Measuring depression and anxiety in medical students: Is HADS an appropriate tool?

Naomi Ruth Marfell

A thesis submitted for the degree of Master of Philosophy, School of Medicine

Cardiff University

September 2019

Summary

INTRODUCTION: There are a number of aspects of the medical course which may make medical students more prone to mental health problems. There is much evidence that suggests that medical students have high levels of common mental health problems, but research results are contradictory. This may be due to methodological differences. Tools validated for other populations may not be suitable for use with medical students. There is a need to accurately measure depression and anxiety in medical students to provide support. This study aimed to investigate the suitability of the Hospital Anxiety and Depression Scale (HADS) for measuring anxiety and depression in a medical student population by comparing HADS to a structured clinical interview.

LITERATURE REVIEW: A review was carried out to review what tools have been used for measuring depression and anxiety in studies with students, medical students and the general population.

METHODS: Medical students from Cardiff University were recruited. Students completed HADS and undertook a clinical interview using the Schedules for Clinical Assessment in Neuropsychiatry (SCAN). HADS data was compared to the SCAN data. Sensitivity, and specificity for HADS were calculated and optimal cut-offs ascertained. HADS and SCAN individual item responses were compared.

RESULTS: 50 medical students were recruited. Three (6%) students met the ICD-10 diagnostic criteria for depressive disorder and four (8%) reached the threshold for generalised anxiety disorder (GAD) as determined by SCAN. In comparison, eight (16%) students had a HADS-D score of 8+ and 25 (50%) students had a HADS-A score of 8+ (indicative of caseness). A cut-off of 7 for HADS-D and 13 for HADS-A was calculated as more accurate in identifying 'caseness' within this population.

CONCLUSIONS: HADS is an appropriate tool for a medical student cohort. The cut-off for 'caseness' for both subscales should be reviewed due to interpretation of some HADS items.

Acknowledgements

To my supervisors, Debbie Cohen and Liz Forty, thank you for your support and guidance throughout my MPhil and for finding funds to help with recruitment. Debbie I am very grateful for all your advice, and encouragement, and for believing in me when I didn't believe in myself. Liz, thank you for taking the time to train Sarah and I in carrying out and interpreting the SCAN interviews and for providing expert advice when needed. Thank you to the medical school for supporting me.

To my colleague and former co-student (now Dr!) Sarah Rees. Thank you for helping me with the interviews/ratings/diagnoses. But thank you even more for your support, friendship and advice, without which this would have been a much harder and lonelier process.

This work would not have been possible without the study participants. Thank you to all of the medical students, who kindly gave up their time and were willing to share and talk about their mental health experiences.

I will always be grateful to my family and friends who have keep me going during this project. To my parents, thank you for your unwavering love, support and encouragement and for all the childcare which has helped enormously. To my children thank you for giving me some quiet time when really you would rather have been out having fun, and thanks for being a wonderful distraction when we were out having fun! Finally, to my husband Mike, thank you so much for all your love, and encouragement but also for your patience, for giving me some perspective and for your ability to make me laugh when I needed it!

Summary of other's contributions to this work

Debbie Cohen and Liz Forty contributed to the study design.

Debbie Cohen reviewed draft versions of this thesis.

Liz Forty provided training in carrying out and interpreting SCAN interviews and provided expert advice for interpreting some of the interview data.

Sarah Winstanley carried out 10% of the SCAN interviews. Sarah also rated and determined diagnoses for 20% of the interviews to determine inter-rater reliability.

Contents

List of figures & tables

Abbreviations

Abbreviation	
ADDQ	Anxiety Disorder Diagnostic Questionnaire
AKUADS	Aga Khan University Anxiety and Depression Scale
BAI	Beck Anxiety Inventory
BMA	British Medical Association
BSI	Brief Symptom Inventory
BSI-18	Brief Symptom Inventory 18
CES-D – short versions	The Center for Epidemiological Studies-Depression scale
CESD-R	The Center for Epidemiological Studies-Depression scale – revised
CIDI	Composite International Diagnostic Interview
CIDI	Composite International Diagnostic Interview
CIS-R	Clinical interview schedule revised
CMD	Common Mental Disorders
CRSD	Carroll Rating Scale for Depression
DASS-21	Depression Anxiety Stress Scale
DASS-21	Depression Anxiety Stress Scale
DASS-42	Depression Anxiety Stress Scale
DASS-42	Depression Anxiety Stress Scale
DEPI	Rorschach Comprehensive System Depression Index
DESC	Rasch-Based Depression Screening
DSM	Diagnostic and Statistical Manual of Mental Disorders
GAD-2	Generalised Anxiety Disorder-2
GAD-7	Generalised Anxiety Disorder – 7
GAD-Q-IV	Generalised anxiety disorder questionnaire
GAS	Goldberg anxiety scale
GDS	Goldberg depression scale
GHQ-12	General Health Questionnaire
GHQ-28	General Health Questionnaire
GMC	General Medical Council
HADS	Hospital Anxiety and Depression Scale
HADS	Hospital Anxiety and Depression Scale
HAM-d/HRSD	Hamilton Depression Rating scale
HARS	Hamilton anxiety rating scale
HDI	Human Development Index
ICD	International Classification of Diseases
IDAS	The Inventory of Depression and Anxiety Symptoms
IDS	Inventory of Depressive symptomatology
ISR	ICD-10 Symptom rating
K10	Kessler 10 item distress scale
K6	Kessler 6 item distress scale
KADS - 6	Kutcher adolescent depression scale - 6
LAPS	the Leuven Affect and Pleasure Scale
MACL	Mood adjective checklist
MADRS	Montgomery–Åsberg Depression Rating Scale (MADRS)
MASQ	Mood and Anxiety Symptoms Questionnaire
MASQ	Mood and anxiety symptom questionnaire

M-BDI	Modified Beck Depression Inventory
MDI	Major depression Inventory
MHI-5	Mental health inventory
MINI	The Mini International Neuropsychiatric Interview
MINI	The Mini International Neuropsychiatric Interview
NICE	National Institute of Clinical Excellence
OASIS	Overall Anxiety Severity and Impairment Scale
ODSIS	Overall Depression Severity and Impairment Scale
PHQ-2	Patient health questionnaire 2
PHQ-4	Patient health questionnaire 4
PHQ-8	Patient health questionnaire 8
PHQ-9	Patient health questionnaire 9
PRIME MD	PRIMary care Evaluation of Mental Disorders
PSWQ	Penn State Worry questionnaire
QIDS	Quick Inventory of Depressive symptomatology
RRS – 10	Rumination Response Scale
SAS	Zung self-rating anxiety scale
SCAN	Schedules for Clinical Assessment in Neuropsychiatry (SCAN)
SCAT	Sinha's comprehensive anxiety test
SCID	Structured clinical interview
SCID	Structured clinical interview
SCL-90	Symptom checklist
SDS	ZUNG self-rating depression scale
SF-12	Short-Form Health Survey
SRE-20	Self-reporting questionnaire
SRQ	The Self-Reporting Questionnaire (SRQ)
STAI	State-Trait Anxiety Inventory
STAIS	State-Trait Anxiety Inventory – State scale
ST-DEP	State trait depression scale
USDI	University student depression inventory
WDQ	Worry domains questionnaire
WHO-5	World health organisation well-being index
ZDS	Zagazig Depression Scale

Chapter 1: Introduction

1.1 Mental health of university students

The mental health of university students has been prominent in both the news and academic literature in recent years. Some suggest that there is a mental health crisis in UK universities (mentalhealth.org). One mental health organisation reported a large increase in the number of first year students reporting they had a mental health problem from 3000 students in 2006 to 15,000 in 2015/16 (Yap 2018). The head of the university's regulator (Nicola Dandridge) is quoted by the evening standard newspaper as say it is "impossible not to be concerned" at the scale of anxiety and depression among students (Davies writing for evening standard Wed 31Oct 2018).

Academic studies have also regularly reported on the mental health of students. Many suggest that the prevalence of mental health problems in university students is significantly higher than in the general population (Stallman, 2010; Ibrahim et al 2013). The peak onset for mental health problems falls before the age of 25 (Macaskill, 2012). Roughly half of all lifetime mental disorders in most studies start by mid-teens and three-fourths by mid-20s (Kessler et al, 2007). For many students this is the time they are attending university.

Some studies suggest that mental health of students may be no worse than the general population of similar age. One study looking at the global public health challenge of the mental health of young people reviewed a range of academic papers looking at the mental health in young people (Patel et al, 2007). The study concluded that "one out of every four to five young people in the general population will suffer from at least one mental disorder in any given year".

Studies comparing student's mental health with age-matched peers are more scarce and less definitive than those comparing to general population data. Blanco et al (2008, cited in Royal College of Psychiatrists, 2011 report) found no difference in overall rates of psychiatric disorders compared to non-students. A study looking closely at the type and severity of mental health problems suggested that tertiary students in Australia had a greater prevalence

of moderate, but not high distress than non-students (Cvetkovski et al, 2012) (Tertiary education in Australia is any education a student receives after final compulsory schooling, including university). This supported a previous study in Australia which suggested that tertiary students were four times more likely to be classified as psychologically distressed than age matched peers (Leahy, et al 2010).

In the UK similar results are seen, the 2016 HEA student academic experience survey which compared student's wellbeing to data from the Office for National Statistics, found that 43% of young people aged 20-24 rate themselves as having very low anxiety compared to only 21% of students (Neves & Hillman, 2016). In addition, a report by the Royal College of Psychiatrists (2003) concluded that on balance British university students report more mental health symptoms than age-matched controls.

Concern over the mental health of UK students has been affected by the increase in suicides reported across the UK in university students. In an 18 months period prior to September 2018, 11 Bristol University students had taken their own life (BBC news, 2018). The BBC also reports that according to statistics from Universities UK, 146 students had taken their own lives in 2016. This is in contrast to reports by the Office of National Statistics (ONS) that between 2013 and 2016 higher education students in England and Wales had a significantly lower suicide rate compared with the general population of similar ages.

1.2 Reasons for increased rates of mental health problems among university students

Attending university is considered to be synonymous with exploring new ideas, having new experiences, meeting new people, and trying to find oneself (Rosenberg, 2018). Others see university as a place where young adults, can 'let their hair down' for three (or more) years, drink, party and generally have a good time (Hirst, 2007). Whilst for many this may be the case, there are many aspects of university which may put those that attend at higher risk of mental health problems.

A number of possible reasons or risk factors for why students may be more likely to experience mental health problems have been suggested, including moving away from home, financial

difficulties, social media, academic demands and worries about future careers. These will be considered in more detail below.

Many students move away from home to attend university and as such move away from what is often the stability of both family life and established friendships (RCP, 2011). Student life may seemingly be based around socialising, however some students may find it difficult to make new friends and may feel isolated and lonely (RCP, 2011).

In addition to the isolation that can accompany moving away from home, students may for the first time be faced with managing the practical and financial aspects of their life. Students today are faced with larger amounts of debt than previous generations (Johnson & Crenna-Jennings, 2018). Some of these debts are in the form of student loans which can be paid off alongside tax, once working. Research has shown that students feel strained financially managing their day to day finances (Johnson & Crenna-Jennings, 2018) A survey carried out by one UK bank (HSBC) found that one fifth of students spent all of their student loan within the first 100 days of receiving it, with the average student spending almost three fifths of their loan in this time period (Johnson & Crenna-Jennings, 2018). The relationship between finances and mental health problems in students are complex. Several studies (Cooke et al 2004; Ross et al 2006) have suggested that there is no significant correlation between levels of debt and mental wellbeing. Other have suggested that students reporting financial difficulties are more at risk of mental health problems (Eisenberg et al 2010). Andrews and Wilding (2004) reported that financial difficulties can increase British students' levels of anxiety and depression and can affect academic performance. Others have suggested that it is worrying about finances that is linked to mental health problems rather than the level of financial difficulty or debt (Ross et al 2006; Jessop et al 2005).

Financial problems or worrying about one's finances may encourage students into finding employment whilst at university (RCP 2011). Whilst this might help with financial worries the time this takes up along with the additional stress of working can affect the mental health of students (RCP, 2011). It is estimated that approximately 40% of students are employed alongside their studies (Robotham & Julian, 2006). In this study two thirds of working students felt that it had affected their studies (Robotham & Julian, 2006). In addition, one study found that poor mental health in British students was related to working long hours alongside studying (Roberts et al, 1999).

A relatively new risk factor affecting the mental health of students is social media, internet and technology use. Whilst this may also affect age matched peers some have suggested that students are particularly vulnerable to pathological internet use (Kandell, 2009). A number of reasons are suggested for this vulnerability including, ready access to the internet and the expectation of computer and internet use (Kandell, 2009). Students are expected to use the internet as a source of information for their research, to support their university education (downloading course notes, webinars etc) but also as a means of communication (Loan, 2011). Depression and anxiety have been found to be higher in groups of people with high smartphone use (Demirci et al, 2015). In addition, use of multiple social media platforms is associated with increased symptoms of depression and anxiety, even when the effects of excessive time spent of social media is controlled for (Primack et al, 2017). Excessive use of social media and technology has been found to impair social interactions and increase feelings of isolation (Rosenberg, 2018). In students who may already be suffering from poor mental health due to isolation and loneliness, this may exacerbate these feelings.

The final factors which might negatively affect student mental health are the academic demands of university and feeling the need to achieve high grades for future career prospects (Johnson & Crenna-Jennings, 2018). In one survey 71 percent of respondents reported that 'performing well in tests and coursework' caused them stress, and 65 percent felt stressed trying to keep up with their studies (Unite Students study, reported in Johnson & Crenna-Jennings, 2018). In a 2016 YouGov poll 71 per cent of all respondents stated that university work was one their main causes of stress (Aronin & Smith, 2016). Not only might students feel pressure from the amount of university work required of them but first year students in particular may have to adapt to different ways of leaning. Students are expected to carry out much more self-directed learning, than students may have experienced at school and as such students need to learn to manage their time effectively (RCP, 2011).

Students may worry about their job prospects following university as the graduate job market is highly competitive and there are not as many graduate roles as there are graduates (Johnson & Crenna-Jennings, 2018). Having a degree is no longer a guarantee of a job (RCP, 2011). In the 2016 YouGov poll cited previously, 77% of all respondents reported that they have a fear of failure, one in five of these students reported that this fear is 'very prevalent' in their day to day life. In addition, the second largest concern (after university work as cited above) for students who completed the poll was finding a job after university (Aronin & Smith, 2016). There is also added pressure on students to demonstrate to employers following graduation

that they also have experience of extra-curricular activities such as work experience and voluntary work (RCP, 2011). Students therefore feel the need to complete these activities alongside their studies whilst at university.

1.3 Mental health of medical students

There is considerable evidence that medical students have high levels of common mental health disorders (Schneider et al, 1993; Firth-Cozens, 1987; Bellini et al, 2002). In a recent study of 760 UK medical students at one medical school, 18% reported having received treatment for a mental health disorder (Korszun et al, 2012). In a wider UK study carried out by the student British Medical Journal in 2016, of the 1,122 respondents, 30% (343) said they had experienced or received treatment for a mental health condition (Student BMJ, 2017). However, the research on mental health in medical students is at times contradictory. This is not only in relation to the prevalence of mental health disorders but also how this varies for different groups such as year of training and gender.

When considering the research around mental ill health and year of training there are differences across studies. Dyrbye suggests the mental health of students deteriorates during training, (Dyrbye et al, 2005). This is also reported in a German study where they found the prevalence of common mental health disorders in newly enrolled medical students were higher than reported in the general population, but lower than in medical students already studying for their medical training. In contrast, some studies have suggested that first year medical students have the highest level of distress over all 5 years of study (Guthrie, 1998). A more recent study also supports this finding stating that the 'first few years' are the most difficult with less mental health problems seen the longer the student is in education (Knipe et al 2018).

There is also disparity between studies looking at levels of distress of medical students and the effect of gender. Some studies suggest female medical students are more distressed than their male peers (Dyrbye et al, 2006). Others suggest levels of distress are the same for both male and female medical students (Moffat et al, 2004; Niemi and Vainiomaki, 2006). The current study aimed to look at depression and anxiety in medical students, these specific disorders will now be considered in turn.

1.3.1 Depression

Depression is described as a common but serious mood disorder (NIMH, 2018). In England in 2017/2018, 4.5 million adults had a diagnosis of depression, which accounted for around 10% of all adults registered with a GP (NICE, 2019). The 2014 *Adult Psychiatric Morbidity Survey 2014* in England found 3.3% of general population respondents were identified as having depression (McManus et al, 2016). Depression rates tend to be higher in women than men (in the 2014 study above 2.9% of men were identified as having depression compared to 3.7% in women). Some studies have suggested prevalence rates are between 1.5 and 2.5 times higher in women than men (NICE, 2019).

There are many definitions of depression but central to most is the presence of depressed mood and/or the loss of pleasure (Mian et al, 2014). Whilst many individuals may experience low or depressed mood, this does not in itself mean that individual has depression. Depression is considered a syndromal disorder, Calvo et al (2003) define a syndrome as "a recognizable complex of symptoms and physical findings which indicate a specific condition for which a direct cause is not necessarily understood".

There are a wide variety of symptoms related to depression and as such depression can affect how you think and feel and can impact on an individual ability to carry out daily activities (NIMH, 2018). Not all symptoms related to depression will be present in all cases. For a clinical diagnosis of depression, the Diagnostic and Statistical Manual of Mental Disorders (DSM-V) requires that five out of nine symptoms need to be present (APA, 2013). The International Classification of Diseases (ICD-10) classification system requires that four out of 10 symptoms are present (WHO, 1992). For both systems these symptoms must include at least one of the following symptoms low mood, loss of interest or loss of energy to be present. The symptoms required for each classification system can be seen in table 1. Both classification systems require that symptoms are present for at least two weeks and symptoms must have been present for a significant proportion of each day (APA, 2013: WHO, 1992) . The more symptoms that are present the more severe the episode of depression. Whilst 4 symptoms are required for a clinical diagnosis of depression (according to the ICD-10 classification), the National Institute for Health and Care Excellence (NICE) in the UK recognise that depressive symptoms below the threshold for depression may still be distressing and

disabling (NICE, 2009). As such the NICE guidelines for management of depression also covers 'subthreshold depressive symptoms' (NICE, 2009).

Table 1 DSM-IV and ICD-10 diagnostic criteria for depression

DSM-IV Diagnostic criteria Depression	ICD-10 diagnostic criteria Depression
The individual must be experiencing five or more symptoms during the same 2-week period and at least one of the symptoms should be either (1) depressed mood or (2) loss of interest or pleasure	At least one of these, most days, most of the time for at least 2 weeks: ▪ persistent sadness or low mood ▪ loss of interests or pleasure ▪ fatigue or low energy
1. Depressed mood most of the day, nearly every day. 2. Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day. 3. Significant weight loss when not dieting or weight gain, or decrease or increase in appetite nearly every day. 4. A slowing down of thought and a reduction of physical movement (observable by others, not merely subjective feelings of restlessness or being slowed down). 5. Fatigue or loss of energy nearly every day. 6. Feelings of worthlessness or excessive or inappropriate guilt nearly every day. 7. Diminished ability to think or concentrate, or indecisiveness, nearly every day. 8. Recurrent thoughts of death, recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide.	Associated symptoms: ▪ disturbed sleep ▪ poor concentration or indecisiveness ▪ low self-confidence ▪ poor or increased appetite ▪ suicidal thoughts or acts ▪ agitation or slowing of movements ▪ guilt or self-blame The above 10 symptoms then define the degree of depression and management is based on the particular degree ○ **not depressed** (fewer than four symptoms) ○ **mild depression** (four symptoms) ○ **moderate depression** (five to six symptoms) ○ **severe depression** (seven or more symptoms, with or without psychotic symptoms)

When considering the evidence around depression in medical students, a 2016 meta-analysis of studies suggested a global prevalence of depression amongst medical students of 28.0% (Puthran et al, 2016). Another 2016 meta-analysis of the literature suggested on average, 27.2% of medical students experience depression or depressive symptoms (Rotenstein et al 2016). The review included one hundred and ninety-five studies involving a total of 129 123 individuals in 47 countries. Across these studies, Rotenstien et al identified prevalence rates from 1.4% to 73.5%. Of the seven studies based in the UK, prevalence ranged from 4.4 % (Newbury-Birch et al, 2001) to 48.8%. (Honney et al, 2010). Possible reasons for these differences will be explored in Section 1.5 below.

There are a number of studies that explore depression in relation to year of study. In their meta-analysis Rotenstein et al (2016) included nine longitudinal studies which assessed depressive symptoms before and during medical school. Analysis of this data suggested an increase in symptoms of 13.5% from before to during medical school. Rosal et al (1997) found medical students begin university with similar rates of depression as their non-medical peers. However, as they progress through medical school medical students' depression scores rise to a greater extent than their non-medical peers (Rosal et al, 1997). In contrast, a meta-analysis of studies using standardised questionnaires suggested, students in their first year of medical school had the highest rates of depression (at 33.5%) (Puthran et al, 2016). The meta-analysis also found rates of depression gradually decreased over the course of medical school, with rates of 20.5% at Year 5 (Puthran et al, 2016). This is supported by Silva et al (2017) who carried out a longitudinal study at a medical school in Portugal and found that medical students' scores on the Beck Depression Inventory (BDI) decreased during their time at medical school. Nearly 20% of students in this longitudinal study reported high BDI scores which were sustained over time (Silva et al, 2017).

In terms of gender differences some have suggested that female medical students are more depressed than their male counterparts (Goober et al, 2009). In contrast, in one study looking at medical students in one UK university, no significant gender differences were seen between males and females in rates of transient or persistent depression (Quince, 2012). In addition, no differences were found in BDI scores of males and females in the longitudinal study cited above (Silva et al, 2017). In one of the 2016 meta-analyses females were more likely to be depressed but the results were not statistically significant (Puthran et al, 2016).

1.3.2 Anxiety

There are a number of anxiety disorders including generalised anxiety disorder (GAD), panic disorder, post-traumatic stress disorder, obsessive–compulsive disorder, social phobia, specific phobias and acute stress disorder (NICE, 2011). Generalised anxiety disorder is defined by the ICD-10 as "Anxiety that is generalised and persistent but not restricted to, or even strongly predominating in, any particular environmental circumstances (i.e. it is "free-floating")" (WHO, 1992). NICE state that the key feature of GAD is "worry and apprehension that is out of proportion to the circumstances" (NICE, 2011). The DSM classification also defines GAD as "excessive and difficult to control anxiety about several different events or activities" (APA, 1994). The American Psychological Association (APA) defines the symptoms of anxiety rather than the disorder itself. It defines anxiety as "an emotion characterized by apprehension and somatic symptoms of tension in which an individual anticipates impending danger, catastrophe, or misfortune" (APA, 2018). The APA describe how the body may react to the anticipated threat with muscles becoming tense, faster breathing, and increased heart rate (APA, 2018).

There are a number of studies looking at prevalence of anxiety both nationally and internationally. In the UK prevalence of anxiety in a 2010 sample of primary care patients was 7.2% (Martin-Merino et al, 2010). In 2013 it was estimated there were 8.2 million cases of anxiety in the UK (Fineberg et al, 2013). The studies cited above, do not distinguish between types of anxiety disorder, but rather report on 'anxiety' in general. Generalised anxiety disorder (GAD) was found to be the most common mental disorder in England in 2014 (Adult Psychiatric Morbidity Survey, 2014). The survey found 5.9% of respondents were identified as having Generalised anxiety Disorder (McManus et al, 2016).

Both the Martin-Marino study and the Adult Psychiatric Morbidity Survey identified more women as having GAD than men (6.8% compared to 4.9% in the Martin Marino study and 6.8% to 4.9% in the Adult Psychiatric Morbidity Survey). In relation to age, prevalence was highest in those aged 45-54 (McManus et al, 2016).NICE also state that studies suggest GAD is less common in both older age groups (Over 55 years) and younger groups (Below 35 years).

Whilst many individuals may experience feelings or symptoms of anxiety or worry, they may not fall within the diagnostic criteria for GAD. Commonly people talk about feeling anxious about events such as change in occupation or exams. However, for a diagnosis of GAD, specific symptoms need to be present for a significant amount of time. For example a diagnosis of

generalised anxiety disorder, the ICD-10 classification system requires prominent tension or worry to have been present for at least six months with at least four other symptoms (see table 2) to be present, of which at least one needs to be palpitations or pounding heart, sweating, trembling or shaking and dry mouth (WHO, 1992). The DSM classification system requires excessive anxiety and worry to be present with at least three other symptoms (see table 2). Symptoms must have been present for at least six months and impair daily functioning. In common with the self-report scales for depression, many self-report scales for anxiety provide a measure of symptoms experienced by the individual but are not designed to provide a diagnosis of an anxiety disorder.

Table 2 DSM-IV and ICD-10 diagnostic criteria for Anxiety

DSM Diagnostic criteria Anxiety	ICD-10 diagnostic criteria Anxiety
A. Excessive anxiety and worry (apprehensive expectation), occurring on more days than not for at least 6 months, about a number of events or activities (such as work or school performance)	▪ A. A period of at least six months with prominent tension, worry and feelings of apprehension, about every-day events and problems
The anxiety and worry are associated with three (or more) of the following six symptoms (with at least some symptoms present for more days than not for the past 6 months). Note that only one item is required in children 1. Restlessness or feeling keyed up or on edge 2. Being easily fatigued 3. Difficulty concentrating or mind going blank 4. Irritability 5. Muscle tension 6. sleep disturbance (difficulty falling or staying asleep, or restless unsatisfying sleep)	At least four symptoms out of the following list of items must be present, of which at least one from items 1 to 4 *Autonomic arousal symptoms* 1. palpitations or pounding heart, or accelerated heart rate 2. sweating 3. trembling or shaking 4. dry mouth (not owing to medication or dehydration) *Symptoms concerning chest and abdomen* • 5. difficulty breathing • 6. feeling of choking • 7. chest pain or discomfort • 8. nausea or abdominal distress (e.g. churning in stomach) *Symptoms concerning brain and mind* • 9. feeling dizzy, unsteady, faint or light-headed • 10. feelings that objects are unreal (derealisation), or that one's self is

	distant or 'not really here' (depersonalisation) • 11. fear of losing control, going crazy or passing out • 12. fear of dying *General symptoms* • 13. hot flushes or cold chills • 14. numbness or tingling sensations *Symptoms of tension* • 15. muscle tension, or aches and pains • 16. restlessness and inability to relax • 17. feeling keyed up, or on edge, or of mental tension • 18. a sensation of a lump in the throat, or difficulty with swallowing *Other non-specific symptoms* • 19. exaggerated response to minor surprises or being startled • 20. difficulty in concentrating, or mind going blank, because of worrying or anxiety • 21. persistent irritability • 22. difficulty getting to sleep because of worrying
The focus of the anxiety and worry is not confined to features of an Axis I disorder	The disorder does not meet the criteria for panic disorder, phobic anxiety disorders, obsessive–compulsive disorder or hypochondriacal disorder

The literature that describes anxiety in medical students has some similarities to that of the literature around depression but is less extensive. Dyrbye et al, in their 2006 review, highlighted that "few studies examined specific anxiety disorders such as generalised anxiety disorder in any formal way". There are fewer studies reviewing or carrying out meta-analyses of studies looking at generalised anxiety in medical students (Hope & Henderson, 2014; Dyrbye et al, 2006) in comparison to those looking at depression (Puthran et al, 2016; Rotenstein et al, 2016, Hope & Henderson, 2014; Dyrbye et al, 2006).

The prevalence of anxiety amongst medical students is described as being higher than the general population (Schneider 1993; Firth Cozens 1987 Kasha 2000; Bellini 2002). In a study of 6658 medical students in Australia the prevalence of students who reported having a diagnosis of anxiety (including current diagnosis and if they have ever had a diagnosis) was higher than for the general population but similar to other students (BeyondBlue, 2013). The prevalence of anxiety in the BeyondBlue study is low compared to others. Hope and Henderson's (2014) systematic review of studies looked at anxiety amongst medical students in Europe and 'English speaking countries outside North America', they found that prevalence of anxiety ranged from 7.7.% to 65.5%.

In relation to gender and anxiety amongst medical students, one study looking at medical students in Brazilian medical schools, found prevalence of state-anxiety was 81.7 % (Mayer et al, 2016). Female students in the study were more likely to score highly for state anxiety but there was no difference in anxiety of students from different academic years (Mayer et al, 2016). This is supported by the Beyond Blue study which found that 8.8% of female medical students in Australia who responded had a current diagnosis of anxiety compared to 5.2% of male medical students.

1.3.3 Comparison with other students

Some studies have compared the mental health of different groups of students, with some comparing medical students and other students. A Brazilian study found a high prevalence of common mental disorders (CMD) among medical students (42.6%) compared to dental students (33.3%), nursing students (31.8%) and physical education students (25%) (Facundes & Ludermir, 2005). A Norwegian study that compared medical students to law, mechanical engineering and psychology students found a prevalence of 44% for psychological distress in medical students Prevalence of psychological distress among the other student groups were 58%, 55% and 40% respectively, suggesting medical students had lower prevalence than both law and mechanical engineering students (Midtgaard et al 2008).

A study in a London university compared levels of depression in medical and non-medical life sciences students (this included students studying social or pure science) (Honney et al, 2010). This study found that the non-medical students showed a higher prevalence of moderate and severe depressive symptoms than medical student peers but medical students reported more

symptoms of mild depression (Honney et al, 2010). In their 2016 meta-analysis, Puthran et al found no significant difference in prevalences of depression between medical and non-medical students in the six studies included in their analysis.

Two studies have compared the prevalence of depression and anxiety between medical and dental students. Knipe et al (2017) compared medical, dental and veterinary students. They suggested that there are similarities between these groups of students such as longer course duration, longer contact hours and exposure to potentially distressing clinical situations (Knipe et al, 2017). Knipe et al's study found that medical students had lower levels of moderate depression and anxiety compared to dental students. There was no difference in depression and anxiety scores between medical and veterinary students. Although veterinary students had lower levels of wellbeing (Knipe et al, 2018). In this study prevalence of depression and anxiety for medical and dental students was higher in the first few years of study with this declining in later years of study (Knipe et al, 2018). This differed from results from a 2002 study comparing medical and dental students (Newbury Birch et al, 2002). Newbury-Birch et al found that during the second year of the course a higher prevalence of dental students reached the HADS cut off for depressive symptoms (8+) 15% (dentistry) compared to (4%) medical students. During the final year 14% of dental students reached the cut-off compared to 5% of medical students.

For anxiety, 47% of both medical and dental students reached the cut-off for anxiety during their second year. For medical student this decreased to 26% by their fifth year, but for dental students, prevalence increased to 67% in their fifth year, (Newbury-Birch et al, 2002).

In summary, results of studies that looked at depression and anxiety amongst medical students and those that compared medical students with other students are mixed. Depression studies suggest prevalences ranging from 1.4% to 73.5%. In studies comparing medical students with other students, some suggest that medical students have similar rates of depression to other students, whereas others suggest that medical students have lower rates of depression. Similar results are found in relation to anxiety. Anxiety prevalence in studies ranged from 7.7.% to 65.5%. One study suggested no difference between medical and dental studies, with another suggesting higher rates of anxiety for dental students compared to medical students. These differences may reflect the different student groups to which medical students are being compared. Alternatively, other methodological differences may affect the results such as the screening tool used, or the timing of the study (eg during exam time).

1.3.4. Resilience

A growing area of research in relation to mental health and wellbeing is that of resilience. 'Resilience is a measure of the ability to cope when faced with adversity' (Houpy et al, 2017 p1). High levels of resilience are linked to higher levels of well-being. Some studies have found that medical students are less resilient than the general population (Houpy et al, 2017; Rahimi et al, 2014). Although Rahimi suggested that in their study it is possible that medical students' lower levels of resilience reflects their higher levels of perceived stress. It has been suggested that resilient medical students are less likely to experience depression (Dyrbye et al, 2010). One study found an association between medical students' higher resilience scores and lower scores of anxiety and depression (Tempski et al, 2015). Tempski et al suggest the relationship between resilience and anxiety and depression may be bi-directional; Resilience may protect against depression and anxiety symptoms but also an anxious or depressed person may be less able to use their coping skills. Many medical schools now incorporate resilience into the medical curriculum (Eg Wright & Mynett, 2019).

1.4 How medical students differ from other students

Research described in section 1.3 above highlighted possible differences in the levels of mental health problems of medical students in comparison to other students. This section will consider ways in which medical students may differ from some other students. There are a number of areas in which differences might occur. One area could relate to the length of the course particularly the contrast to those students studying 3-year non-vocational degrees which generally have less contact hours than medical students. Section 1.2 highlighted the issues facing students in general which may make them more prone to mental health problems compared to their peers who do not attend university. These issues included moving away from home, financial difficulties, social media, academic demands and worries about future careers. Whilst these issues also affect medical students, there are also a number of other factors which medical students face which may affect their mental health. Some of these factors are linked to those discussed above, there are also factors unique to studying medicine which may cause medical students additional mental health problems such as exposure to death and human suffering. Factors which will be considered include isolation, financial

difficulties, academic demands, career worries, personality, ethical conflicts, role models and stigma and support.

As discussed in section 1.2 above, for many students going to university means moving away from their family home and established friends. This is also the case for medical students, many of which move away from home to attend medical school. However, some aspects relating to the course may exacerbate feelings of homesickness or loneliness. In particular, the lack of available free time and the shorter holidays. Lee and Graham (2001) state that "During their first year of training, medical students may have marked lifestyle changes such as diminished leisure and recreational activity, decreased physical activity and sleep deprivation, which can result in a general decrease in physical health and emotional wellbeing" (Lee & Graham, 2001, p. 652) The long course contact hours for medical students allow for less time to explore leisure activities or join clubs, where new friendships might be formed. Shorter holidays make visiting family and established friendships back home more difficult and less frequent. In addition to the greater number of contact hours at medical school, medical students are also expected to attend placements which may mean they have to move to different locations for several weeks, taking them away from university friends. These factors may mean there is less time for personal interests and socialising (Yiu, 2005; Holm et al, 2010) and may lead to medical students feeling a greater sense of isolation and loneliness.

In addition to potentially contributing to a sense of isolation, long contact hours and attending placements may also have a financial implication for medical students (GMC, 2013). The long contact hours and the need for additional studying outside of these hours impact on the students' ability to supplement their income. Attending placements away from their main residence may make any type of employment impossible. The financial burden is considerably higher for medical students undertaking a 5 year course in comparison to many standard university courses which are often only three years. Some have suggested that this leads to many medical students facing significant financial debt (Dyrbye et al, 2005), One UK study suggested that if a medical student graduating in 2014 had accepted all student loan and maintenance loans available, they could amass debts of £82,000 when they leave medical school (Ercolani et al, 2015). This is compared to students in England studying for a standard degree whose debt was predicted to be about £59,000 for students who started in 2012

(Richardson & Roberts, 2015). As stated in section 1.2, financial difficulties and in particular worrying about financial difficulties can negatively affect mental health.

Therefore, the length of the course and the need to attend placements may increase the financial burden on medical students and may affect medical student's mental health and their performance at medical school (Ross et al, 2006).

The academic demands of university can also put a strain on students and for courses which may be seen as more academically challenging this strain may be greater. The academic nature of medical school training is intense (Yiu, 2005). Medical students are required to learn, retain and recall a wealth of knowledge (Holm et al, 2010, Mahajan, 2010). Malik (2000) suggests that much of medical students' stress is course related suggesting common problems include workload, fear of falling behind and exam failure worries. In addition to the academic nature of the course, medical students are required to be seen as competent doctors/clinicians (Chew-Graham et al, 2003) and to uphold the ethics and standards required of a doctor (GMC 2016). Medical school also has a competitive nature, students feel the pressure to not only pass the course but to be seen as one of the top in the class, their peers being those whom they will most likely be up against in the future for jobs (Mahajan 2010). This may be compounded by the shift in position compared to current peers; many medical students are likely to have been top in their class at school, in comparison at university they may be average or even below average, this in itself may cause stress and anxiety (Dunn et al, 2008). Holm et al (2010) suggest that students may be left feeling anonymous and isolated by the competitive nature of the medical course, particularly if the medical school environment is also seen as rigid and authoritarian.

The need to be seen as a competent clinician and to be thriving compared to peers is also linked to the nature of personal tutors, supervisors, lecturers and other medical school staff. These staff may become the people who may be involved in the employment or career progression of students in the future (Chew-Graham et al, 2003). This not only increases the pressure of 'wanting to be seen as having the potential to be a great doctor' but can also increase the worry about admitting that you may be struggling. This along with concerns such as stigma, insufficient time to attend appointments (Dunn et al), service issues (long waiting lists, access issues, scheduling problems (Stecker 2004)), lack of awareness of the service (BMA 2019; Stecker, 2004), and confidentiality concerns (Hillis et al , 2010) may prevent medical

students from seeking support early. One study suggests that less than a quarter of first and second year medical students who were depressed have used mental health services (Hillis et al, 2010). Seeking support early for mental health problems is beneficial as it may improve the prognosis of the mental health problem and can reduce the risk of it becoming a chronic condition or leading to more severe disorders (Mitchel et al, 2017). As such some have suggested that many medical students may not feel comfortable receiving support from their own medical school and may prefer to seek help elsewhere (Dyrbye et al, 2005).

Some have suggested that the personality of medical students may put them at greater risk of mental health problems. Historically research has focused on individual personality traits (Firth-Cozens et al, (1999), which may predispose medical students to mental health problems, such as neuroticism (Mokros et al, 2017). More recent research has suggested that the relationship between personality and wellbeing in medical students is more complex and it is the combinations of personality traits along with circumstances which may predispose an individual to mental health problems (Eley et al, 2016). Combinations suggested have included high neuroticism and high conscientiousness as traits considered to put medical students at high risk of stress (Tyssen at al, 2007). Henning et al (2002) found strong associations between current psychological distress, perfectionism and impostor syndrome.

Other studies have suggested that personality factors which relate to work are more important than broad personality types (Chow et al, 2018). Chow et al suggested that students who score high on the 'work-related personality Stability factor' may be less likely to develop depressive symptoms and suicidal ideation (Chow et al, 2018). Chow et al suggested that the students with this 'stability factor' may be more resilient either naturally or by having developed a resilient mindset or attitude. The Chow study also found that medical students who scored high on the dominance factor were more likely to develop depression, possibly because they may experience more interpersonal problems due to being more likely to be argumentative and confrontational (Chow et al, 2018). Whilst these personality factors may affect medical students' mental health, it is unclear whether these personality factors are more prevalent in medical students than other students.

Eley et al suggest that two main personality types of medical students can be found and that the link between personality and mental health is linked to the resilience of the different personality types. The two types are described as:

"Students with personality Profile 1 are described as *resilient* because they combine being *vigorous* (high in Persistence and low in Harm Avoidance), *industrious* (high in both Persistence and Self-Directedness), and *versatile* (high in Self-Directedness and low in Harm Avoidance).

Students with personality Profile 2 are referred to as *conscientious* and may be more anxious and less resilient than those with the resilient Profile 1, but both profiles are more mature, responsible, and well-organized than the average person in the general population."

Other factors which may affect the research looking at personality and mental health include differences seen in personality types of different genders (Scott et al, 2007), and also the suggestion that the personality of medical students is changing over different generations of students (Twenge, 2009). Consequently, results from research looking at links between personality types and mental health problems in medical students, may change over time, along with the different generations of students.

In summary, the personality of medical students and its relation to mental health problems is complex. Although there is some evidence to suggest that some aspects of personalities often seen in medical students such as high conscientiousness which may predispose students to greater risk of mental health problems.

