Family Medicine and Care

Review Article ISSN: 2516-824X

Systematic review and Meta-analysis on the Patient Health Questionnare-9 (PHQ-9) for depression screening in Chinese primary care patients

CHIU Chi Fai Billy[1]*, CHIN Weng Yee[2]

[1]Family Medicine and Primary Care Centre, Hong Kong Sanatorium & Hospital, Hong Kong

[2]Department of Family Medicine and Primary Care, The University of Hong Kong, Hong Kong

Abstract

Background: Depression is a chronic illness with mounting burden to the society. The 9-item Patient Health Questionnare-9 (PHQ-9) is a commonly used screening instrument in primary care settings to help enhance the detection of depression. There is a need to evaluate the effectiveness of the PHQ-9 in Chinese primary care settings.

Methods: A systematic review and meta-analysis was undertaken to evaluate PHQ-9 amongst Chinese primary care patients. MEDLINE, PubMed and PsycINFO databases, and the Cochrane Library were searched between 2006 and 2016 for any report of studies that evaluated performance of PHQ-9. Four screening studies (n = 8,403) met the inclusion criteria. Quality of studies were assessed according to the accepted guidelines. Meta-analyses were conducted using a random effects model.

Results: The pooled estimate of area under the receiver operator characteristic curves for PHQ-9 reported from four studies was 0.885 (95% CI: 0.805 to 0.965; SE: 0.0408; P < 0.001), which indicated a good performance of the screening instrument. Heterogeneity between studies were observed for screening studies (I^2 statistics: 96.83%, P < 0.0001). Meta-analyses were limited by the small number of studies.

Conclusions: The PHQ-9 instrument is a sensitive screening tool highly predictive of depression among Chinese population in primary care setting. Further studies are however required to increase the power of meta-analysis.

Introduction

The prevalence of depressive disorders has been estimated to be around 10% to 20% in primary care [1]. One in 10 patients are estimated to have mild-to-moderate depressive symptoms at the time of a primary-care consultation, of whom around one in four are identified as having depression by the doctor [2]. However, most of the epidemiological data to date has been retrieved from Western countries, whereas epidemiological data on mental illness in low- or middle-income countries have been lacking [3]. In view of the socio-cultural differences between Western and Eastern societies, results from Western societies may have limited generalizability to Asian populations. This issue is also expected to be of increasing importance in China as an increasing trend of depression is observed in many Asian including Chinese populations.

Depressive disorders are associated with significant morbidity, disability and healthcare utilization in primary-care settings [4,5]. In many countries, treatment for depression is mainly provided in primary-care settings [6,7], in line with the World Health Organization recommendation that common mental illnesses should be treated in primary care [8]. However, up to 50% of depressive disorders were unrecognized in a primary care setting [9,10]. Moreover, Chinese patients tend to underutilize mental health services. According to a local study in 10,179 adult patients from primary-care settings in Hong Kong, only 24.3% of 518 patients who screened positive for depression received services from healthcare professionals [11]. This indicates that depressive disorders are commonly undiagnosed and undertreated in primary-care settings, which may subsequently pose a significant medical burden on the community.

The use of appropriate screening tool is essential to identify and manage patients with depressive disorders in primary care. The 9-item Patient Health Questionnare-9 (PHQ-9), which scores each of the 9 Diagnostic and Statistical Manual of Mental Disorders — Fourth Edition (DSM-IV) criteria on a scale of "0" (not at all) to "3" (nearly every day), has been well validated in primary-care settings in Western populations [12-15]. However, its performance in Chinese populations has not been well established. We therefore aimed to performed a systematic review to evaluate the diagnostic performance of the PHQ-9 as a screening instrument for depression in Chinese primary care patients [16-21].

Methods

Study selection

Search strategy: Literature search was performed in the MEDLINE, PubMed and PsycINFO databases (between 2006 and 2016) and the total database of the Cochrane Library, using "primary care",

Correspondence to: CHIU Chi Fai Billy, Family Medicine and Primary Care Centre, Hong Kong Sanatorium & Hospital, Hong Kong, Tel: +852 2572 0211; Email: hkuchiucf@gmail.com

Received: April 13, 2018; **Accepted:** April 27, 2018; **Published:** April 30, 2018

"screening", "depression" and "Chinese" as search terms to identify candidate articles for systematic review of the screening instrument.

