suicidal ideation	Straightforward administration. Minimal training time. High subject acceptance. Urban samples may not be representative of rural population.
Chen, S., 2013	100 primary care clinics	Self-report	Chinese	1) Random selection of Primary Care Clinics 2) Nurse training 3) Screening: PHQ-9 4) Diagnostic interview on 10% Pts in 10% PCCs	Urban primary care settings are not representative of rural areas.
Chen, I. P., 2016	Primary care and hospital-based outpatient clinics	Self-report	Chinese	1) Recruitment in waiting room 2) PHQ-9 filled out 3) Research staff diagnostic interview	Psychometric measures need to be validated according to different cultural and age contexts. This should be emphasized when relating to a specific cutoff score.

Table 2. Characteristics of PHQ-9 screening in included studies.

Reference	Setting	Administration	Language	Study Stages	Highlights
Cheng, C. M., 2007	14 general practices	Self-report	Chinese	1) MHP training of PCPs 2) PHQ-9 filled out 3) PCP diagnostic interview -MHP available for support	Two-stage screening proposed: PHQ-2 -> PHQ-9.
Chowdhury, A. N., 2004	A general hospital and an outdoor clinic	Self-report	Bengali	1) PHQ-9 filled out 2) MHP diagnostic interview	Training of physician would require little time.
Fogarty, C. T., 2008	Urban family medicine practices	Self-report or assisted	English	1) PHQ-9 filled out in waiting rooms 2) Data analysed	Mental health disorders were associated with increases in primary care visits.
Gelaye, B., 2013	Outpatient General Hospital	Interview	Ethiopian	1) Nurse PHQ-9 interview 2) MHP diagnostic interview	Educational level may affect the accuracy of PHQ-9. It would be useful to determine the minimal clinical modifying factors for PHQ-9.

Table 2. Characteristics of PHQ-9 screening in included studies.

Reference	Setting	Administration	Language	Study Stages	Highlights
Gilbody, S., 2007	Primary care setting	Self-report	English	1) PHQ-9 and other scales filled out 2) Trained researcher diagnostic interview	
Hanlon, C., 2015	Urban, semi-urban and rural primary health care facilities	Interview	Amharic	1) Data-collector PHQ-9 interview 2) MHP diagnostic interview	Cut-off may not be the same in low income countries as in high income countries.
Harriss, L. R., 2018	Annual Young Person's Health Check	Staff-assisted self-report	Adapted for Aboriginal communities	1) PHQ-9 filled out 2) Referral of positive cases (cutoff >10) and identification of self-harm identification to an onsite physician	Little available information about prevalence of depression in checked communities.
Hong, C. L. C. 2018	Private primary care clinic	Self-report	English	1) Recruitment in waiting room 2) PHQ-9 filled out 3) MHP diagnostic interview	PCPs should be adequately trained in diagnosis and treatment of depression.

Table 2. Characteristics of PHQ-9 screening in included studies.

Reference	Setting	Administration	Language	Study Stages	Highlights
Husain, N., 2007	General Practice	Self-report or staff assisted as needed	English Urdu	1) PHQ-9 filled out in waiting room 2) Diagnostic interview	
Inagaki, M., 2013	Outpatient clinic within a rural hospital	Self-report	Japanese	1) PHQ-9 filled out 2) Psychiatric interview	Stigma and prevalence of somatic symptoms may lead to underestimation of depressive disorder.
Indu, P. S., 2018	Primary health center	Staff-assisted	Malayalam	1) PHQ-9 filled out 2) MHP interview	Different settings may need different cut-off points.
Karekla, M., 2012	University clinic	Self-report	Greek		
Kohrt, B.A., 2016	Primary care rural facilities	Interview	Nepali	1) Researcher screening interview: local idiom of distress, PHQ-9 2) MHP diagnostic interview	Combination of local idiom analysis reduced PHQ-9 completion by 50%. Questionnaires developed in high income countries

Table 2. Characteristics of PHQ-9 screening in included studies.