One aspect of the medical school course which greatly differs from most students' university experience is the ethical conflicts, death and human suffering which medical students are exposed to (Mahajan, 2010; Tyssen et al, 2000). The British Medical Association (BMA) claim that medical students often find themselves in traumatic clinical situations that they may be unprepared for (BMA, 2018). One study found that in one UK university 72% of final year medical students surveyed had been involved in end-of-life care at some point during their course (Jones and Finlay, 2013). For some students this had taken part overseas during their elective period. The students questioned reported feeling shocked, upset and sad over the deaths in the UK. Some students who witnessed deaths overseas reported feeling angry or frustrated. Only 13% of the students surveyed thought that their medical training had prepared them to deal with death (Jones and Finlay, 2013).

For some students exposure to death at medical school may be their first experience of death and that can leave them feeling sad and shocked. Some students may feel guilty or that they

should have done more for the patient (BMA, 2018). For others who have experienced the death of a family member or friend, treating or managing dying patients and their families may remind them of their own previous experiences, intensifying emotions and bringing back previous trauma (Whyte et al, 2013).

In terms of ethical dilemmas, a recent study (Monrouxe et al, 2015) suggests that medical students and other healthcare students frequently encounter ethical or moral dilemmas. For some students, particularly female students these dilemmas cause them distress (Monrouxe et al, 2015). Monrouxe et al suggest the most common professional dilemmas encountered were student abuse, patient dignity and safety dilemmas. Dyrbye et al suggest that along with the formal curriculum of medical school there is an informal or hidden curriculum, based on what medical students see doctors doing rather than what they teach (Dyrbye et al, 2005). This informal curriculum can cause ethical dilemmas for students contradicting lessons from the formal curriculum and can cause medical students distress (Dyrbye et al, 2005). Dyrbye suggests ethical challenges from the informal curriculum includes 'demands to write notes about patients not personally examined and a dehumanised approach to patients' (Dyrbye et al, 2005 p1615).

The informal or hidden curriculum can also teach medical students unhealthy behaviours in relation to managing their own mental health with some tutors and supervisors providing poor role models (Cohen et al 2015). Some doctors self-manage, prescribing for themselves whilst delaying getting external help (Cohen et al, 2015). Hooper et al (2005) claim self- management of illness is learnt early in a medical students career having seen others doing the same. Other have suggested that coping mechanisms learnt at medical school affect students' long-term health, either positively or negatively (Hillis et al, 2010). The hidden curriculum message of the need to self-manages illness rather than seeking external support only serves to reinforce the view that if you are suffering you must hide it so as not to be seen as weak (Dyrbye & Shanafelt, 2011). This supports the idea that there is a stigma in relation to mental health problems which may prevent medical student seeking help early (Dyrbye & Shanafelt, 2011).

In summary, there are many different aspects which relate to the medical course which may impact on medical students' mental health, over and above the factors which relate to many other university students. The extent to which these affect medical students is not fully understood. In order to understand the effect of these on the mental health of medical

students, researchers must be able to accurately measure prevalence of common mental health problems such as depression and anxiety in medical students.

1.5 Measuring depression and anxiety in medical students

As highlighted in section 1.3 the prevalence of mental health problems amongst medical students varies in different studies. A number of reasons have been suggested for this, ranging from differences in the characteristics of the samples being tested to differences in aspects relating to the methodology used in the studies (Dyrbye, et al, 2006*)*. These will be considered in turn.

1.5.1 Differences in sample characteristics

Section 1.3 has already highlighted some of the factors which could affect the prevalence of mental health problems in medical student studies including gender and year of study. For example, if females experience more depression or anxiety (or are more likely to report having these symptoms), then the gender balance of the sample being tested may affect the results seen. Other differences in the specific samples used in studies which might affect prevalence, include the medical school which students attend and the age of students.

A recent study suggests that the prevalence of depression in medical students varies by different university and different countries (Gan & Hue, 2019). Prevalence of depression and anxiety in the general population varies across different countries. One study looking at depression across 30 different countries suggested that prevalence of depression was highest in countries with a medium human development index (HDI) (Lim et al, 2018). The HDI is described as “a tool developed by the United Nations to measure and rank countries' levels of social and economic development” (Kenton, 2018). Lim et al suggested that people living in medium HDI countries may have more stressors such as higher cost of living and higher expectations and this may affect their mental health (Lim et al, 2018). It is also likely then that medical students attending medical schools in these different countries may also have a higher prevalence of depression and other mental health problems.

In relation to the differences in universities or medical schools. Different universities carry out their medical training in different way which may have an impact on the mental health of students. For example, some studies have found that medical students report less stress when clinical time is increased (Reed et al, 2011). Other medical schools use small group learning, as it encourages students to work collaboratively with their peers which can reduce isolation, encourage sharing of problems and reduce competition between medical students (Holm et al; Kiessling et al, 2004). One medical school in the US has made extensive changes to their medical course such as changing contact hours, scheduling, grading, electives, learning communities, and making resilience/mindfulness compulsory (Slavin et al, 2014). These changes were made as the medical school had identified that they were associated with significantly lower levels of depression symptoms, anxiety symptoms, and stress (Slavin et al, 2014). Other have also suggested that wellbeing programmes should be compulsory for medical students as they are associated with lower reported levels of stress and anxiety, and a higher tolerance of distress (Greeson et al 2015). Therefore, students from medical schools where some of these interventions and changes have been made may experience less mental health problems. As such studies using these medical students may report lower prevalence of mental health problems. In addition, medical schools around the world differ in terms of the hidden curriculum as described in section 1.4. The level of stigma students feel about mental health problems may affect what they are prepared to disclose (Dyrbye, et al, 2006).

Dyrbye, et al (2006) suggested that in addition to the location of the students included in studies, prevalence can be affected by the age of medical students and the stage of medical training. Section 1.3 highlighted the differences in mental health of medical students as they progress through the course, although results from this area of study are mixed. In relation to age, some have suggested that older students may not necessarily suffer from fewer mental health problems. One study researching anxiety and stress amongst graduate entry students suggested that “anxiety and stress were higher, not lower, with increasing age” (Casey et al, 2016). Therefore, whilst age and year of study may affect the prevalence found in studies, the effects are not clear.

Whilst differences between study samples are inevitable and can make comparisons difficult. Some differences, such as age or year of study may be more easily quantified and incorporated into analysis. The effect of medical school or country is more likely to be difficult to quantify as most studies are only based in one medical school or one country.

1.5.2 Differences in methodology

Prevalence rates may be influenced by study methodology as well as subject characteristics (Dyrbye, et al, 2006). Some have suggested that this may be the result of the different instruments used (Chow et al, 2018). A vast number of instruments are available for measuring depression and anxiety in the general population and in student populations, a review of these is outlined in section 2. These tools vary in terms of whether they are self-report or clinician lead tools. Tools also vary in terms of sensitivity and specificity for diagnosing or screening depression and anxiety (Rotenstein et al, 2016). For an explanation of sensitivity and specificity see section 3.1.6.4. Not all tools used to measure depression and anxiety have been validated for use with different groups.

The most commonly used method to measure depression and/or anxiety in research studies is via self-report tools. Self-report tools are more time and cost effective than tools which require a clinician or researcher to administer (Anderson et al 2002). Self-report tools vary on a number of factors, some of these are described in the literature review (see chapter 2). One way in which self-report tools can vary is the point of cut-off for determining caseness. Caesness being defined as 'the probably presence of the disorder'. Cut off scores for depression and/or anxiety scales are generally determined by comparing scale scores to a standard, such a clinical interview. Participants are asked to complete both the scale and the standard. The standard identifies those participants who reach diagnostic levels of the disorder under consideration. The scale scores of those with and without the disorder are then compared. Sensitivity and specificity scores can then be calculated for different cut-off scores for the scale. Sensitivity is the ability of a test to correctly identify positives whereas specificity is the ability of a test to correctly identify negatives (Parikh, et al, 2008). The cut-off score which gives optimum balance of sensitivity and specificity is the cut-off score chosen. Different groups may require slightly different cut-off points to achieve the optimum sensitivity/specificity balance. For some groups the optimum cut off points might not have been explored.

An example of how cut-off scores may affect the prevalence of different studies can be seen with the use of the Beck Depression Inventory version 2 (BDI-II). The BDI-II does not have cut-off scores which are suitable for all purposes (Smarr & Keefer, 2011). However, it has been suggested that scores of 14–19 suggest mild depression, 20–28 for moderate depression, and 29–63 severe depression (Smarr & Keefer, 2011). Some studies with different groups have suggested different cut off scores, for example in a study looking at cardiac outpatients it was

suggested that the optimum cut off score was 10 for the total sample but for women 13 would achieve better sensitivity and specificity (Moullec et al, 2014). In Rotenstien et al's (2016) meta-analysis of studies looking at depression in medical students 9 studies used the BDI-II. In these 9 studies, 5 different cut-off scores were used (10, 14, 17, 20, 21). Whilst these may reflect differences in the BDI-II used in different countries the differences make it difficult to compare results across these studies. The cut-off of 10 was used with Israeli medical students (Lupo & Strous, 2011) and the 21 with US medical students (Chandavarkar et al, 2007). As such comparing prevalence rates across studies which have used different cut-off points is difficult and may explain some of the differences in prevalence seen in studies looking at the mental health of medical students. One way to ensure that studies are more comparable methodologically is to ensure that instruments used are validated with appropriate cut-off points for the sample being tested.

1.5.3 The need to validate self-report tools for use with medical students

It has been highlighted above that different groups sometimes require different cut-off scores for determining possible cases of depression or anxiety when using self-report tools. In the example in section 1.5.2 different cut off scores were suggested for patients attending a cardiac outpatient clinic. Possible reasons in this case may be that the physical symptoms related to heart related conditions may also be linked to physical symptoms which might relate to depression, for example tiredness and difficulty sleeping.

Other groups have been seen to require different cut-off scores which do not relate to physical symptoms, but which relate to the interpretation of some of the items which are used in self-report tools. In some cases, these are linked to cultural factors when the tools are used in different countries (Kerr & Kerr, 2001). Kerr and Kerr (2001) suggest that when translated into some different languages the BDI has limited predictive power and validity due to cultural factors. Kerr and Kerr cite the example of Latino populations who have a strong work ethic often taking several jobs to support themselves and their families. This strong work ethic linked with a high stigma associated with mental health problems in this population, lead to this group scoring very low on many items including one item in the BDI which asks about an individual's ability to work. The idea of cultural aspects affecting the validity of self-report tools due to interpretation of items is supported by a systematic review of validated screening tools for common mental disorders in low and middle income countries (Ali et al, 2016). The

review highlighted that although common mental disorders are prevalent around the world, how they present clinically varies. This affects the use of some tools and in some cases optimum cut-off points differ for different countries and cultures (Ali et al, 2016). Because of this, Ali et al recommended that local validation of tools should be carried out, comparing the screening tool against a gold standard diagnostic interview. This validation can also determine optimum cut-off scores for the population in question.

This effect of culture on interpretation of self-report items may also apply to medical students. Section 1.4 highlighted the differences seen in medical students compared to some other students. These differences included some personality traits often seen in medical students such as conscientiousness. Section 1.4 also highlighted the difficulties medical students may have in admitting they are struggling. Many students fail to seek support due to the hidden curriculum (Dyrbye & Shanafelt, 2011) and the fact that educational tutors and supervisors may be the same people who could be involved in their future recruitment or career choices. These factors may create a culture amongst medical students which differs from other student cultures. In contrast the culture may be similar to some of the cultures seen in different countries where there is a very strong work ethic and stigma associated with mental health problems (Ali et al, 2016). Therefore, it is possible that tools which have been validated for students in general, with optimum cut-off scores identified, may not be appropriate for medical students. As such, it is suggested that, as advised by Ali et al above, local validation of screening tools against a diagnostic interview should be carried out prior to tools being used for medical students.

1.6 The need for accurate measurement of depression and anxiety in medical students

Accurately measuring depression and anxiety in medical students is important for identifying and supporting students who may require help. A number of consequences of poor mental health amongst medical student have been suggested including impaired academic performance, academic dishonesty, substance abuse and suicide (Dyrbye et al, 2005). It is therefore important that medical students who require support are identified and provided with the help they require. As such in 2013 (and updated in 2015) the General Medical Council (GMC) in the UK issued guidance for medical schools on supporting medical students with mental health conditions. In the guidance the GMC states

"*A supportive environment in which medical schools openly discuss mental health conditions will help students feel happy to ask for support. However, not all students will recognise that they have a mental health condition, so schools need processes to identify students who are struggling with the course and might need support*." (GMC, 2015, p36)

The GMC suggests factors such as a drop in academic performance or non-attendance at lectures might indicate developing mental ill health. An alternative would be to screen medical students to individually identify students in need, or surveying all students to provide information for planning of support.

1.6.1 Screening to identify those in need of support

Screening for mental health problems is a contentious issue. Some have suggested that the high prevalence of mental health problems in medical students, along with the potentially fatal consequences of not receiving treatment for serious mental health problems, warrants screening (Dyrbye et al , 2013). Others have suggested that screening is not effective and can cause other problems (Gilbody et al, 2006).

The reluctance of medical students to seek support and treatment for mental health problems drives the view by some that screening is important. Some have suggested this increases the responsibility of medical schools to identify students who may need help (Silva et al, 2017). As such some advocate the screening of medical students (Dyrbye et al, 2013, Goebert et al, 2009). Screening is often seen as a relatively cost and time efficient method of identifying students who may be in need. One study found that having an actual diagnosis of depression improves the likelihood of treatment and therefore suggested periodic mental health screening of medical students (Tija et al, 2005). Some have suggested for screening to be effective a number of requirements should be fulfilled including amongst others that the condition is common in the screened population, that effective interventions should be available, that screening results in the condition being recognised early and that the tool used must have high specificity and very high sensitivity (Willacy, 2019).

In contrast others claim that screening is not effective and therefore of no use. A 2005 Cochrane review found that routine screening for depression had little effect on its detection,

management or outcome and as such suggested that it should not be carried out to improve the quality of healthcare (Gilbody et al, 2006). A 2010 report by the National Institute for Health and Clinical Excellence (NICE) highlighted that there was a lack of evidence to support depression screening in general (NICE, 2010). Others have acknowledged that screening for depression can identify individuals who might otherwise go undetected but have highlighted that it could also lead to other issues such as misdiagnosis and over diagnosis (Thombs et al, 2011). In the case of medical students, misdiagnosis or over diagnosis (where individuals with a mild form of mental health problem are identified as suffering more severe mental health problems) could result in labelling and issues related to stigmatisation which could cause medical students additional stress, as identified previously.

1.6.2 Planning for support

In addition to screening, another possible reason why medical schools or research studies may wish to accurately measure levels of depression or anxiety is to assist with planning for support and treatment of these mental health problems. It is clear that medical schools have a responsibility to support their medical students' mental health needs, both in terms of prevention and in treatment (GMC, 2015). In order for medical schools to plan this support it is important that they have reliable estimates of prevalence of mental health problems within their student body (Rotenstein et al, 2016). Reliable estimates ensure that sufficient funding is in place to provide sufficient amounts of support that may be required. As highlighted in section 1.3 current estimates of mental health problems vary considerably in the different studies, and therefore do not assist medical schools in planning for support. Medical schools may therefore wish to anonymously survey their students to determine prevalence within their specific sample. To ensure this is as accurate as possible, tools validated for use and appropriate cut-off points for medical students need to be determined.

1.7 HADS and its use in screening Medical Students

The current study therefore aims to validate one commonly used tool for identifying or screening for depression and or anxiety with a medical student sample. Hope and Henderson

(2014) conducted a systematic review of articles looking at depression and anxiety in medical students. The most commonly used tools for identifying depression were the Beck Depression Inventory (BDI) (used in 4 studies) and the Hospital Anxiety and Depression Scale (HADS) (used in 3 studies). Similarly, for anxiety the most commonly used scales were HADS (used in 3 studies) and the Beck Anxiety Inventory (BAI) (used in 1 study). HADS was therefore one widely used scale in the studies reviewed.

HADS was developed to identify caseness of anxiety and depressive symptoms in non-psychiatric settings (Zigmond & Snaith, 2015). Caseness being defined as 'probable presence of the mood disorder'. HADS is a brief, easy to use screening tool which can be used to detect depression and anxiety separately or together as a measure of psychological distress (Julian, 2011). HADS has been criticised for reduced validity with some populations (Bjelland et al, 2002).

An important area related to validity of HADS is the cut-off scores used to determine caseness. The original authors of the instrument proposed cut-off scores between 8 and 10 for 'possible' cases, and scores of 11 or more for 'probable' cases in both the anxiety and depression scales (Zigmond & Snaith, 2015)

Since the publication of the original scale many studies have reviewed the cut-off scores which give the greatest sensitivity and specificity for detecting clinical cases of anxiety or depression symptoms (Mitchell et al, 2010). In a literature review looking at the validity of HADS, cut off points ranged from 3+ to 11+ for anxiety and 4+ to 11+ for depression (Bjelland et al, 2002). Kendrick et al (2009) suggested that a cut-off point of 9+ is appropriate for GPs to use to diagnose depression within the UK.

As highlighted in the discussion of tools in general in section 1.5 above, this disparity in optimum cut-offs between studies may be due to methodological differences such as the instrument or interview to which the tool is being compared, alternatively it may also be due to characteristics which are inherent in the sample being tested. Different groups may require slightly different cut-off points to achieve the optimum sensitivity/specificity balance when compared to a clinical diagnostic interview. It is therefore important that tools are validated with the population in question and optimal cut-off scores should be determined.

1.8 Andrews et al's 2006 study.

The author is not aware of any studies comparing HADS to a clinical diagnostic interview with a medical student sample for the purpose of validation and determining optimal cut-off scores. One study has compared HADS to a clinical interview with a general student sample (Andrews et al, 2006). Andrews et al (2006) surveyed second year undergraduate students at a university in London UK. As the main focus of the study was student depression, priority was given to respondents who scored 8 and above on the HADS depression subscale. 90 students were interviewed using the 'structured interview for DSM-IV axis disorders' (SCID). Results suggested that for the sample tested a cut-off score of 10 for the HADS depression gave optimum sensitivity/specificity balance (Andrews et al, 2006). The best sensitivity/specificity balance for the HADS anxiety scale was at a cut off of 13, although sensitivity and specificity at this cut-off were both below .8. The authors thus concluded that students have a tendency to over-report levels of anxiety using HADS (Andrews et al, 2006). The current study aimed to carry out a similar study with a medical student sample, comparing HADS to a clinical interview. The current study aimed to address one of the main limitations of Andrews et al's study, this being the time between completing HADS and the diagnostic interview taking place. In Andrews et al study there was 'some weeks' between the two being undertaken which may have affected the results of the study.

1.9 Summary

The prevalence of mental health problems including depression and anxiety in medical students is thought to be higher than aged matched peers who are not at University. A number of possible reasons or risk factors for why students may be more likely to experience mental health problems have been suggested including moving away from home, financial difficulties, social media, academic demands and worries about future careers. Many studies have suggested that medical students experience high levels of depression and anxiety, with some finding higher levels than other student groups. Possible reasons for higher levels of depression and anxiety include the length of the course and the resulting financial impact, academic demands, career worries, exposure to death, human suffering and ethical conflicts. However, the evidence surrounding the mental health of medical students is contradictory. Differences found may be methodological and could include variability in the sample being tested or the tool used to measure mental health. A vast number of instruments are available

for measuring depression and anxiety in the general population and in student populations, a review of these is outlined in section 2. Not all tools used to measure depression and anxiety have been validated for use with different groups. In addition, cut-off scores used for determining caseness have not been determined for different populations. One study has compared HADS to a clinical interview with a general student sample (Andrews et al, 2006).

The author is not aware of any studies comparing HADS to a clinical diagnostic interview with a medical student sample for the purpose of validation and determining optimal cut-off scores.

This thesis aimed to carry out a similar study to the Andrews et al study, making a comparison between HADS and clinical interviews with a medical student sample. This project does not aim to address the issue of whether medical students should be screened for mental health problems. Instead it aims to validate a tool to determine if it would be effective for screening. In addition, it aims to explore whether a tool could fulfil the criteria highlighted previously of having high specificity and sensitivity with a specific medical student sample (Willacy, 2019).

1.10 Aims of the current study

1.10.1 Main aim

This project aims to investigate the suitability of HADS as a screening tool for anxiety and depression in a medical student population by comparing HADS to a structured clinical interview (Schedules for Clinical Assessment in Neuropsychiatry) (SCAN).

1.10.2 Research questions

In order to address the main aim of this project, this project has three research questions outlined below:

- Research question 1 – Is HADS an accurate measure of anxiety and depression in medical students?

- Research question 2 – Are the standard cut off scores of 8+ on the HADS-D and HADS-A subscales appropriate for a medical student population?

- Research question 3 – Do the responses to individual items within the HADS subscales truly reflect the presence of anxiety and depression.

In summary, this chapter has discussed the literature around mental health in students and medical students and has set out the aims of the current study. Before describing this study in more details (in chapters three to five) chapter two will review the literature specifically around tools for measuring depression and anxiety.

Chapter 2: Literature review

2.1 Introduction and rationale

This chapter presents the findings of a structured review of depression and anxiety tools found in psychological and medical literature looking at measuring depression and or anxiety within the last 10 years. An overview of these tools is provided alongside a closer review of the more commonly used tools found in the literature. In addition, differences between tools used with students, medical students and the general population will be discussed.

Chapter one has highlighted the difficulties students and medical students face when attending university. The higher prevalence of depression and anxiety in, student groups and specifically amongst medical students compared to the general population was discussed. The introduction also highlighted the need for universities and medical schools to identify those students who may require support and discussed some of the difficulties and benefits of screening students. Screening may include the whole student population or just those who present for support by other means. Services offering help with well-being, such as University support services may also wish to screen individuals for the possible presence of depression and/or anxiety to allow for them to signpost individuals to the right place. If students are to be screened, appropriate tools for carrying out screening must be identified. There is therefore a need for accurate and consistent measurement of the presence or severity of depression and/or anxiety.

A vast number of tools have been developed to measure the presence or severity of depression and or anxiety. These include clinician-administered scales and clinical interviews or self-reported scales or questionnaires which are completed by the individual. These tools vary in their number of items or questions, their time to complete, their recall period (the time period in which the individual completing them is asked to consider symptoms over), how they are scored and their cut-off periods. In addition, the scales vary in their reliability and validity and in their ability to determine diagnosis or severity of depression or anxiety. The literature on this topic is vast and wide ranging. There are a number of papers comparing tools, but these tend to concentrate on a few tools looking at their reliability or validity, there is less

literature providing an overview of the available tools. In addition, there are few papers which compared tools used for different populations groups including the general population, students and specifically medical students.

2.2 Aim

This review aimed to review the literature to answer the following questions:

1. What tools have been used for measuring depression and anxiety in studies with students, medical students and the general population in the last 10 years?
2. How do these tools identified differ?
3. Is there a difference between which tools are used for measuring depression and anxiety in studies with students, medical students and general population samples?

2.3 Method

2.3.1 Design

This review used a structured literature review methodology to provide a qualitative summary (Higgins & Green, 2011), of the literature on depression and anxiety scales used with the general population, students and medical students. The reviewer chose to look at the literature involving the general population, students and medical students to allow for any differences in the tools used for these populations to be explored.

The aim of a systematic review is described as "*to identify, appraise and synthesise all the empirical evidence that meets pre-specified eligibility criteria to answer a given research question*" (Cochrane definition, 2013). Often a systematic review involves the use of two researchers who both conduct the review and assessment of the data. In addition, a systematic review aims to collect precise information and evaluate studies in relation to methods used, rigour of conduct of research and strength of evidence (Robinson & Lowe 2015). The current review used a structured approach in relation to the search and identification of papers included. In contrast to a systematic review the review was carried out by the author alone and aimed to provide an overview of the studies carried out including identification of the tools included, rather than evaluating the quality of the research in the

included studies. As such the methodology used is considered a structured review rather than a systematic review.

2.3.1.1 Overview of method

An initial search of the literature was carried out using the data sources and search strategy outlined below. Following the searches, papers were reviewed by their title following the inclusion and exclusion criteria also set out below. If any titles were ambiguous as to whether or not they might be relevant, they were kept in at this stage. Following the initial removal of papers which did not meet the inclusion criteria, the remaining papers were reviewed by reading their abstracts and further papers thought not to be relevant to the aims of the review were discarded. The final set of papers were then resourced and read to identify the tools included in the papers.

2.3.2 Data sources and search strategy

The electronic databases Psycinfo, Medline and Embase were searched. Table 3 outlines the search terms for this review.

Table 3: Literature review search terms.

Mental health related search terms in title and combined with **OR**		Tool related search terms in title and combined with **OR**		Population search terms in keywords and combined with **OR**
Depres*		Measure*		General population
Anx*		Self-report*		Student*
Distress*		Questionnaire*		Medical student*
Mental ill-health	AND	Scale*	AND	
Mental Illness		Clinical interview*		
'Common mental health problem'		Screening*		
CMHP		Diagn*		
Psych*		Tool		
		Instrument		
		Assessment.		
		Prevalence		

Search terms relating to mental health were searched for in titles of publications and the searches were then combined together using the OR option meaning that any titles with depression or anxiety or distress or mental ill-health etc would be included. The same was done with the tool search terms and similar with the population search terms, although these were searched for in the keywords of a publication and not in the title. Once these three separate searches had been carried out they were combined with the AND function, meaning any publication would need to include a mental health search term and a tool related search term in their title and a population search term in their keywords.

Only English language references from 2007 onwards were included. Only papers from 2007 onwards were included to allow the review to consider which tools are being used in current research, rather than considering all tools which have been developed but which may not still be in use.

Once the searches had been carried out, papers were initially reviewed by their title using the inclusion and exclusion criteria outlined below.

2.3.3 study selection

2.3.3.1 Inclusion Criteria

The inclusion criteria for studies included:

- Original studies looking at measuring depression or anxiety in general or student populations (including medical students)
- Studies published between 2007 and 2019
- English language studies

2.3.3.2 Exclusion Criteria

A number of exclusion criteria were used, these were:

- Studies looking at validation or use of the tool with specific medical groups –for example cancer, diabetes, pregnancy etc
- Reviews and meta-analyses
- Measures for specific demographic groups – for example the elderly, children, pregnant
- Tools used to identify patients in remission from depression or anxiety

- Tools used to assess the functional ability/capabilities of people with depression or anxiety
- Tools for measuring other psychological conditions, other than depression or generalised anxiety, unless these tools were specifically being used for the purpose of screening or diagnosing depression and anxiety.

For example, distress scales were not included if the research was looking at distress, but where these tools were reviewed for screening for depression they were included.

2.3.4 Data extraction

Following the searches, papers were reviewed by their title following the inclusion and exclusion criteria also set out above. If any titles were ambiguous as to whether or not they might be relevant, they were kept at this stage. Following the initial removal of papers which did not meet the inclusion criteria, the remaining papers were reviewed by reading their abstracts and further papers thought not to be relevant to the aims of the review were discarded. The final set of papers were then resourced and read to identify the tools included in the papers. Once the tools had been identified, any details about them were noted. For some tools, not all the information which the review aimed to record was available in the literature resourced from the searches. In these cases, specific searches were carried out to find out more specific details about the tool. This included looking at the following types of papers/sources of information:

- original papers describing the development of the tool/s
- papers looking at the organisations who hold copyright permissions for the tool
- looking at papers describing the use of the tool with different groups
- looking at older research papers not included in the current search.

Any papers found which provided additional information on the tools but are included in the references but not in the 'number of studies found' column of tables 5-8 and 13-15.

In addition to identifying the tools used in the included studies, information was recorded about the nature of the research, which group of participants it involved.

2.3.5 Analysis

The review aimed to provide an overview about the studies included in the review and as such, basic descriptive information about different tools and their uses is presented and discussed. In addition, a narrative overview is included for some of the more commonly used tools found in the literature.

2.4 Results

The current review had three main review questions, these being:

1. What tools have been used for measuring depression and anxiety with students, medical students and the general population in the last 10 years?
2. How do these tools identified differ?
3. Is there a difference between which tools are used for measuring depression and anxiety in studies with students, medical students and general population samples?

A structured review methodology was used to search, identify and review studies measuring depression and anxiety in the general population, in student populations and specifically in medical student samples.

The numbers of studies identified will be outlined below, following which the results for the depression and anxiety reviews will be presented separately as to how they answer the aims of the review outlined above.

2.4.1 Number of studies identified

The initial search yielded 7975 studies. An initial review of these studies using title alone according to the inclusion and exclusion criteria reduced the number to 1000 studies. The abstracts for these were then reviewed and further studies removed. The final set of studies included 299 studies, of which 261 were related to measuring depression and 83 related to

measuring anxiety (45 of these included the measurement of both depression and anxiety). See appendix 1 and 2 for a list of all studies included in the review.

The depression and anxiety reviews will now be explored separately, and results presented under the review questions outlined at the start of this chapter.

2.4.2 Depression review

261 studies relating to the measurement of depression in the three population groups (general population, students and medical students) were identified. These were made of:

- 79 studies relating to the measurement of depression in the general population
- 116 studies relating to the measurement of depression in student populations
- 74 studies relating to measurement of depression in medical student populations

Included in these numbers are one study which used both general population and student population samples and 3 studies which used both medical student and other student samples.

The studies identified in the searches were then resourced and the tools included in them were identified. Information about the tools were collected and compiled into tables (see tables 5-8). This included information about:

- when the tool was developed,
- how many items the tool contains,
- how long it takes to complete,
- if the tool is in the public domain or whether it is copyrighted
- if cut off scores are provided for its use as a screening tool.
- how many times each tool appeared in the references identified as relevant for this review.
- information about how the tool was used in the identified studies; papers were coded against the categories in table 4 according to the stated aims of the studies included.

Table 4 Categories for coding the uses of tools within the identified studies

Category	Description
Comparing groups	Studies comparing two or more population groups in relation to depression and anxiety
Comparing tools	Studies comparing two or more tools for measuring depression or anxiety
Diagnostic	Studies testing the diagnostic ability of tools
Longitudinal	Studies measuring depression or anxiety over two or more time periods
Prevalence	Studies measuring the prevalence of depression or anxiety
Psychometric	Studies testing the psychometric properties (often reliability and validity) of a tool
Screening	Studies using depression and/or anxiety tools to screen individuals for depression or anxiety
Service evaluation	Studies evaluating services/interventions for improving wellbeing (and therefore looking at the effect on test scores)
Technology	Studies looking at using technology to assist in measuring depression or anxiety eg online screening
Tool development	Studies describing the development of new scales.
Translation	Studies translating tools into different languages or evaluating tools which have already been translated.

It was decided that information regarding the reliability and validity of the tools would not be included in the review. This was in part due to the variability in the reliability and validity of tools and its dependence on the population with which the tool is being used. A more in-depth discussion of the tools which were used most frequently in the identified studies will also be presented which will include information of reliability and validity for these tools.

2.4.2.1 Review question 1: What tools have been used for measuring depression in studies with students, medical students and the general population in the last 10 years?

54 different tools were identified in the literature for use in determining presence or severity of depression itself or symptoms of depression. The tools could broadly be grouped into 4 separate categories:

1. Self-report scales, questionnaires, inventories or checklists – 45 tools identified (see table 5)
2. Clinician administered scales or inventories – 3 tools identified (see table 6)
3. Clinical interviews – 5 tools identified (see table 7)

4. Alternative tools – 1 tool identified (see table 8)

Whilst all of the self-repot scales, clinician administered scales and clinical interviews require an individual to report on their symptoms and therefore all use self-report. For the purpose of this review self-report tools will refer to those tools which can be completed by a person with no input from an administrator.

The most common type of tool use in the studies identified were self-report tools and the most commonly used tools in the studies were all self-report tools. Of the 261 studies found in the current review, 64 relate to the Beck Depression Inventory (BDI-II). Other tools which were most commonly used in the identified studies included the Physician Health Questionnaire (PHQ-9) which featured in 50 studies, The Centre for Epidemiological Studies-Depression scale (CESD) which appeared in 36 studies and the Hospital Anxiety and Depression Scale (HADS) in 24 studies.

A brief description of each of the four types of tools is presented below, along with tables of the identified tools in each category. This will be followed by a discussion of the 4 most commonly used depression tools as identified above (BDI-II, PHQ, CESD, HADS).

2.4.2.1.1 Self-report scales questionnaires, inventories or checklists

45 tools were identified as being self-report scales, questionnaires, inventories or checklists for measuring depression in the literature, a list of these tools is found in table 5. Whilst there may be subtle differences between self-report scales questionnaires, inventories or checklists it is difficult to get a clear definition of what these differences are.

The oxford dictionary defines a scale as:

"1. In statistics and measurement theory, a rule governing the relationship between numerical scores and the magnitudes of the attributes or quantities being measured. 2.A test or measuring instrument for implementing a scale (1). scaling *n.* The development or construction of scales (1, 2), often by aggregating or ordering responses of individuals." (Oxford dictionary of English, 2010)

A questionnaire is defined as

“In psychometrics, a set of questions specially designed to provide objective information about some characteristic of a respondent, such as attitudes, preferences, interests, values, or personality. Also called an inventory, particularly when used to measure abilities, aptitudes, or intelligence, such instruments seldom being called questionnaires.” (Oxford dictionary of English, 2010)

An inventory is defined as:

“Any list or schedule of items. In psychometrics, a questionnaire or checklist usually functioning as a self-report test of abilities, aptitudes, or intelligence. Tests designed to measure interests, attitudes, personality traits, preferences, and psychological attributes are also sometimes called inventories, but tests designed to measure abilities, aptitudes, and intelligence are usually called questionnaires.” (Oxford dictionary of English, 2010)

The dictionary does not define a checklist, but papers using the checklists also refers to them as a scale (Lundin et al, 2015).

From these definitions a scale maybe considered to have a measure of magnitude, which may not be found in a questionnaire or inventory. However, some of the inventories also include a scale. For example, the Beck Depression inventory (BDI) provides scores attached to its items which are referred to as a scale (Jackson-Koku, 2016). in practice the terms scales, questionnaires, inventories and checklists are used interchangeably and therefore for the purpose of this review we will review them together as self-report tools.