Inclusion and exclusion criteria

Screening studies were included if the following criteria were met: use of PHQ-9 to identify depressive disorders, Chinese subjects recruited from primary-care settings, and outcome variables which included at least an area under the receiver operator characteristic (ROC) curve with either standard error or 95% confidence interval (CI) to demonstrate the performance of the screening tool. Studies published in a language other than English were not included due to the incapability of translation.

Quality assessment of studies

The studies identified in the search were assessed for inclusion based on their methodological quality. Four screening studies were finally selected for inclusion in the systematic review and meta-analysis. The selected screening studies were validation studies of which blinding is a potential source of bias [22], and therefore blinding of assessors and outcomes was the criteria for assessment of study quality [23].

Data extraction

Following the preliminary search and quality assessment of selected studies, data were extracted from the full text of journal articles and captured in a spreadsheet, which included data fields to collect information on the year of study, type of study design, randomization method, study population, sample sizes, types of screening instrument, and study outcomes.

Data analysis

Screening studies that reported the ability of PHQ-9 to detect depressive disorders were included for meta-analysis. As ROC curves are representative of the inherent trade-offs between sensitivity and specificity for a screening instrument [24], we mainly performed meta-analysis of area under curve (AUC). To standardize deviation of distribution of the statistics, standard error (SE) for AUC was derived from 95% confidence interval (CI) according to the formula "SE = (upper limit – lower limit) / 3.92" for all included studies [16]. Meta-analysis of ROC curves was performed using MedCalc version 16.4.3, MedCalc Software, Ostend, Belgium [25]. All statistical tests were two-sided and the level of significance was set at 5%. As only a small number of studies could be selected and pooled, funnel plot was not performed to assess for publication bias.

Results

Search results

A total of 80 related articles were identified from PubMed/Medline database (N = 31), Cochrane Library (N=17) and PsycINFO database (N=32) according to the preset search criteria. Following the selection of studies in accordance with the inclusion and exclusion criteria, only five full-text articles with complete data were retrieved for further methodological evaluation [26,27].

Of the five eligible studies that investigated the accuracy of the PHQ-9 instrument in Chinese populations [28-32], one study validated the Chinese/English bilingual version of PHQ-9 for depression screening among immigrant Chinese Americans in primary care [32]. Since this study focused on immigrants who attended a community health center in Boston, USA, and the patients there might be influenced by Western culture, the study might not be suitable for meta-analysis to evaluate the overall estimate of the AUC value for PHQ-9 in Chinese populations. Therefore, four studies with a total of 8,403 Chinese subjects were included for evaluation of the performance of the PHQ-9 instrument for screening of depressive disorders in primary-care settings. The flowchart of study selection, from initial search results to final inclusion for meta-analysis, is shown in Figure 1.

Characteristics of included studies

Four included studies [28-32] mainly aimed to examine the reliability and validity of the screening instrument for depression in Chinese primary-care setting. Three studies were performed to determine the reliability and validity of PHQ-9 instrument while one by Chin, *et al* [28] focused on validating the use of Center for Epidemiologic Studies Depression Scale (CES-D) instrument for screening Chinese primary care patients. In the study by Chin et al, the PHQ-9 was used to assess the convergent validity of the screening instrument CES-D as they were both depression instruments measuring a similar construct. As the study also reported an outcome variable of AUC for PHQ-9 instrument, it was not excluded from the subsequent meta-analysis after the final in-depth review of the included studies. A summary of the included studies for meta-analysis of PHQ-9 instrument for screening of depression in Chinese primary-care setting is presented in Table 1 [33-37].