Reference	Setting	Administration	Language	Study Stages	Highlights
					have limited application for population with low literacy.
Kroenke, K., 2001	General Internal Medicine and Primary Care Clinics	Self-report	English	1) PHQ-9 filled out in waiting room 2) MHP diagnostic phone interview	Using PHQ as a severity measure need a deep analysis of its sensitivity to change. This requires longitudinal studies.
Kujawska-Danecka, H, 2016	Rural cardiovascular prevention campaign		Polish		Patients willing to participate to screening could be the ones who care most about their health.
Liu, S. I., 2011	Community-based primary care facilities	Self-report	Chinese	1) PHQ-9 filled out in waiting room 2) Researcher diagnostic interview	

Table 2. Characteristics of PHQ-9 screening in included studies.

Reference	Setting	Administration	Language	Study Stages	Highlights
Lotrakul, M., 2008	Primary care hospital	Self-report	Thai	1) PHQ-9 filled out in waiting room 2) Researcher diagnostic interview	Screening without clear care protocols is not effective and can increase the burden on GPs. Need to consider financial and institutional constraints.
Löwe, B., 2004	Outpatient clinics and General Practices	Self-report	German	1) PHQ-9 filled out 2) Diagnostic interview on a subset of participants	A two-stage approach is desirable for clinical use, whereas a one-stage is more fit for research and epidemiological studies.
Muñoz-Navarro, R., 2017	Primary care clinics	Self-report or assisted	Spanish	1) Individual meeting for PHQ-9 completion 2) Diagnostic interview scheduled within two weeks	Patients diagnosed with depression need to be referred to specialists promptly.

Table 2. Characteristics of PHQ-9 screening in included studies.

Reference	Setting	Administration	Language	Study Stages	Highlights
Muramatsu, K., 2007	Primary care facilities and a General Hospital	Self-report	Japanese	1) PHQ-9 filled out at home and returned to PCP in 48 hours 2) Researcher diagnostic interview	Validity and utility like that in other countries.
Pilowsky, D. J., 2006	Primary care practice	Interview	English Spanish	1) PHQ-9 screening interview in waiting room 2) MHP Diagnostic interview	Using PHQ as an interview rather than a screening instrument may have affected the results.
Rancans, E., 2018	Primary care clinics	Self-report	Latvian Russian	1) PHQ-9 administrated in waiting rooms 2) Interview with socio-demographic questionnaire 3) Diagnostic interview within two weeks	Established cut-off scores and risk factors for depression should be taken into account.
Sherina, M. S., 2012	Primary care clinic	Self-report	Malay	1) PHQ-9 filled out in waiting room -supervision of research assistant 2) Diagnostic interview on a weighted sample of participants	

Table 2. Characteristics of PHQ-9 screening in included studies.

Reference	Setting	Administration	Language	Study Stages	Highlights
Spitzer, R. L., 1999	5 general internal medicines and 3 family practices	Self-report	English	1) PHQ-9 filled out in waiting room 2) PCP clinical examination and score review 3) Questionnaire filled out about PHQ-9 perceived value (Pts.) and impact on decision making (PCP) 4) MHP diagnostic interview	
Sung, S. C., 2013	Peace Family Clinic	Self-report	Chinese Indian Malay	1) PHQ-9 filled out in waiting room 2) PCP diagnostic interview -Supervision of a senior MHP	The optimal cut-off was lower than other studies and did not allow to distinguish between major and minor depression.
Vrublevska, J., 2017	Primary care facility	Self-report	Latvian Russian	1) PHQ-9 filled out in waiting room -MHP available for support 2) MHP diagnostic interview	Larger and longitudinal studies are needed to confirm the effectiveness of screening.
Wulsin, L., 2002	5 rural clinics	Interview	Spanish	1) PHQ-9 interview by PCPs and medical students 2) Diagnostic interview on a weighted sample	

Table 2. Characteristics of PHQ-9 screening in included studies.