Table 5 – Self-report tools identified for measuring depression

Tool	Long name	When developed	Time to complete	Number of items	Recall period	Cost/copyright	Suggested Cut offs	Number of articles found	Uses of the tool in the review studies
ADI	Adolescent Depression Inventory	2003	Not reported	20	Past month	Not reported	≤7 minimal 8–12 mild depression 13–18 moderate 19+ severe depression,	1	Prevalence = 1
AKUADS	Aga Khan University Anxiety and Depression Scale	1998	Not reported	25	2 weeks	Not reported	Cut-off of 19 for depression	3	Prevalence =3
BADS	Behavioural Activation for Depression Scale	2006	Not reported	25 9 for short	past week	In public domain	Not reported	6	Psychometric = 4 Translating = 1 Longitudinal = 1
BDI – FS	The Beck Depression Inventory-Fast Screen	2000	Less than 5 minutes	7	2 weeks	Available at a cost: $67, including manual and 50 score forms	0-3 minimal, 4-8 mild, 9-12 moderate, and 13-21 severe	1	Prevalence =1
BDI-II	Beck Depression Inventory	1961 revised in 1978 and 1996	5 minutes	21	2 weeks	Available at a cost: manual and 25 record forms = £108 ec VAT, Pack 25 forms = £56.50 ec VAT	0–13: minimal depression 14–19: mild depression 20–28: moderate 29–63: severe	64	Prevalence = 40 Psychometric = 4 Screening = 1 Compare groups = 8 Translating = 2 Compare tools = 8 Longitudinal = 1 Other = 1

Tool	Long name	When developed	Time to complete	Number of items	Recall period	Cost/copyright	Suggested cut offs	Number of articles found	Uses of the tool in the review studies
M-BDI	Modified Beck Depression Inventory	2000	Not reported	21	2 weeks	In public domain	≥35	1	Prevalence = 1
BSI	Brief Symptom Inventory	1983	8-10 minutes	53 items	Past 7 days	Costs $132.85 for manual and 50 answer sheets	Not reported	2	Prevalence = 2
CES-D – short versions	The Center for Epidemiological Studies-Depression scale – 10/7/4	Not reported	1-2 minutes	10/7/4	Last week	In public domain	Originally – 8-10 optimal for risk of depression	3	Prevalence =1 Psychometric = 2
CESD-R	The Center for Epidemiological Studies-Depression scale - revised	1977 Revised 2004	2-5 minutes	20	Last week	In public domain	Originally – 16 A meta review suggested cut off of 20 – (Vilagut, 2016)	36	Prevalence = 11 Screening = 1 Psychometric = 12 Comparing groups = 4 Translation = 1 Comparing tools = 3 Tool development = 1 Service evaluation = 2 Other = 2

Tool	Long name	When developed	Time to complete	Number of items	Recall period	Cost/copyright	Suggested cut offs	Number of articles found	Uses of the tool in the review studies
CRSD	Carroll Rating Scale for Depression	1981	Not reported	17	Past few days	Via eprovide website	10 for mild depression	1	Comparing tools = 1
DASS-21	Depression Anxiety Stress Scale	1995	3 mins	21 – 7 for each scale	Past week	In public domain	Depression normal 0-9, mild 10-13, moderate 14-20, severe 21-27, extremely severe 28+	8	Prevalence = 4 Psychometric = 2 Comparing tools = 1 Longitudinal = 1
DASS-42	Depression Anxiety Stress Scale	1995 – 2nd Edition	7 minutes	42 – 14 items for each scale (depression, anxiety, stress)	Past week	In public domain	Depression normal 0-9, mild 10-13, moderate 14-20, severe 21-27, extremely severe 28+	14	Prevalence = 6 Psychometric = 5 Comparing tools = 1 Longitudinal = 1 Comparing tools = 1
DESC	Rasch-Based Depression Screening	2009	Not reported	10 items	Last 2 weeks	The DESC is available from the principal author.	Not reported	1	Comparing tools = 1
DSM	Diagnostic and Statistical Manual of Mental Disorders – Self-Rated Level 1 Cross-Cutting Symptom Measure	2013	Not reported	8	Past two weeks	Free for clinicians	Less than 55 = None to slight 55.0—59.9 = Mild 60.0—69.9 = Moderate 70 and over = Severe	1	Prevalence = 1

Tool	Long name	When developed	Time to complete	Number of items	Recall period	Cost/copyright	Suggested cut offs	Number of articles found	Uses of the tool in the review studies
GDS	Goldberg depression scale	1998	Not reported	Answer yes or no to 9 symptoms	Past 4 weeks	Not reported	Higher scores = greater symptoms	1	Comparing tools = 1
GHQ-12	General Health Questionnaire	Not reported	2 minutes	12	Recently	Royalty fees - commercial users 500 euros per language, non-commercial 100 euros per language.	2/3 points	9	Prevalence = 3 Screening = 2 Psychometric = 2 Comparing tools = 2
GHQ-28	General Health Questionnaire	Not reported	5 minutes	28	Recently	Not reported	5/6 points	4	Prevalence = 3 Comparing groups = 1
HADS	Hospital Anxiety and Depression Scale	1983	2–5 minutes	14 items: 7 depressio n, 7 anxiety	last week	historically freely available. Royalty - commercial 600 euros per language. Non commercial 100 euros per language	mild 8-10, moderate 11-14, Severe 15-21	24	Prevalence = 6 Psychometric = 8 Compare groups = 6 Translating = 1 Compare tools = 5
IDAS	The Inventory of Depression and Anxiety Symptoms (IDAS),	2007	Developer claim takes no longer than the BDI	64 items	Not reported	Not reported	Not reported	2	Comparing groups = 1 Tool development = 1

Tool	Long name	When developed	Time to complete	Number of items	Recall period	Cost/copyright	Suggested cut offs	Number of articles found	Uses of the tool in the review studies
IDS	Inventory of Depressive symptomatology	1986	Not reported	IDS - 30 items short version: QIDS - 16 items	past 7 days	IDS and QIDS are copyright protected. available through Mapi's ePROVIDE online platform	Rush (1996) In the self-rated version (IDS-SR) a cut-off-point of 18 or above indicates the presence of clinically relevant depressive symptomatology.	3	Screening = 1 Psychometric = 1 Comparing groups = 1
ISR	ICD-10 Symptom rating	2008	Not reported	29 items total, 4 for depressio n	Past 2 weeks	In the public domain - free	Not reported	1	Psychometrics = 1
KADS - 6	Kutcher adolescent depression scale - 6	2006	Not reported	6	Past week	In public domain	Cut off of 6	1	Prevalence = 1
K10	Kessler 10 item distress scale	2003	Not reported	10 items	Past 4 weeks	In public domain	Cut off of 20+ for mild 25-29 moderate, 30-50 severe	6	Prevalence = 3 Screening = 2 Psychometrics = 1
K6	Kessler 6 item distress scale	2002	Less than 2 minutes	6 items	Past 4 weeks	In public domain	Cut off of ≥13	4	Prevalence = 2 Screening = 2
LAPS	the Leuven Affect and Pleasure Scale	2017	Not reported	16	Past week	Development/validati on ongoing	Development/validati on on-going	1	Tool development = 1

Tool	Long name	When developed	Time to complete	Number of items	Recall period	Cost/copyright	Suggested cut offs	Number of articles found	Uses of the tool in the review studies
MACL	Mood adjective checklist	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported	1	Psychometrics = 1
MASQ	Mood and Anxiety Symptoms Questionnaire (short version ADDI)	1995	Not reported	Original – 90 items MASQ-D30 MASQ-AD has 22 items	Not reported	Not reported	Not reported	3	Screening = 1 Psychometrics = 1 Comparing groups = 1
MDI	Major depression Inventory	2001	Not reported	12	Last 2 weeks	Freely available	Diagnosis – 1st 3 symptoms must be present most of the time and 5 of the 9 symptoms must have been present in past 2 weeks – These 5 must include symptoms 1 or 2. Symptom severity: Mild 20-24, moderate 25-29, severe 30+	2	Psychometrics = 1 Longitudinal = 1
MHI-5	Mental health inventory	1993	Not reported	5 items	Last 4 weeks	Not reported	Scores from 0-100, no definitive cut off of 70 suggested	1	Psychometrics = 1

Tool	Long name	When developed	Time to complete	Number of items	Recall period	Cost/copyright	Suggested cut offs	Number of articles found	Uses of the tool in the review studies
ODSIS	Overall Depression Severity and Impairment Scale	2010	abbreviated version - 2 minutes	5	past week	Not reported	A cut score of 8 correctly classified 82% of outpatients as with or without a mood disorder;	2	Tool development = 1 Comparing tools = 2
PHQ-2	Patient health questionnaire 2	2003	Not reported	2	Previous 2 weeks	In public domain, no permissions required	A cut-off of 3 is optimal for screening purposes	6	Screening = 3 Psychometrics = 3
PHQ-4	Patient health questionnaire 4	Not reported	Not reported	4	Previous 2 weeks	Not reported	Not reported	3	Screening = 1 Psychometrics = 1 Diagnostic = 1
PHQ-8	Patient health questionnaire 8	Not reported	Not reported	8	Previous 2 weeks	Not reported	Not reported	3	Psychometrics = 2 Diagnostic = 1
PHQ-9	Patient health questionnaire -9	2001	Not reported	10 (9 & 1 on impairment)	Previous 2 weeks	In public domain, no permissions required	PHQ-9 scores of 5, 10, 15, and 20 represents mild, moderate, moderately severe and severe depression	50	Prevalence = 26 Screening = 9 Psychometric = 8 Comparing groups = 3 Technology = 3 Comparing tools = 2
RRS – 10	Rumination Response Scale	2003	Not reported	10	Not reported	Freely available if work by Hoeksma et al is acknowledged	Not reported	2	Psychometrics = 2

Tool	Long name	When developed	Time to complete	Number of items	Recall period	Cost/copyright	Suggested cut offs	Number of articles found	Uses of the tool in the review studies
SCL-90	Symptom checklist	1994	12-15 minutes	90 items – provides scores for 9 domains including depression	Past week	Costs £131.50 for manual and 50 answer sheets	Not reported	1	Psychometrics = 1
SDS	ZUNG self-rating depression scale	1965	about 10 minutes (WHO website)	20	past several days	in public domain	•25-49 Normal •50-59 Mildly Depressed •60-69 Moderately Depressed •70 and above Severely Depressed	10	Prevalence = 3 Screening = 1 Psychometric = 3 Comparing tools = 3
SF-12	Short-Form Health Survey	1996	Not reported	12 items	Past 4 weeks	Contact Optum for a survey licence	No generally accepted 45.6 suggested by Vilagut	3	Screening = 2 Comparing tools = 1
SRE-20	Self-reporting questionnaire	1994	10 minutes	20 items	Past 30 days	Freely available	7 White Europeans 6 British Pakistanis	1	Screening = 1
SRQ	The Self-Reporting Questionnaire (SRQ)	1994	Not reported	20 items	Past month	In public domain	A range of cut offs have been used in studies	1	Prevalence = 1

Tool	Long name	When developed	Time to complete	Number of items	Recall period	Cost/copyright	Suggested cut offs	Number of articles found	Uses of the tool in the review studies
ST-DEP	State trait depression scale	1995	Not reported	20 : 10 state 10 trait	Not reported	Not reported	Five items in each scale represent the presence depression.	3	Psychometrics = 1 Translation = 1 Comparing groups = 3
USDI	University student depression inventory	2006	Not reported	30	past 2 weeks	Not reported	low – 20-73 moderate 74-95 high 96-118 very high 119-147 Scores do not represent severity of depressive disorder or a diagnosis, but rather the amount depressive symptoms weighted by frequency of occurrence	1 •	Psychometrics = 1
WHO-5	World health organisation well-being index	1998	Not reported	5 items	Last 2 weeks	Free no permissions needed	≤50	1	Psychometrics = 1
ZDS	Zagazig Depression Scale	2010	Not reported	52 in Arabic version 46 in UK version	Not reported	Not reported	< 10 absence of depression symptoms, 10-19 mild, 20-29 moderate, ≥30 severe depressive symptoms	2	Psychometrics = 2

2.4.2.1.2 Clinician administered scales or inventories

Three clinician administered depression scales or inventories were identified in the literature, these were the Hamilton depression scale (HAM-d, HRSD), the quick inventory of depressive symptomatology (QIDS) and the Montgomery-Asberg depression rating scale (MADRS) (see table 6). Clinician administered depression scales are similar to self-report scales, but they require a clinician who has received training in their use to complete them with the participant or patient (in clinical settings).

Table 6 - Clinician administered scales or inventories for measuring depression

Tool	Long name	When developed	Time to complete	Number of items	Recall period	Cost/copyright	Cut offs	Number of articles found	Uses
HAM-d/HRSD	Hamilton Depression Rating scale	1960	20 minutes	17-29 (depending on version)	Past week	Public domain	For the 17-item version, a score of 0–7 is considered to be normal while a score of 20 or higher (indicating at least moderate severity)	6	Prevalence = 1 Screening = 1 Psychometrics = 1 Technology = 1 Comparing tools = 1
MADRS	Montgomery–Åsberg Depression Rating Scale (MADRS)	1979	15-20 minutes	10	Past week	Freely available	0 to 6 – normal/symptom absent 7 to 19 – mild depression 20 to 34 – moderate depression >34 – severe depression.	3	Psychometrics = 1 Technology = 1 Comparing tools = 1
QIDS	Quick Inventory of Depressive symptomatology	2003	Not reported	QIDS – 16 items	Past 7 days	copyright protected. Available through Mapi's ePROVIDE online platform	mild (6 –10), moderate (11–15), severe (16 –20), and very severe (≥21) depression	1	Screening = 1

2.4.2.1.3 Clinical Interviews

Five Clinical interview tools were identified in the literature, the Composite International Diagnostic Interview (CIDI), the *Mini*-International Neuropsychiatric *Interview* (MINI) , the primary care evaluation of mental disorders (PRIME-MD), the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) and the Structured Clinical Interview (SCID) (See table 7). There are fewer clinical interviews than self-report scales use in studies measuring depression. Clinical Interviews are considered by some to be the gold-standard of depression measurement (Gelaye et al, 2014).

Table 7 – Clinical interviews for measuring depression

Tool	Long name	When developed	Time to complete	Number of items	Recall period	Cost/copyright	Clinician administered only?	Number of articles found	Uses
CIDI	Composite International Diagnostic Interview	1990	Not reported	276 – although due to screening not all patients will be asked all questions	Past 30 days	Training required and a software licence purchased.	Lay can be trained	8	Prevalence = 3 Screening = 2 Psychometrics = 1 Comparing tools = 2
MINI	The Mini International Neuropsychiatric Interview	1998	15 -30 min	The MINI consists of 19 modules that explore 17 disorders of Axis I of the DSM-IV	Past 2 weeks for initial symptoms	Any use of the paper/PDF versions of the MINI must first be licensed by Dr David Sheehan. A License Agreement must be signed beforehand by all users with Dr Sheehan and fees may be requested.	Lay can be trained	5	Prevalence = 4 Longitudinal = 1

Tool	Long name	When developed	Time to complete	Number of items	Recall period	Cost/copyright	Clinician administered only?	Number of articles found	Uses
PRIME MD	PRIMary care Evaluation of Mental Disorders	1994	11 minutes (Loerch et al 2000)	Screening questionnaire – 26 questions, interview 17 questions	Past month	The PRIME-MD PQ and CEG are accompanied by a 13-page manual, which includes information regarding reimbursement for the evaluation (which is under copyright of Pfizer, Inc).	Designed for primary care (GPs) – nurses have been trained (Preville et al 2004)	1	Prevalence = 1
SCAN	Schedules for Clinical Assessment in Neuropsychiatry (SCAN)	1992	Variable	1,872 items, spread out over 28 sections. Most patients, will only need parts of the interview	Past month	Training required	Lay can be trained	1	Comparing tools = 1
SCID	Structured clinical interview	1997	30 -60 minutes	Not reported	2 weeks	Fee payable based on intended use.	Clinicians or trained mental health professional	1	Prevalence = 1

2.4.2.1.4 Alternative tools

Only one alternative tool for measuring depression was identified in the literature (see table 8), this is the Rorschach comprehensive system depression index (DEPI) which was included in one study. This category was used for any tools identified which did not fit into the criteria of self-report, clinically administered scale or clinical interview.

Table 8 – Alternative tools for measuring depression

Tool	Long name	When developed	Time to complete	Number of items	Recall period	Cost/copyright	Cut offs	Number of articles found	Uses
DEPI	Rorschach Comprehensive System Depression Index	1991	Not reported	15 variables	n/a	Not reported	A DEPI value of 5 indicates depression likely, 6 or 7 provide more certainty	1	Testing diagnostic ability

In summary 45 self-report tools, 3 clinician administered scales, 5 clinical interviews and one alternative tool were identified in the studies included in this review. The most common type of tool use in the studies identified were self-report tools. The most commonly used tools identified in the studies included in this review were the BDI-II (Beck Depression Inventory) (in 64 studies), the PHQ-9 (Physician Health Questionnaire)(in 52 studies), the CESD (The Center for Epidemiological Studies-Depression scale) (in 38 studies) and HADS (Hospital Anxiety and Depression Scale) (in 25 studies). A description of each of these tools is provided below.

2.4.2.1.5 Beck Depression Inventory (BDI-II)

64 studies identified in the review cited use of the Beck Depression Inventory (BDI-II).

The (BDI-I) was developed in 1979 to measure symptoms and severity of depression (Smarr and Keefer, 2011). After several iterations, the BDI-II (from now on referred to as the BDI) was developed in 1996 omitting items relating to weight loss, body image, hypochondria and working difficulty in order to better fit the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition's (DSM-IV) criteria for depression (Kung, 2013). The BDI includes both cognitive (eg feelings of guilt, low mood) and somatic (eg tiredness) aspects of depression. Although developed as a measure of symptom severity in already depressed patients it is often used as a screening tool in some clinical settings (Aalto, 2012). The BDI-II has reported high levels of reliability and validity in a range of settings (Kjaergaard et al, 2013; Uher, 2012; Aalto,2012). The BDI is one of the most widely used depression self-rating scales. and is one of 3 instruments (along with HADS and the Patient Health Questionnaire (PHQ-9)) that are endorsed by the National Institute for Health and Clinical Excellence for use as a formal rating scale for in assessing symptom severity for depression (NICE, 2011).

THE BDI has 21 items and is quick to administer but comes at a cost of £104 for a manual and 25 forms (from http://www.pearsonclinical.co.uk) . The need to obtain the BDI from the publisher at cost has been seen as a major obstacle against the recommendations of its widespread use (Wang & Gorenstein, 2013).

The initial structure of the BDI received heavy criticism, though the methods used in the evaluation of the structure have since come under fire (Skorikov & Vandervoort, 2003). More recent research suggests that the BDI has a three-factor structure which correspond to Beck's initial subscales of Negative Attitude, Performance Difficulty, and Somatic Elements (Skorikov & Vandervoort, 2003). There has also been much discussion regarding the BDI and it's association with the Beck's Anxiety Inventory (BAI). In the same vein, as with the HADS scale (see section 2.2.1.5.4), some question whether anxiety and depression are measuring two separate constructs or the same construct with anxiety and depression representing different measurements on a continuum. Stultz & Crits-Christoph, (2010) carried out an exploratory factor analysis of the 42 BAI and BDI-II items and identified two separate factors indicating that the BAI and BDI reflect separate dimensions of anxiety and depression respectively. However, Stultz & Crits-Christoph found that in the factor analysis some items affected the factor representing the other measure, and therefore they proposed (as has been suggested with the HADS questionnaire) that a 3-factor structure is present, these being 1 - anxiety symptoms (mostly arousal symptoms such as feelings of tension), 2 - depression symptoms (mostly anhedonic symptoms such as loss of interest) and 3 - negative affect. Others have concluded that the correlations between depression and anxiety scales might be linked to underlying constructs and characteristics of the instruments or may be due to the overlap of symptoms and comorbidity of these conditions (Wang & Gorenstein, 2013).

2.4.2.1.6 Physician Health Questionnaire (PHQ-9)

The PHQ-9 was used in 50 studies identified in this review. The PHQ was developed as a self-administered version of the PRIME-MD interview, a diagnostic interview developed for primary care (Kroenke et al, 2010). Although the PRIME-MD can be administered relatively quickly for an interview (11 minutes Loerch et al, 2000) this time can still be a barrier to it's use in a 15-minute consultation (Kroenke et al, 2010). Containing only 9 items the PHQ-9 takes less time to complete than the PRIME-MD and can be completed independently. Despite the PHQ-9 being originally developed for screening in primary care (Kocalevant et al,2013) it has been validated extensively in medical settings (Martin et al, 2006) and has been found to be valid and reliable (Soltani et al, 2015)

Whilst there has been much support for the PHQ-9 in medical settings and for determining severity of depression, some have questioned it's use in identifying depression in the general population (Eack et al, 2006). Others suggest that there is strong support for the PHQ-9 for discriminating between depressed and non-depressed individuals (Martin et al, 2006). One study with 5000 participants found the PHQ-9 to be reliable and valid in both healthcare settings and in the general population (Kocalevent et al, 2013). Furthermore, Lowe et al (2010) suggest that the PHQ-9 is able to detect depression outcomes and changes over time and encourages its use in both research and clinical practice.

One reason for the discourse, may be due to the different methods of scoring the PHQ-9. Manea et al (2015) describe 2 methods for scoring the PHQ-9, either using a scale and a cut-off score, or by the use of an algorithm, looking at the number of symptoms present. Manea et al suggest that the algorithm method leads to low sensitivity for diagnosis.

As with other depression tools (see BDI, CES-D and HADS), there is some discourse concerning whether the PHQ-9 is measuring one uni-dimensional factor, or whether it measures two or more factors (Kocalevent et al, 2013; Cameron et al, 2008). However, the PHQ-9 seems to attract much less attention in this regard than HADS or the BDI. Granillo (2012) and Elhai et al (2012) suggests a two-factor structure can be identified. Many others suggest that the PHQ-9 is unidimensional (Cameron et al, 2008; Kocalevent et al 2013). This may be because the 9 symptoms measured by the PHQ-9 map directly onto the 9 criteria required for a DSM-IV depression diagnosis (Smarr and Keefer, 2011).

The PHQ-9 has be translated into many languages and as such is used around the world. Along with the BDI and HADS, the PHQ-9 is endorsed by the National Institute for Health and Clinical Excellence for use as a formal rating scale for assessing symptom severity for depression (NICE, 2011). Of these three tools (HADS, BDI, PHQ-9) the PHQ-9 is the only one freely available in the public domain, making it a favourable tool for use in research and clinical settings.

2.4.2.1.7 Center for Epidemiological Studies-Depression scale (CES-D)

In the current literature review the Center for Epidemiological Studies-Depression scale (CES-D) was the 3rd most frequently cited depression scale, cited as being used in 36 studies. The CES-D was developed in 1977 as a short self-report scale for measuring depressive symptoms

in the general population (Radloff, 1977). The CES-D was initially developed for research purposes and has been found to have high levels of reliability and validity for a range of populations and languages (Smarr & Keefer, 2011). It has been suggested that the CES-D provides only a rough indication of clinical depression) and that there is a high correlation between CES-D and anxiety (Smarr & Keefer, 2011). As highlighted previously in relation to the BDI, the correlations between depression and anxiety scales might be due to the overlap of symptoms and comorbidity of these conditions (Wang & Gorenstein, 2013).

The CES-D is not intended as a diagnostic tool. (Smarr, & Keefer, 2011). Vilagut et al suggests that it had acceptable levels of accuracy for screening depression in the general population or in primary care but suggests it should not be used as an isolated measure of depression. One reason for this is that when originally developed the CES-D was not designed to reflect the diagnostic criteria as determined by the DSM (Carleton et al, 2013). Coupled with this, some have questioned the validity and psychometric properties of some of the 20 items contained in the CES-D and have questioned the factor structure underpinning the CES-D (Carleton et al, 2012).

When developed, the original model identified 4 factors, these being depressed mood, positive affect, somatic and interpersonal symptoms (Shean and Baldwin, 2012). In addition to the 4 factors originally proposed, studies have suggested a number of different factor structures, from one factor to four factors (Arbona et al, 2017). Carleton et al found support for a 14 item (compared to original 20 items), three factor model which is more in line with the diagnostic criteria as determined by the DSM (Carleton et al, 2013). Gomez and Mclaren (2015) found good support for the original four factor model but also suggest that if researchers wish to use a total unidimensional score, that if the PA items are removed, their results support a unidimensional structure. This is supported by Edwards et al (2010) who suggested that if required although a one factor model may not be the best fit for the data, it is plausible and can effectively be used. However, Edwards et al also advocated dropping some of the 20 items to find a better one factor model. Fong et al (2016) also suggested that although a four-factor model was the best fit, the fact that the correlations between the 4 factors was strong, suggesting 'substantial overlapping among the dimensions'. Fong et al concluded that the overall CES-D score is an approximate unidimensional measure and that the use of CES-D overall score as a screening measure of depression is justified.

2.4.2.1.8 Hospital Anxiety and Depression Scale

24 studies were identified that used the Hospital Anxiety and Depression Scale (HADS). HADS was developed by Zigmond and Snaith in 1983. The scale was designed to assess the presence and severity of anxious and depressive symptoms in hospital patients (Caci et al, 2003) and as such omitted any somatic symptoms such as dizziness or headaches (Brennan et al 2010). Whilst the tool was designed for screening for depression and anxiety in hospital settings, the tool has been used in the general population and many studies have validated it for this purpose (Mykletun et al, 2001). In addition, the tool has been validated for use in general practice (Olsson, et al 2015) and is one of 3 instruments (along with the Beck depression Inventory (BDI) and the Patient Health Questionnaire (PHQ-9)) that are endorsed by the National Institute for Health and Clinical Excellence for use as a formal rating scale for in assessing symptom severity for depression (NICE, 2011). HADS has been validated for use in many different languages and countries (Stern, 2014)

Despite the popularity of HADS there has been much disagreement in the literature over its use. Many studies have found satisfactory levels of internal consistency, concurrent validity and diagnostic ability (Fong et al, 2013), though other studies have questioned the structure of HADS (Cosco et al, 2012). Some studies looking at HADS suggest that the two-factor structure (depression and anxiety) is valid (Fong et al, 2013; Bjelland et al, 2002), other suggest that a three-factor model fits the data more closely (Cosco et al, 2012). These inconsistencies and disparities have led to some suggesting that HADS should no longer be used (Coyne & Van Sonderen, 2012; Zakrewska 2012). It has been suggested that the differences seen in the dimensionality results is due to differences in the methods used (and the specific analytic strategy employed) to evaluate them rather than different psychological traits (Straat, 2013; Cosco et al; Norton 2013).

A recent article reporting the method and protocol for an upcoming meta-analysis looking at the diagnostic accuracy of HADS suggested flaws in many of the existing studies looking at the accuracy of HADS (Thombs et al, 2016). These problems include small sample sizes, inclusion of diagnosed and treated patients and selectively publishing accurate results from cut-offs that perform well (Thombs et al, 2016).

Other criticisms of HADS include the suggestion that it does not sufficiently distinguish between anxiety symptoms and depressive symptoms (Andrea, 2004). Some argue that this is in part to the comorbidity of the two conditions (Langvik et al, 2016; Cosco et al 2012). The insufficient distinction between anxiety and depression has led some to argue that HADS be used as a general measure of distress rather than as depression and anxiety subscales (Cosco et al 2012 ; Norton 2013). Whilst the structural underpinnings of HADS may be in question, the tool has been seen to be a valid screening tool and many still consider it to be a valuable clinical assessment tool (Cosco et al, 2012)

In summary, 54 different tools were identified in the literature for use in determining presence or severity of depression itself or symptoms of depression. The most commonly used were self-report tools (BDI-II, PHQ-9, CES-D, HADS). Whilst all four of these tools have received criticisms regarding their factor structure, the BDI-II and PHQ-9 have much support in the literature. Three of the tools are endorsed by NICE for assessing symptom severity (BDI-II, PHQ-9 and HADS).

2.4.2.2 Aim/Review question 2: How do the depression tools identified differ?

Tables 5-8 above include Information for each of the tools identified in the review. This included information about, when the tool was developed, how many items it contains, how long it takes to complete, if the tool is in the public domain or whether it is copyrighted, if cut off scores are provided for its use as a screening tool. Finally, it was noted how many times each tool appeared in the studies identified as relevant for this review and how the tools were used in the studies.

Each of the types of tools identified above (Self-report, clinician administered scales, clinical interview and other) will be considered in turn.

2.4.2.2.1 Self-report scales questionnaires, inventories or checklists

45 self-report tools were identified in the studies included in this review. The different characteristics of these studies (eg age, length of time of administration, recall period etc) will now be considered.

2.4.2.2.1.1 When were the tools developed?

Of the 45 self-report tools identified, the oldest was the BDI (Beck depression inventory) originally developed in 1961 and the newest was the LAPS (Leuven Affect and Pleasure scale) developed in 2017 and still undergoing validation and testing with different populations.

2.4.2.2.1.2 Completion time

There was limited information on the length of time it takes to complete the different self-report scales and inventories with many scales or related articles not providing details. Of those where a completion length was provided the SCL-90 (symptom checklist) was stated to be the longest at 12-15 minutes. The shortest was the short versions of the CES-D (The Center for Epidemiological studies depression scale) at 1-2 minutes and the K6 (Kessler distress scale) at less than 2 minutes.

2.4.2.2.1.3 Number of items

Whilst a number of factors may influence the time it takes to complete the self-report scales, the number of items is the predominant factor. The SCL-90 had the highest number of items of all self-report scales (at 90 items) along with the MASQ which also had 90 items, as noted above the SCL-90 had the longest report completion length. The PHQ-2 had the fewest number of items, containing only 2 items, no information was found on time to complete the PHQ-2. Once the number of items are reviewed it gives an understanding about the completion time which also varied and seemed from the review to correlate with number of items.

2.4.2.2.1.4 Recall period

The recall period, the time period which individuals are asked to consider their symptoms over, varied for the 45 self-report tools found from ‘the past few days’ (CRSD (Carroll Rating Scale for Depression) & SDS (Zung Self-rating Depression Scale)) to the past 30 days (eg SRE-20

(Self-reporting Questionnaire)). The recall periods varied across tools as shown in table 9 below

Table 9: Number of tools for different recall periods

Recall period	Number of Scales/inventories
Past few days	2
Past week/ 7 days	12
Last 2 weeks	14
Last 30 days/ 4 weeks	8
Recently	4

The DSM-IV and ICD-10 classification of depression require symptoms to have been present for at least 2 weeks (American Psychiatric Association, 2013; World Health Organisation, 1992). The scales which asked for symptoms over the last 4 weeks were mostly scales which were not originally developed to specifically screen for depression (eg K6 and K10, SF-12) but which may be used to screen for other mental health problems such as anxiety which requires symptoms to be present for a much longer period of time (WHO, 1992). The General-Health Questionnaires which asks for 'recent' symptoms were developed to be sensitive to short-term psychiatric disorders and are not designed for long-term assessment.

2.4.2.2.1.5 Public domain/copyrighted tools

Of the 45 self-report tools, 24 were in the public domain. Some of these (eg RRS (Rumination Response scale) require specific references to be cited for their use (eg RRS – 10). 11 of the tools identified required permissions and/or payment for their use. Payments vary for some of the tools (HADS, GHQ-12) dependent on their use. For example, the GHQ-12 commercial royalty fees are €500 (approximately £440) per language for commercial uses and €100 (approximately £88) per language for non-commercial users.

2.4.2.2.1.1 How the tools were used

The majority of the self-report tools identified were developed to assess symptoms relating to depression. Some of these tools only measure depression symptoms (for example the BDI, PHQ), other measure depression symptoms along with other symptoms such as anxiety (eg HADS) and/or stress (Eg Depression anxiety and stress scale DASS). However, some of the tools were developed to measure other symptoms but have been found to correlate with depression and as such are used for measuring depression in some studies (eg the Kessleer distress scales and the World Health organisation wellbeing index – WHO-5).

In the studies using self-report inventories included in the current review, the tools were used to measure depression for a range of purposes. These include:

- Measuring prevalence of depression in different populations – 121 studies
- Review of the psychometric properties of the tool (reliability, validity, factor structure) – 49 studies
- Comparing different population groups (eg gender, ethnicity, year of study)- 19 studies
- Evaluating the effectiveness of the self-report tool for screening depression – 22 studies
- Comparing different self-report tools – 14 studies
- Development or evaluation of a tool in a different language – 7 studies
- Other studies included, incorporating tools into technology (4 studies) tool development (4 studies), longitudinal studies (5 studies)

Table 10 shows how the 4 most commonly used self-report depression tools were used in the studies

Table 10 How the 4 most commonly used self-report depression tools were used in the studies

	BDI (in 64 studies)	CES-D (in 36 studies)	HADS (in 24 studies)	PHQ-9 (in 50 studies)
Prevalence studies	40 (63% of BDI studies)	11 (31% of CESD studies)	6 (25% of HADS studies)	26 (52% of PHQ studies)
Psychometric properties	4 (6% of BDI studies)	12 (33% of CESD studies)	7 (29% of HADS studies)	8 (16% of PHQ studies)
Comparing populations	8 (13% of BDI studies)	4 (11% of CESD studies)	6 (25% of HADS studies)	3 (6% of PHQ studies)
Screening evaluation	1 (2% of BDI studies)	1 (3% of CESD studies)	0 (0% of HADS studies)	9 (18% of PHQ studies)
Comparing tools	8 (13% of BDI studies)	3 (8% of CESD studies)	5 (21% of HADS studies)	2 (4% of PHQ studies)
Different languages	2 (3% of BDI studies)	1 (3% of CESD studies)	1 (25% of HADS studies)	0 (0% of PHQ studies)

The BDI and the PHQ-9 were most frequently used in prevalence studies and the CES-D and HADS were most frequently used in studies evaluating their psychometric properties.

2.4.2.2.2 Clinician administered scales or inventories

Three clinician administered depression scales or inventories were identified in the literature, these were the Hamilton depression scale (HAM-d, HRSD), the quick inventory of depressive symptomatology (QIDS) and the Montgomery-Asberg depression rating scale (MADRS). The oldest of these was the HAM-D developed in 1960 and the newest was the QIDS developed in 2003. Number of items varied from 10 for MADRS to 29 for the longer version of the HAMS. The HAMS takes about 20 minutes to complete and the MADRS about 15-20 minutes, no information was found on the administration time of the QIDS. All the scales have a recall period of one week and whilst the HAMD and MADRS are freely available in the public domain

the QIDS is copyrighted. Of the three scales the HAMD was the most prevalent in the literature, being used in 11 studies.

2.4.2.2.3 Clinical Interviews

Five Clinical interview tools were identified in the literature, the Composite International Diagnostic Interview (CIDI), the *Mini*-International Neuropsychiatric *Interview* (MINI), the primary care evaluation of mental disorders (PRIME-MD), the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) and the Structured Clinical Interview (SCID). They were all developed in the 1990's with the earliest being the CIDI in 1990. All five have a large number of questions to cover a range of mental disorders, but they acknowledge that not everyone would need all questions. The CIDI, MINI and PRIME-MD have screening questions to determine which parts of the interview to use. The SCAN and SCID have initial questions asking about what kinds of problems or symptoms have been experienced to allow the interviewers to determine which parts of the interview are required. Given the all or part of the interviews may be used during any given interview, the administration time for the interviews vary from individual to individual. The interview requiring the least amount of administration time is the MINII, which on averages takes between 15 and 30 minutes to administer. The MINI is structured so that responders only answer yes or no to questions, allowing for a quicker interview. All of the identified interviews except the SCAN are copyrighted and require some permissions and or payment for their use. Training is required for all the interviews. The CIDI, MINI and SCAN can be carried out by individuals who are not trained clinicians but who have received training. Although the PRIME-MD is designed for use in general practice and as such designed for use by general practitioners, some studies (eg Preville et al, 2004) have trained other health professionals such as nurses to use the tool. The SCID is designed for use by a clinician or trained health professional.

2.4.2.2.4 Alternative tools

The final tool identified in the literature is the Rorschach comprehensive system depression index (DEPI) identified in one study. The DEPI is a structured version of the Rorschach inkblot test (Mondal, 2017). The inkblot test is described by Mihura and Meyer (2015) in their book 'The inkblot test' as follows:

“The test including 10 inkblots: 5 are achromatic and 5 include chromatic colors. Respondents are asked what the inkblots “might be.” Their responses are coded, tabulated, and compared with norms for interpretation. “

Use of the Rorschach test in general (not just the depression index) has been controversial and had resulted in much discussion in the literature. In 1999 Garb called for a moratorium on the use of the Rorschach Inkblot Test in clinical and forensic settings until the validity of different scores had been determined (Garb, 1999). In 2015 a meta-review provided a more detailed look at the Rorschach and determine it still had some use in clinical settings. Specifically looking at the DEPI, little support has been given for its use and Mondal et al concluded that “DEPI scores should be interpreted with considerable caution when applied for diagnostic purposes.”

In summary the 54 tools identified from the studies included in this review vary on a number of aspects. Differences occur not only between the different categories of tool (eg self-report versus clinical interview) but there are also differences in some of the characteristics seen within categories, for example in the recall-period of the different self-report tools which vary from the past week to the past month.

2.4.2.3 Aim/Review question 3: Is there a difference between which tools for measuring depression are used in studies with students, medical students and the general populations

Out of the 261 studies identified in this review 79 studies used a general population sample, 116 studies used a student sample and 74 studies used a medicals student sample. Due to the low number of studies using clinician administered scales, clinical interviews and other tools, this section will concentrate only on the self-report scales.

The 79 studies which included a general population sample used 28 different self-report tools. The 116 studies which included a student sample used 25 different self-report tools. The studies which included a medical student sample used 15 different tools. Therefore, in the

studies included in the review, fewer tools were used with medical students than with the general population or other students.

The studies identified in this review were split into categories based on the participants which took part and the percentage of studies using the different tools in each category was calculated. The following table highlights for each category the percentage of studies using the most commonly used tools overall (as identified in section 2.4.1 above). I.e. For the studies which included a general population sample 7% used the BDI, 25% the PHQ, 17% the CES-D and 16% HADS.

Table 11 the percentage of studies in each participant category using the most commonly used tools

	Percentage of all General Population studies identified which used this tool (n=79)	Percentage of all student studies identified which used this tool (n=116)	Percentage of all medical student studies identified which used this tool (n-74)
BDI	8%	35%	34%
PHQ-9	24%	16%	19%
CES-D	18%	17%	3%
HADS	17%	5%	9%

Table 11 highlights that whilst the BDI was used in about 1/3 of all studies in both the student (n=35) and medical student (n=25) participant categories only 8% (n=6) of studies with a general population sample used the BDI. The most commonly used tool with a general population sample was the PHQ-9 which was used in 24% (n=19) of studies with a general population sample. HADS was used least in studies with a student sample (5%, n=6) and most in studies with a general population sample (17% n=13). Out of the four most commonly used

tools overall, the tools used fewest with a medical student sample was the CESD used in only 3% (n=2) of medical student studies.

How the tools were used in the studies with the different populations was also evaluated, results are shown in table 12

Table 12 – Uses of the tools in the studies with different populations

	General population (79 studies)	Students (116 studies)	Medical students (74 studies)
Prevalence studies	22 (28% of general population studies)	50 (43% of student studies)	57 (77% of medical student studies)
Psychometric properties	25 (32% of general population studies)	27 (23% of student studies)	1 (1% of medical student studies)
Comparing groups of participants	5 (6% of general population studies)	15 (13% of student studies)	7 (9% of medical student studies)
Screening evaluation	10 (12% of general population studies)	10 (9% of student studies)	3 (4% of medical student studies)
Comparing tools	9 (11% of general population studies)	7 (6% of student studies)	1 (1% of medical student studies))
Different languages	2 (3% of general population studies)	4 (3% of student studies)	1 (1% of medical student studies)

In studies with the general populations, the most common use of self-report tools was to measure their psychometric properties. For students and medical students, the most common use of the self-report tools was to measure the prevalence of depression, with 77% of medical student studies identified being used to measure prevalence of depression among medical students.

2.4.2.4 Summary of the depression review

54 different tools for use in determining presence or severity of depression itself or symptoms of depression were identified in the literature. The most common type of tool used in the studies identified were self-report tools with 45 different self-report tools being identified. Of the 261 studies which measured depression found in the current review, 64 used the BDI, 50 the PHQ, 36 the CES-D and 24 HADS

The 45 self-report tools varied in relation to a number of characteristics including the purpose for which they were developed and the way in which they are used. The majority of the self-report tools were developed for measuring depression, however 4 were originally developed for measuring other mental health problems. In relation to the use of the tools, the tools were most commonly used for measuring prevalence, reviewing the psychometric properties of the tool, comparing different groups, evaluating the effectiveness of the tool for screening and comparing different self-report tools. The most common use was for measuring the prevalence of depression in different populations which occurred in 121 studies. In addition, the tools varied according to the length of time they take to complete, the number of items they included and the recall period. 3 clinician administered scales for measuring depression were identified. The clinician administered scales take longer to complete than the majority of the self-report scales at around 20 minutes and include between 10 and 17 items. 5 clinical interviews for diagnosing depression were identified in the literature. They all have a large number of items but 3 have screening questions to determine which parts of the interview are required. The clinical interviews between 11 minutes and 60 minutes to complete. One alternative tool was identified for measuring depression which was the DEPi a structured version of the Rorschach inkblot test. Limited information was available for this test and some have suggested its use should be treated with caution.

Some differences were observed between the tools used by studies with general population samples and those with student or medical student samples. The PHQ-9 was more commonly used with general population samples and the BDI with students and medical students. Differences were also seen in the use of self-report tools with the different populations; In studies with the general population, the most common use of self-report tools was to measure

psychometric properties. For students and medical students, the most common use was to measure the prevalence of depression.

2.4.3 Anxiety review

83 studies relating to the measurement of anxiety in the three population groups (general population, students and medical students) were identified. These were made of:

- 26 studies relating to the measurement of anxiety in the general population
- 42 studies relating to the measurement of anxiety in student populations
- 16 studies relating to measurement of anxiety in medical student populations

Included in these numbers is one study which included both student and medical student samples.

The studies identified in the searches were then resourced and the tools included in them were identified. Information about the tools were collected and compiled into tables 13-15. As with the depression review information collected included:

- when the tool was developed,
- how many items the tool contains,
- how long it takes to complete,
- if the tool is in the public domain or whether it is copyrighted
- if cut off scores are provided for its use as a screening tool.
- how many times each tool appeared in the references identified as relevant for this review.
- Information about how the tool was used in the identified studies, using categories outlined in table 4 above.

Again, as with the depression review a closer look at the literature for the most frequently used tools for measuring anxiety was carried out and a discussion of these tools will be presented.

2.4.3.1 Aim/review question 1: What tools have been used for measuring anxiety in studies with students, medical students and the general population in the last 10 years?

27 different tools were identified in the literature for use in determining presence or severity of anxiety itself or symptoms of anxiety (see tables 13-15).

For the anxiety review, the tools could broadly be grouped into 3 separate categories:

- Self-report scales questionnaires, inventories or checklists – 21 tools identified
- Clinician administered scales or inventories – 2 tools identified
- Clinical interviews – 4 tools identified

A brief description of each of the three type of tools can be found in the depression review above.

2.4.3.1.1 Self-report scales questionnaires, inventories or checklists

The most common type of tool use in the studies identified were self-report tools. Of the 83 studies found in the current review, the most commonly used tools were the Beck Anxiety Inventory (BAI) used in 15 studies, followed by the Generalised Anxiety Disorder (GAD-7) scale used in 12 studies, the Depression, Anxiety and Stress Scale – 42 items version (DASS-42), used in 12 studies and the DASS-21 used in 11 studies. Table 13 provides an overview of the identified self-report scales, questionnaire and inventories for measuring generalised anxiety.

