

IS THERE GOOD EVIDENCE THAT THE TWO QUESTIONS IN PHQ-2 ARE USEFUL QUESTIONS TO USE IN ORDER TO SCREEN FOR DEPRESSION?

Rachel Wilson¹ & Mark Agius^{1,2}

¹Clare College Cambridge, Cambridge, UK

²Department of Psychiatry, University of Cambridge, Cambridge, UK

SUMMARY

The PHQ-2 has been presented as an ultra-short screening questionnaire to detect depression in several healthcare settings. This paper will outline its current uses, evaluate its accuracy as a screening method, and suggest potential uses.

Key words: depression - PHQ-2 - screening

* * * * *

INTRODUCTION

Depression is a mental illness characterized by depressed mood and anhedonia (Arroll 2010). Multiple questionnaires and criteria have been proposed which may be used to enable a diagnosis of depression to be made. Such Diagnostic questionnaires include the Composite International Diagnostic Interview (CIDI) (Arroll 2010), the Structured Clinical Interview for DSM-IV (SCID) (Bhana 2015), the Hospital Anxiety and Depression Scale (HAD) (Al-Qadhi 2014), the Hamilton Depression scale, the Montgomery and Asberg Depression Rating Scale (MADRS), the Beck Depression Inventory (Gjerdingen 2009), as well as the Patient Health Questionnaire (PHQ).

Family physicians often do not identify depressive symptoms; it is reported that over half of the cases of MDD are missed in primary care settings (Arroll 2010, Al-Qadhi 2014, Gjerdingen 2009, Choi 2015, Zhang 2013). This may be because physicians concentrate on primary conditions rather than underlying mental health issues (Margrove 2011). Alternatively, this might be because of a lack of a systematic screen. For example, only 5.9% of staff in cancer, palliative care, or related disciplines in the UK reported using a formal questionnaire to screen for depression, since there was no national guidance on which depression screens to use (Mitchell 2008).

WHAT IS THE PHQ-2?

The PHQ-2 comprises the first two questions of the PHQ-9. The PHQ-9 is a questionnaire made up of 9 questions derived from the DSM-IV criteria for major depressive disorder (MDD) (Arroll 2010). The answers are scored on a Likert scale. A cumulative score of above ten or eleven usually results in a diagnosis of depression, but there are also algorithmic methods of working out PHQ-9 scores. The PHQ-9 has become the

preferred diagnostic questionnaire in non-psychiatric settings because it is easy to administer and score (Zuithoff 2010).

The first question of the PHQ-2 asks how often the patient has felt down, depressed, and hopeless over the past two weeks, with answers from 0 (not at all) to 3 (nearly every day). The second question asks how often the patient has lost pleasure or interest in doing things over the same time scale, with the same answers (Arroll 2010). The cumulative score is added up, with threshold score for suspecting the possibility of depression usually of 3 or higher (Arroll 2010, Zhang 2013).

HOW IS PHQ-2 CURRENTLY USED?

The PHQ-2 is not a way of diagnosing depression. This is emphasised in almost all studies that research the accuracy and validity of the questionnaire, however there are some reports in which it is incorrectly used as a definitive diagnosis (Schlosser 2016).

The PHQ-2 can be used to screen patients and find those who might be suffering from depression. These patients should then be given more comprehensive diagnosis questionnaires, such as the PHQ-9 (Arroll 2010, Gjerdingen 2009, Zhang 2013, Corson 2004, Fuchs 2015, Inagaki 2013, Li 2007, Loeb 2015, Thombs 2008).

Unfortunately, the PHQ-2 is not often recommended in healthcare settings around the world. The exception is in the USA. The American Heart Association (AHA) recommends the PHQ-2 followed by PHQ-9 as a systematic screen for depression in cardiovascular patients (Elderon 2011). The US Preventive Services Task Force (USPSTF) recommends routine screening for depression provided that accurate systems are in place (Gjerdingen 2009, Richardson 2010). The American Academy of Paediatrics recommends screening amongst adolescents (Sudhanthat 2015). Despite recommendations for systematic screening, in the USA,

a positive PHQ-2 only currently leads to a PHQ-9 in 5% of cases, instead, physicians tend to use prior knowledge, judgement, and patient history to diagnose depression (Fuchs 2015).