Summary of the results

Preliminary appraisal of study quality indicated that all investigators of the selected studies were blinded to the screening score results and therefore all four studies were determined to be of good quality for inclusion in the meta-analysis. Since most of the studies presented an estimate with 95% CI, calculation of the SE was performed for all studies to standardize the measurement of outcome deviations. Pooled estimates of AUC were calculated to illustrate the overall estimates. Considering the heterogeneity between studies with I^2 statistics of 96.83% ($P < 0.0001$), a random effect model was adopted. Figure 2 shows the pooled AUC estimates for PHQ-9 for depression screening. The pooled area under the ROC curves was 0.885 (95% CI: 0.805 to 0.965; SE: 0.0408; $P < 0.001$) (Table 2) [38].

Discussion

Depression is a chronic illness commonly affecting the quality of life of patients and posing a burden on the society globally [39]. The non-communicable disease has a significant impact in primary-care settings [4,5] where subthreshold depressive symptoms are prevalent [40,41]. It is therefore essential to identify an appropriate screening tool to help tackle the under detection of the problem from the point of entry to the healthcare system for most people requiring health services.

In a local study that estimated the 12-month cumulative incidence and predictors of a positive screen for depressive symptoms using PHQ-9 among primary-care patients [42], the cumulative incidence of PHQ-9-screened depressive symptoms were found to be higher than the incidence of depressive disorders reported in systematic reviews [3,43]. The screening instrument has been translated into Chinese and validated in Chinese primary-care patients [30,31,44], suggesting that it may be an appropriate tool to identify patients with depressive symptoms in local primary-care settings in view of its high sensitivity of 0.81 (95% CI: 0.68 to 0.89) and specificity of 0.88 (95% CI: 0.84 to 0.91) as shown in a previous systematic review [45]. However, data about its screening ability in Chinese populations are lacking. In this

