THE USE OF THE PHQ9 SELF-RATING SCALE
TO ASSESS DEPRESSION WITHIN PRIMARY CARE

Ankita Sahni¹ & Mark Agius¹,²
¹Clare College Cambridge, Cambridge, UK
²Department of Psychiatry, University of Cambridge, Cambridge, UK

SUMMARY
The PHQ-9 is effective in screening patients for depression as well as monitoring progress in a variety of situations. Using the PHQ-9 after a pre-assessment with the PHQ-2 increases its specificity, as well as preventing under-diagnosis. Although it is not suitable as a stand-alone tool for a diagnosis, it is a cost-effective, efficient method of screening patients in primary care.

Key words: depression – screening - Patient Health Questionnaire-9 - primary care

Introduction
Depression is one of the most prevalent mental disorders (The WHO World Mental Health Survey Consortium 2004). It is described by the DSM-IV-TR as a loss of interest or pleasure in daily activities resulting in impaired functioning. Nine specific symptoms of depression are described: depressed mood, interest; change in appetite, sleep, activity; fatigue, feelings of worthlessness, decrease in concentration, and suicidality (The American Psychiatric Association 2000).

Depression has many risk factors. These include chronic medical conditions, stress, a family history of depression, gender, and substance abuse. There are also several social risk factors which contribute to the development of depression: low income, job loss, lack of social support, and being single, divorced, or widowed (Maurer 2012).

PHQ-9
The Patient Health Questionnaire (PHQ-9) is an instrument designed for making diagnoses of depressive disorders based exclusively on the criteria identified by the Diagnostics and Statistical Manual of Mental Disorders. It consists of nine, multiple-choice items. Each criterion is scored between 0 and 3 denoting “not at all”, “some days” and “nearly every day”, respectively. The PHQ-9 is the self-administered version of the mood module of the PRIME-MD. It has diagnostic validity comparable to the clinician-administered PRIME-MD (Spitzer 1999). Studies have also validated its reliability when used as a telephone assessment tool (Pinto-Meza 2005). The PHQ-9 is specifically designed to be used in primary care. It is short, and has been shown to be effective in a variety of patient groups.

In the UK, the PHQ-9 has been included in the NICE guidelines as an assessment for the identification of common mental health disorders. It is frequently used as a screening and diagnostic tool for mental health disorders of depression, anxiety and eating in primary care in the United Kingdom. It is used to screen patients with a history of depression, and at-risk groups. These include individuals with chronic medical conditions. The current guidelines recommend a two-stage process of identification and diagnosis. Patients are first screened for depression using the diagnostic tool, such as the PHQ-2 or PHQ-9. Those screened positive should then be interviewed before a formal diagnosis is made and given (National Collaborating Centre for Mental Health 2010).

Cut off values
The proposed cut-off points for the scores are 5, 10, 15, and 20 for mild, moderate, moderately severe, and severe depression, respectively (Kroenke 2001). The optimal cut-off score is, however, disputed. A 2012 meta-analysis concluded that the PHQ-9 had acceptable diagnostic properties with cut-off scores between 8 and 11 (Manea 2012). Other meta-analyses, however, have identified a score of 10 in order to fulfil the minimum criteria for sensitivity and specificity (Moriarty 2015, Pettersson 2015).

The PHQ-9 is a useful screening tool in many different settings. It does not have a specified target scenario or audience. An optimal cut-off score may differ depending on the setting. Care must be taken to ensure that the optimal cut-off scores are being used for the specific settings. An example is in the screening of adolescent depression. The PHQ-9 has been validated for use in adolescents. Its sensitivity and specificity are similar to those for adults. However, the optimal cut-off is, higher for adolescents (Richardson 2010).

Benefits of the PHQ-9
The PHQ-9 has been validated for use as a clinical and research tool (Kroenke 2001). It is shown to have overall high specificity. The sensitivity is suboptimal in certain conditions, such as when using the algorithm scoring method (Manea 2015). In selected subgroups of patients with a high prevalence of depressive disorder (30-40%), the PHQ-9 has a positive predictive value of 85-90% (Wittkampf 2007).
The PHQ-9 is a quick and short assessment, with easy to remember cut-off values. The scores are calculated additively, decreasing the room for error. As well as being useful for the identification of depression, the PHQ-9 can also be used to grade the severity of depressive symptoms (Kroenke 2001, Kroenke 2002, Löwe 2004). Grading severity allows it to be effective for monitoring patients (Löwe 2004, Beard 2016). The PHQ-9 is used to monitor responses to treatment by observing changes in the scores. End-of-treatment cut-off points are used to support clinical decision-making (Schueller 2015). The PHQ-9 has potential efficacy in identifying non-response to psychiatric treatment (Fowler 2015).

The PHQ-9 is available in a variety of languages, including Arabic, Chinese (Cantonese, Mandarin), English, French, Hindi, and Spanish. It has been found to be effective and is in use in several countries worldwide. These countries include China (Xiong 2015), East Africa (Gelaye 2013), and Malaysia (Sherina 2012). The PHQ-9 responses of different racial and ethnic groups have also been analysed. It has been shown to be an effective screening tool in diverse populations (Huang 2006).