Reference	Setting	Administration	Language	Study Stages	Highlights
Yeung, A., 2008	Community Health Centre	Self-report	Chinese English	1) PHQ-9 filled out in waiting room 2) MHP telephonic interpretation of results 3) MHP diagnostic interview	PHQ-9 functions well in trans-cultural settings.
Zuithoff, N.P., 2010	7 general practices	Self-report	Dutch	1) PHQ-9 filled out at home and returned to PCP by mail 2) MHP diagnostic interview	PHQ-9 scores were consistent with functional status, sick days, number of GP consultations.

MHP: Mental Health Professional; GP: General Practitioner; PCP: Primary Care Physician; PHQ-9: 9-item Patient Health Questionnaire.

Table 3. Operating Characteristics of PHQ-9 against reference diagnostic interviews.

Reference	Study Design	Sample Size	Diagnostic Interview	Cut-off	Sensitivity *n, 95% CI*		Specificity *n, 95% CI*		PPV *n, 95% CI*		NPV *n, 95% CI*		LR(+)[$]	LR(-)[$]
3.1 Studies performing a fully structured diagnostic interview, divided by interview														
Hanlon, C., 2015	Cross-sectional	306	MINI	5	0.83		0.75		0.17		0.99		3.32	0.23
Hong, C. L. C. 2018	Cross-sectional	400	MINI	6	0.92		0.72		0.09		1		3.29	0.11
Inagaki, M., 2013	Cross-sectional	511	MINI	10	0.45		0.99		0.72		0.96		45.00	0.56
				5	0.86		0.85		0.32		0.99		5.73	0.16
Indu, P. S., 2018	Cross-sectional	238	MINI	9	0.83	0.72 - 0.93	0.90	0.85 - 0.96	0.73	0.62-0.84	0.94	0.90 - 0.98	8.30	0.19

Table 3. Operating Characteristics of PHQ-9 against reference diagnostic interviews.

Reference	Study Design	Sample Size	Diagnostic Interview	Cut-off	Sensitivity *n, 95% CI*	Specificity *n, 95% CI*	PPV *n, 95% CI*	NPV *n, 95% CI*	LR(+)[$]	LR(-)[$]
Lotrakul, M., 2008	Cross-sectional	279[%]	MINI	10	0.74	0.85	0.27	0.98	4.93	0.31
				9	0.84	0.77	0.21	0.99	3.65	0.21
Muramatsu, K., 2007	Cross-sectional	131	MINI-Plus	10	0.84	0.95	0.87	0.94	16.80	0.17
Rancans, E., 2018	Cross-sectional	1467	MINI	10	60.8 [Latvian: Russian: 67.7]	91.1 [Latvian: Russian: 89.7]			6.84 [Latvian: Russian: 6.59]	0.43 [Latvian: Russian: 0.36]
Sung, S. C., 2013	Cross-sectional	400	MINI	6	0.92	0.72	0.92	1	3.29	0.11

Table 3. Operating Characteristics of PHQ-9 against reference diagnostic interviews.

Reference	Study Design	Sample Size	Diagnostic Interview	Cut-off	Sensitivity *n, 95% CI*	Specificity *n, 95% CI*	PPV *n, 95% CI*	NPV *n, 95% CI*	LR(+)[$]	LR(-)[$]
Vrublevska, J., 2017	Cross-sectional	272[%]	MINI	10	0.86	0.89			7.82	0.16
				8	0.97	0.75			3.88	0.04
Azah, N., 2005	Cross-sectional	265	CIDI depression of any severity	9	0.39	0.83	0.73		2.85	0.72
				10	0.37	0.87	0.77		3.21	0.48
			CIDI severe depression	9	0.63	0.78	0.35		2.29	0.73
				10	0.61	0.81	0.38		2.85	0.72

Table 3. Operating Characteristics of PHQ-9 against reference diagnostic interviews.