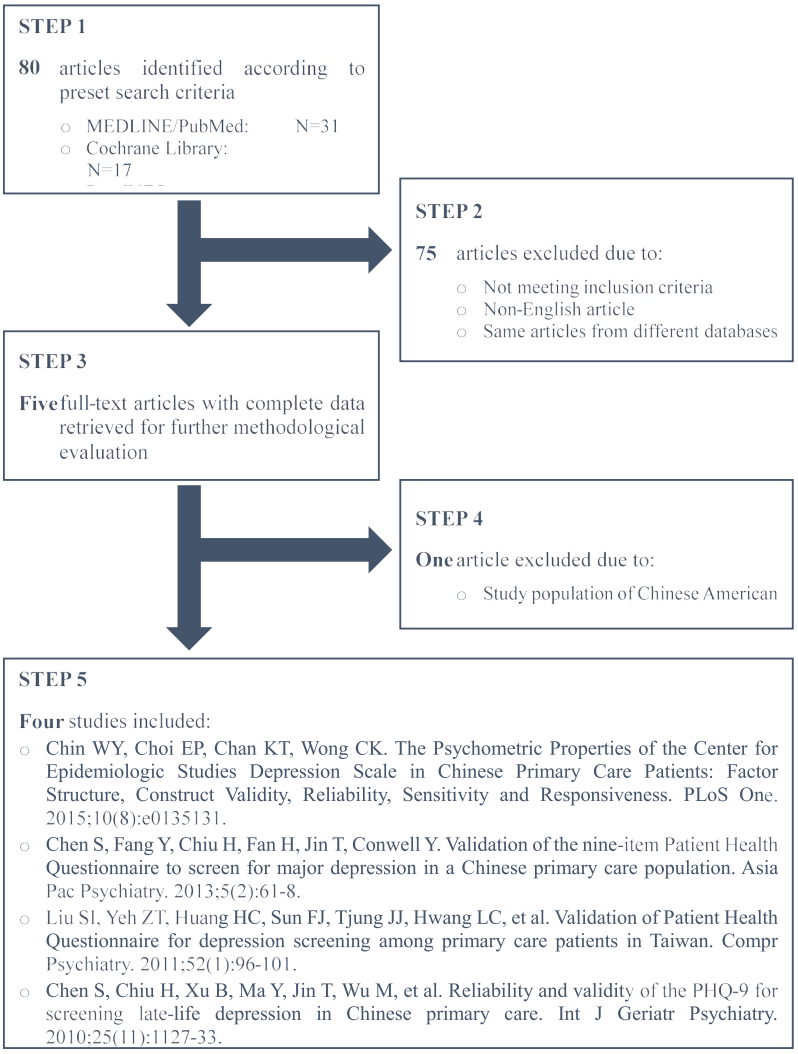

Figure 1. Flowchart of screening study selection

Table 1. Summary of included studies. AUC: Area Under the Receiver Operator Characteristic Curve; CES-D: Center for Epidemiologic Studies Depression Scale; HAMD: Hamilton Rating Scale for Depression; PHQ: Patient Health Questionnaire; Q-LES-Q SF: Short Form of the Quality of Life Enjoyment and Satisfaction Questionnaire; SE: Standard Error; SF-12 (v2) MCS: Short Form-12 Health Survey (version 2) Mental Component Summary

Study	Population	Sample Size	Age Range	Screening Instruments	Study Endpoints for PHQ-9 Instrument
Chin WY [28]	Chinese	3868	≥ 18 yrs	PHQ-9, CES-D and SF-12 (v2) MCS	AUC: 0.747, 95% CI: 0.71-0.78
Chen S [29]	Chinese	2639	18-60 yrs: 84% > 60 yrs: 16%	PHQ-9	AUC: 0.91, 95% CI: 0.87-0.94 Sensitivity: 87% Specificity: 81%
Liu SI [30]	Taiwanese	1532	≥ 18 yrs	PHQ-9, PHQ-2, PHQ-1, Q-LES-Q SF and HAMD-17	AUC: 0.96, 95% CI: 0.93-0.98 (SE = 0.01) Sensitivity: 86% Specificity: 93.9%
Chen S [31]	Chinese	364	≥ 60 yrs	PHQ-9 and PHQ-2	AUC: 0.92, 95% CI: 0.88-0.96 Sensitivity: 86% Specificity: 85%

Table 2. Meta-analysis of the Area under the Receiver Operator Characteristic Curves for PHQ-9 Instruments.

Study	ROC Area	Standard Error	95% CI	Z-value	P-value	Weight (%)	
						Fixed	Random
Chin WY, 2015 [28]	0.747	0.0180	0.712 to 0.782			21.15	24.92
Chen S, 2013 [29]	0.910	0.0180	0.875 to 0.945			21.15	24.92
Liu SI, 2011 [30]	0.960	0.0130	0.935 to 0.985			40.56	25.51
Chen S, 2010 [31]	0.920	0.0200	0.881 to 0.959			17.14	24.64
Total (fixed effects)	0.898	0.00828	0.881 to 0.914	108.409	<0.001	100.00	100.00
Total (random effects)	**0.885**	**0.0408**	**0.805 to 0.965**	**21.697**	**<0.001**	**100.00**	**100.00**