Table 13 – Self-report tools identified for measuring generalised anxiety

Tool Self-report	Long name	When developed	Time to complete	Number of items	Recall period	Cost/copyright	Suggested cut offs	Number of studies found	How the tool is used in the studies
ADDQ	Anxiety Disorder Diagnostic Questionnaire	2010	Not reported	4 parts	Past month	Public domain	27.5	1	Tool development = 1
AKUADS	Aga Khan University Anxiety and Depression Scale	1998	Not reported	25	2 weeks	Used in Pakistan - Urdu language	Cut-off of 19 for depression	1	Prevalence = 1
BAI	Beck Anxiety Inventory	1988	5-10 minutes	21	Past month	Copyrighted $138.25 for a manual and 25 forms	0-9 normal 10-18 mild 19-29 moderate 30+ severe	15	Prevalence = 8 Psychometrics = 3 Comparing groups = 1 Comparing tools = 2 Tool development = 1
BSI-18	Brief Symptom Inventory 18	1993	4 minutes	18	Past week	Copyrighted $132.85 for manual and 50 sheets	T score of 63 or above suggests clinical levels anxiety	1	Tool development = 1
Cottle anxiety test	No information found for this tool	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported	1	Prevalence

Tool Self-report	Long name	When developed	Time to complete	Number of items	Recall period	Cost/copyright	Suggested cut offs	Number of studies found	How the tool is used in the studies
DASS-21	Depression Anxiety Stress Scale	1995	3 mins	21 – 7 for each scale	Past week	In public domain	Not reported	11	Prevalence = 2 Psychometrics = 5 Comparing groups = 1 Translation = 2 Comparing tools = 1 Longitudinal = 1
DASS-42	Depression Anxiety Stress Scale	1995 – 2nd Edition	7 minutes	42 – 14 items for each scale (depression, anxiety, stress)	Past week	In public domain	Depression normal 0-9, mild 10-13, moderate 14-20, severe 21-27, extremely severe 28+	12	Prevalence = 5 Psychometrics = 4 Comparing groups = 1 Tool development = 1
GAD-7	Generalised Anxiety Disorder - 7	2006	Not reported	7	Last 2 weeks	Public domain	• 0-5 = Mild anxiety • 6-10 = Moderate anxiety • 11-15 = Moderately severe anxiety • 15-21 = Severe anxiety	12	Prevalence = 3 Screening = 1 Psychometrics = 3 Comparing groups = 1 Technology = 1
GAD-Q-IV	Generalised anxiety disorder questionnaire	2002	Not reported	10	6 months	Public domain	5.7 suggested by authors	1	Tool development = 1

Tool Self-report	Long name	When developed	Time to complete	Number of items	Recall period	Cost/copyright	Suggested cut offs	Number of studies found	How the tool is used in the studies
HADS	Hospital Anxiety and Depression Scale	1983	2–5 minutes	14 items 7 depressio n, 7 anxiety	last week	historically freely available. Royalty - commercial 600 euros per language non commercial 100 euros per language	mild 8-10, moderate 11-14, Severe 15-21	9	Prevalence = 4 Psychometrics = 1 Comparing groups = 3 Technology = 2
K10	Kessler 10 item distress scale	2003	Not reported	10 items	Past 4 weeks	In public domain	Cut off of 20+ for mild 25-29 moderate, 30-50 severe	2	Prevalence = 2
K6	Kessler 6 item distress scale	2002	Less than 2 minutes	6 items	Past 4 weeks	In public domain	Cut off of ≥13	2	Prevalence = 2
MASQ	Mood and anxiety symptom questionnaire	1995	Not reported	77	Not reported	Not in public domain available from Prosetta stone	Not reported	1	Psychometrics = 1
OASIS	Overall *Anxiety* Severity and Impairment Scale	2006	2-3 minutes	5	Past week	Public domain	8	1	Tool development = 1

Tool Self-report	Long name	When developed	Time to complete	Number of items	Recall period	Cost/copyright	Suggested cut offs	Number of studies found	How the tool is used in the studies
PHQ-4/GAD-2	Generalised Anxiety Disorder-2	2009	Not reported	2	Past 2 weeks	Public domain	Greater than or equal to 3	3	Screening = 1 Psychometrics = 1 Diagnosis - 1
PSWQ	Penn State Worry questionnaire	1990	Not reported	16	No time period	Public domain	45	1	Tool development = 1
SAS	Zung self-rating anxiety scale	1971	Not reported	20	Past 2 weeks	Public domain	Raw score of 36 cut off for clinically significant anxiety	4	Prevalence = 2 Psychometrics = 1 Tool development = 1
SCAT	Sinha's comprehensive anxiety test	1995	15-20 minutes	90	N/A	Available in India	Not reported	1	Prevalence = 1
STAI	State-Trait Anxiety Inventory	1983	10 minutes	40	Not reported	Copyrighted $125 for 50 inventories	No cut off in administration materials.	1	Screening = 1
STAIS	State-Trait Anxiety Inventory – State scale	1983	5 minutes	20	Not reported	Copyrighted	No cut off in administration materials.	1	Translation = 1
WDQ	Worry domains questionnaire	1992	Not reported	25	No time period	Public domain	Not reported	1	Psychometrics = 1

2.4.3.1.2 Clinician administered scales or inventories

Two clinician administered scales for measuring generalised anxiety were identified in the literature, as seen in table 14

2.4.3.1.3 Clinical Interviews

Four Clinical interview tools were identified in the literature, the Composite International Diagnostic Interview (CIDI), the *Mini*-International Neuropsychiatric *Interview* (MINI), the Clinical interview schedule revised (CIS-R), and the Structured Clinical Interview (SCID) (See table 15). As with the depression review, there were much fewer clinical interviews than self-report scales use in studies measuring generalised anxiety.

Table 14 – Clinician rated tools for measuring generalised anxiety

Tool - Clinician rated	Long name	When developed	Time to complete	Number of items	Recall period	Cost/copyright	Cut offs	Number of articles found	Uses
GAS	Goldberg anxiety scale	1987	Not reported	9 items	Past month	public	5+	1	Comparing tools = 1
HARS	Hamilton anxiety rating scale	1959	10-15 minutes	14	Not reported	Public	<17 mild severity, 18–24 mild to moderate severity 25–30 moderate to severe	2	Service evaluation = 1

Table 15 – clinical interviews for measuring generalised anxiety

Tool - Interview	Long name	When developed	Time to complete	Number of items	Recall period	Cost/copyright	Training	Number of articles found	Uses
CIDI	Composite International Diagnostic Interview	1990	15 mins	276 – although due to screening not all patients will be asked all questions	Past 30 days	Copyrighted and a software licence required	Training required	6	Prevalence = 4 Comparing tools = 1 Diagnosis = 1
CIS-R	Clinical interview schedule revised	1992	30 mins	14 symptom groups	Previous month	Not reported	Training required	1	Diagnosis = 1
MINI	The Mini International Neuropsychiatric Interview	1998	15 -30 min	The MINI consists of 19 modules that explore 17 disorders of Axis I of the DSM-IV	Past 2 weeks for initial sympto ms	A License Agreement must be signed and fees may be requested.	Lay can be trained	3	Prevalence = 2 Psychometrics = 1
SCID	Structured clinical interview	1997	30 -60 minutes	Not reported	Not reporte d	Fee payable based on intended use.	Clinicians or trained mental health professiona l	3	Diagnosis = 1 Comparing tools = 1 Psychometrics = 1

2.4.3.1.1 Beck Anxiety Inventory (BAI)

The BAI was one of the most prevalent anxiety tools used in the literature reviewed, appearing in 15 studies. The BAI was developed as a brief measure of anxiety which aimed to discriminate between depression and anxiety (Julian, 2011). As such the author included items specific to the physiological and cognitive symptoms of anxiety, which do not overlap with symptoms of depression. Leyfer et al (2006) suggest this focus on physiological symptoms extends its use to non-clinical samples and as such the BAI has been used widely. The BAI has been translated into a number of languages and has been used with a range of different samples including students (and medical students) and community samples (Julian, 2011). The BAI is considered to have good convergent validity with other measures of anxiety and good internal consistency (Julian, 2011). Although there is some discussion in the literature regarding the factor structure of the BAI and whether the BAI is a better measure of symptoms of panic rather than general anxiety (Leyfer et al, 2006). Since its development different factor structures have been suggested for the BAI, ranging from 2 to 5 (De lima et al, 2011). Some of these suggested factor structures better reflect the nature of panic disorders rather than anxiety disorders (Leyfer et al 2006). In addition, the BAI also has been found to have better convergent validity with measures of panic disorder than with measures of generalised anxiety (Leyfer et al, 2006). Leyfer et al suggest this discord is due to the original aim of the BAI as a measure of anxiety which could also discriminate between depression and anxiety. This aim resulted in the exclusion of anxiety symptoms which overlap with symptoms of depression, leaving more symptoms which link to panic disorders. Muntingh et al (2011) suggested that the testing of the BAI in settings such as treatment centres might result in participants with a greater severity of disorder than in primary care. Muntingh et al however, concluded that the BAI does measure general anxiety in primary care. Others conclude that if it is to be used as a measure of general anxiety, any user should be aware of its limitations, in particular users should consider that the BAI may provide an assessment of one aspect of anxiety and therefore may require other forms of assessment alongside (Leyfer et al, 2005).

2.4.3.1.2 Generalised Anxiety Disorder (GAD-7) scale

GAD-7 (Spitzer et al. 2006) was the second most commonly used tool identified from the studies include in this review, used in 11 studies. The GAD-7 was developed from the PRIME-MD diagnostic interview and measures a range of anxiety disorders including generalised anxiety disorder, social anxiety and panic disorder (Maideen et al, 2015). The tool has 7 items which are each scored from 0-3, giving a total score of 21. A cut off of 10 is suggested for detecting generalised anxiety using the GAD-7 scale (Johansson et al, 2013), however a cut-off of 8 has been shown to have a better balance of sensitivity and specificity (Kroenke et al., 2007). The GAD-7 has been translated into a number of different languages including German (Wiltink et al), Swedish (Johansson et al, 2013) and Chinese (Lu et al). The GAD-7 has been found to be a reliable and valid tool for measuring anxiety in the general population (Lowe et al, 2008). Comparisons with the Beck Anxiety Inventory shown a high correlation between the 2 tools (Kroenke et al., 2010).

2.4.3.1.3 Depression, Anxiety and Stress Scale – 42 (DASS-42)

The DASS-42 was one of the most prevalent anxiety tools cited in the literature reviewed, appearing in 12 studies. The DASS is a set of three scales designed to measure depression, anxiety and stress. Each scale has 14 items, which are scored from 0–3, with higher scores indicating greater levels of depression, anxiety, and/or stress (Gale, 2015). Similarly, to the development of the BAI, the DASS-42 was developed with the aim of discriminating between depression and anxiety (Ediz et al, 2017). Items which were considered for inclusion, but which were not closely related to depression and anxiety formed a third group which became known as the stress scale (Ediz et al, 2017). studies included in this this review, which discussed the factor structure of the DASS-42, supported the three-factor structure (Bilgel & Bayram, 2010; Chan et al, 2012). Both of the DASS scales have been validated against other measures and have been found to possess good reliability in both clinical and community samples (Beesdo-Baum et al, 2014). Dunstan et al (2017) compared the DASS-42 with two Zung scales; the Zung depression scale (SDS) and the Zung anxiety scale (SAS). Dunstan et al found the DASS-42 to be a stronger predictor of depression and anxiety but also found a high correlation between the depression and anxiety subscales. Dunstan et al suggested this was due to the co-morbidity of the two conditions.

2.4.3.1.4 Depression, Anxiety and Stress Scale – 21 (DASS-21)

The DASS-21 is a shortened version of the DASS-42. Whilst the DASS-42 contains 14 questions for each subscale, the DASS-21 contains 7 questions for each subscale. Both of the DASS scales have been validated against other measures and have been found to possess good reliability in both clinical and community samples (Beesdo-Baum et al, 2014). Antunez & Vinet (2012) found that the DASS-21 was valid for use with Chilean students and supported a 3-factor structure. Alfonsson et al (2017) were looking at the Swedish version of the DASS-21, whilst they found adequate support for a 3 factor model their results also support a bi-factor structure with depression, anxiety, stress and a general factor. Alfonsson et al suggest that the stress and anxiety symptoms overlap which can make specific screening difficult. Osman et al (2012) also found evidence of a general distress factor and stated their findings are consistent with the co-morbidity of mood and anxiety disorders. Ruiz et al (2017), investigated the Spanish version of the DASS-21 and found support for a hierarchical factor structure, containing 3 'first-order' factors (depression, anxiety and stress) and one 'second order' factor which they stated were 'emotional symptoms'. Whilst the discord between the factors may make screening difficult Alfonsson et al conclude that "In practice, it seems that the DASS-21 can be used to measure both specific psychiatric domains and general psychological distress".

In summary, 27 different tools were identified in the literature for use in determining presence or severity of generalised anxiety itself or symptoms of generalised anxiety. The most common type of tool used in the studies identified were self-report tools (BAI, GAD-7, DASS42/21). Whilst the GAD-7 is considered by most to be a reliable and valid tool. The BAI and DASS scales have been criticised for their structure and overlap between the depression and anxiety scales (the BAI being compared to the BAI). The BAI has also received criticism for measuring more aspects of panic disorder than generalised anxiety.

2.4.3.2 Aim/Review question 2: How do these anxiety tools identified differ?

Information for each of the tools was collated and can be found in tables 13-15 above. A discussion of the differences between tools will be presented below, each of the types of tools identified above (Self-report, clinician administered scales and clinical interview) will be considered in turn.

2.4.3.2.1 Self-report scales questionnaires, inventories or checklists

23 self-report tools were identified in the studies included in this review. The different characteristics of these studies (eg age, length of time of administration, recall period etc) will now be considered.

2.4.3.2.1.1 When were the tools developed?

Of the 23 self-report tools identified, the SAS has been in use the longest, having been developed in 1971. The tool most recently developed was the SQ-48 developed in 2012.

2.4.3.2.1.2 Completion time

There was limited information on the length of time it takes to complete the different self-report tools. Of those where a completion length was found the SCAT was stated to be the longest at 15-20 minutes and the K6 was found to be the quickest to complete, taking less than 2 minutes. As with the depression review, one factor affecting the length of completion is the number of items contained within the tool.

2.4.3.2.1.3 Number of items

The SCAT contains the greatest number of items, at 90 items which may explain why it takes longer to complete than the K6 which contains only 6 items. The tool with the fewest number of items was the GAD-2 (part of the PHQ-4) which contains only 2 items.

2.4.3.2.1.4 Recall period

The recall period, the time period which individuals are asked to consider their symptoms over, varied for the 23 tools found, from the past week (BSI, DASS-21&42, HADS, OASIS) to the past 6 months (GAD-IV). 4 tools asked about symptoms over the last 2 weeks (GAD-7, GAD-2, SAS), 4 tools asked about symptoms over the past month or 4 weeks (ADDQ, BAI, K10 & K6, AKUADS). 6 did not mention any time frame but just asked how often the if the individual had these symptoms for example: always, sometimes or never (MASQ, PSWQ, STAI, STAIT, STAIS, WDQ). The DSM-IV and ICD-10 classification of generalised anxiety requires symptoms to have been present for at least 6 months (American Psychiatric Association, 2013; World Health Organisation, 1992). Only 1 tool asked about symptoms going back that far. There may be a number of reasons why this might be the case for example many of these tools are developed for screening for anxiety and not diagnosing generalised anxiety disorder. In the case of screening for the purposes of providing help, it may not be necessary to wait for 6 months for a diagnosis of generalised anxiety disorder prior to starting support.

2.4.3.2.1.5 Public domain/copyrighted tools

Of the 21 self-report anxiety tools identified, 12 are in the public domain and 6 require some permission and/or payment for their use. For some of these tools payment varies depending on use for example, research verses commercial use (see HADS). Of the tools which require payments, these varied from 600 euros (£539) for HADS commercial use to $138.25 (approx. £107) for a BAI manual and 25 forms.

In summary, the 21 self-report scales for measuring anxiety varied on a range of characteristics including number of items, completion time and the recall period. The clinician administered scales and clinical interview will now be considered.

2.4.3.2.1.6 How the tools were used

The 23 self-report anxiety tools which were identified from the studies included in this review differed in terms of their purpose and usage. 19 of the tools were developed to measure anxiety, 2 of the tools were developed as distress scales (K6 & K10) and 2 as worry scales (PSWQ, WDQ). These 4 tools however have since been used to measure general levels of anxiety in studies.

In the studies using self-report inventories included in the current review, the tools were used to measure anxiety for a range of purposes. These include:

- Measuring prevalence of anxiety in different populations – 40 studies
- Review of the psychometric properties of the tool (reliability, validity, factor structure) – 17 studies
- Comparing different population groups (eg gender, ethnicity, year of study)- 6 studies
- Evaluating the effectiveness of the self-report tool for screening depression – 4 studies
- Comparing different self-report tools – 3 studies
- Development or evaluation of a tool in a different language – 4 studies
- Other studies included, incorporating tools into technology (1 study) tool development (4 studies), testing diagnostic ability (2 studies).

The following table shows how the 4 most commonly used self-report anxiety tools were used in the studies

Table 16 How the 4 most commonly used self-report anxiety tools were used in the studies

	BAI (in 15 studies)	GAD-7 (in 12 studies)	DASS-21 (in 11 studies)	DASS-42 (in 12 studies)
Prevalence studies	8 (53% of BAI studies)	6 (50% of GAD7 studies)	5 (45% of DASS21 studies)	2 (17% of DASS42 studies)
Psychometric properties	3 (20% of BAI studies)	3 (25% of GAD7 studies)	4 (36% of DASS21 studies)	5 (42% of DASS42 studies)
Comparing groups	1 (7% of BAI studies)	1 (8% of GAD7 studies)	1 (9% of DASS21 studies)	1 (8% of DASS42 studies)
Screening evaluation	0 (0% of BAI studies)	1 (8% of GAD7 studies)	0 (0% of DASS21 studies)	0 (0% of DASS42 studies)
Comparing tools	2 (13% of BAI studies)	0 (0% of GAD7 studies)	0 (0% of DASS21 studies)	1 (8% of DASS42 studies)
Different languages	0 (0% of BAI studies)	0 (0% of GAD7 studies)	0 (0% of DASS21 studies)	2 (17% of DASS42 studies)

The BAI, GAD-7 and DASS-21 were most frequently used in prevalence studies, in contrast the DASS-42 was most frequently used in studies reviewing the psychometric properties of the tool.

2.4.3.2.2 Clinician administered scales or inventories

Two clinician administered anxiety scales were identified in the literature, these were the Hamilton anxiety rating scale (HARS), and the Goldberg anxiety scale (GAS) The oldest of these was the HARS developed in 1959 and the most recently developed was the GAS in 1987. The HARS takes around 10-15 minutes to complete, no information was found for the administration time of the GAS. The number of items included in the scales range were 9 for the GAS, and 14 for the HARS. No information was found on the recall periods of the HARS,

the GAS asks about symptoms during the past month. Both of the clinician-administered scales are available in the public domain.

2.4.3.2.3 Clinical Interview

Four clinical interview tools were identified in the literature for use in diagnosing anxiety these being the Composite International Diagnostic Interview (CIDI), the Clinical Interview Schedule (CIS-R), the Mini International Neuropsychiatric Interview, (MINI) and the Structured Clinical Interview (SCID). All the interviews were developed in the 1990s with the earliest being the CIDI in 1990 and the most recent the MINI in 1998. The CIDI contains 276 items but has screening questions to determine which parts of the interview to use. The CIS-R contains questions asking about 14 symptom groups and the MINI included 19 modules that explore 17 disorders. The administration time for the different clinical interviews varied from 15 minutes for the CIDI up to 60 minutes for the ADIS. All of the identified interviews are copyrighted and require some permissions and or payment for their use. Training is required for all the interviews, but the training and administration can be completed by non-clinicians.

In summary 27 different tools were identified in the literature for use in determining presence or severity of generalised anxiety. 23 of these were self-report tools, 3 were clinician administered scales and 4 were clinical interviews. The anxiety tools identified varied on a number of aspects. Differences occur not only between the different categories of tool (eg self-report versus clinical interview) but there are also differences in some of the characteristics seen within categories, such as the mental health condition for which the tools were developed.

2.4.3.3 Aim/Review question 3: Is there a difference between which tools for measuring depression are used in studies with students, medical students and the general populations

The 27 anxiety tools identified in the literature have been developed and validated for use with different populations. This review searched for papers which had included the general

population, students and/or medical students as keywords. Out of the 83 studies identified in this review which used anxiety measures, 26 studies used a general population sample, 42 studies used a student sample and 16 studies used a medical student sample. Due to the low number of studies using clinician administered scales, clinical interviews and other tools, this section will concentrate only on the self-report scales.

The 26 studies which included a general population sample used 18 different self-report tools. The 42 studies which included a student sample used 20 different self-report tools. The 16 studies which included a medical student sample used 9 different tools. Therefore, in the studies included in the review, fewer tools were used with medical students than with the general population or other students.

The studies identified in this review were split into categories based on the participants which took part and the percentage of studies using the different tools in each category was calculated. The following table highlights for each category the percentage of studies using the most commonly used tools (as identified in section 2.4.1 above).

Table 17 the percentage of studies in each participant category using the most commonly used tools

	Percentage of all General Population studies identified which used this tool (n=26)	Percentage of all student studies identified which used this tool (n=42)	Percentage of all medical student studies identified which used this tool (n-16)
BAI	0%	26%	31%
GAD-7	27%	10%	6%
DASS-42	9%	19%	13%
DASS-21	4%	19%	13%
HADS	8%	10%	19%

Table 17 highlights the difference between the tools used with general population samples and those with student and medical student samples. 27% (n=7) of studies which involved general population samples used the GAD-7 scales, no studies with a general population sample used the BAI. In contrast the BAI was the most commonly used tool in studies which involved students (11 studies – 26%) and medical students (5 Studies – 31%). HADS was used in 2 general population studies (8%), 4 student studies (10%) and 3 medical student studies (19%).

How the tools were used in the studies with the different populations was also evaluated, results are shown in table 18.

Table 18 – Uses of the tools in the studies with different populations

	General population (26 studies)	Students (42 studies)	Medical students (16 studies)
Prevalence studies	11 (42% of general population studies)	18 (43% of student studies)	12 (75% of medical student studies)
Psychometric properties	6 (23% of general population studies)	11 (26% of student studies)	0 (0% of medical student studies)
Comparing groups of participants	1 (4% of general population studies)	3 (7% of student studies)	2 (13% of medical student studies)
Screening evaluation	0 (0% of general population studies)	2 (5% of student studies)	2 (13% of medical student studies)
Comparing tools	1 (4% of general population studies)	1 (2% of student studies)	1 (6% of medical student studies))
Different languages	2 (8% of general population studies)	2 (5% of student studies)	0 (0% of medical student studies)

For all study populations the most common use of self-report tools was to measure the prevalence of anxiety.

2.4.3.4 Summary of the anxiety review

27 different tools were identified in the literature for use in determining presence or severity of anxiety itself or symptoms of anxiety. 23 self-report anxiety tools, 3 clinician administered scales and 4 clinical interviews were identified for measuring anxiety. Within the 83 studies

which measured anxiety in the current review, 15 used the BAI, 12 the GAD-7, 12 the DASS-42 and 11 the DASS-21.

The 23 self-report anxiety tools differed in terms of their purpose and usage. 19 of the tools were developed to measure anxiety, however 2 of the tools were developed as distress scales (K6 & K10) and 2 as worry scales (PSWQ, WDQ). In 40 studies the anxiety self-report tools were used to measure prevalence of anxiety with 17 studies evaluating the psychometric properties of some of the anxiety self-report tools. Time to complete the scales varied from 2 minutes to 15-20 minutes and number of items varied from 2 to 90. As with the depression review the recall period varied greatly for the different tools from the past week to the past month. 12 of the anxiety self-report tools are available in the public domain and 6 require permissions and/or charge for use.

Three clinician administered anxiety scales were identified in the literature, (HARS, GAS & BAS). The BAS and HARS both take around 10-15 minutes to complete, and number of items included in the scales ranged from 9 to 14. No information was found on the recall periods of the BAS and HARS, the GAS asks about symptoms during the past month. All three of the clinician-administered scales are available in the public domain.

Four clinical interview tools were identified in the literature for use in diagnosing (CIDI, CIS-R, MINI & SCID). The CIDI contains 276 items but has screening questions, the CIS-R contains questions asking about 14 symptom groups and the MINI 17 disorders. The administration time for the different clinical interviews varied from 15 minutes up to 60 minutes. All of the identified interviews are copyrighted and require some permissions and or payment for their use. Training is required for carrying out all the interviews however the training and administration can be carried out by non-clinicians.

Some differences were observed between the tools used in studies with general population samples and those with student or medical student samples; The GAD-7 was more commonly used with general population samples and the BAI with students and medical students. With each of the different population groups, the tools were most frequently used to measure the prevalence of anxiety.

2.5 Discussion

This review asked three questions. These were:

1. What tools have been used for measuring depression and anxiety in studies with students, medical students and the general population in the last 10 years?
2. How do these tools identified differ?
3. Is there a difference between which tools are used for measuring depression and anxiety in studies with students, medical students and the general population samples?

Three databases were searched and studies which involved the measurement of depression and anxiety with general population, students and medical student samples were identified. 299 studies were identified, 261 of which included the use of depression measuring tools and 83 included the use of generalised anxiety measuring tools, 45 studies included the measurement of both depression and anxiety. The tools used in these studies were identified and their characteristics noted.

The results found in the review will now be summarised and discussed in relation to the review questions above

2.5.1 Review question 1: What tools have been used for measuring depression and anxiety in studies with students, medical students and the general population in the last 10 years?

2.5.1.1 Summary of findings

54 different tools for use in determining presence or severity of depression itself or symptoms of depression were identified in the literature. The most common type of tool used in the studies identified were self-report tools with 45 different self-report tools being identified. In addition, 3 clinically administered scales were identified along with 5 clinical interviews and

one alternative tool. Of the 261 studies which measured depression found in the current review, 64 used the BDI, 50 the PHQ, 36 the CES-D and 24 HADS.

27 different tools were identified in the literature for use in determining presence or severity of anxiety itself or symptoms of anxiety. As with the depression review the most common type of tool used in the studies were self-report tools with 23 self-report anxiety tools identified. 3 clinician administered scales for anxiety were identified along with 5 clinical interviews. Within the 83 studies which measured anxiety in the current review, 15 used the BAI, 12 the GAD-7, 12 the DASS-42 and 11 the DASS-21. HADS was the 5th most commonly used tool to measure anxiety in the studies identified, having been used in 9 studies.

2.5.1.2 Discussion of findings

This review has identified a number of different depression and anxiety tools. The main difference seen between the depression and anxiety tools available is the greater number of self-report depression tools compared to generalised anxiety self-report tools. 45 depression self-report tools were identified compared to 23 anxiety self-report tools. 7 tools were identified that were used to measure both depression and anxiety symptoms and, 3 other sets of scales were identified which whilst not combined scales were developed by the same author or following similar principles, and are often considered in conjunction(eg BDI/BAI, OASIS/ODSIS, SDS/SAS. Despite the greater number of self-report tools, the number of clinician-administered tools and clinical interviews identified were similar for both depression and anxiety (3 clinician-administered tools were identified for both depression and anxiety; 5 clinical interviews were identified for depression and 4 for anxiety).

One reason for the differences may be due to this review only searching for anxiety tools which measure generalised anxiety, rather than tools for measuring specific types of anxiety disorder, such as those looking at panic disorder or social anxiety or those tools measuring state anxiety. Anxiety is a multifaceted condition which can include a range of different anxiety disorders with include generalised anxiety, social anxiety, phobias and panic disorders. In addition, anxiety can relate to different situations and tools have been developed for purposes such as test anxiety, for example Spielberger's Test Anxiety Inventory (TAI) (Spielberger, 2010). Whilst there are some different types of depression such as post-natal

depression for which tools have been developed (for example the Edinburgh Postnatal Depression Scale (Cox et al, 1987)), the main condition is not made up of as many subtypes as is the case with anxiety.

Differences were also seen in the number of self-report scales compared to clinician-administered and clinical interviews. For depression 45 self-report tools were identified compared to 3 clinician administered scales and 5 clinical interviews. For anxiety 23 self-report tools were identified compared to 3 clinician administered scales and 4 clinical interviews

It is unclear from the literature why there are so many self-report rating scales for both depression and anxiety. Clinical interviews are seen by many as the gold standard of assessment due to their thorough assessment and relatively positive psychometric evaluations (Gelaye et al, 2014). Therefore, it may be that these tools are accepted, particularly in the research community and as such are not seen as requiring significant improvement. This may not be the case with self-report tools. Researchers may be trying to find or create 'better' self-report tools which have higher predictive values, which are quicker and easier to complete, and which could be used for diagnosis. In 1960 when developing his depression scale Hamilton wrote

"The appearance of yet another rating scale for measuring symptoms of mental disorder may seem unnecessary, since there are so many already in existence and many of them have been extensively used. Unfortunately, it cannot be said that perfection has been achieved, and indeed, there is considerable room for improvement" (Hamilton, 1960 p56).

Another reason for the increase use of self-report tools in the literature may be that they are easier and cheaper to use. The use of tools which require clinician or researchers time are more time consuming and costly as such they may be a less popular option in research studies.

2.5.1.3 Summary

In summary, a large number of tools were identified in the literature for measuring both depression and anxiety. More depression tools were identified and for both depression and anxiety many more self-report tools were identified in the literature. Reasons for these are unclear but may be due to the current review only including studies looking at generalised

anxiety, rather than other types of anxiety. The large number of self-report tools which have been developed may be due to their ease of use, that they are cheaper to use, or due to researchers trying to improve on reliability and validity of these types of tool.

2.5.2 Review question 2: How do these identified tools differ?

2.5.2.1 Summary of findings

2.5.2.1.1. Depression

54 different tools for use in determining presence or severity of depression itself or symptoms of depression were identified in the literature. 45 different self-report tools were identified. These 45 tools varied in relation to a number of characteristics. Firstly, the different self-report tool varied in both the purpose for which they were developed and the way in which they are used. The majority of the self-report tools were developed for measuring depression, however 4 were originally developed for measuring distress (K6, K10), general mental health (SF-12) and wellbeing (WHO-5). In addition, 7 tools were identified that were used to measure both depression and anxiety symptoms within the same tool these included BSI-18, DASS21/42, HADS, K10, K6 and MASQ. In addition, 3 other sets of scales were identified which whilst they are not combined scales were developed by the same author or following similar principles, and are often considered in conjunction. These were the BDI and the BAI developed by Beck, the OASIS and the ODSIS (the ODSIS was adapted from the OASIS) and the SDS and SAS Zung scales.

In relation to the use of the tools, in 121 studies the tools were used for measuring the prevalence of depression in different populations, compared to 49 studies reviewing the psychometric properties of the tool and 22 studies evaluating the effectiveness of tools for screening. The BDI and PHQ were most commonly used for measuring prevalence and the CES-D and HADS more commonly used in studies reviewing the psychometric properties of the tools. The self-report tools differed in the number of items they include (from 2 items (PHQ-2) to 90 items (SCL-90)). The SCL-90 has the longest reported completion length at 12-15 minute. The recall period for which symptoms are considered varies from 7 days to 4 weeks.

24 of the self-report tools are available in the public domain, 11 require payment or for specific requirements to be met for their use.

The 3 clinician administered scales which were identified for measuring depression also varied on the recoded characteristics. The clinician administered scale take longer to complete than the majority of the self-report scales at around 20 minutes. The clinician administered tools had between 10 and 17 items. 2 of the clinician administered tools are in the public domain and one is copyrighted.

The 5 clinical interviews for diagnosing depression were identified in the literature. All of the clinical interviews were developed in the 1990s and have a large number of questions/items. 3 of the clinical interviews have screening questions to determine which parts of the interview is required. The clinical interviews take the longest time to complete of all the tools with the quickest being the PRIME-MD interview taking about 11 minutes. The longest interview to administer is the SCID which can take up to 60 minutes.

One alternative tool was identified for measuring depression which was the DEPi a structured version of the Rorschach inkblot test. Limited information was available for this test in relation to the characteristics being compared. Its use is controversial, and authors have suggested its use should be treated with caution.

2.5.2.1.2. Anxiety

27 different tools were identified in the literature for use in determining presence or severity of anxiety itself or symptoms of anxiety. As with the depression review the most common type of tool used in the studies were self-report tools with 23 self-report anxiety tools identified.

The 23 self-report anxiety tools which were identified from the studies included in this review differed in terms of their purpose and usage. 19 of the tools were developed to measure anxiety, however 2 of the tools were developed as distress scales (K6 & K10) and 2 as worry scales (PSWQ, WDQ). These 4 tools however have since been used to measure general levels of anxiety in studies. In 40 studies the tools were used to measure the prevalence of anxiety compared to 17 studies reviewing the psychometric properties of tools. The BAI, GAD-7 and

DASS-21 were all most frequently used to measure prevalence in studies whereas DASS-42 was most commonly used in studies reviewing it's psychometric properties.

Less information was available on the length of time to complete the anxiety self-report scales, those with a suggested time ranged from 2 minutes to 15-20 minutes. The SCAT self-report tool contains the greatest number of items, at 90 items which may explain why it takes longer to complete (15-20 minutes) than the K6 which contains only 6 items and takes less than 2 minutes to complete. The tool with the fewest number of items was the GAD-2 (part of the PHQ-4) which contains only 2 items. As with the depression review the recall period varied greatly for the different tools from the past week to the past month. The DSM-IV and ICD-10 classification of generalised anxiety requires symptoms to have been present for at least 6 months (American Psychiatric Association, 2013; World Health Organisation, 1992). Only 1 tool asked about symptoms going back that far. There may be a number of potential reasons for this difference. Many of these tools were developed for screening for anxiety and not diagnosing generalised anxiety disorder. In the case of screening for the purposes of providing help, it may not be necessary to wait for 6 months for a diagnosis of generalised anxiety disorder prior to starting support. 14 of the anxiety self-report tools are available in the public domain and 9 require permissions and/or charge for use.

Three clinician administered anxiety scales were identified in the literature, (HARS, GAS & BAS). The BAS and HARS both take around 10-15 minutes to complete, no information was found for the administration time of the GAS. The number of items included in the scales range were 9 for the GAS, 10 for the BAS and 14 for the HARS. No information was found on the recall periods of the BAS and HARS, the GAS asks about symptoms during the past month. All three of the clinician-administered scales are available in the public domain.

Four clinical interview tools were identified in the literature for use in diagnosing anxiety these being the Composite International Diagnostic Interview (CIDI), the Clinical Interview Schedule (CIS-R), the Anxiety Disorders Interview Schedule (ADIS) and the Structured Clinical Interview (SCID). Limited information was found regarding the number of items included in the ADIS or the SCID. The CIDI contains 276 items but has screening questions to determine which parts of the interview to use. The CIS-R contains questions asking about 14 symptom groups. The administration time for the different clinical interviews varied from 15 minutes for the CIDI up to 60 minutes for the ADIS. All of the identified interviews are copyrighted and require some permissions and or payment for their use. Training is required for all the interviews however

all can be carried out by individuals who are not trained clinicians but who have received training.

2.5.2.2 Discussion of findings

These findings highlight the wide variation in the tools available for measuring depression and anxiety. The tools vary according to a number of factors including whether they are self-report, clinician administered or clinical interviews. In addition, the measures varied according to length of time they take to administer, number of items included, recall period, whether they are copyrighted (so need to be purchased) and whether cut-off scores are provided for screening purposes. When determining which tool to use, clinicians or researchers will need to consider these differences and make a decision based on their individual need.

A number of discussion points arise from these differences and these will be considered below. These are

- Criticisms of combined depression and anxiety scales
- How tools are used
- Time and cost versus quality of information gathered
- Recall periods used

2.5.2.2.1 Criticisms of combined depression and anxiety scales

7 tools were identified that were used to measure both depression and anxiety symptoms within the same tool these included BSI-18, DASS21/42, HADS, K10, K6 and MASQ. In addition, 3 other sets of scales which are often considered in conjunction these being BDI/BAI, OASIS/ODSIS, SDS/SAS. It may be considered that the use of these combined tools allow for quick and easy screening of both depression and anxiety (and stress in the case of the DASS scales) at the same time. In addition, some of these combined tools (eg BAI, BDI) were developed to allow clinicians and researchers to measure one disorder irrespective of the other. However, as discussed previously in the case of the BAI this has possibly led to some suggesting that the BAI more closely measures panic disorder rather than anxiety. The combining of depression and anxiety measurement has also contributed to the criticism and evaluation of the factor structure of some of these tools, for example HADS (see section

2.4.2.1.8). However, the factor structure of other independent depression and anxiety scales have also been questioned/evaluated (PHQ-9 – see section 2.4.2.1.6). In addition, some have suggested that the correlations between depression and anxiety scales might be linked to underlying constructs and characteristics of the instruments or may be due to the overlap of symptoms and comorbidity of these conditions (Wang & Gorenstein, 2013).

2.5.2.2.2 How tools are used

The most common use of the self-report tools identified in the studies included in this review was to measure the prevalence of depression or anxiety. 121 studies reported on the prevalence of depression in studies with general population, student or medical student samples.

The time frame chosen for the current review may have had an impact upon the nature of the studies in which scales were included. Only papers from 2007 onwards were included to allow the review to consider which tools are being used in current research, rather than considering all tools which have been developed but which may not still be in use. The tools included in the study were developed between 1959 (Hamilton anxiety rating scale) and 2017 (Leuven Affect and Pleasure Scale). It is likely that scales included in studies in the current review, which were developed many years ago will be included in studies of a different nature to those which have been developed more recently. Scales developed many years ago may have already been included in lots of studies prior to 2007 in which their psychometric properties are tested. If their psychometric rigour has already been extensively tested prior to 2007, then these scales may be more likely to be used for other purposes in more recent studies. For example, these scale maybe used for researching prevalence of depression or anxiety in different populations. Alternatively, these well tested studies, may be used as a standard for comparing newly developed tools against.

In the depression review, the BDI and PHQ-9 were most commonly used to measure prevalence of depression whereas the CES-D and HADS where more frequently used in studies which reviewed their psychometric properties. These differences may reflect the weight of studies supporting the use of the BDI and PHQ-9. Many studies have validated these tools for screening depression in different populations (Smarr & Keefer, 2011). In contrast HADS and

the CES-D have some support for their use as screening tools (Vilagut et al 2016, Mykletun et al, 2001). However, their use has received more criticism than the BDI and PHQ-9 (Coyne & van Sonderen, 2012 ; Zakrewska 2012). Therefore, the BDI or PHQ-9 may be considered a more reliable choice than HADS or the CES-D as a tool for screening for, or measuring prevalence of, depression. In addition, researchers may continue to test the psychometric properties of HADS and the CES-D to provide more support for their continued use.

The tools used in the anxiety review were also most commonly used to measure the prevalence of anxiety with 40 studies reporting on the prevalence of anxiety in studies with general population, student or medical student samples. The BAI, GAD-7 and DASS-21 were most commonly used to measure prevalence, whereas the DASS-42 was more commonly used in studies investigating the psychometric properties of the tool. The less frequent use of the DASS-42 in prevalence studies may be explained by the more common use of the DASS-21. The DASS-21 has less items and is therefore quicker to administer than the DASS-42.

Whilst the tools mostly commonly used in the studies included this review (BDI, PHQ-9, CES-D, HADS, BAI, GAD-7, DASS-21, DASS-42) have all been suggested to be valid for screening depression or anxiety (Smarr, 2011; Julian, 2011; Thombs, 2016; Nilges & Essau, 2015), they were not developed as diagnostic tools (Smarr, 2011; Wang & Gorenstein, 2013). For a true diagnosis, individuals identified by screening tools should be followed up with another method of diagnosis, for example a clinical interview (Thombs & Zieggelstein, 2014). This has implications for studies measuring prevalence. Many studies do not confirm diagnoses with a clinical interview but rather report prevalences based on the findings of self-report tools. Levis et al (2019) suggest that this use of screening tests in prevalence studies is problematic. Levis et al state:

"Theoretically, based on sensitivity and specificity estimates, screening tools would be expected to exaggerate prevalence compared to rates based on diagnostic criteria, although the degree to which one would expect this to be the case would depend on the specific screening tool and cut-off used. Because the false positive rate of screening tools is disproportionately high in lower prevalence populations, such as primary health care, estimated prevalence based on screening tools would be expected to be exaggerated most when true prevalence is lowest"

Therefore, whilst prevalence studies provide an indication of possible numbers of individuals within a population who may suffer from a condition, it is likely that true numbers may be lower.