Reference	Study Design	Sample Size	Diagnostic Interview	Cut-off	Sensitivity *n, 95% CI*		Specificity *n, 95% CI*		PPV *n, 95% CI*		NPV *n, 95% CI*		LR(+)[$]	LR(-)[$]
Kohrt, B.A., 2016	Cross-sectional	125	CIDI	10	0.94		0.8		0.42		0.99		4.70	0.08
Sherina, M. S., 2012	Prospective Cross-sectional	895	CIDI	10	0.87	0.71 - 0.95	0.82	0.74 - 0.88	0.15		0.99		4.83	0.16
Zuithoff, N.P., 2010	Cross-sectional	1338	CIDI	10	0.49	0.42 - 0.56	0.95	0.94 - 0.96	0.59	0.51-0.67	0.93	0.92 - 0.94	9.80	0.54
				5	0.86	0.81 - 0.91	0.75	0.73 - 0.77	0.34	0.3-0.38	0.97	0.96 - 0.98	3.44	0.19
3.2 Studies performing a semi-structured diagnostic interview, divided by interview														
Becker, S.,	Cross-sectio	173[%]	SCID-R		0.6		0.9						12.4	0.40

Table 3. Operating Characteristics of PHQ-9 against reference diagnostic interviews.

Reference	Study Design	Sample Size	Diagnostic Interview	Cut-off	Sensitivity *n, 95% CI*	Specificity *n, 95% CI*	PPV *n, 95% CI*	NPV *n, 95% CI*	LR(+)[$]	LR(-)[$]
2002	nal				2	5			0	
Chen, S., 2010	Cross-sectional	77[%]	SCID	10	0.75	0.89			6.82	0.28
				8	0.9	0.8			4.50	0.13
Chen, S., 2013	Multi-center, cross-sectional	280[%]	SCID	10	0.87	0.81			4.58	0.16
				9	0.95	0.75			3.80	0.07
Gilbody, S., 2007	Cross-sectional	96	SCID	10	0.92	0.78			4.18	0.10
Kroenke, K., 2001	Cross-sectional	580[%]	SCID	10	0.88	0.88	0.35	0.99	7.33	0.14

Table 3. Operating Characteristics of PHQ-9 against reference diagnostic interviews.

Reference	Study Design	Sample Size	Diagnostic Interview	Cut-off	Sensitivity *n, 95% CI*		Specificity *n, 95% CI*		PPV *n, 95% CI*	NPV *n, 95% CI*	LR(+)[$]	LR(-)[$]
Löwe, B., 2004	Cross-sectional	501[%]	SCID	11	0.98	0.92 - 1.00	0.8	0.76 - 0.83			4.90	0.03
Muñoz-Navarro, R., 2017	Cross-sectional	178[%]	SCID-I	10	0.95	0.89 - 0.97	0.67	0.53 - 0.79	0.88	0.83	2.90	0.08
Spitzer, R. L., 1999	Cross-sectional	585[%]	SCID	10	0.73	0.59 - 0.87	0.98	0.96 - 1.00			36.50	0.28
Wulsin, L., 2002	Cross-sectional	199	SCID	10	0.77		0.99		1**	0.88**	77.00	0.23
Yeung, A., 2008	Cross-sectional	184[%]	CB-SCID-I/P	15	0.81		0.98		0.92	0.95	40.50	0.19
Chen, I. P.,	Cross-sectio	634	SCAN	10	0.7		0.9		28.	99.5	19.7	0.22

Table 3. Operating Characteristics of PHQ-9 against reference diagnostic interviews.

Reference	Study Design	Sample Size	Diagnostic Interview	Cut-off	Sensitivity *n, 95% CI*	Specificity *n, 95% CI*	PPV *n, 95% CI*	NPV *n, 95% CI*	LR(+)[$]	LR(-)[$]
2016	nal				9	6	2		5	
				6	1	0.85			6.67	0.00
Gelaye, B., 2013	Cross-sectional	363[%]	SCAN	10	0.86	0.78-0.92	0.67	0.61-0.73	0.48	0.40-0.56	0.93	0.89-0.96	2.61	0.21
				11	0.79	0.7-0.87	0.74	0.53	0.54	0.45-0.63	0.89	0.85-0.93	3.04	0.28
Liu, S. I., 2011	Cross-sectional	1532	SCAN	10	0.86		0.94						14.33	0.15
3.3 Studies performing other diagnostic interviews														
Cheng, C.	Multi-center	357	CHDS	10	0.8		0.9		48		98		10.0	0.22

Table 3. Operating Characteristics of PHQ-9 against reference diagnostic interviews.