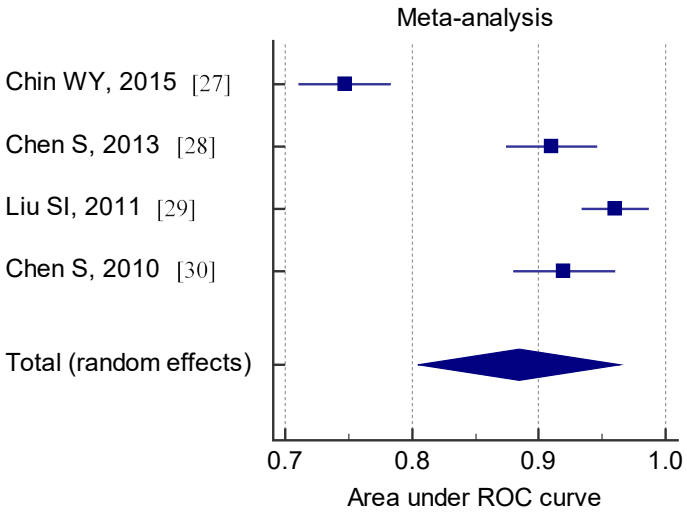

Figure 2. Pooled Area under the ROC Curve

current systematic review and meta-analysis, statistical details of sensitivity and specificity were not available for the included studies, but data on the area under the ROC curve were obtained for meta-analysis. The pooled AUC value of PHQ-9 was 0.885 (95% CI: 0.805 to 0.965), which was greater than 0.7, the value for an instrument to be considered as a sensitive screening tool [46]. It is, important to note that the AUC value conveys little about the diagnostic properties of a screening modality, and only a limited number of studies was included in this meta-analysis.

Limitations

The number of studies included in the systematic review was not adequate for meta-analysis. The sample size was too small to draw a conclusion on how well the PHQ-9 instrument could identify depressive patients in primary-care settings. This might also mask publication bias and heterogeneity between studies, and the current review is likely to be vulnerable to publication bias, a threat to the validity of systematic reviews [47]. However, the current study provided preliminary information for primary-care practitioner's reference, and further research in this area is required. The selection of papers for the systematic review was conducted by one investigator and therefore the review process and selection of studies could be biased.

Implications and future directions

There is currently no local guidelines on the use of screening instruments for identifying patients with depressive symptoms in primary-care setting in Hong Kong. The present systematic review has demonstrated that the PHQ-9 is a sensitive screening instrument which is highly predictive of depression among Chinese patients in primary care settings, suggesting primary care physicians could utilize the widely available instrument to help identify patients with depressive symptoms and provide treatment earlier as a gatekeeper of psychiatric and other specialist services. Due to the limitation in number of relevant studies, further evaluation studies are still needed to build the evidence for the utility of the PHQ-9 and a future meta-analysis needs to be performed again when more studies are available.

Conclusion

The systematic review has demonstrated a high predictive ability of the PHQ-9 as a depression screening instrument in Chinese patients. The review and analyses were limited by the small number of studies that met the inclusion criteria. A further systematic review to compare different screening modalities are required when more studies become available.

Acknowledgements

This study was undertaken as part of a Master's in Public Health degree at the University of Hong Kong. The full dissertation is available on line at https://hub.hku.hk/handle/10722/237273 (Chiu and 趙志輝 2016). The authors would like to thanks and acknowledge Dr Dennis Ip for his supervision and expert advice in guiding this project.

Conflict of Interest Statement

The authors declare no conflicts of interest

Authors and Contributors

BC conceived the research question, conducted the data extraction and analysis and drafted the original thesis.

WC contributed to the conception of the research question, advised on the methods and made critical revisions to the manuscript for publication.

Funding

This study was unfunded.

References

1. Mitchell AJ, Vaze A, Rao S (2009) Clinical diagnosis of depression in primary care: a meta-analysis. *Lancet* 374: 609-619. [Crossref]
2. Chin W, Chan K, Lam C, Wong S, Fong D, et al. (2014) Detection and management of depression in adult primary care patients in Hong Kong: a cross-sectional survey conducted by a primary care practice-based research network. *BMC Family Practice*.15: 30.
3. Ferrari AJ, Somerville AJ, Baxter AJ, Norman R, Patten SB, et al. (2013) Global variation in the prevalence and incidence of major depressive disorder: a systematic review of the epidemiological literature. *Psychol Med* 43: 471-481.
4. Roy-Byrne PP, Stein MB, Russo J, Mercier E, Thomas R, et al. (1999) Panic disorder in the primary care setting: comorbidity, disability, service utilization, and treatment. *J Clin Psychiatry* 60: 492-499.
5. Simon G, Ormel J, VonKorff M, Barlow W (1995) Health care costs associated with depressive and anxiety disorders in primary care. *Am J Psychiatry* 152: 352-357.
6. Regier DA, Goldberg ID, Taube CA (1978) The de facto US mental health services system: a public health perspective. *Arch Gen Psychiatry* 35: 685-693.
7. Miller BF, Druss B (2013) The role of family physicians in mental health care delivery in the United States: implications for health reform. *J Am Board Fam Med* 26: 111-113.
8. WHO (2010) mhGAP intervention guide for mental, neurological and substance use disorders in non-specialized health settings. Secondary mhGAP intervention guide for mental, neurological and substance use disorders in non-specialized health settings. World Health Organization.
9. Depression Guideline Panel (1993) Depression in Primary Care: Detection, Diagnosis and Treatment. Journal of Psychosocial Nursing and Mental Health Services 31: 19-28.
10. Simon GE, VonKorff M (1995) Recognition, management, and outcomes of depression in primary care. *Arch Fam Med* 4: 99-105. [Crossref]
11. Chin WY, Chan KT, Lam CL, Lam TP, Wan EY (2015) Help-seeking intentions and subsequent 12-month mental health service use in Chinese primary care patients with depressive symptoms. *BMJ Open* 5: e006730.
12. Spitzer RL, Kroenke K, Williams JB (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.
13. Arroll B, Goodyear-Smith F, Crengle S, Gunn J, Kerse N, et al. (2010) Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. *Ann Fam Med* 8: 348-53.