2.5.2.2.3 Time and cost versus quality of information

In many cases administration time and cost must be weighed against the quality of the information gained from the tool. This affects not only the choice of type of tool (interview, clinician-administered or self-report) but also which individual tool might be chosen. For example, the clinician administered HAMD has long been seen as the standard for clinical trials (Cusin et al, 2009), however, its use, (and the use of clinician rated scales in general) in routine clinical practice is seen as costly and requiring additional clinician time (Uher et al, 2012). The main argument in the literature relating to clinician administered scales relates to whether the time taken for their administration provides greater benefits in terms of their use and ability to diagnose or determine the severity of depressive symptoms. Cusin et al (2009) claim that there is disagreement in research as to the concordance rates of the two methods (Clinician rated and self-rated), concluding that the clinicians and patients rate symptoms differently. Uher et al (2012) also conclude that clinician rated, and self-rated scales are not equivalent but suggest that each provide information that is relevant to clinical prognosis.

In terms of self-report questionnaires, it is argued that in general practice, during appointments which last on average 9 minutes (Irving et al, 2017), unless key 'sign-post' symptoms of depression are mentioned by patients, GP may have a low suspicion of depression (Michell & Coyne, 2007). If depression is suspected a screening tool may be used, however, even a tool which takes 2 minutes may be considered too lengthy in a 9-minute appointment. This has led to some authors endorsing scales with low item numbers for initial screening purposes in general practice (Mitchell & Coyne, 2007).

Research studies, particularly those paying participants, may not be so concerned with the amount of time participants have to spend completing a questionnaire and may be more concerned with gaining rich data for analysis. Clinical interviews are often seen as the gold standard to which other tools (eg self-rating scales) are compared (Gelaye et al, 2014). However, there is limited research surrounding their use in clinical practice. In one small study (with only 40 respondents) it was found that the majority of psychiatrists (72.5%) did not use

structured interviews (Aboraya, 2009). Of the 27.5% that did use structured interviews the MINI or SCAN were the most used (Aboraya, 2009). The time factor involved in their use was one reason why clinical interviews were not used. Other reasons included they are designed as research tools, they interfere with establishing rapport with the patients and psychiatrists feel they do not need to use any structured interviews to diagnose or manage patients. (Aboraya, 2009). One US study suggests that clinical interviews using SCID resulted in more accurate diagnoses for new patients and that these could be carried out by trained nurses to help manage time and budget constraints (Kashner et al, 2003).

Linked to the issues of time a measure takes to administer, is the cost of administration. As suggested previously, measures which take longer to administer are considered most costly. In addition to the administration time cost, users must consider the potential cost of purchasing of some of the measures identified. Whilst some of the tools which require payment, such as the BDI, have been found to have high scores for reliability and validity (Jackson-Koko, 2016), other tools which are in the public domain have been found to be equivalent (Smarr & Keefer, 2011), The cost of the BDI has led to some questioning if the cost of the BDI is "prohibitive given less expensive public domain assessments are readily available" (Smarr & Keefer, 2011, page S136). However, the cost of the BDI does not appear to have restricted the research interest in it, of the 261 depression related studies found in the current review, 64 used the BDI.

2.5.2.2.4 Recall periods used

Another factor, which varies for the different measures is the recall period during which symptoms are considered. The DSM-IV and ICD-10 classification of depression require symptoms to have been present for at least 2 weeks (APA, 2013; WHO, 1992), however the recall period, the time period which individuals are asked to consider their symptoms over, varied for the depression measures from 'the past few days' to the past 30 days. In terms of generalised anxiety, the DSM-IV and ICD-10 classification of depression require symptoms to have been present for at least 6 months (American Psychiatric Association, 2013; World Health Organisation, 1992). The recall period of the anxiety tools identified varied from the past week to the past 6 months. These miss-matches are often due to how the tools are developed, for example HADS, as it is assessing both depression and anxiety, asks for both subscale for

individuals to consider how they have been feeling over the last week. Some of the tools were developed for other purposes (eg K6/K10) but have since been used to measure depression and as such the recall period may not be in line with the classifications for diagnosis. In addition, many of the tools were developed for screening and not for diagnosing depression or anxiety. The time frame for recall of symptoms, particularly for the diagnosing of anxiety is under debate in the wider literature. Many claim a duration of one month should be sufficient for the diagnosis of generalised anxiety (Angst et al, 2006; Lee at al, 2008). Another reason why tool developers may not use longer recall periods may be due to recall bias. Recall bias is defined as "a systematic error that occurs when participants do not remember previous events or experiences accurately or omit details" (Spencer et al, 2017). The over-estimating of past emotions is also referred to as the memory-experience gap (Urban, 2018). It is thought that the over-estimation of negative emotions tend to be more pronounced than positive emotions (Urban, 2018). Recall bias is greater when recalling events over a long time period (Spencer et al, 2017). Clinical interviews allow participants to consider their symptoms and provide prompt questions to try and reduce recall bias, this is not as easy with self-report scales. This may be why self-report scales only ask about symptoms up to a month previously.

2.5.2.3 Summary

In summary, the identified tools vary not only on the administration type but also on a range of characteristics. A number of issues arise out of these differences which have been discussed. These include how the tools are used and in particular the usage of screening tools in prevalence studies. The criticisms of tools which measure both depression and anxiety have been highlighted particularly in relation to the overlap in the factor structures of the depression and anxiety subscales. In addition, the need for balancing the cost and time of administration of the tests against the quality of information gained has been discussed. Finally, aspects relating to the recall period used in the tools have been discussed and reasons for the disparity of some of the recall periods with the time periods required for an ICD-10 diagnosis considered

2.5.3 Review question 3: Is there a difference between which tools are used for measuring depression and anxiety in studies with students, medical students and the general population samples?

2.5.3.1 Summary of findings

2.5.3.1.1 Depression findings

Out of the 261 depression related studies identified in this review 78 studies used a general population sample, 116 studies used a student sample and 74 studies used a medical student sample. The 78 studies which included a general population sample used 28 different self-report tools. The 116 studies which included a student sample used 25 different self-report tools. The studies which included a medical student sample used 15 different tools.

Not only were differences seen in the range of tools used in different samples, differences were also found in the most commonly used tools with the different samples. Whilst the BDI was used in about 1/3 of all studies in both the student (n=35) and medical student (n=25) participant categories only 8% (n=6) of studies with a general population sample used the BDI. The most commonly used tool with a general population sample was the PHQ-9 which was used in 24% (n=19) of studies with a general population sample. HADS was used least in studies with a student sample (5%, n=6) and most in studies with a general population sample (17% n=13).

Studies with students and medical student populations most commonly used the tools to measure the prevalence of depression. The majority (57 studies, 77%) of studies with a medical student sample used self-report tools to measure prevalence of depression compared to only 1 study (1%) reviewing the psychometric properties of self-report tools with medical students. Studies with general population samples most commonly used the tools to evaluate the psychometric properties of the tool.

2.5.3.1.2 Anxiety findings

Out of the 83 anxiety related studies identified in this review 26 studies used a general population sample, 42 studies used a student sample and 16 studies used a medical student sample. As with the depression review there was a difference in the variety of tools used in the studies with these different groups, the 26 studies which included a general population sample used 18 different self-report tools. The 42 studies which included a student sample used 20 different self-report tools. The 16 studies which included a medical student sample used 9 different tools.

Not only were differences seen in the range of tools used in different samples, differences were also found in the most commonly used tools with the different samples. 27% (n=7) of the studies which involved general population samples used the GAD-7 scales, with no studies involving a general population sample using the BAI. In contrast the BAI was the most commonly used tool in studies which involved students (11 studies – 26%) and medical students (5 Studies – 31%). In relation to HADS, HADS was used in 2 general population studies (8%), 4 student studies (10%) and 3 medical student studies (19%).

The most common type of study across all three populations were studies measuring the prevalence of anxiety. This type of study was most common with medical student populations with 75% (12 studies) measuring prevalence of anxiety with medical students.

2.5.3.2 Discussion of findings

These findings suggest that studies with general population samples, use a wider variety of tools relative to the number of studies identified, than students and medical students.

In addition, the most commonly used tools varied between the studies with a general population sample and those with student or medical student samples. The PHQ-9 for depression and GAD-7 were most commonly used for general population samples. Whereas the BDI for depression and BAI for anxiety were most commonly used for student and medical student samples.

While there may be a number of reasons for these differences, it is possible that these differences may be explained by the initial development and subsequent evaluation of these scales. The PHQ-9 was first developed for use in primary care, having been designed to be a self-report version of the Primary Care Evaluation of Mental Disorders (PRIME-MD) (Kroenke,et al, 2001). Since it's development it has been widely used in primary care and medical settings. However, studies have also supported its use in the general population with some suggesting it has high reliability and validity when used in the general population (Kocal event et al, 2013). The GAD-7 was developed by the same authors as the PHQ-9 and was also developed for use in primary care (Spitzer et al, 2006*)*. As with the PHQ-9 the GAD-7 has received support as a measure of anxiety in the general population (Lowe et al, 2008).

In contrast the BDI-II when developed was validated with college students and psychiatric outpatients (Smarrr & Keefer, 2011). The BAI has been extensively tested with students (eg Smith et al, 2007). However, in studies with older populations, some research suggests that the discriminant validity of the BAI may be lower than in younger populations (Julian, 2014). This may explain why it is not used as frequently with general population samples which may contain older individuals than in student samples.

The studies using medical student population included in the current review used similar self-report tools to studies involving other students. Whilst the use of the same tools across studies with different populations allows for more appropriate comparisons of results. Only one study was identified across both the depression and anxiety review which investigated the psychometric properties of these tools with medical students (In contrast to 33 studies which investigated the psychometric properties of these tools with other students). The validation of the BDI and BAI for use with general student populations may suggest that these tools are validated for use with medial students. However, no studies have been identified (in this study, or in wider literature searches) which validated these tools (or any of the self-report) tools against a clinical interview to confirm their use with medical students.

The introduction Section 1.4 highlighted the differences seen in medical students compared to some other students, including differences in the individual (eg personalities), the course (greater isolation, academic demands, ethical conflicts) and the culture in which medical students study (eg role models, career worries). Therefore, it is possible that tools which have been validated for students in general, with optimum cut-off scores identified, may not be appropriate for medical students.

2.5.4 Strengths and weaknesses

This review benefited from a structured literature review methodology to provide a qualitative summary of the literature on depression and anxiety scales used with the general population, students and medical students. This methodology allowed for a wide-reaching review and comparison of the different depression and anxiety measures used in studies over the last 10 years. The review provides an overview of the different features of the tools available to allow for comparison at an administration level. Reviewing both depression and anxiety scales allowed for a comparison of measures across the disorders to identify similarities and differences in the issues raised by the review.

One weaknesses of the current review is that within the scope of this review, an in-depth look at the psychometric properties of the different measures was not possible. This review aimed to provide an overview of a large number of different measures available, rather than an in-depth look at a few. In addition, papers were not reviewed for their quality, but rather the measures used were identified and the basic scope/aims of the studies considered and reported. This review may have benefitted from the use of two researchers shortlisting the articles, to allow for consensus and to ensure that all relevant articles were kept, however this was not possible due to the nature of the project.

Finally, more information regarding the countries in which the studies took place would again provide more details surrounding the use of different tools, for example are different tools used in the US to the UK or to China or India.

2.5.5 Conclusions

The current review has highlighted the vast number of depression and anxiety tools available for use in research and clinical practice. They vary greatly with regards to many aspects of their use and administration. Whilst all of the most commonly used tools were used in studies with students, medical students and the general population, the frequency in which they were used with this different groups varied. The tools used with medical student samples were similar to those used with other students, however no validated studies emerged that validated any of the self-report tools against a clinical interview to confirm their use with medical students. Given the differences highlighted in the introduction between the groups of students and the need for local validation of tools as suggested by Ali et al, (2016), this is still an area which requires further research.

Chapter 3: Method

3.1 Overview

This chapter outlines the research methodology for the present study and explains how this methodology supports the aims of the study. The chapter starts with the rationale for the study, an overview of the method, and an explanation over the choice of tools used. Details of the training received by the author and support provided by other researchers will be outlined. Finally, a detailed description of the methodology will be presented.

3.1.1 Rationale

Chapter two described a structured literature review, looking at what is already known about measuring anxiety and depression in students generally and specifically in medical students.

The review highlighted the large number of self-report scales used in studies compared to clinical interviews. Whilst the use of self-report scales provides benefits in terms of the time and cost taken to administer, clinical interviews are seen as the gold standard for diagnosing depression and anxiety (Anderson et al, 2002). Before screening tools can be used to screen individuals for a mental health condition it is important to evaluate the accuracy of the tool compared to a gold standard interview. If the tool is able to correctly determine those people with a condition compared to those without, the screening tool's use can be advocated.

Many different scales have been used to measure depression and anxiety (separately or together) in students, however, results from different scales and different populations vary. One area where differences may occur, are the cut-off points used to determine caseness within different populations for different tools. Caseness being defined as 'probable presence of the mood disorder'. Some tools have been developed with one population, with norm-based cut-offs advised, but then following validation with another population revised cut-offs have been suggested (Singer et al, 2009).

As outlined in the literature review in chapter 2, HADS is one of the four most commonly used tools for measuring and screening depression and anxiety in students both in the UK and internationally. Some have questioned if HADS and other similar assessments can provide valid approximations of mental health conditions in students (Andrews et al, 2006). Andrews et al validated HADS against clinical interview (using the SCID clinical interview) in a group of university students and concluded that the HADS depression scale was an accurate indicator of depression in this group using the recommended cut-off of 8 and above. However, in this group the HADS anxiety scale overestimated the levels of anxiety (Andrews et al, 2006). Andrews concluded that for the sample tested both depression and anxiety optimum cut off scores differed from the recommended 8. Andrews et al suggested an optimal cut-off of 10 for depression and 11 for anxiety.

HADS has been used with medical students in a number of studies however there are limited studies looking at comparison of HADS against clinical interview in medical students.

To address this, this study aimed to investigate the suitability of HADS as a screening tool for anxiety and depression in a medical student population by comparing HADS to a structured clinical interview (Schedules for Clinical Assessment in Neuropsychiatry) (SCAN). Three research questions were posed:

- Research question 1 – Is HADS an accurate measure of anxiety and depression in medical students?

- Research question 2 – Are the standard cut off scores of 8+ on the HADS-D and HADS-A subscales appropriate for a medical student population?

- Research question 3 – Do the responses to individual items within the HADS subscales truly reflect the presence of anxiety and depression.

3.1.2 Methodology overview

The methodology was to undertake a pilot study, recruiting medical students from all year groups to take part in a face to face meeting where they completed HADS and undertook a clinical interview. HADS was chosen due to it's use in previous projects with medical students and to allow for comparison with the Andrews et al (2006) study. The clinical interviews used the SCAN (Schedules for Clinical Assessment in Neuropsychiatry) interview. Following the face to face meeting a mixed methodology was used for analysis. Firstly, medical students HADS scores were compared to a diagnosis via clinical interview, using quantitative analysis to understand if HADS was able to distinguish between medical students who did have diagnosable levels of depression and/or anxiety and those who do not. In addition, optimum cut off scores could be calculated. Secondly, the interview responses of those students who had either false positive or false negative results on HADS (using the standard cut off scores of 8 and above) in comparison to clinical interview were qualitatively reviewed. This aimed to provide a more in-depth review of possible reasons for disparity between HADS scores and clinical interview.

3.2 Choice of Tools

The current study used two tools; HADS and SCAN. Reasons behind the use of these tools are outlined below along with a description of each tool.

3.2.1 HADS

HADS (see appendix 3) was chosen for use in the current project due to a number of factors. Firstly, the literature review outlined in chapter 2 found HADS was one of the more commonly used tools overall and with medical students. In the review HADS was used in 7 studies measuring depression with a medical student sample and in 4 studies measuring anxiety among medical students. In addition, use of HADS allowed for a comparison with the Andrews et al (2006) study which had compared HADS to a clinical interview diagnosis with a general student sample (see section 3.1.1).

The hospital anxiety and depression scale was developed by Zigmond and Snaith (1983) to identify 'caseness' of anxiety and depressive symptoms in non-psychiatric settings. It is a brief, easy to use screening tool which can be used to detect depression and anxiety separately or together as a measure of psychological distress (Bjelland et al, 2002). HADS is designed for measuring depression and anxiety separately to other medical conditions and as such focuses on non-physical symptoms of depression and anxiety (Stern, 2014).

HADS comprises two subscales HADS-A (anxiety) and HADS-D (depression) each with seven questions. Each question has four possible answers rated from zero to three. Higher scores indicate higher levels of symptoms. The scales were developed to ensure that the scale could distinguish between anxiety and depression (Zigmond and Snaith, 1983). In addition, to prevent symptoms from other conditions such as somatic disorders affecting the depression and anxiety scores, all physical symptoms eg dizziness and headaches were excluded from the scales (Bjelland et al, 2002).

HADS has been used extensively with a range of different populations but has been criticised due to inconsistencies in its factor structure (Coyne & van Sonderen 2012). These inconsistencies and disparities have led to some suggesting that HADS should no longer be used (Coyne & van Sonderen 2012; Zakrewska 2012). However, a comprehensive review of HADS stated that HADS has good levels of reliability and validity and is effective at screening for caseness of anxiety disorders and depression in patients (Bjelland et al, 2002). It claimed that "HADS seems to have at least as good screening properties as similar, but more comprehensive, instruments used for identification of anxiety disorders and depression" (Bjelland et al, 2002). In addition, it was concluded that HADS has the same properties when used in different samples including the general population, general practice, and with psychiatric patients (Bjelland et al, 2002). Since the publication of the original scale many studies have reviewed the cut-off scores which give the greatest sensitivity and specificity for detecting clinical cases of anxiety or depression symptoms (Mitchell et al, 2010). In one review optimal cut offs were concluded to be 8 for each subscale with a total HADS score of 15+ being generally considered to be strongly indicative of psychological distress (Bjelland et al, 2002). However, there may be different optimal cut off scores for caseness in different populations. For example, Kendrick et al (2009) suggested that a cut-off point of ≥9 is appropriate for GPs to use to diagnose depression within the UK.

3.2.2 SCAN

SCAN was chosen for use in the current project as it can be administered by a trained lay person and does not require the purchase of a licence for its use (See literature review section 2.4.2.2.3).

The Schedules for Clinical Assessment in Neuropsychiatry (SCAN) clinical interview technique is a set of instruments and manuals aimed at assessing, measuring and classifying psychopathology and behaviour associated with the major psychiatric disorders in adult life (Wing et al, 1990). SCAN was developed by the World Health Organisation and was a development of the Present State Examination (Wing et al, 1990).

Scan is a semi-structured interview technique which allows the interviewer to explore an individual's symptoms whilst providing a glossary and guidance as to the clinical relevance of the symptom (Janca et al, 1994). The technique allows the interviewer to decide the severity of a symptom and the length of time that symptom has been present (Janca et al, 1994). Scan has been found to have good reliability and validity and has been used as a standard against which the validity of other instruments is tested (Nienhuis et al, 2010). Although SCAN may not be used as frequently as might be expected given its strengths, (Nienhuis et al, 2010), it has been suggested this may be due to the length of time it takes to interview using SCAN. Alternatively it may be because historically it was designed for use by in clinical settings for example by psychiatrists and clinical psychologists (Janca et al, 1994). More recent studies however have suggested that SCAN can give reliable results even when used by less experienced (but well trained) interviewers (Rijnders et al, 2000) as in the current study. In the current study SCAN was used to determine clinical diagnoses of depression and generalised anxiety disorder (GAD)

3.3 Training and support

To assist with carrying out the interviews, Sarah Winstanley, a PhD student based at Centre for Psychosocial Research, Occupational and Physician Health was recruited.

Sarah Winstanley and the author received training from an expert (Dr Liz Forty) in administering the SCAN interview. Dr Liz Forty is a psychologist with extensive experience at using and training in the use of the SCAN instrument. The training included instruction on how to carry out the semi-structured interviews and also in how to interpret the responses given to make a diagnosis. Dr Liz Forty also provided support during the scoring of the interviews and determining diagnoses.

3.4 Ethical approval

The project received approval from Cardiff University, School of Medicine Research Ethics Committee. Permission was granted in April 2014 (MREC 14/19).

3.5 Participant selection

The project was time limited and had financial constraints therefore a maximum of 50 medical students were recruited to take part in the face to face meetings for this project. Whilst it was recognised that to comprehensively test the effectiveness and validity of HADS for a medical student sample would require recruitment of more students this was not possible in the present study. However, as the author was not aware of any similar studies carried out with medical students it was considered that this study could form a useful pilot project.

3.5.1 Inclusion and exclusion criteria

Medical students were recruited using the following inclusion and exclusion criteria

3.5.1.1 Inclusion Criteria

Inclusion criteria for the project were that participants needed to be current medical students studying at Cardiff University, School of Medicine.

3.5.1.2 Exclusion Criteria

Exclusion criteria included:

- Anyone not currently studying at Cardiff University
- Students not studying medicine

3.5.3 Sampling

Calculating the sensitivity and specificity of different HADS scores compared to a clinical diagnosis of depression and or anxiety required the recruitment of medical students with a range of HADS scores. Two methods of sampling were considered to achieve this.

Firstly, consideration was given to sampling students according to their HADS scores prior to interviewing. However, to eliminate any bias within the interviews it was decided that the interviewers should not know the HADS scores until the student had completed the entire face to face meeting. To sample students according to their HADS scores would have required additional staff (who were not undertaking the interviews) and would have required more students to be recruited (as not all students recruited would have taken part in the interviews). Due to time and financial constraints this was not possible.

Secondly, in an attempt to recruit medical students with a range of HADS scores, consideration was given to sampling according to a medical diagnosis of depression and or anxiety. However, it was acknowledged that not all individuals with depression and anxiety will have had a formal diagnosis. Research has highlighted that many medical students and doctors are reluctant to seek support (Hillis et al, 2010). In addition, a previous diagnosis of depression and or anxiety would not guarantee that individuals had higher HADS scores. Those who had a previous diagnosis may also have had some treatment for their depression and or anxiety and thus may no longer be symptomatic. Therefore, it was decided that medical students would not be sampled according to a diagnosis.

Instead different recruitment strategies were considered, with the aim to recruit medical students with a range of HADS scores. This included specifically targeting recruitment of medical students from Medic Support who were likely to be experiencing more mental health

problems than the general student population along with recruitment via general email to all students (see section 3.1.5.2).

3.5.4 Recruitment

Recruitment took place between June 2014 and November 2015, this allowed for recruitment to take into account timetabling, exam and placement constraints. Students were offered a £10 gift voucher for participating as an incentive to assist recruitment. Due to a budget of £500 this allowed for the recruitment of 50 students.

Medical students were recruited via three different sources to try and recruit students with a range of HADS scores including those with and without a history of mental ill health. The details of the recruitment methods are outlined below.

1. At the end of an online questionnaire

Between November 2013 and June 2014 Barts and the London Medical School and School of Medicine, Cardiff University sent out an online questionnaire to all their medical students. The purpose of the questionnaire was to look at prevalence of mental health problems in the medical student populations. The questionnaire included demographic details, questions to assess 'lifetime' prevalence of Common Mental Disorders (CMDs) along with two validated questionnaires; HADS and the Warwick Edinburgh Mental Well-being Scale (WEMWBS) (Tennant, et al, 2007). The questionnaires were sent out at two time points. At the end of the second questionnaire, medical students from School of Medicine, Cardiff University were told that a further study was being carried out looking at comparing self-report and clinical interview diagnosis of depression and anxiety. Students were asked to follow a link if they would like more information or to express an interest in taking part. The link took the students to a separate form which was not linked to their questionnaire data. The form provided some more information and also asked for their email address so that further information could be sent to the provided email address.

2. Via Medic Support at School of Medicine, Cardiff University

Medic Support in Centre for Medical Education, School of Medicine, Cardiff University provides support for medical students who have been identified as having issues relating to performance or health. Students referred to the unit may be experiencing physical and mental health problems and/or performance related issues such as behavioural or attitudinal problems, language or written skills or difficulty with passing exams. During their initial assessment with Medic Support students are asked to sign a consent form asking if they are willing to be contacted in future with information about research projects. Only those students who had consented and were still attending the unit were provided with information about the present study. Those who were interested in receiving more information or in taking part were asked to complete an additional consent form providing contact details to be passed to the researcher. Further details were then sent to the contact details provided.

3. Via an email

All medical students at Cardiff Medical School were invited to take part via email. The email was sent out via the C4ME Medical School office and provided the student with information about the project. Students who were interested in receiving more information or in taking part were asked to contact the researcher directly.

Medical students who responded to the emails were sent a participant information sheet (see appendix 4) and invited to a face to face meeting at a convenient time in a confidential setting at Cardiff University.

3.6 Face to face meeting

The face to face meetings took place between July 2014 and November 2015. The face to face meetings were undertaken by one of two researchers (Naomi Marfell and Sarah Winstanley).

The purpose of the face to face meetings was to allow comparison between diagnoses derived from clinical interview data with self-reported data using the HADS questionnaire. After providing consent, students were asked to first complete the HADS self-report scale before undertaking the SCAN interview. Student's HADS questionnaires were not studied or scored prior to the commencement of the clinical interview. The HADS scale and the clinical interview data were linked using responses to three unique identifier questions.

The unique identifier questions were taken from ones previously used in another study (Reisbig et al, 2007). The students were asked to give the first two letters of the answers to the three questions:

1. What was the name of your primary (or first) school?
2. What is your mother's first name?
3. Name of town where you were born?

This produced a six letter identifier which enabled the matching of the HADS questionnaire and the interview response whilst maintaining anonymity. The six-letter code was stated at the start of audio recording of the interview.

3.6.1 Consent

At the start of the face to face meeting, the researchers reiterated the purpose of the research and the meeting which had been included in the information sent to all medical students who had shown an interest in taking part. Medical students attending the face to face meeting were also told that the interviews would include asking them questions about symptoms relating to depression and anxiety and that they were not obliged to answer all questions. Participants were given the opportunity to withdraw from the study. If the students were happy to proceed, they were asked to sign a consent form (see appendix 5).

3.6.2 Time frames

Consideration was given for the time frames students would need to reflect on their symptoms over in the SCAN interview.

HADS required respondents to consider how they have been feeling over the previous week.

For the clinical interviews data, a clinical diagnosis of depression ICD-10 classification requires that symptoms are present for at least a month.

For a clinical diagnosis of GAD the ICD-10 classification requires that symptoms are present for at least six months. However, this requirement of a period of six months has been debated in the literature with many claiming a duration of one month should be sufficient (Angst et al, 2006; Lee at al, 2008). The National Comorbidity Survey carried out in the US suggested that in terms of onset, persistence, impairment, and comorbidity, patients who experience anxiety symptoms of between 1 and 5 months do not differ greatly from those with symptoms of greater than six months (Kessler et al, 2005). As such it is claimed that there is little evidence for excluding those with symptoms of at least a month from a diagnosis of GAD (Kessler et al, 2005).

It was decided that asking participants to consider their symptoms over the last month for both depression and anxiety would prevent confusion due to switching time frames during the SCAN interview. Especially given that in the present study student's diagnoses were being compared to their HADS scores and HADS required respondents to consider how they have been feeling over the previous week.

3.6.3 Protocol for distressed students

Given the subject matter of the face to face meeting and specifically the SCAN interview it was recognised that students may experience psychological distress whilst talking about their symptoms and past experiences. The researchers took every precaution to ensure students experienced as little psychological distress as possible during the interview. In order to

minimise psychological distress and to fully support students the researchers followed the protocol outlined below.

- The face to face meetings were held in a suitable and confidential room and where possible at Medic Support Unit to ensure further support would be available if needed.
- The face to face meetings were held at a convenient time for the students.
- The interviews lasted no longer than 60 minutes.
- The interviews were carried out by Naomi Marfell and Sarah Winstanley, researchers at the Centre for Psychosocial Research, Occupational and Physician Health (based at the same premises as Medic Support). The researchers received training in carrying out the SCAN interviews.
- Before students agreed to take part in an interview, and at the start of the interview the researchers ensured that the students were fully aware of what the study was about and that they would be asked questions about symptoms relating to depression and anxiety during the interview.
- Students were informed and reminded that they were not committed to answer any questions that they did not feel comfortable answering.
- If there was any sign of distress the interview was paused, and the student asked whether they would like to stop the interview completely.
- At the end of the interview students were provided with a debrief sheet which outlined where they could seek support if they had queries or were distressed about the study. This included information about Medic Support and the student counselling service. Students were also reminded that they should contact their GP if they have any physical or mental health concerns.

If during or following the interview the researcher had serious concerns about the mental health of a student, the following protocol was followed:

- The researcher stopped the interview.
- The researcher informed the student that they were concerned about them.

- The researcher suggested that the student should seek appropriate help. The researcher also suggested they may want to see their GP or contact student counseling service or another psychiatric service if appropriate.
- If necessary, the researcher would support the student in contacting appropriate help.
- If the researcher felt that there was immediate risk of harm, the researcher would accompany the student to an emergency service such as their GP, Accident and Emergency or local psychiatric crisis services (e.g. via Crisis Resolution and Home Treatment Teams). The researcher would ensure that the student is not left on their own and at risk.

3.6.4 Debrief.

At the end of the face to face meeting participants were thanked for taking part and given a debrief sheet (see appendix 6). This sheet thanked the participants for taking part and outlined places where they could seek support if they felt they needed it following the face to face meeting. The Interviewers explained to participants what the sheet included to ensure participants were fully aware of where they could receive support should it be required.

3.7 Data collection

All interviews were carried out face to face. The interviews were audio recorded to allow for scoring of the SCAN interview after the interviews. Interviews were carried out by Naomi Marfell and Sarah Winstanley. Neither interviewer had any prior relationship with any of the participants.

HADS questionnaires were completed via paper and pen and then inputted into excel.

A number of different verification checks are suggested in the literature to ensure that the data is entered accurately. The most popular check for reducing inputting errors, found in the literature is ‘double entry’. This is the process of entering the data twice into two separate databases and comparing the two data sheets, any discrepancies are highlighted, and the

original data referred to and the correct response recorded. Many suggest that double entry can result in a data set with fewer errors compared to single entry data (Reynolds-hartle et al, 1992) and as such this type of verification is considered important with particularly high value data, where a mistake in the data input could affect patient care or safety (Goldberg et al, 2015). Double entry is often criticised for being time consuming, however for the present study, only 50 HADS scales were required to be inputted and as such double entry was achievable.

There was no missing HADS data and therefore no calculations for missing data were required.

3.7.1 Data Storage

The audio recordings of the interviews were stored under the students' unique identifiers on password protected computers. After transfer to the computer, the audio files were deleted from the audio-recorder and the audio-recorder's 'format' function used to ensure that data could not be retrieved. The HADS data was inputted into excel stored under the students' unique identifiers on password protected computers. The original HADS sheets and consent forms were stored separately in a locked cabinet at Centre for Psychosocial Research, Occupational and Physician Health Cardiff University School of Medicine. Names and email addresses used to arrange the interviews, was stored in a password protected file for the duration of the study and for three months afterwards.

3.8 Data analysis

3.8.1 Scoring HADS

Once the raw HADS data had been inputted into excel, student's HADS scores were calculated. Each HADS statement is categorised as either HADS-A for statements which relate to symptoms of anxiety or HADS-D which relate to symptoms of depression. Each statement has 4 possible responses, answers to which are scored from 0 to 3. HADS had both positively and negatively scored items to avoid acquiescence bias (Wouters et al, 2012), where a responder may just agree or disagree with all statements. Therefore, each statement must be carefully scored according to the instructions.

For example, one positively scored (ie the more agreement the higher the score) HADS-A items is:

I feel tense or 'wound up;

Possible responses include:

- Most of the time – this scores 3 points
- A lot of the time – this scores 2 points
- From time to time occasionally – this scores 1 point
- Not at all – this scores 0 points.

An example of a negatively scored (the more agreement, the lower the score) HADS-D items is:

"I still enjoy the things I used to enjoy

Possible responses include:

- Definitely as much – this scores 0 points
- Not quite so much – this scores 1 points
- Only a little – this scores 2 points
- Hardly at all – this scores 3 points.

Once responses to each item have been determined, subscale totals (HADS-A and HADS-D) were calculated by adding responses to all the relevant subscale items.

3.8.2 Scoring SCAN

Following the face to face meeting, the clinical interview data was scored as set out in the SCAN manual. Each symptom is given a score as shown in figure 1.

Figure 1 – scoring for SCAN symptoms as outlined in the SCAN manual

0 This is a positive rating of absence. It does not mean 'not known' or 'uncertain whether present or not'. It can only be used if sufficient information is available to establish its accuracy.

1 This is a positive rating of presence, but presence of such a minor degree that it is not appropriate for use in classification. Like (0) it does not mean 'not known' or 'uncertain'. Rating of (1) count in scores (but not for diagnostic purposes), which in turn influence the level allocated on the Index of Definition.

2 This rating means that the item is present at a level sufficient to use in classification. For this purpose it is equivalent to (3), but it contributes less to scores. In general, it is used when symptoms are of moderate severity during most of the period being assessed.

3 A rating of (3) is similar to (2) except that the symptom is present in severe form for most of the period under review.

5 The presence of psychotic symptoms can make the rating of Part One items very difficult, because of problems in interpreting the meaning of what R says, or because the symptoms (for example, anxiety or a phobia about leaving one's house) may themselves be based in psychotic experiences. The rating should be only made when there is doubt about the nature of the symptom or the balance is in favour of the symptom being psychotic.

8 If, after an adequate examination, the interviewer is still not sure whether a symptom is present (rated 1-3) or absent (rated 0), the rating is (8). This is the only circumstance in which (8) is used. It should not be used to indicate a mild form of the symptom.

9 This rating is only used if the information needed to rate an item is incomplete in some respect, for example because of language or cognitive disorder, or lack of co-operation, or because the interviewer forgot to probe sufficiently deeply. It is distinguished from (8) because the examination was not, for whatever reason, carried out adequately.

For duration ratings, duration less than 1 week/month should be rated 1. A rating of 0 thus means that the phenomenon has been totally absent.

Once all symptoms had been scored as per the SCAN manual, these were used to establish whether each of the symptoms required for an ICD-10 diagnosis of depressive disorder and generalised anxiety disorder were present at clinically significant levels and for the required duration as outlined in the ICD-10 classifications (see figures 2 & 3). Those students who had the required number of symptoms present were considered to have a clinical diagnosis of either depression or generalised anxiety disorder.

SCAN can be used to make clinical diagnoses using the Diagnostic and Statistical Manual of Mental Disorders (currently in its fifth edition, DSM-V) or the International Statistical Classification of Diseases, (currently in its 10th edition, ICD-10) (Janca et a, 1994). DSM was developed by the American Psychological Association (APA, 2009) to offer standard criteria for the classification of mental disorders. ICD was developed by the World Health Organisation (WHO) to improve the diagnosis and classification of mental health disorders (WHO, 1992). When first developed the two classification systems were quite different, however as approaches to diagnosing mental disorders has changed over the years and the two systems have been updates, they have become very similar with some claiming that "there is little justification for maintaining the DSM as a separate diagnostic system from the ICD in the long run" (APA, 2009, p63). The DSM is popular in the US and produces much revenue for the APA (Khoury et al, 2014), whereas the ICD-10 can be produced at much less cost and is used more widely in Europe and in the rest of the world (APA, 2009). Given this and the location of the present study the SCAN clinical interview data was used to derive clinical diagnoses according to ICD-10 (WHO, 1992).

Figure 2 – Symptoms required for an ICD-10 classification of generalised anxiety disorder

Generalised anxiety disorder classification

Prominent tension or worry along with at least 4 of the symptoms listed below at least one of which must be from items 1-4

Category	Symptoms
Autonomic arousal At least 1 of these	1 - Palpitations pounding heart
	2 - Sweating
	3 - Trembling/shaking
	4 - Dry mouth
Chest & abdomen	Difficulty breathing
	Feeling of choking
	Chest pain or discomfort
	Nausea or abdominal distress
Mental state	Dizzy, unsteady, faint, lightheaded
	Derealisation/depersonalisation
	Fear losing control, going crazy
	Fear dying
General	Hot flushes, cold chills
	Numbness or tingling
tension	Muscle tension/aches and pains
	Restlessness & inability to relax
	Feeling keyed up on edge, tense
	Sensation of a lump in the throat
Other non-specific	Exaggerated response to minor surprise
	Difficulty concentrating
	Persistent irritability
	Difficulty getting to sleep

Figure 3: Symptoms required for an ICD-10 classification of depressive disorder

ICD-10 Classification – Depressive disorder

Main symptoms, for diagnosis at least 2 of these symptoms must be present most days, most of the time for at least 2 weeks

- Persistent sadness or low mood
- Loss of interest or pleasure
- Fatigue or low energy

Associated symptoms

- Disturbed sleep
- Poor concentration or indecisiveness
- Low self-confidence or self-esteem
- Poor or increased appetite
- Suicidal thoughts or acts
- Agitation or slowing of movements
- Guilt or self-blame

Diagnostic Criteria:

- Sub clinical - Fewer than 4 symptoms in total
- Mild – 4-5 symptoms (2 of which must include main symptoms below)
- Moderate –6-7 symptoms (2 of which must include main symptoms below)
- Severe – 8+ (2 of which must include main symptoms below)

Once any diagnoses had been determined, student's scores on the anxiety and depression sub-scales of the HADS were compared with clinical diagnosis (ICD-10) of depressive or anxiety disorder obtained via the SCAN clinical interview.

Basic descriptive statistics were calculated along with sensitivity, specificity and positive and negative predictive value calculations.

3.8.3 Inter-rater reliability

To ensure reliability of the interviewer's scoring and diagnoses, a sample of 10 interviews were randomly selected for review by both interviewers and inter-rater reliability was then calculated. Inter-rater reliability was calculated for the main symptoms for depression and anxiety and for the diagnoses.

The symptoms compared are outlined in table 19 below

Table 19 – symptoms compared for IRR

Depression symptoms	Persistent sadness or low mood
	Loss of interest or pleasure
	Fatigue or low energy
Anxiety symptoms	Prominent tension worry
	Palpitations pounding heart
	Sweating
	Trembling/shaking
	Dry mouth

For those where there was a discrepancy diagnoses were determined via discussions between the interviewers and a supervisor (LF) to ensure accurate diagnoses

3.8.4 Sensitivity/Specificity

For any screening test to be of use it needs to be able to identify as accurately as possible the presence or absence of a condition (Trevethan, 2017). When determining the ability of a test to predict the presence or absence of a condition, one method is to compare the screening

test results to a diagnostic test result (Trevethan, 2017). In the present study the screening test – HADS, was compared to the diagnostic test – SCAN, at different cut-off points for HADS. The comparison of these tests produces four possible responses in individuals. These are:

1. True positive (HADS score above cut off and positive diagnosis according to SCAN)
2. False positive (HADS score above cut off but no diagnosis according to SCAN)
3. True negative (HADS score below cut off and no diagnosis according to SCAN)
4. False negative (HADS score below cut off but positive diagnosis according to SCAN)

The following table highlights these different responses.