Downfalls of the PHQ-9

A common drawback of the PHQ-9 is that, although it has high specificity, it has a sub-optimal sensitivity. This shortcoming is averted through the use of the PHQ-2. The PHQ-2 is an ultra-short screening questionnaire designed to be used as a first line measure. If a patient screens as positive on PHQ-2, they are followed up with the PHQ-9. The PHQ-2 has a very high sensitivity at a cut-off value of “2”. This identifies patients that may be depressed. The specificity of the PHQ-9 prevents false positives (Richardson 2010).

Some findings have suggested that PHQ-9 assessment may lead to unnecessary diagnosis and antidepressant recommendation. It may be more likely to identify a depressive episode than a major depressive disorder (Jerant 2014). Caution should be taken to ensure that the benefits of screening far outweighs the risk of over treatment with antidepressants.

The DSM-IV-TR specifies several exclusion criteria, including the absence of manic or hypomanic episodes. The PHQ-9 does not include these exclusion items. This can result in the misdiagnosis of bipolar disorder patients as having major depressive disorder (Inoue 2012). This can be avoided if clinicians rule out other causes of depression, including bereavement, and history of a manic episode (Kroenke 2001).

Detecting depression in patients with chronic medical conditions

It is known that chronic medical conditions are risk factors for depression. These conditions can be debilitating and cause mental distress to patients.

Research shows that people with diabetes have an increased risk of developing depression (Roy 2012). It is therefore important that they are assessed for depressive symptoms. Present guidelines are ambiguous, leading to low rates of screening (Willborn 2015). The validity of screening tools used in these situations must also be assessed. It has been pointed out that several questions listed in the PHQ-9, screen for symptoms that are also categorised as diabetes-related symptoms. This decreases the specificity of the method, and results in over-identification of depression in the diabetic population (Twist 2013). The PHQ-9 performs well for identifying high risk patients. It is not, in its current state, however, suitable to be a stand-alone diagnostic tool for patients with diabetes (Van der Zwaan 2016). In these patients, the PHQ-9 identifies more patients as moderate to severe than HADS-D (Reddy 2010). The HADS-D may be a more suitable tool for screening patients with diabetes.

Depression is common in patients with coronary heart disease. It has independent associations with increased cardiovascular morbidity and mortality (Lichtman 2009). The PHQ-8 may, however, be a more suitable tool for coronary heart disease patients. Item 9 is not an accurate suicide screen. It is omitted in the PHQ-8 (Razykov 2012). The PHQ-9 is used to screen for depression in patients with cardiovascular disease. It has a high specificity for depression, but poor sensitivity (Thombs 2008). Lowering the cut-off scores improves the sensitivity of the PHQ-9. The specificity is retained. This way of using PHQ-9 may be a more useful screen for cardiovascular disease patients (Stafford 2007).

Why choose the PHQ-9?

There are several other diagnostic tools for depression that have been validated for use. The Beck Depression Inventory, and the Hospital Anxiety and Depression Scale (HADS) are commonly used. Unlike these tools, the PHQ-9 is free to use. The PHQ-9 has equivalent, if not superior, specificity and sensitivity to these other scales (Thapar 2013, Kung 2013, Löwe 2004). It is therefore cost effective to use the PHQ-9.

The PHQ-9 is shorter than many other diagnostic tools for depression, such as the HADS and the Composite International Diagnostic Interview (CIDI). A CIDI must be performed by a clinician, while the PHQ-9 and HADS can be administered in a variety of ways including as self-assessments. This is beneficial considering the time constraints of primary care.

All of these scales demonstrate high internal consistency in their scoring. They differ in how they categorise severity (Cameron 2008). HADS and PHQ-9 do not fully identify the same cases. While they recognise the same prevalence of mild and moderate depression, PHQ-9 identifies more patients as severe. This suggests that the two scales are not fully interchangeable (Hansson 2009).
Limitations

It needs to be taken into account that these conclusions are drawn in reference to previous studies. The accuracy of the results of these studies must be questioned. Depression screening tools must be able to accurately identify depressed patients that would not otherwise be recognised. Inclusion of currently diagnosed and treated patients may increase bias in studies by inflating estimates of screening accuracy (Rice 2016). It is difficult to evaluate the accuracy of studies that do not report samples size calculations, or have small sample sizes. Very few of the studies quoted meet the criteria needed to determine precise estimates of their accuracy (Thombs 2016).

Closing thoughts

It is also important to note that psychiatric illnesses are thought to have a causal heterogeneity (Maung 2016). Although the PHQ-9 has efficacy in screening for depression, it is unable to determine predisposing, precipitating, and perpetuating factors. It would be unwise to make a diagnosis using solely the PHQ-9. However, using it in conjunction with a full bio-psycho-social history could allow a more accurate and comprehensive assessment of a patient, allowing better diagnosis and management in primary care.

Acknowledgements: None

Conflict of interest: None.

Contribution of individual authors:

Mark Agius suggested the project, and supervised and edited the project, Ankita Sahni researched and wrote the text.

References