14. Kroenke K, Spitzer RL, Williams JB (2001) The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 16: 606-613. [Crossref]

15. Huang FY, Chung H, Kroenke K, Delucchi KL, Spitzer RL (2006) Using the Patient Health Questionnaire-9 to measure depression among racially and ethnically diverse primary care patients. *J Gen Intern Med* 21: 547-552.

16. Cape J, Barker C, Buszewicz M, Pistrang N (2000) General practitioner psychological management of common emotional problems (II): A research agenda for the development of evidence-based practice. *Br J Gen Pract* 50: 396-400.

17. Kennedy SH, Lam RW, Morris B (2003) Clinical guidelines for depressive disorders. *Can Fam Physician* 49: 489-491.

18. Arean P, Hegel M, Vannoy S, Fan MY, Unuzter J (2008) Effectiveness of problem-solving therapy for older, primary care patients with depression: results from the IMPACT project. *Gerontologist* 48: 311-23.

19. Archer J, Bower P, Gilbody S, Lovell K, Richards D, et al. (2012) Collaborative care for depression and anxiety problems. *Cochrane Database Syst Rev* 10: CD006525.

20. Katon WJ, Lin EH, Von Korff M, Ciechanowski P, Ludman EJ, et al. (2010) Collaborative care for patients with depression and chronic illnesses. *N Engl J Med* 363: 2611-2620. [Crossref]

21. NICE clinical Guidance (2009) Depression in adults with a chronic physical health problem: recognition and management. National Institute for Health and Care Excellence.

22. Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, et al. (2004) Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med* 140: 189-202.

23. Higgins JPT, Green S (2011) Cochrane Handbook for Systematic Reviews of Interventions. Cochrane Collaboration, London, UK.

24. Knottnerus JA (2002) The evidence base of clinical diagnosis. BMJ, London UK

25. MedCalc (2016) MedCalc Software, Version 16.4.3. Ostend, Belgium.

26. Comprehensive Meta-Analysis (2014) Version 3.3. Biostat, Inc, New Jersey, USA.

27. Higgins JP, Thompson SG, Deeks JJ, Altman DG (2003) Measuring inconsistency in meta-analyses. *BMJ* 327: 557-560. [Crossref]

28. Chin WY, Choi EP, Chan KT, Wong CK (2015) The Psychometric Properties of the Center for Epidemiologic Studies Depression Scale in Chinese Primary Care Patients: Factor Structure, Construct Validity, Reliability, Sensitivity and Responsiveness. *PLoS One* 10: e0135131.

29. Chen S, Fang Y, Chiu H, Fan H, Jin T, et al. (2013) Validation of the nine-item Patient Health Questionnaire to screen for major depression in a Chinese primary care population. *Asia Pac Psychiatry* 5: 61-68.

30. Liu SI, Yeh ZT, Huang HC, Sun FJ, Tjung JJ, et al. (2011) Validation of Patient Health Questionnaire for depression screening among primary care patients in Taiwan. *Compr Psychiatry* 52: 96-101.