Table 20: Possible responses when comparing HADS scores to diagnosis using SCAN data

		Diagnosis according to SCAN	
		Diagnosis according to SCAN	**No diagnosis according to SCAN**
Screening via HADS	**HADS above cut off**	**True positive** **a**	**False positive** **b**
	HADS below cut off	**c** **False negative**	**d** **True Negative**

Any screening test aims to have high numbers of true positives and true negatives and aims to have few false positives and false negatives. Sensitivity is the ability of a test to correctly identify positives – reported as a proportion of actual positives that are correctly identified as such (Parikh, et al, 2008). Using the above table sensitivity is calculated as **a/(a+c)*100**

Specificity is the ability of a test to correctly identify negatives, reported as the proportion of negatives that are correctly identified (Parikh, et al, 2008). In the above table specificity is calculated as **d/b+d*100.**

In addition to sensitivity and specificity the negative predictive value (NPV) must also be considered. The NPV estimates the probability that subjects with a negative screening test truly do not have the disease (Maxim et al, 2014). In the above table NPV is calculated as **d/(c+d)*100.**

Finally, positive predictive value (PPV) can also be considered. The PPV estimates the probability that subjects with a positive test result, will truly have the disease. In the above table PPV is calculated as **a/(a+c)*100**

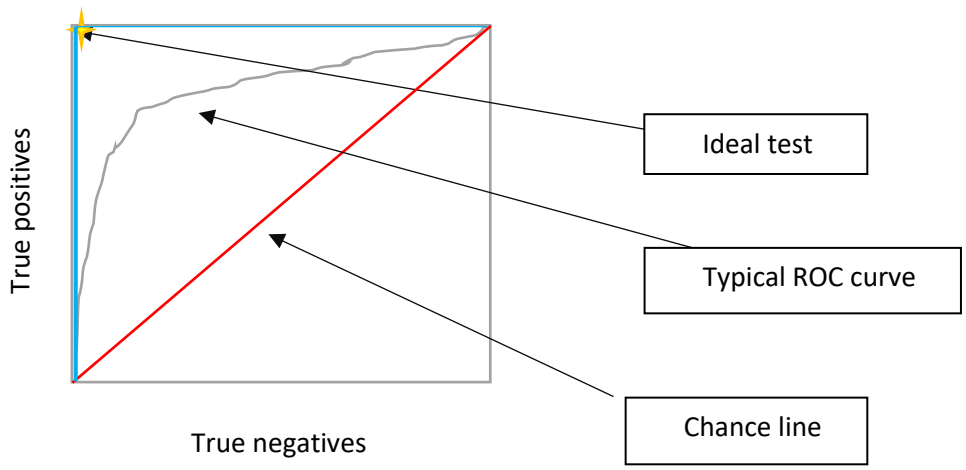
Sensitivity, specificity PPV and NPV were calculated for different cut-off scores.

Any screening tool should have high sensitivity and specificity (Maxim et al 2014). Optimum specificity, sensitivity and NPV levels would all be 100 % (or 1.0) where all cases were identified, there were no false negatives and all negative results were truly negative. This however, is rarely achieved and as such studies need to consider an optimum balance. This is particularly difficult with sensitivity and specificity. As general rule you cannot increase both simultaneously, one will be increased at the expense of the other (Hazibzadeh et al, 2016). This relationship between sensitivity and specificity, of a specific test can be represented graphically as a receiver operating characteristic curve.

3.8.5 Receiver operating characteristic curve (ROC curve)

A ROC curve was produced using the data from the present study. A ROC curve is a graphical representation of the trade-off between sensitivity and specificity (Hazibzadeh et al, 2016). The curve is a plot of true positives (sensitivity) against true negatives (1-specificity). An example of a ROC curve can be seen below in figure 4

Figure 4: Example of a ROC curve



The red line would represent a test in which at each cut off point there were as many false positives as false negatives and as such is a test which is only as good as chance (eg flipping a coin). The blue line would represent a test in which there was 100% accuracy ie when it has any positives there are no negatives and vice versa. In reality, a typical ROC curve falls somewhere between these – an example might be the grey curve.

As such the ROC demonstrates the ability of a test to discriminate between those with a disease and those without a disease (Akobeng, 2007). The curve also allows for determining the optimal cut off point for the screening test (see below).

The use of ROC curves have been used extensively in research. Advantages of using a ROC curve in addition to calculating sensitivity and specificity are its ability to provide graphical representation of the accuracy of all cut-off values. In addition, the ROC curve is not affected by the prevalence of the disease/condition in question. Finally, it allows for comparison of different screening tests (Halligan et al, 2015).

However, some have criticised the use of ROC curves, in particular some have suggested that the data produced is not understood or effectively interpreted by clinicians (Halligan et al, 2015). In addition, it assumes that sensitivity and specificity are of equal importance (Halligan et al, 2015). Sensitivity and specificity are not always of equal importance, for example sensitivity may be more important, particularly if there are serious consequences of not

identifying someone with the disease (Hazibzadeh et al 2016). Alternatively, specificity may be more important if the consequence of treating an individual might itself bring risks (e.g. side effects of treatment). In addition, ROC curves require a broad spectrum of responses on the screening test to ensure the curve is accurate. Despite these criticisms some have concluded that the ROC curve is useful in the early stages of diagnostic test assessment (Halligan et al, 2015). Given this project aimed to test the effectiveness of HADS as a screening tool with a previously untested sample and it aimed to review a range of possible cut-off values it was decided that a ROC curve would be of help in this situation.

There are a number of methods for identifying optimal cut off points using the curve (Akobeng, 2007). Two of the most commonly used are:

- **Identifying the point on the curve closest to the ideal test point (**Also known as the Euclidian's index). As described above a perfect test would have all positives and no negatives or vice versa and as such would only have one threshold which would fall at the o,1 point on the plot (at the star), (0 negative and 1 positive). The optimal cut off point for any other screening test using this method, is to identify the point on the curve closest to this ideal test point.
- **The Youden Index.** This method involves identifying the cut-off point furthest away from the line of chance (ie the red line on the above curve).

Whilst both of these methods provide similar optimal cut off point and receive similar criticism and recommendations for use. It has been suggested that the Youden index provides a slightly more sensitive cut-off threshold (ie a slightly lower cut-off) (Hajian-Tilaki, 2017). Given the sensitive nature of the conditions being tested, it was decided that a test which produced a slightly more sensitive threshold may be of more value and as such the Youden Index method for calculating the optimum cut-off score was used.

In addition to calculating the optimum cut off score the ROC was also used to calculate an area under the curve value (AUC). The AUC is a calculation of how good the test is at distinguishing between patients with disease and those without. It is a single measure which allows for comparison of different tests. AUC values can vary between 0.5 which represents a poor test and 1 which is a perfect test.

3.8.6 Qualitative analysis

Following the quantitative analysis of the data, the interview responses of those students who had either false positive or false negative results on HADS (using the standard cut off scores of 8 and above) in comparison to clinical interview were qualitatively reviewed. This aimed to provide a more in-depth review of possible reasons for disparity between HADS scores using the standard cut off and clinical interview.

3.8.7 False positives

Those students who scored 8 and above on HADS but who did not reach the threshold for clinically significant levels of depression or anxiety as determined by SCAN, were identified as false positives. Responses to the HADS items of these participants were then explored further. In particular, items in which these participants scored particularly highly were identified. Responses to equivalent SCAN questions were reviewed and compared to their HADS responses.

3.8.8 False negatives

Those students who scored 7 and below on HADS but who reached the threshold for clinically significant levels of depression or anxiety as determined by SCAN, were identified as false negatives. Responses to the HADS items of these participants were then explored further. In particular, items in which these participants scored particularly low were identified. Responses to equivalent SCAN questions were reviewed and compared to their HADS responses.

3.9 Summary

In summary, this study set out to investigate the suitability of HADS as a screening tool for anxiety and depression in a medical student population by comparing HADS to a structured clinical interview (Schedules for Clinical Assessment in Neuropsychiatry) (SCAN). The methods used for

this study have been described. Medical students were recruited via a range of methods and invited to attend a face to face meeting. During the face to face meeting students completed HADS and took part in a clinical interview using SCAN. Students HADS scores were calculated and any depression or anxiety diagnoses determined by scoring the SCAN data. Individual students' results from HADS and SCAN were then compared. The results of these comparisons will be outlined in the next chapter.

4. Results

This chapter outlines the results from the study in which medical students were recruited and asked to attend a face to face meeting where they completed HADS and then undertook a SCAN interview to assess clinical levels of anxiety and depression. The face to face meetings took place between July 2014 and November 2015.

The interviews lasted between 6 minutes 15 seconds and 61 minutes 50 seconds, with a mean time of 21 minutes 28 seconds. The students were engaged and willing to discuss their experiences and aspects relating to their mental health. One interview was stopped due to the student being upset but was then restarted at the request of the student, once they had composed themselves. Four students talked about feeling very down and the interviewer checked that they had support in place and ensured they were aware of the support offered by Medic Support. One student was interested in finding out more about Medic Support and was introduced to the Medic Support administration team following the face to face meeting. The interviewers were not worried about the safety of any of the students interviewed.

The demographic information collected from the students will be described to provide an overview of the sample who took part. Results of the inter-rater reliability calculated to ensure reliability of the interviewer's scoring and diagnoses will also be reported. The research questions for the current study will then be answered by reporting firstly on the depression related data and then on the anxiety related data. The analysis of both the depression and anxiety data included two steps.

Firstly, quantitative analysis was carried out comparing the HADS data with SCAN diagnoses. This quantitative analysis included calculating the sensitivity and specificity, negative and positive predictive values. In addition, the optimum subscale cut-offs for this cohort will be determined, by plotting a ROC curve and the area under the curve will be calculated to give an indication of how good the test is at distinguishing between patients with disease and those without. Results from this quantitative analysis will answer research questions 1 and 2.

The second step of analysis involved reviewing the interview responses of those students who had either false positive or false negative results on HADS (using the standard cut off scores of

8 and above) in comparison to their clinical interview results. Those students who scored 8 and above on HADS but who did not reach the threshold for clinically significant levels of depression or anxiety as determined by SCAN, were identified as false positives. Those students who scored 7 and below on HADS but who reached the threshold for clinically significant levels of depression or anxiety as determined by SCAN, were identified as false negatives.

The interview responses of students identified as false positives or natives at the original cut-off were compared to the responses given to related HADS items. This aimed to provide a more in-depth review of possible reasons for disparity between HADS scores (using the standard cut off) and clinical interview diagnoses. This analysis aimed to answer research question 3.

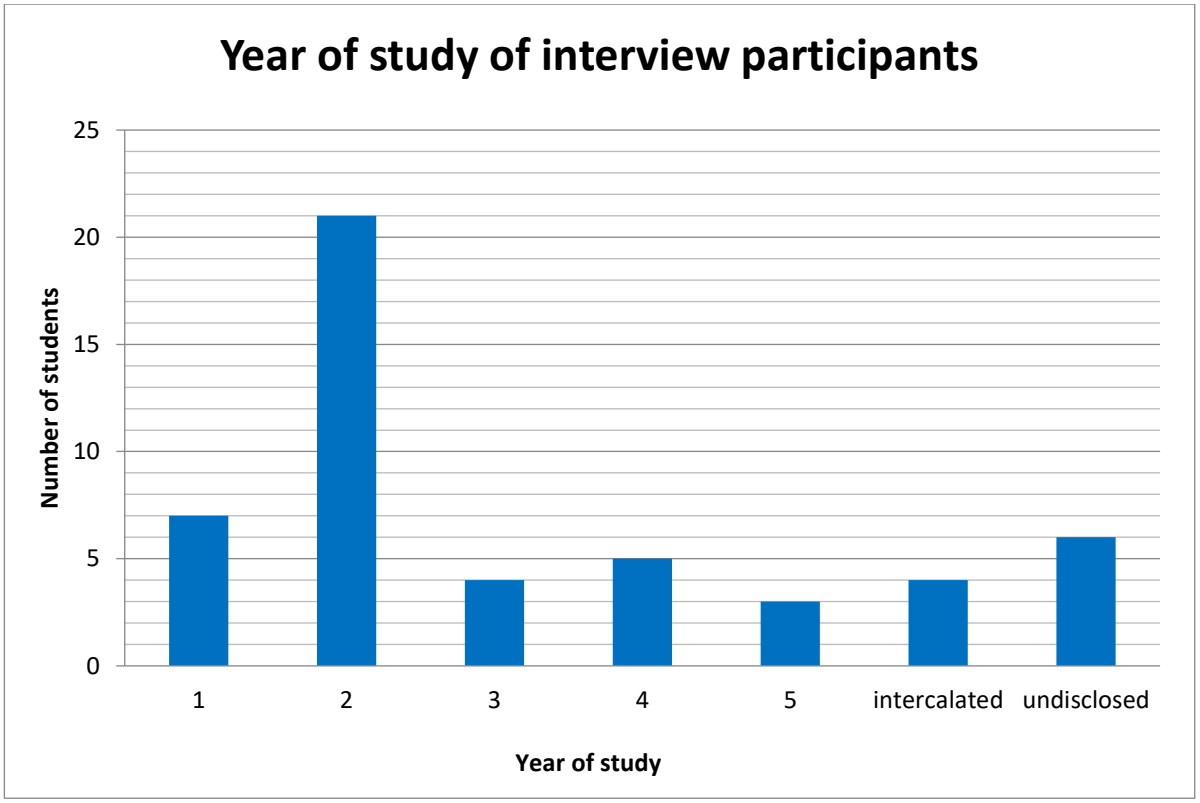
4.1 Demographics

50 medical students from School of Medicine Cardiff University were recruited to the study and took part in a face to face meeting between July 2014 and November 2015. To increase recruitment and respond to concerns regarding the need for anonymity of medical students taking part, minimal demographic details were collected about the students. In addition, these demographic details were not linked to participants' data but were only collected to provide an overview of the sample. The only demographic information reported will be gender and year of study.

Of the 50 students who participated, 7 (14%) were male and 43 (86%) were female. In C4Me in the School of Medicine Cardiff University at the time of the project total student population were 37% male and 63% female.

The following graph shows the years of study of participants.

Figure 5: Distribution of years of study of participants.



42% (21) of students were from year 2. This was believed to be due to many medical students in years 3-5 being away on placement during the year and as such may have found it harder to attend a face to face interview in Cardiff.

4.2 Inter-rater reliability

The interviews used the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) clinical interview technique. From the SCAN interview symptoms were rated as to whether they were not present, present but to not clinical significance or present to a clinically significant level. The SCAN clinical interview ratings were used to establish whether each of the symptoms required for an ICD-10 diagnosis of depressive disorder and generalised anxiety disorder were present.

10 interviews (20%) were randomly sampled from the 50 interviews and independently rated by the two researchers who had carried out the interviews (Naomi Marfell & Sarah Winstanley).

Each symptom was rated either not present (0) or present (1) or unknown (9) or unsure (8). Cohen's Kappa was then calculated for the main symptoms for depression and anxiety to test inter-rater reliability for the symptoms.

The symptoms compared are outlined in table 19 below.

Table 1 – Symptoms used for IRR

Depression symptoms	Persistent sadness or low mood
	Loss of interest or pleasure
	Fatigue or low energy
Anxiety symptoms	Prominent tension worry
	Palpitations pounding heart
	Sweating
	Trembling/shaking
	Dry mouth

For the depression symptoms 30 symptom ratings were compared (3 main symptoms X 10 interviews). The cross tabulation for the comparison of the depression symptoms are shown in table 21.

Table 21 – Cross tabulation of the comparison of ratings of the depression symptoms

		Rating of depression symptoms – Rather 2		Total
		Not present	Present	
Rating of depression symptoms – ratter 1	Not present	25	0	25
	Present	1	4	5
Total		26	4	30

Kappa statistics for the level of agreement between the rating of depression symptoms are shown in table 22.

Table 22 – Kappa statistics for inter-rater reliability of rating of depression and anxiety symptoms

	Kappa Value	Approx. Sig.
Depression symptoms	.765	0.000

For anxiety symptoms 50 symptom ratings were compared (5 main symptoms X 10 interviews). The cross tabulation for the comparison of the anxiety symptoms are shown in table 21.

Table 23 – Cross tabulation of the comparison of ratings of the anxiety symptoms

		Rating of anxiety symptoms – Rater 2		Total
		Not present	Present	
Rating of anxiety symptoms – Rater 1	Not Present	41	1	42
	Present	2	6	8
Total		43	7	50

Kappa statistics for the level of agreement between anxiety rating of symptoms are shown in table 24.

Table 24 – Kappa statistics for inter-rater reliability of rating of anxiety symptoms

	Kappa Value	Approx. Sig.
Anxiety Symptoms	.870	0.000

There was a high level of inter-rater reliability between the two raters with respect to symptom rating.

Cohen's Kappa statistic were also calculated for the overall ICD-10 diagnostic ratings of depressive disorder and generalised anxiety disorder to determine level of agreement

between raters (20 diagnoses in total). Cohen's kappa was significant There was moderate agreement between the two rater's judgements, κ = .643, $p < .0005$. Cross tabulation for this calculation is shown in table 25.

Table 25 - Cross tabulation of the comparison of ratings of the diagnoses

		Diagnosis ratings – Rater 2		Total
		Not present	present	
Diagnosis ratings – Rater 1	Not present	18	0	18
	Present	1	1	2
Total		19	1	20

The remaining interviews were rated by the interviewer who conducted the interview (NM & SW) and the symptoms scored according to the ICD-10 criteria for diagnosis for depression and anxiety.

These diagnoses were then compared to students' HADS depression and anxiety scores. Results relating to depression and anxiety will be presented separately.

4.3 Depression

As outlined above analysis of the depression data involved two steps. The first step involved the quantitative analysis of the HADS and SCAN data. Step two took a qualitative approach to reviewing the interview responses of those students who had either false positive or false negative results when their HADS scores were compared to their clinical interview data. The results of these two steps will be reported in turn.

4.3.1 Quantitative analysis

The quantitative analysis involved firstly identifying those students who reached the diagnostic criteria for depression. This data was then used to compare individual students HADS scores to their diagnosis and then calculate sensitivity and specificity at different HADS cut-offs. The data was also used to create a ROC curve and the optimum cut-off score for this sample identified. Finally, the area under the curve was calculated to give an indication of how good HADS is at distinguishing between medical students with depression and those without.

Using the ICD-10 criteria 3 students reached the diagnostic criteria for depression

Figure 6 shows the distribution of HADS depression scores as completed by students immediately prior to their interviews, broken down by those who following interview were determined to have reached the ICD-10 criteria for depressive disorder and those who did not reach the ICD-10 criteria for depressive disorder.

Figure 6 - Comparison of individual student's HADS-D scores to ICD-10 criteria for depressive disorder

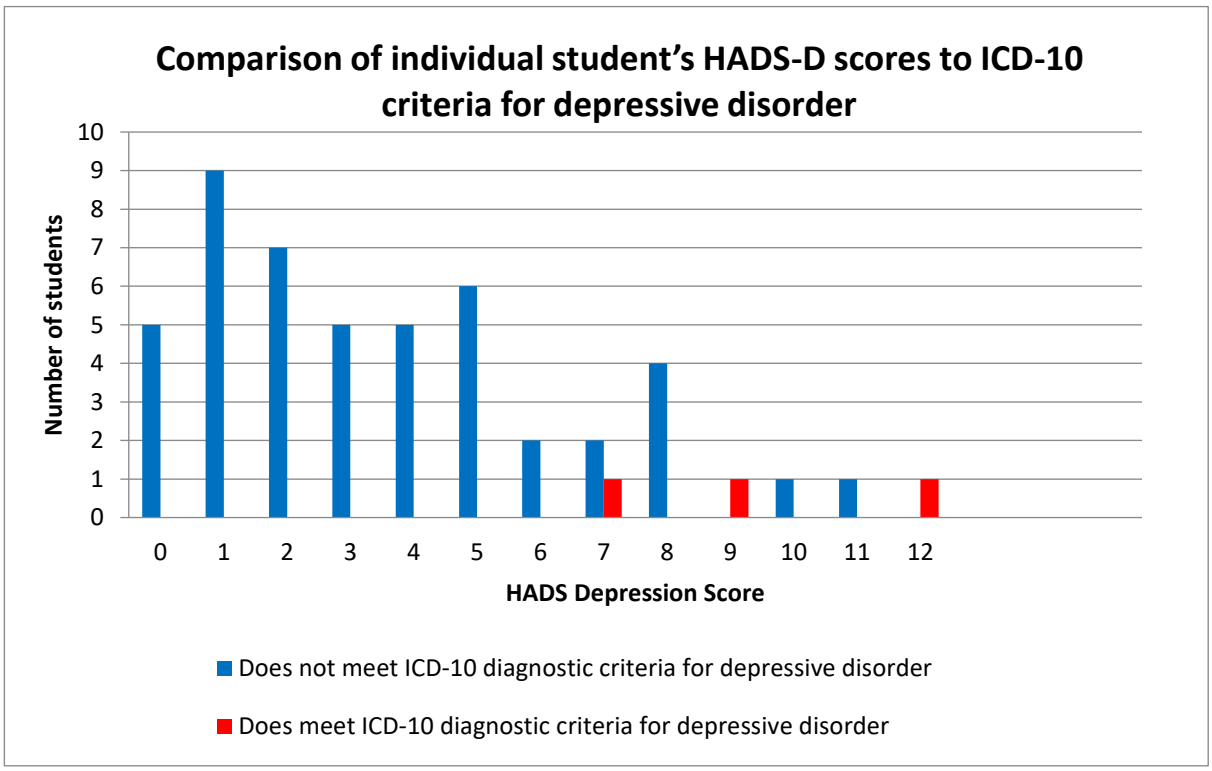


Figure 6 highlights that the HADS-D scores for students who meet the ICD-10 diagnosis for depressive disorder ranged from 7 to 12, whereas the HADS-D scores for those who did not meet the criteria ranged from 0 to 11.

In order to consider how effective HADS-D is at determining whether medical students have depression, the sensitivity and specificity of HADS-D was calculated.

4.3.1.1 Sensitivity and specificity of the HADS-D subscale

The sensitivity, specificity, positive and negative predictive values were calculated for different cut off points.

- Sensitivity being defined as the proportion of people with disease who have a positive test result (Altman & Bland, 1994).
- Specificity is defined as the proportion of people without disease who have a negative test result (Altman & Bland, 1994).
- The negative predictive value estimates the probability that subjects with a negative screening test truly do not have the disease (Maxim et al, 2014).
- Finally, the positive predictive value estimates the probability that subjects with a positive test result, will truly have the disease.

One systematic review suggested that for depression tools used for case finding, sensitivity should be at least 85% and a specificity at least 75% (Gilbody et al, 2006).

Table 26 shows the difference in sensitive and specificity calculations for different cut off points for the HADS depression scores using the ICD-10 depression diagnoses as a basis for comparison.

Table 26 – Sensitivity and specificity values for different HADS-D cut off points using the ICD-10 depression diagnoses.

	HADS cut off ≥7	HADS cut off ≥8	HADS cut off ≥9	HADS cut off ≥11	HADS cut off ≥12
Sensitivity	1.00	0.67	0.67	0.33	0.33
Specificity	0.83	0.87	0.96	0.98	1.00
Positive predictive value	0.27	0.25	0.50	0.5	1.00
Negative predictive value	1.00	0.98	0.98	0.96	0.96

Using the accepted HADS-D cut-off point of 8, the sensitivity in the current sample is 0.67 with a specificity of 0.87. This implies that of those students with a HADS score of 8 or more, 67% were correctly classified as having a diagnosis of depressive disorder. 87% of cases were true negatives, meaning that 87% of students who scored below the cut off of 8, did not have a diagnosis of depressive disorder. The NPV would be high at 0.98.

If the cut-off was reduced to 7, the sensitivity of the test would rise to 1.00. The specificity however would reduce to .83. The NPV would be optimum at 1.00. By applying a cut off to 7 would provide a test with higher level of accuaracy although not quite 'optimum' levels with specificty of .83 being below optimum of .9

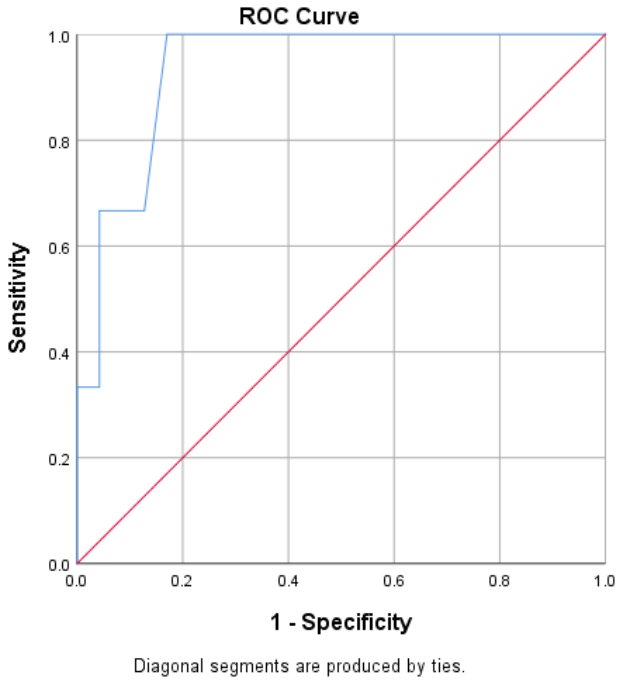
An ROC curve was then plotted and youden's index calculated for different cut offs to determine the optimum cut-off for this sample.

4.3.1.2 ROC curve

The ROC curve for the HADS-D and SCAN data was plotted in SPSS. The ROC curve is a plot of true positives (sensitivity) against true negatives (1-specificity) and is a graphical representation of the trade-off between sensitivity and specificity (Hazibzadeh et al, 2016).

In the graph below (figure 7) the blue curve shows the plot for the depression data for the present study. The red line would represent a test in which at each cut off point there were as many false positives as false negatives and as such is a test which is only as good as chance.

Figure 7 – ROC curve for the depression data



To determine the optimim cut-off point for the current sample the youden index was calauclated. This method involves identifying the cut-off point furthest away from the line of chance. Table 27 shows the different Youden values for different cut-off values.

Table 27 – Youden values for different HADS-D cut-off scores

Cut off	Sensitivity	1 - Specificity	specificity	**Youden** (sens+spec-1)
0	1.000	1.000	0.000	0.000
1	1.000	0.894	0.106	0.106
2	1.000	0.702	0.298	0.298
3	1.000	0.553	0.447	0.447
4	1.000	0.447	0.553	0.553
5	1.000	0.340	0.660	0.660
6	1.000	0.213	0.787	0.787
7	1.000	0.170	0.830	0.830
8	0.667	0.128	0.872	0.539
9	0.667	0.043	0.957	0.624
10	0.333	0.043	0.957	0.291
11	0.333	0.021	0.979	0.312
12	0.333	0.000	1.000	0.333
14	0.000	0.000	1.000	0.000

The youden index is greatest for a cut-off of 7 and above at 0.830. This suggests that for this sample a HADS-D cut-off of 7 would be optimum for determining those students who are likely to suffering fom depression .

4.3.1.3 Area under the curve

In addition to detemining the optimum cut-off score the ROC curve was used to calculate the area under the curve (AUC) to provide an indication of how good the test is at distinguishing between medical students with depression and those without. AUC values can vary between 0.5 which represents a poor test and 1 which is a perfect test.

For the HADS-D data the area under the curve equalled 0.936 ($p<0.05$ Confidence intervals: 0.847-1.0) suggesting that the HADS-D subscale was very good at distingusing between medical students with depression and those without in the current sample.

4.3.1.4 Summary of the quantitiative analysis

In summary the HADS-D quantitative data demonstrated that three students reached the diagnostic criteria for depression. The HADS-D scores for students who met the ICD-10 diagnosis for depressive disorder ranged from 7 to 12, whereas the HADS-D scores for those who did not meet the criteria ranged from 0 to 11. The AUC data suggsted that HADS-D subscale was very good at distingusing between medical students with and without depression. The original cut off score of 8 and above had relatively high levels of specificity at 0.87, however sensitivity at this cut off was lower at 0.67. A lower cut-off of 7 is suggested as it had a better balance of sensitivity and specificity for the current sample, with sensitivity of 1.0 and specificity of 0.83.

To understand why there might be differences between the original cut-off of 8 and the optimal cut-off for the current sample. the items which make up the HADS-D scale were explored and compared to medical students' responses given in the clinical interviews.

4.3.2 Exploration of the HADS-D items

Following the quantitative analysis of the data, the interview responses of those students who had either false positive or false negative results on HADS (using the standard cut off scores of 8 and above) in comparison to clinical interview were qualitatively reviewed. This aimed to provide a more in-depth review of possible reasons for disparity between HADS scores using the standard cut off and clinical interview. These factors may influence why alternative cut-off scores are more appropriate for medical students.

Eight of the 50 students interviewed (16%) scored 8+ on the HADS-D scale, but only 2 reached the diagnostic threshold for depression as determined by SCAN. At the original cut off of 8, the 6 students who scored above 8 but did not reach the threshold for depression as determined by SCAN would be considered false positives. I.e. they scored positive on the test

(HADS-D) but did not have depression. The responses to the HADS-D items of the 6 students who scored highly on HADS-D but did not reach diagnostic threshold for depression as determined by SCAN are explored below.

Only 1 student scored 7 on HADS but reached the diagnostic threshold for depression as determined by SCAN, this would be considered a false negative if the original cut-off score of 8 was being used. With only 1 student in this category patterns and themes in the responses of students identified as false negatives at the original cut-off was not possible.

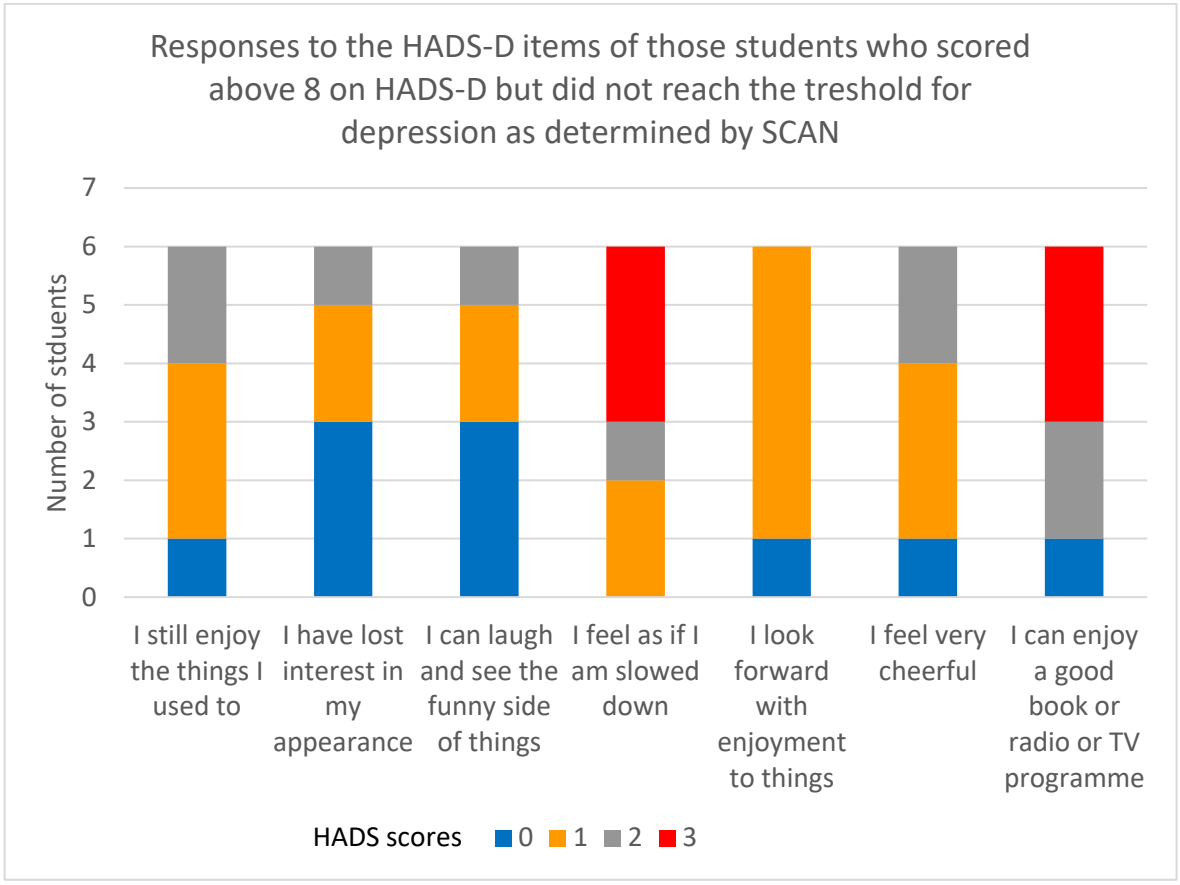
4.3.2.1 HADS-D false positives

Those students who scored 8 and above on HADS but who did not reach the threshold for clinically significant levels of depression or anxiety as determined by SCAN, were identified as false positives. At the original cut off of 8, the 6 students who scored above 8 but did not reach the threshold for depression as determined by SCAN. Responses to the HADS items of these participants were explored further. In particular, items in which these participants scored particularly highly were identified. Responses to equivalent SCAN questions were reviewed and compared to their HADS responses.

Responses to each HADS item are rated between 0-3. 0 suggesting the symptom is not present and 3 suggesting the symptom is present to a significant level.

The responses given to HADS-D items by the 6 students identified as false positives on HADS-D at the original cut off can be seen in figure 8 below.

Figure 8 - Responses to the HADS-D items of those students who scored above 8 on HADS-D but did not reach the threshold for depression as determined by SCAN



This graph highlights that only two HADS-D items elicited a response that scored 3 in this group of 6 medical students who score above 8 on HADS-D but did not reach diagnostic criteria for depression. These items were '**I feel as if I am slowed down**' and '**I can enjoy a good book or radio or TV Programme**'. These items are explored below.

4.3.2.1.1 'I feel as if I am slowed down'

3 students responded on HADS that they felt slowed down '*nearly all the time*' (a score of 3 on the HADS item) and 1 student reported that they felt slowed down '*very often*' (a score of 2 on the HADS item). None of the students said they 'did not feel slowed down at all.'

During the SCAN interviews these students were asked: “Over the past month have you felt as though you were slowed down in your movements, as though everyone and everything else was moving much faster?”. This SCAN question determines whether someone is suffering from "psychomotor retardation" where their thoughts and/or physical movements are slowed down. Psychomotor retardation can be a symptom of depression.

5 of the 6 students stated that *they felt slowed down in their movements*. When asked to be more explicit only 1 student felt that they were moving physically slower to an extent that was clinically significant. Some students answered this question in relation to how they feel about their progress on the medical course.

“Only in the sense of relating it back to work, if you don’t understand something then you feel that you are behind everyone else and everyone’s like oh have you done this and you are like no I haven’t had time for that yet”
“I feel that everyone is doing things faster and more efficient than me.”

4.3.2.1.2 “I can enjoy a good book or radio or TV Programme”

In response to the HADS-D item “**I can enjoy a good book or radio or TV Programme”** 3 of the 6 students under consideration, answered ‘Very seldom’ (HADS score of 3 for this item) and 2 answered ‘not often’ (HADS score of 2 for this item).

During the SCAN interview students were asked if they could still find enjoyment in things, a loss of interest being one symptom of depression. 3 of the 6 students under consideration, mentioned *not enjoying things as much as they normally do*, but this did not reach a significant level for any of the students. For some who felt that they could no longer enjoy activities that they used to enjoy they stated this was due to time restraints and the feeling of guilt if they spent time not working.

"Little things like watching the TV, I just wont do it I'll just go to sleep. …. I see my family and my boyfriend and that is all the free time I have."

"No… lack of time, because I'm on an intensive course I feel bad if I don't work even though I know I shouldn't work all the time.."

"I haven't done them (leisure activities) because I haven't had time"

4.3.2.2 Summary of the HADS-D qualitative analysis

Exploration of the HADS-D items of those students who scored above 8 on HADS-D but did not reach diagnostic criteria for depression (n=6) highlighted that these medical students scored very highly on two items. These items were '**I feel as if I am slowed down**' and '**I can enjoy a good book or radio or TV Programme**'. Although 4 students scored high on the HADS-D item 'I feel as if I am slowed down', when retardation was explored via SCAN only 1 of these students felt that they were moving physically slower to an extent that was clinically significant. Some of these students talked instead about feeling behind on the course. In terms of 'I can enjoy a good book or radio or TV programme', 5 of the 6 students scored high on this HADS-D item. However, none of the students who talked about not enjoying activities during their SCAN interviews had a loss on interest to a clinically significant level. Some talked about not having the time to do these activities, suggesting their lack of enjoyment of activities may be due to their circumstances and not their mental health.

4.4 Anxiety

As with the depression data, the analysis of the anxiety data involved two steps. The first step involved the quantitative analysis of the HADS and SCAN data. Step two took a qualitative approach to reviewing the interview responses of those students who had either false positive or false negative results when their HADS scores were compared to their clinical interview data. The results of these two steps will be reported in turn.

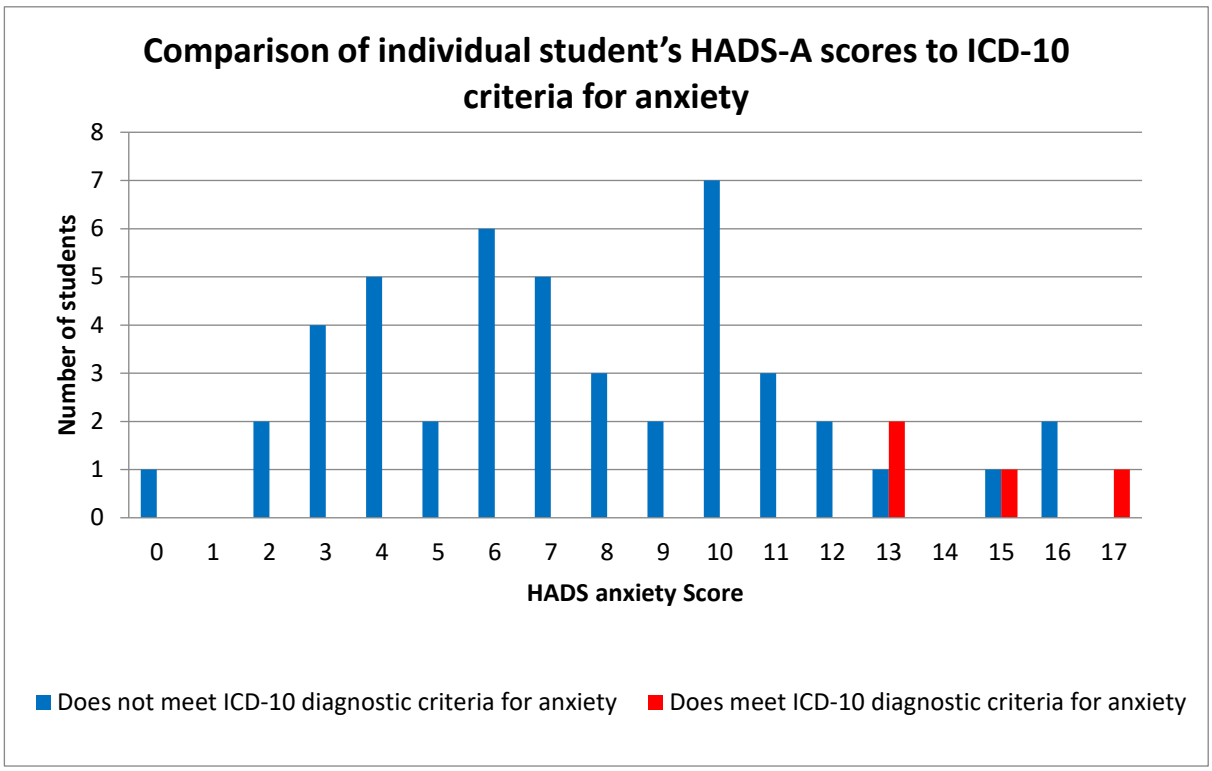
4.4.1 Quantitative analysis

As with the depression data the quantitative analysis involved identifying those students who reached the diagnostic criteria for anxiety, comparing this data with the HADS-A scores and calculating sensitivity and specificity. The data was then used to plot a ROC curve and the optimum cut-off score for this sample identified. Finally, the area under the curve was calculated to give an indication of how good HADS is at distinguishing between medical students with and without generalised anxiety.