It has been suggested that formal implementation of systematic screening, such as PHQ-2 followed by PHQ-9, could help to decrease the number of missed cases of depression (Loeb 2015). Systematic screening improve rates of referral for further psychiatric care (Sudhanthath 2015). Student-run free clinics (SRFCs) in the USA have successfully implemented the PHQ 2 screen and have identified previously undiagnosed depression (Soltani 2015).

DOES PHQ-2 EFFECTIVELY SCREEN FOR DEPRESSION?

PHQ-2 correlates well with PHQ-9 and with other measures of depression (Al-Qadhi 2014, Yu 2011). Adolescents with scores of above three on PHQ-2 also correlated with functional impairment by the Columbia Impairment Scale, and parent-reported psychosocial impairment (Richardson 2011). These data suggest that PHQ-2 can predict depression.

The accuracy of screening may be measured using sensitivity, specificity, and positive predictive values (PPV). Sensitivity is the probability of testing positive if disease is truly present. Specificity is the probability of testing negative if disease is not present. PPV is the proportion of positive screens that are true positives.

Sensitivity, specificity, and PPVs all compare one screen to other diagnostic tools, but as there are not definitive diagnostic criteria for depression, it is difficult to calculate and compare PHQ-2 results. In each of the studies in question, PHQ-2 results are compared with different 'gold-standard' diagnostic questionnaires, so that data are not comparable between studies. Despite the lack of ability to directly compare accuracy, most studies conclude that the PHQ-2 is both reliable and valid as a screening tool (Arroll 2010, Zhang 2013, Kroenke 2003). It is observed that PHQ-2 scores do remain stable in patients over a long period of time, suggesting that it is a reliable measure (Margrove 2011).

One problem to consider is that specificity may be significantly different depending on age, sex, race, and ethnicity. These differences may be due to different presentation of symptoms. For example, adolescents can present with irritability, rather than low mood (Richardson 2010). However, such differences are not highly clinically significant and may be ignored for the purpose of screening (Li 2007). As a consequence, studies from different populations all over the world do show similar results (Figure 1).

Sensitivity and specificity depend on the threshold score used in evaluating the PHQ-2. When a threshold of three is used, the PHQ-2 is highly specific, with lower sensitivity. Lowering the threshold to two increases

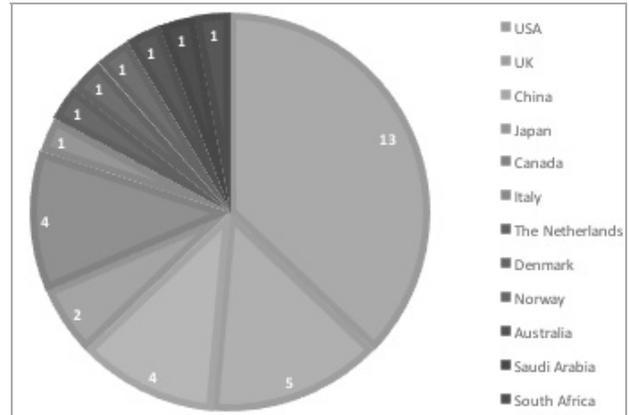


Figure 1. Pie chart showing the country of origin for the studies cited in this paper

sensitivity at the cost of specificity (Arroll 2010, Bhana 2015, Choi 2015, Zuithoff 2010, Inagaki 2013, Thombs 2008, Liu 2016, Suzuki 2016, Thapar 2014). Therefore, selecting the correct threshold for screening, depends on whether the sensitivity or the specificity is considered more valuable.

The main argument for a threshold of two is that sensitivity should be high for PHQ so that patients who could be a danger to themselves are not missed (Inagaki 2013). Specificity does not necessarily have to be high for a screen, since PHQ-9 can rule out false positives. False positives may indicate patients at risk of depression, so lower specificity could actually be quite useful (Richardson 2010).