Reference	Study Design	Sample Size	Diagnostic Interview	Cut-off	Sensitivity *n, 95% CI*		Specificity *n, 95% CI*		PPV *n, 95% CI*	NPV *n, 95% CI*	LR(+)[$]	LR(-)[$]
M., 2007	cross-sectional						2				0	
Chowdhury, A. N., 2004	Cross-sectional	200	DSM-IV criteria	10	0.98		0.42		0.37	0.98	1.69	0.05
				13	0.92		0.67		0.49	0.96	2.79	0.12
Ganguly, S., 2013	Cross-sectional	233	ICD-10 Codes	10	0.58	0.41 - 0.74	0.98	0.94 - 0.99	0.78	0.94	29.00	0.43
				5	0.87	0.71 - 0.95	0.8	0.74 - 0.85	0.4	0.98	4.35	0.16
Husain, N., 2007	Cross-sectional	218	PAS	7	0.70		0.89		0.83	0.81	6.54	0.34
Richardson,	Cross-sectio	442	DISC-IV	10	0.9		0.7		0.1	0.99	3.21	0.14

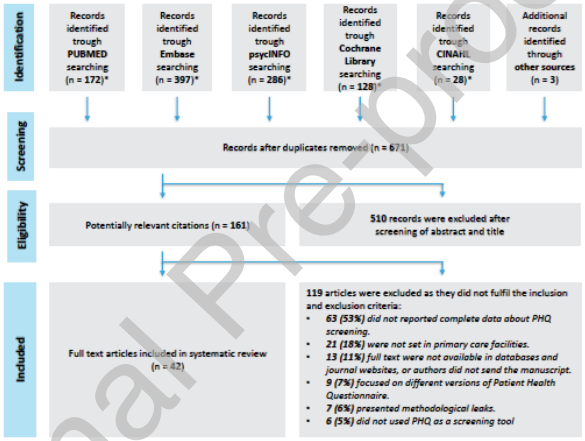
Table 3. Operating Characteristics of PHQ-9 against reference diagnostic interviews.

Reference	Study Design	Sample Size	Diagnostic Interview	Cut-off	Sensitivity *n, 95% CI*	Specificity *n, 95% CI*	PPV *n, 95% CI*	NPV *n, 95% CI*	LR(+)[$]	LR(-)[$]
L.P., 2010	nal					2	3			
				11	0.9	0.78			4.09	0.13

[$]: Measures were added by reviewers, based on available data.
[%]: Diagnostic interviews were carried out in a weighted subset of patients. For whole sample size, see Table 1.
CHDS: Chinese Hamilton Depression Scale - CIDI: Composite International Diagnostic Interview - DISC-IV: Diagnostic International Schedule for Children - DSM-IV: 4th edition of Diagnostic Statistical Manual of Psychiatric Disorders - ICD-10: 10th edition of International Classification of Diseases - LR(+): Positive Likelihood Ratio - LR(-): Negative Likelihood Ratio - MINI: Mini-International Neuropsychiatric Interview - NPV: Negative Predictive Value - PAS: Psychiatric Assessment Schedule - PPV: Positive Predictive Value - SCAN: Structured Clinical Assessment in Neuropsychiatry - SCID: Structured Clinical Interview for DSM-IV.

FIGURES

Figure 1. Flow diagram of selected articles.



[insert Fig.1 here]

[$]Search strategy limited from January 1st, 1995 to October 31st, 2018, English language, human subjects older than 12.