Following the SCAN interviews, 4 medical students' symptoms for the month prior to interview reached the threshold for ICD-10 diagnostic criteria for generalised anxiety disorder.

Figure 9 shows the distribution of HADS anxiety scores as completed by students immediately prior to their interviews, broken down by those who following interview were determined to have reached the ICD-10 threshold for generalised anxiety disorder (if the symptoms had persisted for a period of six month) and those who didn't reach the ICD-10 threshold for generalised anxiety disorder.

Figure 9 - Comparison of individual student's HADS-A scores to ICD-10 criteria for generalised anxiety disorder



25 medical students (50%) of those interviewed had a HADS-a score of 8 and above and 13 students (26%) had a HADS-a score of 11 and above, but only 4 (8%) of these reached the threshold for generalised anxiety disorder as determined by clinical interview (if the symptoms had persisted for a period of six month).

In order to consider how effective HADS-A is at determining whether medical students have generalised anxiety, the sensitivity and specificity of HADS-A was calculated.

4.4.1.1 Sensitivity and specificity of the HADS-A subscale

The sensitivity and specificity for different cut off points for anxiety caseness were calculated for the HADS-a scores. Table 28 shows the difference in sensitive and specificity calculations

for different cut off points for the HADS anxiety scores using the ICD-10 generalised anxiety diagnoses as a basis for comparison.

Table 28 – Sensitivity and specificity calculations for different HADS-A cut off points using the ICD-10 anxiety threshold for comparison.

	HADS cut off ≥8	HADS cut off ≥11	HADS cut off ≥13	HADS cut off ≥15	HADS cut off ≥17
Sensitivity	1.00	1.00	1.00	0.50	0.25
Specificity	0.54	0.80	0.91	0.93	1.00
Positive predictive value	0.16	0.31	0.50	0.40	1.00
Negative predictive value	1.00	1.00	1.00	0.96	0.94

For this sample, when considering the sensitivity of the scale, a cut-off point of ≥8 (the level considered indicative of possible levels of anxiety in the general population) would mean that all students (ie sensitivity of 1.0/ 100%) who have clinical levels of depression as rated by SCAN would have a positive HADS-a test. However, a cut of of ≥8 would results in veru low speficity of the test; at this cut off point 54% of students who did not have anxiety having negative HAD-a scores, but 46% of students without clinical levels of depression, as rated by SCAN, would have a positive HADS-D result (1-0.45).

At a cut off of 11 (the original cut off for probably cases), the sensitivity would remain at 100% (1.0) and specificity would increase to 80%, resulting in 20% of students who did not have anxiety having negative HAD-a scores .

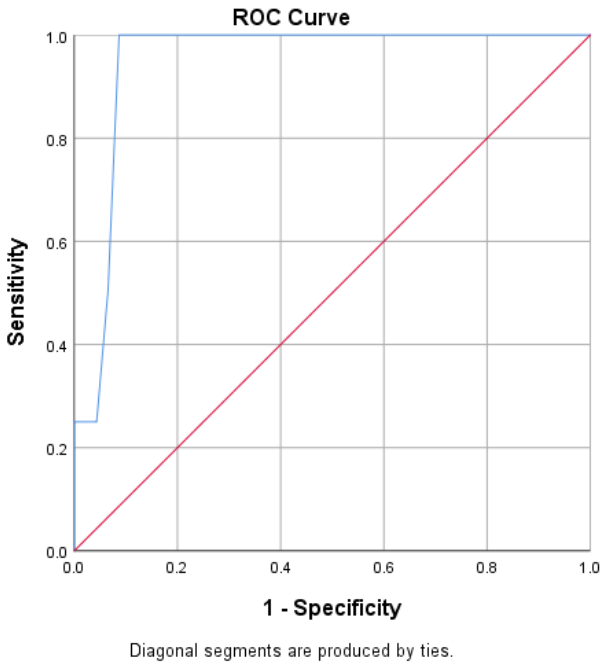
If a HADS-A cut-off of ≥13 is used, again all students who had clinical levels of anxiety as rated by SCAN had a positive HADS-A result, however specificity increases to 91% resulting in only 9% of this group without clinical levels of anxiety as rated by SCAN having a positive result on HADS-A.

An ROC curve was then plotted and youden's index calculated for different cut offs to determine the optimum cut-off for this sample.

4.4.1.2 ROC curve

The ROC curve for the HADS and SCAN data was plotted in SPSS. In the graph below (figure 10) the blue curve shows the plot for the anxiety data for the present study. The red line would represent a test in which at each cut off point there were as many false positives as false negatives and as such is a test which is only as good as chance.

Figure 10 – ROC curve for anxiety data



To determine the optimim cut-off point for the current sample the youden index was calauclated. This method involves identifying the cut-off point furthest away from the line of chance. Table 29 shows the different Youden values for different cut-off values.

Table 29 – Youden values for different HADS-A cut-off scores

cut off	Sensitivity	1 - Specificity	specificity	Youden (sens+spec -1)
0	1.000	1.000	0.000	0.000
1	1.000	0.978	0.022	0.022
3	1.000	0.935	0.065	0.065
4	1.000	0.848	0.152	0.152
5	1.000	0.739	0.261	0.261
6	1.000	0.696	0.304	0.304
7	1.000	0.565	0.435	0.435
8	1.000	0.457	0.543	0.543
9	1.000	0.391	0.609	0.609
10	1.000	0.348	0.652	0.652
11	1.000	0.196	0.804	0.804
12	1.000	0.130	0.870	0.870
13	1.000	0.087	0.913	0.913
14	0.500	0.065	0.935	0.435
16	0.250	0.043	0.957	0.207
17	0.250	0.000	1.000	0.250
18	0.000	0.000	1.000	0.000

The youden index is greatest for a cut-off of 13 and above at 0.913. This suggests that for this sample a HADS-A cut-off of 13 would give the optimum balance of sensitivity and specifity for determining those students who are likely to suffering from anxiety.

4.4.1.3 Area under the curve

In addition to detemining the optimum cut-off score the ROC curve was used to calculate the area under the curve (AUC) to provide an indication of how good the test is at distinguishing

between medical students with generalised anxiety and those without. AUC values can vary between 0.5 which represents a poor test and 1 which is a perfect test.

For the HADS-A data the area under the curve equalled 0.948 ($p<0.05$ std Error 0.032 Confidence intervals: 0.885-1.0) suggesting that the HADS-A subscale is very good at distingusing between medical students with generalised anxiety and those without in the current sample.

4.4.1.4 Summary of the quantitiative analysis

In summary the HADS-A quantitative data demonstrated that four students reached the diagnostic criteria for generalised anxiety disorder. The HADS-A scores for students who met the ICD-10 diagnosis for GAD ranged from 13 to 17, whereas the HADS-D scores for those who did not meet the criteria ranged from 0 to 16. The AUC data suggsted that HADS-A subscale was very good at distingusing between medical students with and without GAD. The original cut off score of 8 and above had very high levels of sensitivity but relatively low levels of specificity at 0.54. The Youden index suggested that the optimum balance of sensitivity and specificity was at a cut-off of 13, significantly higher than the original cut off of 8.

To understand why there might be differences between the original cut-off of 8 and the optimal cut-off for the current sample the items which make up the HADS-A scale were explored and compared to medical students' responses in the clinical interviews. This will be discussed next.

4.4.2 Exploration of the HADS-A items

Following the quantitative analysis of the data, the interview responses of those students who had either false positive or false negative results on HADS (using the standard cut off scores of 8 and above) in comparison to clinical interview were qualitatively reviewed. This aimed to provide a more in-depth review of possible reasons for disparity between HADS scores using the standard cut off and clinical interview.

25 of the 50 students interviewed (50%) scored 8+ on the HADS-A scale, of which only 4 students reached the diagnostic threshold for anxiety as determined by their SCAN interview. At the original cut off of 8, the 21 students who scored above 8 but did not reach the threshold for generalised anxiety as determined by SCAN would be considered false positives. Ie they scored positive on the test but did not have the condition. The responses to the HADS-A items of the 21 students who scored highly on HADS-A but did not reach diagnostic threshold for GAD as determined by SCAN are explored below.

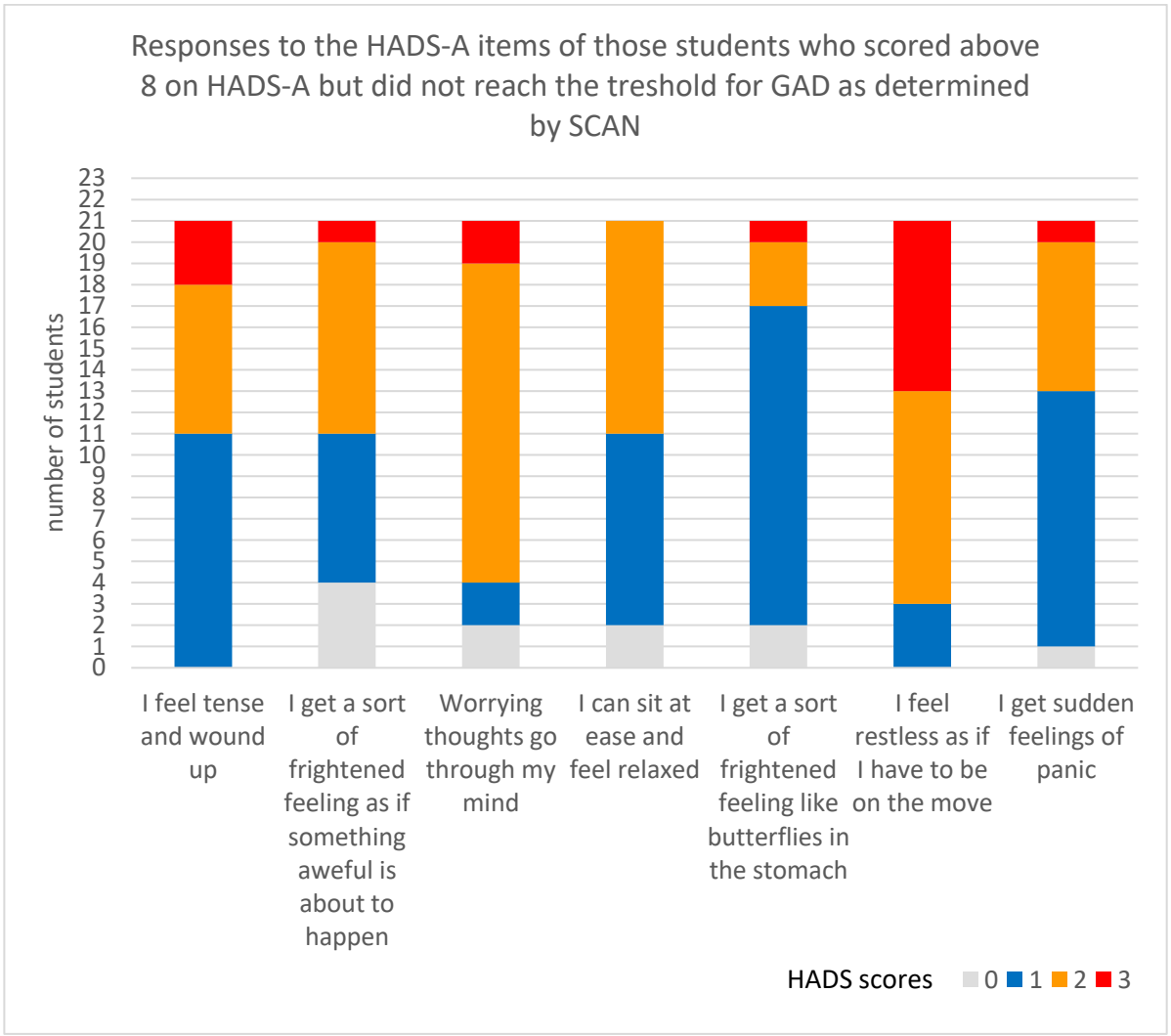
No students who reached the diagnostic threshold for generalised anxiety disorder as determined by SCAN had a HADS-A score below 8, there were therefore no false negatives using the original cut-off score of 8.

4.4.2.1 HADS-A false positives

To explore possible false positives the 21 of the students interviewed who scored 8+ on the HADS-A scale, but did not reach diagnostic threshold for GAD as determined by SCAN were then explored. Items in which these participants scored particularly highly were identified. Responses to equivalent SCAN questions were reviewed and compared to their HADS responses. Responses to each HADS item are rated between 0-3. 0 suggesting the symptom is not present and 3 suggesting the symptom is present to a significant level.

The responses given to HADS-A items by the 21 students identified as false positives on HADS-A at the original cut off can be seen in figure 11 below.

Figure 11 - Responses to the HADS-D items of those students who scored above 8 on HADS-D but did not reach the threshold for depression as determined by SCAN



This graph highlights that those HAD-A items which these 'false positive' medical students responded highly on included '***Worrying thoughts go through my mind***' and '***I feel restless as if I have to be on the move***'. *Items which these medical students scored lower on included "I get a sort of frightened feeling like butterflies in the stomach' which 17 out of the 21 students said they experienced only occasionally or not at all and 'I get sudden feelings of panic' which*

13 of the 21 students experienced not very often or not at all. The items which students scored highly on will be explored in more detail below.

4.4.2.1.1 Worrying thoughts go through my mind

In relation to the statement '***Worrying thoughts go through my mind***' 14 (67%) of the 21 students responded with 'a lot of the time' (HADS score 3 for this item) and 3 responded with 'a great deal of the time' (HADS score 2 for this item). (2 responded with 'from time to time' and 2 responded occasionally).

In the SCAN interviews the medical students were asked about their worrying, this being one of the main symptoms of anxiety. 17 of the 21 students in question talked about having worrying thoughts or feeling anxious but only 4 students were worrying to a clinically significant level where it was affecting their everyday activities.

In their responses to the SCAN questions many medical students talked about feeling worried about their course and the amount of work that they needed to do, but some felt that this was normal for medical students.

"I worry a lot, constantly……. I worry that I won't pass this year, that I won't pass finals, that I am not improving in any way, instead of improving I am degenerating. Also, that I haven't done the things I wanted to do."

*"*not doing as well as everyone else, not working as hard as everyone else, not being as confident as everyone else. Feeling a bit behind. …I worry about them a lot of the time"

"I think I worry more than I should but there is so much pressure on us as medical students, pressure to do this work and hundreds of assessments to do… they make it seem to us like it's a big deal, so then we think it's a big deal then we worry about it, whereas if I stood back and took a look at medical students I would say your worrying about nothing, but we are all

worrying about the same thing….from any other medical student I am no different… You are talking to a ** year medical that is normal"

"I worry about work, I worked really hard last year and did rubbish so I'm working harder. But I couldn't have worked harder last year so I'm worrying a lot about that and also because my degree is one where you never finish the work, it's like how much do I do. When do I give myself a break because I could not sleep and work all night and still wouldn't have done enough. So, I definitely worry about that.

"(I worry about) workload and making enough time to do the things I enjoy."

"(I worry about) Mainly work related things"

Whilst many were worrying about their course, this worry for most (17 out of the 21 students) did not reach clinically significant levels.

4.4.2.1.2 I feel restless as if I have to be on the move

When answering the HADS-A item ***'I feel restless as if I have to be on the move'*** 3 of the 21 students under consideration reported they felt restless 'not very much' (HADS score 1 for this item), 10 said they felt restless 'quite a lot' and 8 said they felt restless 'very much indeed'.

In the SCAN interviews, students were asked if they had been feeling restless over the last month, restlessness being a symptom of generalised anxiety. 13 of the 21 students described feeling restless but only 2 had restlessness to an extent that was clinically significant where they had feelings of restlessness most of the time or were finding themselves often pacing around. During the SCAN interviews some students did talk about being fidgety and restless, many said that this was no different to how they normally behaved, some talked about having lots to do and some said they feel restless when sat working for long periods.

Normal behaviour

"I'm just like that I'm quite an energetic person"

"I like to be physically active"

"Yes but I do that quite a lot, I'm a bit of a fiddler"

"Yes I've always been like that"

Being busy

"I'm just a busy bee… I'd rather just be doing something"

"All the time I feel like I need to be doing something"

"I'm not pacing for no reason I'll get up and be busy and active in a purposeful way"

Having been sat for a long time

"Mainly in terms of concentration, you just want to get up and not… I think it's more of an avoidance thing you don't want to be sat doing it."

"when I'm sat down doing work of an evening"

"for example if I'm sat down for an hour I completely stretch out my legs,… because I'm in lectures do its not appropriate to start running around"

"In the evenings when I haven't got enough exercise in… if I'm wanting to do some exercise.

Therefore, whilst many students felt they had been restless during the previous month, most (19 out of 21) did not experience this to an extent that was clinically significant.

4.4.2.2 Summary of the HADS-A qualitative analysis

Exploration of the HADS-A items of those students who score above 8 on HADS-A but did not reach diagnostic criteria for GAD highlighted that responses to two HADS-A items were particularly high. These items were '***Worrying thoughts go through my mind***' and '***I feel restless as if I have to be on the move***'. Although 17 of these students scored high on the HADS-A item Worrying thoughts go through my mind' when worrying thoughts were explored in the SCAN interview only 4 had clinically significant levels of worrying. Some of the students were worrying about their course and whether they are working hard enough.

When exploring the item 'I feel restless as if I have to be on the move' 18 students scored relatively highly on this HADS-A item. However, when restlessness was explored during the SCAN interviews only 2 students in the 'false positives' group had clinically significant levels of anxiety. Others talked about always feeling restless (their normal state), others that they are just busy, or like to be busy and some talked about not being able to sit for long periods of time without feeling restless.

4.5 Summary of interview data

50 clinical SCAN interviews were carried out with medical students. Prior to starting the clinical interviews, the students were asked to complete HADS. The interviews were rated by the interviewers and a diagnosis determined using ICD-10 criteria for depression and generalised anxiety disorder. Results from HADS and the interviews were compared. To check consistency of ratings inter-rater reliability was calculated. There was a high level of inter-rater reliability when comparing ratings relating to symptoms and a moderate level of agreement relating to diagnosis ratings.

Using the ICD-10 diagnostic criteria for depression, three students had clinically significant levels of depression. Using these cases as clinically significant sensitivity and specificity was calculated for different cut-off thresholds for HADS-D. These calculations found a cut-off of 7 gave optimum levels of sensitivity and specificity with depression in the current sample. At this cut-off threshold 11 students in the current sample would reach this threshold compared to 3 students who had clinically significant levels of depression as determined by clinical interview.

The HADS-D items of those medical students who scored above 8 on HADS but which did not reach the threshold for a depression diagnosis were compared to their responses in the SCAN interview. The HADS-D items which students in this sample scored themselves highest on were '*I feel as if I am slowed down*' and '*I can enjoy a good book or radio or TV Programme'*. Four of the six students scored high on the HADS-D item 'I feel as if I am slowed down', but only 1 was moving physically slower to an extent that was clinically significant as determined by the SCAN interviews. Some of these students talked instead about feeling behind on the course, rather than a physical feeling of moving slower. Five of the six students scored high on 'I can enjoy a good book or radio or TV programme', however, no students in this 'false positives' group had a loss of interest to a clinically significant level as determined by their SCAN interview. Some talked about not having the time to do leisure activities, suggesting their lack of enjoyment of activities may be due to their circumstances/lack of free time and not their mental health.

Using ICD-10 criteria for anxiety 4 medical students in the current sample had clinically significant symptoms for generalised anxiety disorder for the month prior to interview. Sensitivity and specificity was calculated for different cut-off thresholds for HADS-A. It is suggested given these calculation that a HADS-A cut off of greater than or equal to 13 would be optimal in identifying medical students with anxiety in the current sample. At this cut-off threshold, 8 students in the current sample would reach this threshold compared to 4 students who had clinically significant levels of anxiety symptoms as determined by clinical interview.

The HADS-A items of those medical students who scored above 8 on HADS-A but which did not reach the threshold for a GAD diagnosis (n=21) were compared to their responses in the SCAN interview. The items which these students scored high included 'Worrying thoughts go through my mind' and 'I feel restless as if I have to be on the move'. 17 students scored high on the HDAS-A item worrying thought go through my mind but only 4 of this group of students had significant levels of worry. Many of the students were worrying about their course but not to a clinically significant level. 18 students in this 'false positives' group scored highly on 'I feel restless as if I have to be on the move' although only two students in the group have restlessness to a clinically significant level as determined by the SCAN interviews. Students talked about a range of aspects relating to restlessness including being a restless person, being busy or not finding it easy to sit for long periods when working or in lectures.

4.6 Responses to research questions

This project aimed to answer three research questions, these are described below with an explanation of how the above results answer each question.

The research questions for this project were:

Research question 1: Is HADS an accurate measure of depression and anxiety in medical students?

The area under the curve for both the depression (AUC= 0.936) and the anxiety (AUC = 0.948) ROC curves suggest that in the current sample both the HADS depression and anxiety subscales can distinguish between medical students who are likely to be suffering from depression or anxiety and those that are not.

Research question 2: Are the standard cut off scores of 8+ on the HADS-D and HADS-A subscales appropriate for a medical student population?

Results from this study suggest the HADS-D cut off should be 7 and the HADS-A cut off should be increased to 13.

For the HADS-D subscale the cut-off score which resulted in the best balance between sensitivity and specificity was 7 for this medical student sample. Sensitivity at the original cut-off of 8 was 0.67 for this sample with a specificity of 0.87. Thus the original cut-off would have a sensitiviy below what is considered an acceptable level (Gilbody et al, 2006) suggesting that the original cut off may not be appropriate for a medical student population.

For the HADS-A subscale the cut-off score which resulted in the best balance between sensitivity and specificity was 13 for this medical student sample. Sensitivity at the original cut-off of 8 was excellent at 8 for this sample. However specificity was very low at 0.54. Thus the original cut-off would have a specificity below what is considered an acceptable level (Gilbody et al) suggesting that the original cut off may not be appropriate for a medical student population.

Research question 3: Do the responses to individual items within the subscales truly reflect the presence of anxiety and depression?

Exploration of the responses of the students who scored highly on the HADS subscales but which did not reach the threshold for clinical diagnosis, suggests that some HADS items are scored more highly by the medical students in these groups. These were *I feel as if I am slowed down*' and '*I can enjoy a good book or radio or TV Programme' for the depression subscale and* 'Worrying thoughts go through my mind' and 'I feel restless as if I have to be on the move' for the anxiety subscale. For all of these items, only a small number actually experienced these symptoms to a clinically significant level as determined by their SCAN interviews. Some of the medical students talked in their interviews about other aspects broadly relating to these symptoms. These students did not interpret and answer the questions in a way in which would suggest that they were experiencing the symptoms to a clinically significant level as described in the ICD-10. This may explain why, in the case of anxiety, the cut-off for the HADS-A subscale may require a change for this particular cohort.

Chapter 5: Discussion

The preceding chapters have detailed the rationale, design, methodology and results of a project which had the primary aim of investigating the suitability of HADS to screen for depression and anxiety in a medical student population.

This chapter will review the results of the study and discuss these in the context of the wider literature. The strengths and limitations of the study will also be reviewed. Finally, the implications of the study will be highlighted along with conclusions which can be drawn from the study.

5.1 Key findings and contributions

This project had 3 specific research questions:

- Research question 1 – Is HADS an accurate measure of anxiety and depression in medical students?

- Research question 2 – Are the standard cut off scores of 8+ on the HADS-D and HADS-A subscales appropriate for a medical student population?

- Research question 3 – Do the responses to individual items within the HADS subscales truly reflect the presence of anxiety and depression.

To answer these questions 50 medical students were recruited and took part in a face to face meeting. During the face to face meeting the medical students completed the HADS questionnaire and then undertook a clinical interview using the SCAN clinical interview technique. Of the 50 students who participated, 7 (14%) were male and 43 (86%) were female. The proportion of male students who took part in the project was lower than the proportion of male medical students studying at Cardiff at the time which was on average 37% male. A summary of how the results from the study answer each of these research questions is detailed below. The cut-off scores for the depression and anxiety subscales (Research question 2) will be considered separately.

5.1.1 – Is HADS an accurate measure of anxiety and depression in medical students?

This project aimed to determine if HADS is able to distinguish between medical students who are suffering from depression and/or anxiety and those who are not. The current study was concerned with the two separate subscales of HADS, these being the depression subscale and the anxiety subscale. These subscales are used to screen for anxiety and depression. In order to establish how valid these subscales were for screening for depression and generalised anxiety, within a medical student sample, the medical students HADS results were compared to clinical interview data. Clinical interviews are seen as the gold standard for diagnosing depression and anxiety. Interviews allow the interviewer to explore the responses provided to establish if a symptom has been present, for how long and the effect that the symptom has had on the individual's quality of life. To test the validity of HADS, students undertook a SCAN clinical interview to determine if they meet the criteria for depression or generalised anxiety. An individual's diagnosis could then be compared to their HADS subscale scores. Sensitivity and specificity were then calculated for different HADS cut-off points and a ROC curve plotted. The area under the curve (AUC) was then calculated. The AUC score is defined as "the probability that a randomly sampled respondent will be correctly assigned to the appropriate group" (Hanley & McNeil, 1982). AUC is the probability that individuals who score above the cut-off for depression on HADS also reach the threshold for a diagnosis of depression according to SCAN. The AUC value is therefore considered to represent the overall accuracy or validity of the instrument (Balsamo & Saggino, 2014). AUC scores vary from 0.5 to 1, a score of 1 suggests that the test is perfect and a score of 0.5 suggests a worthless test.

In the current sample 3 students reached the diagnostic criteria for depression (from their SCAN interviews). The HADS-D scores for students who meet the ICD-10 diagnosis for depressive disorder ranged from 7 to 12, whereas the HADS-D scores for those who did not meet the criteria ranged from 0 to 11. The area under the curve for the HADS depression subscale was 0.936.

For the anxiety subscale 4 medical students' symptoms for the month prior to interview reached the threshold for ICD-10 diagnostic criteria for generalised anxiety disorder. The

HADS-A scores for students who meet the ICD-10 diagnosis for generalised anxiety ranged from 13 to 17, whereas the HADS-A scores for those who did not meet the criteria ranged from 0 to 16. The area under the curve for the HADS anxiety subscale was 0.948.

The high AUC scores for both subscales suggest that in the current sample both the HADS depression subscale and the HADS anxiety subscale were relatively good at distinguishing between medical students who were and were not suffering from depression or anxiety respectively. Therefore, for the current sample of medical students, HADS would be considered a suitable tool for identifying students who may be suffering from depression and/or anxiety.

5.1.2 Is the standard cut off scores of 8+ on the HADS-D subscale appropriate for a medical student population?

The original authors of the instrument proposed cut-off scores between 8 and 10 for 'possible' cases, and scores of 11 or more for 'probable' cases in both the anxiety and depression scales. Since then studies have used values between 8 and 11 when using HADS, with the majority using a single cut-off value of 8. For the current study comparisons will be made to the standard cut-off score of 8 for ease of comparison.

To establish if the standard cut-off score of 8+ for the HADS-D subscale is appropriate for a medical student population, the sensitivity and specificity of different cut-off scores was calculated. These were then plotted on a ROC curve and Youden's index was calculated to determine the optimum cut-off for the current sample.

Using the standard accepted HADS-D cut-off point of 8 for possible cases, the sensitivity in the current sample was 0.67 with a specificity of 0.87. This implies that of those students with depression, 67% had a HADS score above 8, however, 33% of students with depression would go undetected at this cut-off. 87% of cases who did not have a diagnosis of depressive disorder had a score below 8. Negative predictive value is the probability that a person with a negative (normal) test result is truly free of disease. The NPV for the current sample at a cut

off of 8+ was high at 0.98, suggesting 98% of those whith a negative result would not not have depression but 2% of those with a negative result would have depression.

The youden index values suggested that a cut-off of 7 and above for the HADS depression scale would be optimum for this sample. If the cut-off was reduced to 7, the sensitivity of the test would rise to 1.00, suggesting at this cut-off all students with depression would be idenfiied. The specificity however would reduce to .83, suggesting 17% of students who did not have depression would reach the cut-off for depression.

In summary, results from the current study suggest that the standard cut-off score of 8+ would mean that not all medical students with depression would reach the HADS cut-off, suggesting that this cut-off score is not optimum for medical students. A optimal cut-off would be 7+ for the current sample. Reducing the cut-off for medical students may identify those students who would reach the threshold for depression at an earlier stage. Earlier detection of depression would allow for earlier support and treatment if required.

5.1.3 Is the standard cut off scores of 8+ on the HADS-A subscale appropriate for a medical student population?

As with the depression subscale to establish if the standard cut-off score of 8+ for the HADS-A subscale is appropriate for a medical student population, the sensitivity and specificity of different cut-off scores was calculated. These were then plotted on a ROC curve and Youden's index was calculated to determine the optimum cut-off for the current sample.

At the original suggested cut of point of 8+ for the HADS anxiety subscale, for the current sample, sensitivity was 1.00 suggesting that all students who have clinical levels of depression as rated by SCAN would score positively on the HADS-A subscale. However at this cut-off specificity was very low at 0.54 suggesting only 54% of students who did not have anxiety had a negative HAD-A score (ie below the cut-off), and 46% of students without clinical levels of depression, as rated by SCAN, would have a positive HADS-D result

The Youden Index was calculated for all cut-off and was greatest for a cut-off of 13 and above at 0.913. This suggests that for this sample a HADS-A cut-off of 13 would give the optimum

balance of sensitivity and specifity for determining those students who are likely to suffering from anxiety.

At a HADS-A cut-off of ≥13 , again all students who had clinical levels of anxiety as rated by SCAN had a positive HADS-A result, however specificity increases to 91% resulting in only 9% of this group without clinical levels of anxiety as rated by SCAN having a positive result on HADS-A. If the HADS cut-off was increased beyond 13, whilst specificity would increase the sensitivity would decrease meaning that not all students with anxiety would have a positive HADS-A result.

In summary, results from the current study suggest that the standard cut-off score of 8+ would mean that nearly half of students who did not have anxiety would have a positive HADS-A score and thus this cut-off would not appear to be appropriate for medical students. A more optimal cut-off would be 13+ for the current sample.

5.1.4 Research question 3: Do the responses to individual items within the HADS subscales truly reflect the presence of anxiety and depression.

In order to answer this research question students who scored above the standard cut-off points on the HADS subscales but did not reach the diagnostic criteria for depression or anxiety (false positives) were explored. For the HADS-D subscale 6 students scored above the standard cut-off and did not reach the diagnostic criteria for depression. Exploration of these 6 students HADS-D responses highlighted that these medical students scored very highly on two items. These items were '**I feel as if I am slowed down**' and '**I can enjoy a good book or radio or TV Programme**'. Further exploration was carried out on responses to these two items and responses were compared to responses to similar questions asked in the SCAN interviews. Although 4 students scored high on the HADS-D item 'I feel as if I am slowed down', when retardation was explored via SCAN only 1 of these students felt that they were moving physically slower to an extent that was clinically significant. Some of these students talked instead about feeling behind on the course and feeling that others were doing more work or

were better than them. In terms of the HADS-D item 'I can enjoy a good book or radio or TV programme', 5 of the 6 students scored high on this HADS-D item. Though, none of the students who talked about not enjoying activities during their SCAN interviews had a loss on interest to a clinically significant level. Some talked about not having the time to do these activities, suggesting their lack of enjoyment of activities may be due to their circumstances and not their mental health.

A similar exploration was carried on the HADS-A items of those students who score above 8 on HADS-A but did not reach diagnostic criteria for GAD. 21 students score above the standard cut-off of 8 on HADS-A but did not reach diagnostic criteria for GAD. Exploration of these students' HADS-A responses highlighted that responses to two HADS-A items were particularly high. These items were '***Worrying thoughts go through my mind***' and '***I feel restless as if I have to be on the move***'. Although 17 of these students scored high on the HADS-A item Worrying thoughts go through my mind' when worrying thoughts were explored in the SCAN interview only 4 had clinically significant levels of worrying. Some of the students were worrying about their course and whether they are working hard enough. When exploring the item 'I feel restless as if I have to be on the move' 18 students scored relatively highly on this HADS-A item. When restlessness was explored during the SCAN interviews only 2 students in the 'false positives' group had clinically significant levels of anxiety. Others talked about always feeling restless (their normal state), others that they are just busy, or like to be busy and some talked about not being able to sit for long periods of time without feeling restless.

In summary, exploration of the responses of the students who scored highly on the HADS subscales, but which did not reach the threshold for clinical diagnosis, suggests that there were some patterns particularly in the items in which these students scored particularly high. When these high scoring items were further explored, only a small number of students actually experienced these symptoms to a clinically significant level as determined by their SCAN interviews. In all cases, some of the medical students talked in their interviews about other aspects broadly relating to these symptoms. These students did not interpret and answer the questions in a way in which would suggest that they were experiencing the symptoms to a clinically significant level as described in the ICD-10. This may explain why, in the case of anxiety, the cut-off for the HADS-A subscale may be different to the original cut-off.

5.1.5 Summary of findings

Results from this study found that in the current sample both the HADS depression subscale and the HADS anxiety subscale were relatively good at distinguishing between medical students who were and were not suffering from depression or anxiety respectively. Suggesting that for the current sample of medical students, HADS would be considered a valid tool for identifying students who may be suffering from depression and/or anxiety. However, optimum cut off scores for the depression and anxiety subscales, for the current sample varied from the standard cut-off score of 8+. Optimum cut-off scores were 7+ for the depression subscale and 13+ anxiety subscale,

Exploration of the responses of the students who scored highly on the HADS subscales, but which did not reach the threshold for clinical diagnosis, suggests that there were some patterns particularly in the items in which these students scored particularly high. Items scored highly were I feel as if I am slowed down' and 'I can enjoy a good book or radio or TV Programme' for the depression subscale. For the anxiety subscale Worrying thoughts go through my mind' and 'I feel restless as if I have to be on the move' were scored highly. Only a small number of students experienced these symptoms to a clinically significant level as determined by their SCAN interviews. The remaining students did not interpret and answer the questions in a way in which would suggest that they were experiencing the symptoms to a clinically significant level as described in the ICD-10.

5.2 Findings in the context of wider literature

The results from this study have been summarised in Section 5.1 above. The results for each research question will now be discussed in relation to the wider literature.

5.2.1 Is HADS an accurate measure of anxiety and depression in medical students?

Results from this study suggests that HADS is a valid tool for use with a UK medical student population. Whilst HADS has been criticised for reduced validity with some populations (Bjelland et al, 2002) many studies have used HADS with a medical student sample. Many of the studies using HADS with a medical student population are either using it to determine the prevalence of depression and or anxiety in different medical student populations around the world (eg Rab et al, 2008, Ibrahim et al, 2013). Other studies using HADS with medical students have looked at factors which affect medical student's health wellbeing (eg Voltmer et al, 2012, Kotter et al, 2014). However, the author is not aware of any other studies which compare HADS results to a clinical interview diagnosis for determining its validity for use with this population. Ali et al (2016) suggest that any tools (especially if used for screening for depression and anxiety) should be validated locally against a gold standard diagnostic interview to confirm its validity for the study population.

Therefore, this study provides new support for HADS with this population. This study was concerned with the criterion validity of HADS, determining its ability to identify students who may be suffering from depression or anxiety, again a gold standard diagnostic interview. Nevertheless, other aspects relating to the use of HADS with this population also need to be explored, before users can be fully confident in using HADS as a screening took for depression and anxiety in medical students. These include, exploring the factor structure and construct validity of HADS and ensuring HADS provides reliable results with this population. As outlined in the literature review (see section 2) many studies have questioned the structure of HADS (Cosco et al, 2012). However, it has been suggested that the differences seen in the dimensionality results may be due to methodological differences and the specific analytic strategy employed to evaluate them rather than different psychological traits (Straat, 2013; Cosco et al; Norton 2013). Determining if a two factor (depression and anxiety) structure is found when HADS is used with a medical student sample would provide additional support for its use with this population.

5.2.2 Are the standard cut off scores of 8+ for the subscales appropriate for a medical student population?

The original authors of the HADS instrument proposed cut-off scores between 8 and 10 for 'possible' cases, and scores of 11 or more for 'probable' cases for both the anxiety and depression subscales (Zigmond & Snaith, 1983). Since the publication of the original scale many studies have reviewed the cut-off scores that are likely to give the greatest sensitivity and specificity for detecting clinical levels of anxiety or depressive symptoms (Bjelland et al, 2002). In a literature review looking at the validity of HADS, cut off points ranged from 3+ to 11+ for anxiety and 4+ to 11+ for depression (Bjelland et al, 2002).

Studies using HADS with medical students have used cut-off scores for the depression scale ranging from 7+ to 12+ (Rotenstein et al, 2016). Reasons for these differences are in part due to the interpretation of the original cut-off range of between 8-10 for possible cases and 11+ for probable cases, meaning some use 8+ as a cut off, others 11 and one study 12 (Rab et al, 2008; Prinz et al, 2012; El-Gilany at al, 2008). Other reasons are due to validation of HADS in specific languages, hence Turkish studies used a cut-off of 7 (Akvardar et al, 2004) as this was the cut off suggested when the tool was validated for a Turkish population (Aydemir et al. 1997) The use of different specific cut-off points makes comparison of results across studies difficult. In addition, none of these cut-offs are based on a validation study specifically involving medical students.

Studies comparing self-report to clinical interviews are sparse, no studies of this nature were found for HADS with a UK medical student population. In a study involving a UK general student population, Andrews et al (2006), compared HADS to clinical interviews (using the Structured Clinical Interview for DSM-IV Axis 1 disorders). The authors noted that students tended to over-report levels of anxiety using HADS (11). They concluded that for the HADS-D subscale a cut-off score of 10 gave optimum sensitivity/specificity balance. For the HADS-A subscale the best sensitivity/specificity balance was a cut-off of 13. At this cut off however for both scales the sensitivity and specificity values were both below 8 and only half of students with HADS-A scores of 13 and over had clinical levels of anxiety.

Results from the current study suggest that for the HADS-D subscale the standard cut-off score of 8+ (original score for possible cases) would mean that not all medical students with

depression in the current sampple would reach the HADS cut-off. A optimal cut-off would be 7+ for the current sample. For the HADS-A subscale the standard cut-off score of 8+ (for possible cases) would mean that nearly half of students who did not have anxiety would have a positive HADS-A score. At the cut off of 11+ (the original cut off for probably cases) whilst all those who had anxiety would have a positive score, 20% of those who did not have anxiety would also have a positive score. A optimal cut-off for the HADS-A subscale would be 13+ for the current sample.