However, a threshold of three is most commonly used. This is because specificity is considered a more valuable property of screens (Mitchell 2012). A high specificity ensures faster and more accurate screening in busy healthcare settings, thus minimising the number of false positive patients who will have to go through further tests (Bhana 2015, Yu 2011).

Interestingly, it has been found that, when looking at single questions from PHQ-9 as a screen for depression in cancer patients in the UK, the two used in PHQ-2 are neither the most sensitive nor the most specific when compared with DSM-IV depression criteria (Mitchell 2012). They do have high specificity amongst the others, but the PHQ-2 questions have the lowest sensitivity of all the questions except the question that asks about suicidal thoughts. The question that asks about having little energy has been found to be the most sensitive, but with the least specificity. The question that asks about having trouble concentrating on things such as reading is the question with the most specificity (Mitchell 2012) and has been suggested to be used as a supplementary question for the PHQ-2 to increase its PPV (Thapar 2014).

There is an ongoing meta-analysis looking at the diagnostic accuracy of PHQ-2, but the results have not yet been published (Thombs 2014) at time of writing. Generally, PHQ-2 is considered an effective screening tool, but the results of the meta-analysis will give a more definitive answer.

HOW COULD PHQ-2 BE USED IN THE FUTURE?

Screening for depression with PHQ-2 has low PPV in primary care, but high PPV for high-risk samples (Mitchell 2012). This may be because primary care patients are likely to have depressive symptoms because of their health status (Al-Qadhi 2014). For example, health conditions such as the flu may cause anhedonia because of fatigue. This could also work in the opposite way – people may attribute anhedonia to their health status and then answer negatively in a depression screen because of this attribution (Margrove 2011, Inagaki 2013).

APHQ-2 screen could have useful applications in at-risk groups such as adolescents (Zhang 2013, Richardson 2010, Haugen 2016), women who have recently given birth (Gjerdingen 2009), as well as patients with chronic illnesses (Choi 2015, Margrove 2011, Schlosser 2016, Mitchell 2012), rather than the general population in primary care settings.

The PHQ-2 takes less than 2 minutes (Mitchell 2012), so that it can be quickly administered in busy settings (Li 2007, Liu 2016, Thapar 2014), such as in routine check-ups, or in student clinics. The short length makes electronic administration possible, which is helpful for screening for post-partum depression (Gjerdingen 2009), cardiovascular patients (Wang 2015, Elderon 2011), patients who have suffered trauma (Warren 2016), amongst students, or amongst patients with chronic illness (Smits 2015) including haemodialysis (Cusimano 2015).

It is also necessary to take ethics into consideration. In one study, only 80% of new mothers were comfortable with the idea of screening for postpartum depression (Gjerdingen 2009). It may be considered to be unethical to introduce the PHQ-2 into a routine check-up without first asking for permission to screen for depression. On the other hand, there is the risk that, if the patient is informed, stigma against depression may influence them to answer negatively, especially amongst the elderly (Inagaki 2013).

PHQ-2 is a reliable method of screening, but not diagnosing, depression. It has a wide range of applications, however it must be implemented as part of a systematic screen within healthcare services for it to reach its full potential.

Acknowledgements: None.

Conflict of interest: None to declare.

Contribution of individual authors:

Contribution of individual authors: Mark Agius devised the questions supervised and corrected the text.

Rachel Wilson drafted the text and conducted the literature search.