31. Chen S, Chiu H, Xu B, Ma Y, Jin T, et al. (2010) Reliability and validity of the PHQ-9 for screening late-life depression in Chinese primary care. *Int J Geriatr Psychiatry* 25: 1127-1133. [Crossref]

32. Yeung A, Fung F, Yu SC, Vorono S, Ly M, et al. (2008) Validation of the Patient Health Questionnaire-9 for depression screening among Chinese Americans. Compr Psychiatry 49: 211-217.

33. Chen S, Conwell Y, He J, Lu N, Wu J (2015) Depression care management for adults older than 60 years in primary care clinics in urban China: a cluster-randomised trial. Lancet Psychiatry 2: 332-339.

34. Zhang de X, Lewis G, Araya R, Tang WK, Mak WW, et al. (2014) Prevention of anxiety and depression in Chinese: a randomized clinical trial testing the effectiveness of a stepped care program in primary care. *J Affect Disord* 169: 212-220.

35. Tiwari A, Fong DY, Yuen KH, Yuk H, Pang P, et al. (2010) Effect of an advocacy intervention on mental health in Chinese women survivors of intimate partner violence: a randomized controlled trial. *JAMA* 304: 536-543.

36. Lam CL, Fong DY, Chin WY, Lee PW, Lam ET, et al. (2010) Brief problem-solving treatment in primary care (PST-PC) was not more effective than placebo for elderly patients screened positive of psychological problems. *Int J Geriatr Psychiatry* 25: 968-980. [Crossref]

37. Hou RJ, Wong SY, Yip BH, Hung AT, Lo HH, Chan PH, et al. The effects of mindfulness-based stress reduction program on the mental health of family caregivers: a randomized controlled trial. *Psychother Psychosom* 83: 45-53.

38. Guo T, Xiang YT, Xiao L, Hu CQ, Chiu HF, et al. (2015) Measurement-Based care versus standard care for major depression: A randomized controlled trial with blind raters. *Am J Psychiatry* 172: 1004-1013.

39. Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, et al. (2012) Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 380: 2197-2223.

40. Ansseau M, Dierick M, Buntinkx F, Cnockaert P, De Smedt J, et al. (2004) High prevalence of mental disorders in primary care. *J Affect Disord* 78: 49-55. [Crossref]

41. Blazer D, Swartz M, Woodbury M, Manton KG, Hughes D, et al. (1988) Depressive symptoms and depressive diagnoses in a community population. Use of a new procedure for analysis of psychiatric classification. *Arch Gen Psychiatry* 45: 1078-1084.

42. Chin WY, Wan EY, Choi EP, Chan KT, Lam CL (2016) The 12-Month Incidence and Predictors of PHQ-9-Screened Depressive Symptoms in Chinese Primary Care Patients. *Ann Fam Med* 14: 47-53.

43. Ferrari AJ, Charlson FJ, Norman RE, Patten SB, Freedman G, et al. (2013) Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. *PLoS Med* 10: e1001547.

44. Cheng C, Cheng M (2007) To validate the Chinese version of the 2Q and PHQ-9 questionnaires in Hong Kong Chinese patients. *The Hong Kong Practitioner* 29: 381.

45. Moriarty AS, Gilbody S, McMillan D, Manea L (2015) Screening and case finding for major depressive disorder using the Patient Health Questionnaire (PHQ-9): a meta-analysis. *Gen Hosp Psychiatry* 37: 567-576.

46. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, et al. (2007) Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 60: 34-42.

47. Rosenthal R, Sutton AJ, Bornstein M (2005) Publication bias in meta-analysis: Prevention, assessment, and adjuvstments. Chichester, UK.

48. Lau J, Ioannidis JP, Schmid CH (1997) Quantitative synthesis in systematic reviews. *Ann Intern Med* 127: 820-826. [Crossref]

49. Chiu Cf (2016) A systematic review of screening and treatment for depression among Chinese patients in primary care setting, (HKUTO).