Can the two questions of the PHQ-2 screen effectively for depression?

Rachel Wilson[1,2], Mark Agius[1,2,*]

[1] Clare College Cambridge, Cambridge, UK

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

***Corresponding author email: ma393@cam.ac.uk**

This paper is an updated version of 'Is there good evidence that the two questions in PHQ-2 are useful questions to use in order to screen for depression?' previously published in Psychiatria Danubina, 2017; 29 (Suppl. 3): 232-235

Abstract

The PHQ-2 has been presented as an ultra-short screening questionnaire to detect depression in several healthcare settings. It consists of two questions on low mood and anhedonia, both questions being derived from the more accurate PHQ-9. The PHQ-2 is to be used as an initial screening tool for patients with possible depression, which can then be evaluated in more detail using other tools, such as the PHQ-9. The PHQ-2 is both reliable and valid as a screening tool. Sensitivity and specificity do depend on the threshold score used in evaluating the PHQ-2. When a threshold of three is used, the PHQ-2 is highly specific, but with lower sensitivity. The PHQ-2 is not currently used as widely as it could be and in many situations depression continues to be under-diagnosed. We hypothesise that, if used routinely in several situations, the identification of depression would be much improved.

Keywords: depression, PHQ-2 screening

Introduction

Depression has been described as a mental illness characterised by depressed mood and anhedonia (1). Several criteria and questionnaires have been proposed in order to facilitate the process of diagnosing depression. These diagnostic questionnaires include the Composite International Diagnostic Interview (CIDI) (1), the Structured Clinical Interview for DSM-IV (SCID) (3), the Hospital Anxiety and Depression Scale (HAD) (2), the Hamilton Depression scale, the Montgomery and Asberg Depression Rating Scale (MADRS), the Beck Depression Inventory (4), and the Patient Health Questionnaire (PHQ).

It has been reported that family physicians often do not identify and effectively evaluate depressive symptoms: over half of the cases of major depressive disorder (MDD) are missed in primary care settings (1, 2, 4-6) This may be because physicians concentrate on primary physical conditions rather than underlying mental health issues (7). Alternatively, it might be because of a lack of a suitable systematic screening tool. For example, only 5.9% of staff in cancer care, palliative care, or related disciplines in the UK reported using a formal questionnaire in order to screen for depression, since there was no national guidance on which depression screening tool to use (8).

What is the PHQ-2?

The PHQ-2 consists of the first two questions of the PHQ-9. The PHQ-9 is a questionnaire which consists of nine questions derived from the DSM-IV criteria for MDD (1). The answers are scored on a Likert scale. A cumulative score of above 10 or 11 usually results in a diagnosis of depression, but there are 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 (9).

The first question of the PHQ-2 enquires how often the patient has felt down, depressed, or hopeless over the past two weeks, with answers from 0 (not at all) to 3 (nearly every day).

The second question enquires how often the patient has noted that he/she has lost pleasure or interest in doing things over the past two weeks, with the same possible answers (1). The cumulative score is added up, with a threshold score for considering the possibility of depression of three or higher (1, 6).

How is PHQ-2 currently used?

The PHQ-2 is not meant to be a method for diagnosing depression. Almost all studies that research the accuracy and validity of the questionnaire emphasise this. However, there are some reports in which it has been incorrectly used to make a definitive diagnosis of depression (10).

The PHQ-2 is designed to 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 (1, 4, 6, 11-16).

The PHQ-2 is not often recommended in healthcare settings around the world but it is recommended in the USA. The American Heart Association (AHA) recommends the PHQ-2 followed by the PHQ-9 as a systematic screen for depression in cardiovascular patients (17). The US Preventive Services Task Force (USPSTF) recommends routine screening for depression provided that accurate systems are in place (4, 18). The American Academy of Pediatrics recommends screening amongst adolescents (19). However, 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 (12).