References

1. Al-Qadhi W, Ur Rahman S, Ferwana MS, Abdulmajeed IA. Adult depression screening in Saudi primary care: prevalence, instrument and cost. *BMC Psychiatry* 2014; 14:190.
2. Arroll B, Goodyear-Smith F, Crengle S, Gunn J, Kerse N, Fishman T, Falloon K, Hatcher S. Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. *Ann Farm Med* 2010; 8:4.
3. Bhana A, Rathod SD, Selohilwe O, Kathree T, Petersen I. The validity of the Patient Health Questionnaire for screening depression in chronic care patients in primary health care in South Africa. *BMC Psychiatry* 2015; 15:118.
4. Carey M, Boyes A, Noble N, Waller A, Inder K. Validation of the PHQ-2 against the PHQ-9 for detecting depression in a large sample of Australian general practice patients. *Aust J Prim Health* 2016; 22:262-6.
5. Choi SK, Boyle E, Burchell AN, Gardner S, Collins E, Grootendorst P, Rourke SB, OHTN Cohort Study Group. Validation of Six Short and Ultra-short Screening Instruments for Depression for People Living with HIV in Ontario: Results from the Ontario HIV Treatment Network Cohort Study. *PLoS One* 2015; 10:11.
6. Corson K, Gerrity MS, Dobscha SK. Screening for depression and suicidality in a VA primary care setting: 2 items are better than 1 item. *Am J Manag Care* 2004; 11:10.
7. Cusimano P, Palermo A, Locascio G. Screening for depression in hemodialysis. *G Ital Nefrol* 2015; 3:32.
8. Elderon L, Smolderen KG, Na B, Whooley MA. Accuracy and prognostic value of American Heart Association: recommended depression screening in patients with coronary heart disease: data from the Heart and Soul Study. *Circ Cardiovasc Qual Outcomes* 2011; 5:4.
9. Fuchs CH, Haradhvala N, Hubley S, Nash JM, Keller MB, Ashley D, Weistberg RB, Ueblacker LA. Physician actions following a positive PHQ-2: implications for the implementation of depression screening in family medicine practice. *Fam Syst Health* 2015; 1:33.
10. Gjerdingen D, Crow S, McGovern P, Miner M, Center B. Postpartum depression screening at well-child visits: validity of a 2-question screen and the PHQ-9. *Ann Farm Med* 2009; 7:1.
11. Haugen W, Haavet OR, Sirpal MK, Christensen KS. Identifying depression among adolescents using three key questions: a validation study in primary care. *Br J Gen Pract* 2016; 643:66.
12. Inagaki M, Ohtsuki T, Yonemoto N, Kawashima Y, Saitoh A, Oikawa Y, Kurosawa M, Muramatsu K, Furukawa TA, Yamada M. 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. *Gen Hosp Psychiatry* 2013; 6:35.
13. Kroenke K, Spitzer RL, Williams JB. The Patient Health Questionnaire-2: validity of a two-item depression screener. *Med Care* 2003; 11:41.
14. Li C, Friedman B, Conwell Y, Fiscella K. Validity of the Patient Health Questionnaire 2 (PHQ-2) in identifying major depression in older people. *J Am Geriatr Soc* 2007; 4:55.
15. Liu ZW, Yu Y, Hu M, Liu HM, Zhou L, Xiao SY. PHQ-9 and PHQ-2 for Screening Depression in Chinese Rural Elderly. *PLoS One* 2016; 3:11.