It has been suggested that the formal implementation of systematic screening, such as PHQ-2 followed by PHQ-9, could help to decrease the number of missed cases of depression (15). Systematic screening could improve rates of referral for further psychiatric care (19). Student-run free clinics (SRFCs) in the USA have successfully implemented using the PHQ-2 screen and have led to the identification of previously undiagnosed depression (20).

Does PHQ-2 screen effectively for depression?

It has been shown that PHQ-2 correlates well with PHQ-9 and with other measures of depression (2, 21). In adolescents, scores of above three on PHQ-2 have correlated with functional impairment on the Columbia Impairment Scale, as well as parent-reported psychosocial impairment (18). These data suggest that the PHQ-2 can predict a diagnosis of depression.

The accuracy of screening may be measured using sensitivity, specificity, and positive predictive values (PPV). Sensitivity is the probability of testing positive if a condition is truly present. Specificity is the probability of testing negative if a condition is not present. PPV is the proportion of positive screens that are true positives. Sensitivity, specificity, and PPV can all serve to compare a single screening tool to other diagnostic tools but, as there are no definitively accepted diagnostic criteria for depression, it is difficult to evaluate PHQ-2 results.

In each of the studies mentioned, the PHQ-2 results are compared to different 'gold-standard' diagnostic questionnaires, implying that data from the various studies cannot be compared directly. However, most studies do conclude that the PHQ-2 is both reliable and valid as a screening tool (1, 6, 22). PHQ-2 scores appear to remain stable in patients over a long period of time, suggesting that it is a reliable measure (7).

One problem to consider is that specificity may be significantly different depending on age, sex, race, or ethnicity. These differences may arise because of different presentations of symptoms. For example, adolescents may present with irritability rather than low mood (18). However, such differences are not highly clinically significant and it has been suggested that they may be ignored for the purpose of screening (14). As a consequence, studies from different populations all over the world do appear to show similar results. Countries, which have reported the use of the PHQ-2, include the USA, UK, China, Japan, Canada, Italy, the Netherlands, Denmark, Norway, Australia (34), Saudi Arabia (2) and South Africa. 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, but with lower sensitivity. Lowering the threshold to two increases sensitivity at the cost of specificity (1, 3, 5, 9, 13, 16, 23-25). Therefore, selecting the appropriate threshold for screening depends on whether sensitivity or specificity is considered the more important in the screening program. The main argument for using 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 (13).

Specificity does not necessarily have to be high for a screening tool, since PHQ-9 can then be used to

rule out false positives. False positives may indicate patients at risk of depression, and so lower specificity could actually be quite useful (18). However, in practice, a threshold of three is more commonly used. This is because specificity is considered a more valuable property for screening tools (26, 27). A high specificity does ensure 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 (3, 21).

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 were found to be neither the most sensitive nor the most specific when compared with DSM-IV depression criteria (26, 27). They do have a 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 has the least specificity. The question that asks about having trouble concentrating on things such as reading is the question with the most specificity (26, 27) and this has therefore been suggested as a supplementary question for the PHQ-2 to increase its PPV (25).

There is an ongoing meta-analysis looking at the diagnostic accuracy of PHQ-2; however, the results have not yet been published (28) at time of writing.

Generally, PHQ-2 is considered to be an effective screening tool, but we await the results of the meta-analysis in order to have a more definitive answer.

How could we use the PHQ-2 in the future?

Screening for depression with PHQ-2 has a low PPV in primary care, but a high PPV for high-risk samples (26, 27). This could be because primary care patients are likely to have depressive symptoms because of their health status (2). For example, health conditions such as the flu may cause anhedonia due to fatigue. This could also work in the opposite direction - people may attribute anhedonia to their health status and then answer negatively when filling out a depression screening tool because of this attribution (7, 13). A PHQ-2 screen could have useful applications in at-risk groups such as adolescents (6, 18, 29), women who have recently given birth (4), and also patients with chronic illnesses (5, 7, 10, 26, 27), rather than in the general population in primary care settings.

The PHQ-2 takes less than two minutes (26, 27), which makes it possible to administer it quickly in busy settings (14, 23, 25), including routine check-ups, or in student clinics. The short length also makes electronic administration possible, which could be helpful for screening for postpartum depression (4), cardiovascular patients (30, 17), patients who have suffered trauma (31), amongst students, or amongst patients with chronic illness (32), including haemodialysis (33).

It is also necessary to take ethics into consideration. Some people do not wish to be screened for depression. For example, in one study, only 80% of new mothers were comfortable with the idea of screening for postpartum depression (4). It could 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 of the screening process, stigma against depression may influence them to answer negatively. It is expected that this may be especially so amongst the elderly (13).

Conclusion

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

References

1. Carroll 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. Annals of Family Medicine 2010; 8:4.

2. Al-Qadhi W, Ur Rahman S, Ferwana MS, AbdulmajeedIA. Adult depression screening in Saudi primary care: prevalence, instrument and cost. BMC Psychiatry 2014; 14:190.

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. 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. Annals Family Medicine 2009; 7:1.

5. Choi SK, Boyle E, Burchell AN, Gardner S, Collins E, Goaltenders 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. 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 Pacific Psychiatry 2013; 5:4.

7. Margrove KL, Thapar AK, Mensah SA, Kerr MP. Helpseeking and treatment preferences for depression in epilepsy. Epilepsy & Behavior 2011; 4:22.

8. 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.

9. 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 Family Practice 2010; 98:11.

10. 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 Otolaryngology Head & Neck Surgery 2016; 4:142.

11. Corson K, Gerrity MS, Dobscha SK. Screening for depression and suicidality in a VA primary care setting: 2 items are better than 1 item. American Journal of Managed Care 2004; 11:10.

12. 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. Families Systems and Health 2015; 1:33.

13. 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. General Hospital Psychiatry 2013; 6:35.

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. Journal of the American Geriatric Society 2007; 4:55.

15. 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. American Journal of Medical Quality 2015; 4:30.

16. 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. Journal of General Internal Medicine 2008; 12:23.

17. 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. Circulation: Cardiovascular Quality and Outcomes 2011; 5:4.

18. 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.

19. Sudhanthat S, Thakur K, Sigal Y, Turner J. Improving validated depression screen among adolescent population in primary care practice using electronic health records (EHR). BMJ Quality Improvement Reports 2015; 1:4

20. Soltani M, Smith S, Beck E, Johnson M. Universal depression screening, diagnosis, management, and outcomes at a student-run free clinic. Academic Psychiatry 2015; 3(39).

21. 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. Journal of Affective Disorders 2011; 1-3:134.

22. Kroenke K, Spitzer RL, Williams JB. The Patient Health Questionnaire-2: validity of a two-item depression screener. Medical Care 2003; 11:41.

23. 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.

24. Suzuki K, Kumei S, Ohhira M, Nozu T, OkumuraT. 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.

25. 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. British Journal of General Practice 2014; 618:64.

26. 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 nonsomatic symptoms. Journal of Affective Disorders 2012; 1-2:2012.

27. Mitchell, AJ. Clinical utility of screening for clinical depression and bipolar disorder. Current Opinion in Psychiatry 2012; 1:25.

28. Thombs BD, Benedetti A, Kloda LA, Levis B, NicolauI,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. Systematic Reviews 2014; 124:3.

29. Haugen W, Haavet OR, Sirpal MK, Christensen KS. Identifying depression among adolescents using three key questions: a validation study in primary care. British Journal of General Practice 2016; 643:66.

30. 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 ZaZhi 2015; 5:43.

31. 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. Journal of Trauma and Acute Care Surgery 2016; 2:80.

32. 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. General Hospital Psychiatry 2015; 5:37

33. Cusimano P, Palermo A, Locascio G. Screening for depression in heamodialysis. Giornale Italiano di Nefrologia 2015; 3:32.

34. 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. Australian Journal of Primary Health 2016; 22:262-6.