16. Loeb D, Sieja A, Corral J, Zehnder NG, Guilton G, Nease DE. Evaluation of the role of training in the implementation of a depression screening and treatment protocol in 2 academic outpatient internal medicine clinics utilizing the electronic medical record. *Am J Med Qual* 2015; 4:30.
17. Margrove KL, Thapar AK, Mensah SA, Kerr MP. Help-seeking and treatment preferences for depression in epilepsy. *Epilepsy Behav* 2011; 4:22.
18. Mitchell AJ, Kaar S, Coggan C, Herdman J. Acceptability of common screening methods used to detect distress and related mood disorders-preferences of cancer specialists and non-specialists. *Psychooncology* 2008; 3:17.
19. Mitchell AJ, Lord K, Symonds P. Which symptoms are indicative of DSMIV depression in cancer settings? An analysis of the diagnostic significance of somatic and non-somatic symptoms. *J Affect Disord* 2012; 1-2:2012.
20. Mitchell, AJ. Clinical utility of screening for clinical depression and bipolar disorder. *Curr Opin Psychiatry* 2012; 1:25.
21. Richardson LP, Rockhill C, Russo JE, Grossman DC, Richards J, McCarty C, McCauley E, Katon W. Evaluation of the PHQ-2 as a brief screen for detecting major depression among adolescents. *Pediatrics* 2010; 5:125.
22. Schlosser RJ, Hyer JM, Smith TL, Mace JC, Cortese BM, Uhde TW, Rudmik L, Soler ZM. Depression-Specific Outcomes After Treatment of Chronic Rhinosinusitis. *JAMA Otolaryngol Head Neck Surg* 2016; 4:142.
23. Soltani M, Smith S, Beck E, Johnson M. Universal depression screening, diagnosis, management, and outcomes at a student-run free clinic. *Acad Psychiatry* 2015; 3(39).
24. Smits N, Finkelman MD. Shortening the PHQ-9: a proof-of-principle study of utilizing Stochastic Curtailment as a method for constructing ultrashort screening instruments. *Gen Hosp Psychiatry* 2015; 5:37
25. Sudhanthas S, Thakur K, Sigal Y, Turner J. Improving validated depression screen among adolescent population in primary care practice using electronic health records (EHR). *BMJ Qual Improv Rep* 2015; 1:4
26. Suzuki K, Kumei S, Ohhira M, Nozu T, Okumura T. Screening for major depressive disorder with the Patient Health Questionnaire (PHQ-9 and PHQ-2) in an outpatient clinic staffed by primary care physicians in Japan: a case control study. *PLoS One* 2015; 3:10.
27. Thapar A, Hammerton G, Collishaw S, Potter R, Rice F, Harold G, Craddock N, Thapar A, Smith DJ. Detecting recurrent major depressive disorder within primary care rapidly and reliably using short questionnaire measures. *Br J Gen Pract* 2014; 618:64.
28. Thombs BD, Ziegelstein RC, Whooley MA. Optimizing detection of major depression among patients with coronary artery disease using the patient health questionnaire: data from the heart and soul study. *J Gen Intern Med* 2008; 12:23.
29. Thombs BD, Benedetti A, Kloda LA, Levis B, Nicolau I, Cuijpers P, Gilbody S, Ioannidis JP, McMillan D, Patten SB, Shrier I, Steele RJ, Ziegelstein RC. The diagnostic accuracy of the Patient Health Questionnaire-2 (PHQ-2), Patient Health Questionnaire-8 (PHQ-8), and Patient Health Questionnaire-9 (PHQ-9) for detecting major depression: protocol for a systematic review and individual patient meta-analyses. *Syst Rev* 2014; 124:3.
30. Yu X, Stewart SM, Wong PT, Lam TH. Screening for depression with the Patient Health Questionnaire-2 (PHQ-2) among the general population in Hong Kong. *J Affect Disord* 2011; 1-3:134.
31. Wang L, Lu K, Li J, Sheng L, Ding R, Hu D. Value of patient health questionnaires (PHQ)-9 and PHQ-2 for screening depression disorders in cardiovascular outpatients. *Zhonghua Xin Xue Guan Bing Za Zhi* 2015; 5:43.
32. Warren AM, Reynolds M, Foreman ML, Bennett MM, Weddle RJ, Austin JD, Roden-Foreman K, Petrey LB. Validation of a brief, two-question depression screen in trauma patients. *J Trauma Acute Care Surg* 2016; 2:80.
33. Zhang YL, Liang W, Chen ZM, Zhang HM, Zhang JH, Weng XQ, Yang SC, Zhang L, Shen LJ, Zhang YL. Validity and reliability of Patient Health Questionnaire-9 and Patient Health Questionnaire-2 to screen for depression among college students in China. *Asia Pac Psychiatry* 2013; 5:4.
34. Zuithoff NP, Vergouwe Y, King M, Nazareth I, van Wezep MJ, Moons KG, Geerlings MI. The Patient Health Questionnaire-9 for detection of major depressive disorder in primary care: consequences of current thresholds in a cross-sectional study. *BMC Fam Pract* 2010; 98:11.

Correspondence:

Mark Agius, MD
Department of Psychiatry, University of Cambridge
Cambridge, UK
E-mail: ma393@cam.ac.uk