Therefore, both the current study and Andrews et al suggest that a cut off of 13+ for the HADS-A subscale would give the best sensitivity/specificity balance for students and medical students. In contrast, Andrews et al suggest a cut off of 10+ for the HADS-D subscale whereas in the current study found an optimal cut off of 7+ for the HADS-D subscale.

Andrews study (11) examined anxiety and depression in a general student population whereas the current study specifically looked at a medical student cohort. Both these studies suggest that students when completing HADS report high levels of anxiety which is not supported by clinical interview data. The current study used the Youden Index to determine the optimum cut-off score for probable cases. It has been suggested that the Youden index provides a slightly more sensitive cut-off threshold (ie a slightly lower cut-off) (Hajian-Tilaki, 2017). Given the sensitive nature of the conditions being tested, it was decided that a test which produced a slightly more sensitive threshold may be of more value and as such the Youden Index method for calculating the optimum cut-off score was used. Andrews et al, used a measure of overall efficiency to determine optimum cut off which was the percentage of cases correctly classified as having or not having a diagnosis.

Thus, the disparity in sensitivity, specificity and optimum cut-offs between the current study and Andrews et al study may be due to methodological differences. This could relate to the instrument or interview to which the HADS tool was compared to. Andrews in his study made comparisons using DSM IV for diagnostic criteria whereas our study used ICD 10 criteria. Differences in outcomes across the two studies may also be due to characteristics which are inherent in the sample being tested. Medical students are recognised as having high levels of neurotic perfectionist traits which is a risk factor for common mental health conditions such as anxiety and depression (Lessin, & Pardo, 2017). Medicine also recruits a high percentage of woman to the course. Woman are more likely to self-report anxiety than men. The current sample consisted of 84% female, whereas the sample used in Andrews et al's study was 70%

female It could be that for accurate assessment of self-report symptoms different student groups may require slightly different cut-off points to achieve the optimum sensitivity/specificity balance.

5.2.2.1 Effect of cut-off values on prevalence

The cut-off values which are used for HADS will affect the prevalence found in studies using HADS. For example, in the current study, true prevalence of depression as determined by the scan interviews was 6%. However, prevalence as determined by HADS-D would be 16% at a cut-off of 8+ and 22% at a cut-off of 7+. As cited in the introduction a 2016 meta-analysis of studies suggested a global prevalence of depression amongst medical students of 28.0% (Puthran et al, 2016). Another 2016 meta-analysis identified prevalence rates from 1.4% to 73.5% (Rotenstein et al, 2016). Of the seven studies identified in the Rotenstein analysis, which were based in the UK, prevalence ranged from 4.4 % (Newbury-Birch et al, 2001) to 48.8%. (Honney et al, 2010). Of these seven studies, four used HADS to measure prevalence and all used a 8+ cut off value. Prevalence of these four studies were 4.4% (Newbury-Birch et al, 2001), 9.6% (Picjard et a, 2000), 1.6% (Ashton and Kamali 1995) and 6.6% (Quince at al, 2012). Whilst these values are not dissimilar to the 6% true prevalence (as determined by SCAN interviews), they are lower than the prevalence found in the current sample as determine by HADS-D with a 8+ cut-off. The true prevalence found in the current study is similar to that found in the general population of a similar age: A National survey of mental health in the UK suggests that the prevalence of depression amongst 16-24 year olds is 2.3% (Stansfeld, et al, 2016) suggesting that medical students may have a slightly higher prevalence of depression (6%) than the general population of a similar age (2.3%).

For anxiety in the current sample true prevalence (as determined by the SCAN interviews) of generalised anxiety disorder was 8%. However, prevalence as determined by HADS would be 50% at a cut-off of 8+ and 16% at a cut-off of 13+. Andrews et al, found prevalence of anxiety at a cut-ff of 8+ in their general student sample of 76%. In a systematic review of studies looking at anxiety amongst medical students in Europe and English speaking world outside North America prevalence of anxiety was very variable between studies, ranging from 7.7.% to

65.5% (Hope & Henderson, 2016). In one study using HADS with a UK medical student sample found prevalence of 39.2% for anxiety using an 8+ cut-off value (Ashton and Kamali, 1995). Newbury-Birch et al (2001), found prevalence levels of 47% for second year medical students which dropped to 26% by final year again using 8+ as the cut-off. Given the variable level of anxiety in medical students from these studies it is difficult to determine how typical the prevalence of anxiety found in current study is compared to other samples of medical students. The true prevalence of anxiety in the current sample of medical students (as determined by the SCAN interviews) of 8% is not dissimilar to that seen amongst 16-24 years as seen in the 2014 Adult Psychiatric Morbidity Survey (APMS) (prevalence amongst UK 16-24 year olds was 6.3%) (Stansfeld, et al, 2016). This may suggest that prevalence of anxiety in medical students may only be slightly higher than the general population of similar age.

5.2.3 Do the responses to individual items within the HADS subscales truly reflect the presence of anxiety and depression.

A further area considered in the present project is interpretation of the individual items of the scales or instruments. For both HADS subscale item responses there was disparity, for students who scored highly on the HADS subscales, but which did not reach the threshold for clinical diagnosis, between the responses given on the HADS items and the answers given by the same medical students in clinical interview. The HADS-D subscale items in which it seemed students tended to report highly were '*I feel as if I am slowed down*' and '*I can enjoy a good book or radio or TV Programme'*. Within the HADS A subscale items students seemed to report highly on the items ''*Worrying thoughts go through my mind*' and '*I feel restless as if I have to be on the move'*. When these high scoring items were further explored, for all the items, only a small number of students actually experienced these symptoms to a clinically significant level as determined by their SCAN interviews.

Andrews et al (2006) also explored the interview responses of some of their 'non-cases' (according to their diagnostic interviews) who had high anxiety HADS scores (11+). Andrews et al, do not report qualitatively on the types of responses given, instead they identified the symptoms which were present in these students and the length of time the symptoms were present for. Andrews et al found that a third of these non-cases with high HADS-A scores, had

high levels of worry and difficult controlling worry. In addition, Andrews et al found that these symptoms were not present for more than a month. Andrews et al conclude, given these students' responses that "much of the high anxiety that students endorse on questionnaires may reflect transient feelings and responses to transient situations. Examples of such situations are likely to include coursework deadlines, and relatively short-lived relationship problems".

Analysis of the interview responses in the present study found that students who were non-cases but had high HADS-A scores, as with the Andrews study, reported excessive worrying but not to a clinically significant level. These symptoms did not reach a clinically significant level either due to the symptoms not being present for more days than not and/or for not being present for at least a month. Whilst as Andrews et al suggests this may suggest that students' anxiety is transient or linked to transient situations, qualitative examination of the current students interview responses suggest otherwise. Whilst the medical students in the current study talk about worrying about work, their worry appears to not be about an individual deadline but a greater feeling that they are not doing enough or that they won't pass the course. Whilst this worry may not be present continually or to a significant level, neither does it appear for medical students to be linked to a transient situation such as a coursework deadline. This may reflect the differences between medical students and other students. Medical students have a much more academically demanding course than many other students and may have many deadlines and exams (Yiu, 2005). Wider literature has also suggested that much of medical students' stress is course related suggesting common problems include workload, fear of falling behind and exam failure worries (Malik, 2000). Results from the current study suggest that for many of the medical students interviewed this is true, though, for the majority of students these worries do not reach a clinically significant level.

Also, in contrast to Andrews et al's students, the medical students in the present study who were non-cases but had high HADS-A scores also reported symptoms of restlessness and described finding it difficult sitting working for long periods. Again, these findings may be the result of the demanding nature of the course, combined with the high contact hours experienced by medical students, meaning medical students may end up working into the evening resulting in feelings of restlessness.

Andrews et al did not look at the responses of students who were non-cases for depression but had high HADS-D scores. In the present study some of the medical students in this category interpreted feeling slowed down with feeling behind with work. In the introduction for this study, the competitive nature of the medical course was discussed. Medical students feel the pressure to not only pass the course but to be seen as one of the top in the class, their peers being those whom they will most likely be up against in the future for jobs (Mahajan 2010). In addition, the shift in position compared to current peers (going from being one of the top achievers at school to average at medical school) may cause stress and anxiety (Dunn et al, 2008). This shift in class position compared to being at school , or feeling that you are not performing as well as your peers, may cause students to feel that they are getting behind on the course, which some medical students interpreted as moving more slowly compared to their peers. This feeling may explain the high responses to this HADS-D item. However, on exploration of this symptom with the students many did not feel that they were physically moving more slowly suggesting that the students' interpretation of this symptom is not in line with what the item is designed to measure.

The other HADS-D item scored highly by students who were non-cases for depression but had high HADS-D scores was '*I can enjoy a good book or radio or TV Programme'.* Some students in this category talked about time restraints and the feeling of guilt if they spent time not working. As discussed in the introduction on attending medical school, students may experience many lifestyle changes resulting from a lack of leisure time (Graham, 2001). Spending time on leisure and recreational activities may cause medical students to feel guilty and finding a beneficial work-life balance is something many doctors still haven't achieved (Rich et al, 2016). However, whilst these students are not able to enjoy these things, it does not appear to reach a clinically significant level.

In summary, For both HADS subscale item responses there was disparity, for students who scored highly on the HADS subscales but which did not reach the threshold for clinical diagnosis, between the responses given on the HADS items and the answers given by the same medical students in clinical interview. When these high scoring items were further explored, for all the items, only a small number of students actually experienced these symptoms to a

clinically significant level as determined by their SCAN interviews. In all cases, some of the medical students talked in their interviews about other aspects broadly relating to these symptoms. In many cases these aspects were things relating to their course such as feeling behind with work, not having time for leisure activities, worrying about the course and feeling restless when sitting for long hours. In many cases these students did not interpret and answer the HADS items in a way in which would suggest that they were experiencing the symptoms to a clinically significant level as described in the ICD-10. Whilst these aspects may make medical students more likely to interpret the HADS questions differently it does not necessarily mean that these aspects cause medical students to suffer more depression and anxiety. This may explain why some medical students may score highly on HADS when they are not experiencing anxiety or depression. Andrews et al reach a similar conclusion, although only in regard to anxiety stating that "students' self-reports of anxiety symptoms might have a different significance or meaning to that of other groups in the population" (Andrew et al, 2006 p33).

5.2.4 Summary of findings in the context of wider literature.

This study provides new support for HADS with this population. However, additional psychometric properties of HADS need to be explored with this population, including the factor structure which has been questioned with other populations. A range of HADS subscale cut-off scores have been used in other studies. The optimal anxiety subscale score for medical students as identified in the current study matches that found by Andrews et al (2006) in their study with other students. In contrast, Andrews et al suggest a cut off of 10+ for the HADS-D subscale whereas in the current study found an optimal cut off of 7+ for the HADS-D subscale. The use of different cut-off scores could affect the prevalence of depression and anxiety found in research studies. Prevalence found in the current study using the SCAN data suggests that suggesting that medical students may only have a slightly higher prevalence of depression and anxiety than the general population of a similar age.

The disparity between the responses given on some HADS items and the answers given by the same medical students in clinical interview were discussed. Aspects relating to such as feeling behind with work, not having time for leisure activities, worrying about the course and feeling restless when sitting for long hours, were highlighted. These aspects have been highlighted in other studies as potential factors which may cause increases in depression and/or anxiety. Results from the current study suggest they may not cause increase in depression and/or anxiety but rather influence how medical students interpret some of the items on self-report scales. This is supported by Andrews et al who suggest students may place a different significance to items on self-report scales.

5.3 Strengths and limitations

This study has a number of strengths and limitations, these will now be explored.

5.3.1 Strengths

The main strengths in the present study are the contribution of new data to the study of measurement of depression and anxiety in medical students. This study is the first that the author is aware of which compares medical students' depression and anxiety self-report responses to clinical interview data. In order to make this comparison two validated tools (HADS and SCAN) were used. Whilst HADS has been criticised for use in some studies it has been validated for use in different settings including general practice and community settings and is one of the National Institute for Health and Care Excellence (NICE) recommended tools for diagnosis of depression and anxiety (Stern, 2014). SCAN and has been used as a standard against which the validity of other instruments is tested and has good reliability and validity (Nienhuis et al, 2010). Use of the standardised SCAN technique provides confidence in the diagnoses of depression and anxiety which were used to compare to HADS. In addition, interviewer's data was checked and validated between the two researchers and supported by

a psychologist with extensive experience at using and training in the use of the SCAN instrument. Thus, two validated instruments were used, including a gold standard test and one endorsed by NICE (for determining severity of depression) providing high face validity and more confidence in the results of the study.

The two tools, were completed during the same face to face meeting, providing further strength to the analysis and comparison. This is in contrast to the Andrews et al study where there was 'some weeks' between HADS being completed and the interview being undertaken which may have affected the results of the study. This is particularly important, given Andrew et al's conclusion that students high anxiety as recorded by HADS-A but not by interview, may be due to students' anxiety being transient or linked to transient situations. A delay of several week between the two instruments being completed may result in a change of individuals' situation and therefore there may have been a change in the students' HADS scores during this time. This may have contributed to the lower sensitivity and specificity scores seen in the Andrew et al study in comparison to the present study. In the current study medical students completed HADS and then immediately undertook the clinical interview. The interviewer did not study or score the HADS questionnaires prior to undertaking the interview to ensure that the interview was not influenced by any HADS responses. Scoring of both HADS and the interview data was carried out subsequent to the face to face meeting.

The study also investigated the items and diagnostic decisions within the two tools to understand the potential differences in responses that might be related to interpretation of items and thus the efficacy of HADS as a screening tool for medical students. Whilst Andrews et al studied the type and duration of symptoms and how they related to the students' HADS scores they did not explore qualitatively the responses given in the interviews. Qualitative exploration of the interview responses in the current study provided further information regarding how students might interpret the questions relating to the symptoms of depression and anxiety. This is important as it is widely acknowledged that there are a number of differences between medical students and other students or the general population as discussed in the introduction. The author believes this is the first time this type of analysis has been undertaken and adds strength to interpretation of the results and how these data might be used in the future to support students.

5.3.2 Limitations

This study has identified five main limitations these being the small number of participants, the small number of medical students with depression or anxiety, the self-selection of students to participate, sequencing bias and the limited demographic data. Each of these limitations will be discussed and their possible effect on the results of the study will be highlighted.

The first limitation to the study is the small number of medical students involved in the study; only 50 medical students were recruited to the study. This study was designed as a pilot project to investigate the suitability of HADS as a screening tools for medical students. As this was a pilot project with limited funds, the project did not aim to recruit a 'calculated sample size' of students. Calculation of the sample size needed for an effective validation of a screening tool is difficult. The sample size calculation requires knowledge of a number of factors which are not all know for this sample. These factors include knowing the prevalence of a condition within a population and the sensitivity and specificity of the diagnostic tool used (Bujan & Adnam, 2016). As discussed in the introduction, the prevalence of depression and anxiety in medical students varies significantly in different studies. Of the seven UK studies identified in the Rosenstein et al (2016) meta-analysis, prevalence ranged from 4.4 % (Newbury-Birch et al, 2001) to 48.8%. (Honney et al, 2010). In a systematic review of studies looking at anxiety amongst medical student prevalence of anxiety ranged from 7.7.% to 65.5% (Hope & Henderson, 2016). The higher the prevalence of the condition in the sample being tested the lower the sample size required. Bujan and Adman's (2016) guide for calculating sample size suggests that the minimum sample size required for a screening test validation study for a condition with a prevalence of 5% would be 400 participants with 20 having the condition. For a prevalence of 50% minimum sample size would be 40 with 20 having the condition. This suggests that if the prevalence of depression and/or anxiety in medical students was as high as the highest prevalence seen in some studies than the whole sample size would be sufficient, but the number of students with depression or anxiety would not be sufficient. However, it is unlikely that the prevalence of depression or anxiety is as high as suggested by some studies. True prevalence found in the current study was 6% for depression and 8% for anxiety, at these prevalences the sample size used in the current study falls significantly short. This information could be used to determine sample sizes for a larger scale study.

The low number of students with depression and anxiety in the current sample, is also a limitation to the current study. Whilst recruitment methods aimed to target students who may have had depression and anxiety (e.g. recruiting students via the student support unit), few students reached the threshold for depression and or generalised anxiety disorder. Andrews et al, pre-screened the students recruited for their study to increase the number of students who scored above 8 on HADS. However, particularly with the HADS anxiety scale, a score of 8 and above would not necessarily mean that the individual had anxiety when interviewed. Thus, pre-screening may not have increased the number of students with true depression or anxiety. In addition, this pre-screening in the Andrew et al study resulted in the delay between HADS being completed and the interview taking place, which itself brings about difficulties as highlighted in section 5.2.1 above. The current study aimed to as much as possible ensure that the interviewers were blind to the medical students HADS results and therefore screening the students on the day, prior to interview would have proved difficult.

In addition to the low numbers overall and the low number of medical students with depression and/or anxiety, the current study is limited by the self-selection of students to take part. Inherent in this type of study is selection bias. Selection bias suggests that those who chose to take part in the study may have different characteristics than those who choose not to take part. As discussed in the introduction some medical students may feel that there is stigma attached to mental health conditions. It is possible that students who feel this, may be less likely to take part in a project looking at mental health conditions, particularly if they feel that they may be suffering from one. In addition, medical students with depression or anxiety may be more likely to take part in a project in which they only have to complete a mental health tool or scale anonymously, rather than one in which they are required to undertake an interview. This may explain differences in prevalence between the current study and others in which students have not had to undertake an interview. The limited number of students and particularly those with depression and or anxiety, coupled with the self-selecting nature of the study limits the generalisability of the findings to the full medical student population.

Another form of bias which may have affected the results of the current study is order-effect bias; HADS was always completed prior to the clinical interviews in the face to face meetings. Completing HADS prior to the interviews may have influenced how participants then answered questions in the interview. For example, if respondents when asked to make a quick response

to a HADS item, responded very negatively, they may have felt that they had to provide a similar response in their clinical interview. However, had they been asked the same question in the interview, without having first completed HADS, they may have given the response more consideration and not responded as negatively. The failure to switch the order of completion of the two tools would mean that any order effects were not mitigated. The order of the questionnaire and interview completion was not switched due to the instructions provided for completion of the two tools; HADS asks responders not to think too long about their responses but to try and give their immediate reaction. In contrast the nature of the clinical interview encourages participants to consider their symptoms and explore their severity, duration and impact. Switching the order would have meant that if the clinical interviews were completed first by some participants, they would had given consideration to similar questions to those asked in HADS and therefore may have given a more considered response to HADS items, which is contrary to the HADS instructions. In addition, the follow up questions and for the SCAN clinical interviews, allows for each symptom to be explored in detail to help prevent bias.

The final limitation identified in the current study is the limited demographic information included in the study. Due to the sensitivity of investigating mental ill health in a population of medical students who are known to be concerned about disclosing mental ill health, limited demographic information was collected to try and increase recruitment. Whilst year of study and gender were reported, these were not linked to participants data. This meant that analysis of the data in relation to demographics was not possible. Given the small number of participants, it is likely that breaking down the analysis via demographics would have proved difficult. Collection of further demographics such as ethnicity and age would have provided a more rounded picture of the sample included in the study. However, the more demographic information collected, the more likely it is that someone could be identified as having taken part and may possibly have put students off participating. As student demographic data was not linked to their results, it was not possible to explore the effect in the current study of the large number of students who took part from year 2. If as has been suggested by Newbury-Birch et al (2001) medical student anxiety decreases from year two to year five, high numbers of students recruited from year two in the current study may have increased the prevalence of anxiety in this sample. All years of study were approached in the same way to take part in the project, but a higher number of students from year two responded. This high response may be

due to a number of factors including students in higher years of study being away on placements.

In summary, a number of strengths and weaknesses have been identified for the present study. The study is the first to study to examine rates of depression and anxiety in medical students comparing self-report responses to clinical interviews. The study strengths were that it used well validated tools to make the comparison. The interviews and self-assessment scales were completed within one sitting which gives greater validity to the comparison. Interview data was checked and validated between the two researchers, supported by a psychologist with extensive experience of using and training in the use of the SCAN instrument. The items and diagnostic decisions within the two tools were also examined to understand the potential differences in responses that might be related to interpretation of items and thus the efficacy of HADS as a screening tool for medical students. It is recognised that there are a number of limitations to the study and any results should be interpreted with these in mind. The students self-selected to take part in the study and only 50 students were recruited. This may have resulted in students who have a particular interest in mental health taking part in the study. This limits the generalisability our findings to the full medical student population. We collected limited demographic data on the students. This was to potentially increase recruitment, as there is recognised sensitivity in investigating mental ill health in a population of medical students who may have concerns about disclosing mental ill health. The results are therefore interesting and important but should be interpreted with caution.

5.4 Implications

Whilst the results of this study should be treated with caution due to the limitations outlined above, a number of implications may still be drawn from the results. If the optimum HADS-D cut-off is lower than used in previous studies, prevalence of depression may be higher than previously reported. In contrast if the optimum HADS-A cut-off is higher than previously used,

prevalence of generalised may be lower than previously reported. If amended cut-offs are required for the HADS-D and HADS-A scales, consideration needs to be given to how previous research studies are interpreted. In addition, these results have implications for medical school, in particular ensuring they do not over pathologise normal responses to stress and distress. These implications will now be discussed in more detail under two headings of research implications and implications for medical schools/universities/support services.

5.4.1 Research implications

The results from the current study has implications for both the interpretation of previous research and possible directions for future research, these will be considered in turn.

5.4.1.1 Previous research implications

The current study suggested that whilst HADS may be a valid tool for use with medical students, the original cut-off values were not optimum for the medical students in the current sample. Optimum cut-off values for the current medical student sample were 7 for the HADS-D scale and 13 for the HADS-A scale. As highlighted in section 5.2.2, a range of HADS cut-off values have been used in previous research with medical students. The use of different cut-off values, and the use of cut-off values which may not be optimum may affect the results of previous studies. One way in which different cut-off values may affect the result of studies is the effect on prevalence on different cut off values. For example, in the current study, true prevalence of depression as determined by the scan interviews was 6%. However, prevalence as determined by HADS-D would be 16% at a cut-off of 8+ and 22% at a cut-off of 7+. For other studies using HADS with UK medical students, prevalence ranged were recorded as 4.4% (Newbury-Birch et al, 2001), 9.6% (Pickard et a, 2000), 1.6% (Ashton and Kamali 1995) and 6.6% (Quince at al, 2012). Whilst these values are not dissimilar to the 6% true prevalence (as determined by SCAN interviews), they are lower than the prevalence found in the current sample as determine by HADS-D with a 8+ cut-off (ie the same as used in the studies cited). If a lower HADS-D cut-off value, as suggested by the results of this study, was used in these other studies, it is likely that the prevalence rate would have been higher.

For anxiety in the current sample true prevalence (as determined by the SCAN interviews) of generalised anxiety disorder was 8%. However, prevalence as determined by HADS would be 50% at a cut-off of 8+ and 16% at a cut-off of 13+. One study using HADS with a UK medical

student sample found prevalence of 39.2% for anxiety using an 8+ cut-off value (Ashton and Kamali, 1995). Newbury-Birch et al (2001), found prevalence levels of 47% for second year medical students which dropped to 26% by final year again using 8+ as the cut-off. Conclusions from these studies therefore suggest that medical students have high levels of anxiety compared to their peers. If a higher cut-off values (e.g. 13 as suggested by this study) was used, it is likely that the prevalence found in these studies would be lower.

If the prevalence's from these cited studies are being compared to other groups (such as in comparison to the general population) then incorrect comparisons and conclusions may be made. Therefore, the results and conclusions of previous studies must be read bearing the results of the current study and the possibility that the original cut-off values may not be optimum, in mind.

5.4.1.2 Future research implications

In addition to the implications for interpreting previous research, this study has some implications for future research. In discussion of the limitations for the current project, the difficulties of the small sample size and small number of medical students with depression and/or anxiety were discussed. It was concluded that these limitations affect the confidence that can be placed in the results of the present study. A larger study should be undertaken to verify the results in this study and to confirm the optimum cut offs for caseness. If a larger sample could be recruited and further demographic details collected and linked to results, further analysis of different groups would also be possible. In addition, a larger study would allow for other aspects relating to the validity of HADS with medicals students to be tested. This could include analysis of the factor structure of HADS as suggested in section 5.2.1. In order to achieve this, consideration would need to be given as to how to increase recruitment. Research has suggested that there are a number of things which can increase recruitment these include, increasing potential participants' awareness of the health problem being studied, its potential impact on their health, and their engagement in the learning process (Caldwell et al, 2010). For medical students this might be achieved by spending time going into medical students lectures and highlighting the issues surrounding mental health in medical students and the need for further research.

This study aimed to try and understand potential disparities between interview responses and responses to some HADS items. Some of the medical students talked in their interviews about other aspects broadly relating to their symptoms but which were not specifically linked to the symptom being asked about. In many cases these students did not interpret and answer the HADS items in a way in which would suggest that they were experiencing the symptoms to a clinically significant level as described in the ICD-10. The author is unaware of other studies looking at this aspect of measuring depression and anxiety and these potential differences in how HADS items are interpreted by medical students requires further exploration. Ali et al (2016) advocate that all tools used for screening common mental disorders should be validated locally via a pilot study, such as the current study. Whilst Ali et al are discussing the use of tools with different countries and in different languages, as highlighted in the introduction, this may also be relevant to a medical student population. In addition to a validation study, Ali et al suggest that focus groups should be carried out with representatives of the population in question. The aim of these focus groups would be to ensure that are questions are correctly understood and to understand local experience and expression of mental illness (Ali et al, 2016). Focus group study would allow for medical students understanding of their own mental health and their interpretation of the items that make up HADS to be explored.

Finally, it would be useful to explore if the results found for the HADS self-report scale are also relevant to other depression and or anxiety scales. This suggestion is also made by Andrews et al (2006) who suggest specifically that that an understanding of whether all self-report anxiety scales, over report anxiety in students is needed. A study of this nature would help to clarify if prevalence of anxiety in medical students is not as high as previously thought, or whether this is an anomaly when using the HADS scale. Once this is know, further exploration of anxiety in medical students can be explored. Medicine is a highly stressful career with challenging events and situations. More need to be understood about how medical students learn to manage these situations and the anxiety symptoms that they may provoke.

5.4.2 Implications for medical schools/Universities/support services

Two aspects of the results of this study may have particular implications for medical schools. The first of these is the specific validation and suggestion of optimum cut-offs for a depression and anxiety scale for use with medical students. The second being the conclusion that the original cut-off value for the HADS anxiety scale is possibly too low and the implications of this for the prevalence of medical student anxiety. These aspects will be considered in turn.

The importance of being able to use sensitive and appropriate self-report tools to guide support and advice for students cannot be underestimated. Having a tool specifically validated for use with medical students is therefore important. A directive from the General Medical Council in the UK, who oversee medical education, has acknowledged the increasing level of distress and mental ill health in students (GMC, 2015). Poor mental health can have significant consequences for medical students including impaired academic performance, academic dishonesty, substance abuse and suicide (Dyrbye et al, 2005). It is therefore of utmost importance that medical students who require support are identified and provided with the help they require as early as possible. The GMC (2015) suggests that medical school need to have processes to identify students who might need support. As highlighted previously (section 1.6) screening for mental health problems is seen as controversial. However, the reluctance of medical students to seek support and treatment for mental health problems drives the view by some that screening is important. Some have suggested this increases the responsibility of medical schools to identify students who may need help (Silva et al, 2017). Others have acknowledged that screening for depression can identify individuals who might otherwise go undetected but have highlighted that it could also lead to other issues such as misdiagnosis and over diagnosis (Thombs et al, 2011). Having a sensitive tool which can be used for screening purposes may therefore be of use for those who wish to screen students. For those who wish to use alternative methods to identify students who may need support for depression and anxiety, self-report tools may still play an important role once a student has been identified. Self-report scales may help to highlight the type of support that students may require. However, these tools are only of use if they accurately measure depression and anxiety symptoms. Receiving a high anxiety self-report score when one has subclinical levels of anxiety may cause students additional worry and may lead to signposting to services which are not required. Over-diagnosis of anxiety could put additional strain on often stretched resources (Brown, 2016).

In addition to the benefits of a potentially more accurate tool and cut off for identifying students suffering from depression and or anxiety, the current study has implications for the planning of support for medical students. Support for students can broadly fall into two categories, support for those currently struggling and preventative support. Medical schools need to understand the true prevalence of mental health conditions to ensure that adequate support is in place for those with mental health concerns. As discussed, if prevalence is determined via the use of self-report tool, ensuring these are sensitive and appropriate cut-off are used is important for providing accurate data. In addition to this support, medical schools need to consider and implement training to increase the 'mental resilience' of the medical student population in the UK. However, medical school also need to ensure that they do not over pathologise normal responses to stress and distress but support coping strategies to everyday events that students might encounter. Medicine recruits a population of high achievers with often perfectionist traits, this along with the high anxiety scores on self-report questionnaire, has led to the conclusion that medical students suffer from high anxiety. The current study suggests that whilst medical student prevalence of anxiety may be slightly higher than age-matched peers, it may not be as high a previously thought. Over pathologising through self-assessment scales that do not take into account the specific population could deliver the opposite of what medical education is trying to achieve. A true understanding of the problems facing medical students is required to ensure emotions can be managed appropriately and emotional intelligence can be developed. These aspects have been found to be key to building resilience (Haaland et al, 2015) along with providing mental health aware environments where emotions such as anxiety are discussed openly and without the fear of stigma.

5.4.3 Summary of implications

In summary, there are a number of implications resulting from the current study, these have been categorised as implications for research and implications for medical schools. Implications for research include the need to consider the cut-off scores used in previous studies and the impact of these on the prevalence of depression and anxiety reported for medical students. If medical students have a tendency to over-report anxiety symptoms on self-report tools then prevalence measured by these tools may be inflated, thus results from previous studies must be read with this in mind. A number of suggestions for future research

are suggested including a larger scale validation study and a focus group study to try and further understand the interpretation medical students give to HADS items. The implications for medical schools highlight the benefits of a sensitive and appropriate self-report tool validated for use with a medical student population. These benefits include the ability to be able to identify students who may be suffering and to assist in the support of students. However, the potential for students who do not reach the threshold for anxiety to report high levels of anxiety symptoms via self-report scales, highlights the potential for over-pathologising symptoms which may be normal responses of medical students to events that they might encounter. Medical schools may choose to explore these 'normal' responses with medical students to highlight to students some of the feelings that they may experience (overwhelmed, guilty) and help students understand the need for self-care and leisure time,

5.5 Conclusions

Having highlighted the lack of studies validating self-report tools for use with medical students, the current study compared medical students HADS scores to their clinical interview data. Results suggest that the HADS scale is an appropriate tool to use in a medical student population. However, the cut-off for 'caseness' for both anxiety and depression subscales in this cohort may need reviewing. This could relate to how some of the individual HADS items are interpreted by the medical student population. Whilst the study provides some useful insights, results must be interpreted with caution due to the small number of students recruited and the low numbers of students with depression and or anxiety. A number of implications of the results of the study have been suggested including the implications for both the interpretations of past research particularly research using HADS to determine prevalence of depression and anxiety in medical students. Suggestions for future studies have been highlighted including a larger scale validation study and a focus group study to try and further understand the interpretation medical students give to HADS items. Finally, implications for medical schools were considered including the possible uses for a depression and anxiety self-report tool validated for use with medical students and the need to avoid over-diagnosis or over pathologising anxiety in medical students.

Today's medical students are tomorrows doctors, medicine is a difficult career with many challenges. Helping to support medical students to look after their health and wellbeing is a

vital aspect of medical education. This education must be guided by what is known about mental health in this group. Despite the many studies in the area there is still much contradiction in what is known about medical students' mental health. The current study suggests that the use of tools which are not validated for use with the population may fuel these contradictions particularly given the possibility of students who do not reach the threshold for anxiety to report high levels of anxiety symptoms via self-report scales. More research is required specifically to explore medical students' anxiety and to determine what is normal behaviour for this population and what is true pathology.

6.0 References

Aalto, A., Elovainio, M., Kivimaki, M., Uutela, A. &, Pirkola, S. (2012). The Beck Depression Inventory and General Health Questionnaire as measures of depression in the general population: A validation study using the Composite International Diagnostic Interview as the gold standard. *Psychiatry Research*. 197(1-2): 163-171.

Abbas, A., Rizvi, S., Hassan, R., Aqeel, N., Khan, M., Bhutto, A., Khan, Z. &, Mannan, Z. (2015). The prevalence of depression and its perceptions among undergraduate pharmacy students. *Pharmacy Education*, 15 (1): 57-63.

Abdel Wahed, W. &, Hassan, S. (2017). Prevalence and associated factors of stress, anxiety and depression among medical Fayoum University students. *Alexandria Journal of Medicine*, 53(1):77-84.

Abdulahovic, D., Kurspahic-Mujcic, A., Dzubur, A. &, Loga-Zec, S. (2018). Depressive Symptoms Among Sarajevo University Students: Prevalence and Socio-Demographic Correlations. *Acta medica academica*, 47(2): 155-164.

Aboraya, A. (2009). Do Psychiatrists Use Structured Interviews in Real Clinical Settings? *Psychiatry*, 6(6): 24–28.

Aghakhani, N., Nia, H., Eghtedar, S., Rahbar, N., Jasemi, M. &, Zadeh, M. (2011). Prevalence of depression among students of Urmia University of Medical Sciences (Iran). *Iranian Journal of Psychiatry and Behavioral Sciences.* 5 (2): 131-135.

Ahmad, S., Rehman, T., Masood, I., Abbasi, W., Bilal, M., Ghauri, A. (2017). Evaluation of depression prevalence and associated demographic risk factors among students of a public sector university: A cross-sectional study. *Value in Health. Conference: ISPOR 22nd Annual International Meeting*. United States. 20 (5):298.

Ahmed, I., Banu, H., Al-Fageer, R. &, Al-Suwaidi, R. (2009). Cognitive emotions: depression and anxiety in medical students and staff. Journal of Critical care, 24(3):e1-7.

Akbari, V., Hajian, A. &, Damirchi, P. (2014). Prevalence of Emotional Disorders Among Students of University of Medical Sciences; Iran. *The Open Psychology Journal, 7,* 29-32.

Akin, A &, Bayram, C. (2007). The Depression Anxiety and Stress Scale (DASS): The study of validity and reliability. *Kuram ve Uygulamada Egitim Bilimleri*. Vol.7(1), 2007: 260-268.

Akobeng, A. (2007). Understanding diagnostic tests 3: Receiver operating characteristic curves. *Acta Paediatric*. 96(5):644-7.

Akvardar, Y., Demiral, Y., Ergor, G. &, Ergor, A. (2004). Substance use among medical students and physicians in a medical school in Turkey. *Social Psychiatry and Psychiatric Epidemiology*.39(6):502-6.

Al Rashed, A., Al-Naim, A., Almulhim, B., Alhaddad, M., Al-Thafar, A., Alali, M., Aleem A., Kashif, S. &, Bougmiza, I. (2019). Prevalence and associated factors of depression among general population in Al-Ahsa, Kingdom of Saudi Arabia: A community-based survey. *Neurology Psychiatry and Brain Research.* 31: 32-36.

Al-Alawi, M., Al-Sinawi, H., Al-Qubtan, A., Al-Lawati, J., Al-Habsi, A., Al-Shuraiqi, M., Al-Adawi, S. &, Panchatcharam, S. (2017). Prevalence and determinants of burnout Syndrome and Depression among medical students at Sultan Qaboos University: A cross-sectional analytical study from Oman. *Archives of Environmental and Occupational Health*. 1-10.

AlFaris, E., Irfan, F., Qureshi, R., Naeem, N., Alshomrani, A., Ponnamperuma, G., Al Yousufi, N., Al Maflehi, N., Al Naami, M., Jamal, A. &, van der Vleuten, C. (2016).

Health professions' students have an alarming prevalence of depressive symptoms: exploration of the associated factors. *BMC medical education*. 16 (1): 279.

Al-Faris, E., Irfan, F., Van der Vleuten, C., Naeem, N., Alsalem, A., Alamiri, N., Alraiyes, T., Alfowzan, M., Alabdulsalam, A., Ababtain, A., Aljabab, S., Bukhari, M., Alsinaidi, O. &, Alofaisan, Y. (2012). The prevalence and correlates of depressive symptoms from an Arabian setting: a wake up call. *Medical teacher*. 34 Suppl 1: S32-36.

Alfonsson, S., Wallin,E. &, Maathz, P. (2017). Factor structure and validity of the depression, anxiety and stress scale-21 in swedish translation. *Journal of Psychiatric and Mental Health Nursing*, 24, 154-162.

Al-Ghafri, G., Al-Sinawi, H., Al-Muniri, A., Dorvlo, A., Al-Farsi, Y., Armstrong, K. &, Al-Adawi, S. (2014). Prevalence of depressive symptoms as elicited by Patient Health Questionnaire (PHQ-9) among medical trainees in Oman. *Asian Journal of Psychiatry*. 8 (1):59-62.

Ali, G., Ryan, G. &, De Silva (2016). Validated Screening Tools for Common Mental Disorders in Low and Middle Income Countries: A Systematic Review. PLoS ONE 11(6): e0156939. https://doi.org/10.1371/journal.pone.0156939.

Alpizar, D., Lagana, L., Plunkett, S. &, French, B. (2017). Evaluating the Eight-Item Patient Health Questionnaire's Psychometric Properties With Mexican and Central American Descent University Students. *Psychological Assessment*. December 04.

Altman, D. &, Bland, M. (1994). Diagnostic tests 1: sensitivity and specificity. *BMJ.* Jun 11;308(6943):1552

Alvi, T., Assad, F., Ramzan, M. &, Khan, F. (201). Depression, anxiety and their associated factors among medical students. *Journal of College of Physicians and Surgeons Pakistan*, 20 (2): 122–6.

Amarasuriya, S. D., Jorm, A. F., & Reavley, N. J. (2015). Prevalence of depression and its correlates among undergraduates in Sri Lanka. *Asian Journal of Psychiatry*. 15: 32-37.

American Psychological Association (2009). ICD vs. DSM. October 2009, Vol 40, No. 9. Retrieved from http://www.apa.org/monitor/2009/10/icd-dsm.aspx

APA (2018), APA Dictionary of Psychology. American Psychological Association. Retrieved from https://dictionary.apa.org/anxiety

American Psychological Association (2017). What is depression. Retrieved from https://www.psychiatry.org/patients-families/depression/what-is-depression

American Psychological Association (2018). APA dictionary of Psychology. Retrieved from https://dictionary.apa.org/anxiety

Amir, M. &, Gillany, A. (2010). Self-reported depression and anxiety by students at an Egyptian Medical School. Journal of Pakistan Psychiatric Society, 7(2).

Andersen, L., Grimsrud, A., Myer, L., Williams, D., Stein, J. &, Seedat, S. (2011). The psychometric properties of the K10 and K6 scales in screening for mood and anxiety disorders in the South African Stress and Health study. *International Journal of Methods in Psychiatric Research*. 20(4):215-23.

Anderson, E., Michalek, E. &, Lam, R. (2002*).* Depression in primary care: Tools for screening, diagnosis, and measuring response to treatment. *British Columbia Medical Journal*. 44 (8): 419.

Andrea, H., Bültmann, U., Beurskens, A., Swaen, G., Schayck, C. &, Kant, I. (2004). Anxiety and depression in the working population using the HADS scale. *Social Psychiatry and Psychiatric Epidemiology,* 39: 637-646.

Andrews, B. &, Wilding, J. (2004). The relation of depression and anxiety to life-stress and achievement in students. British Journal of Psychology, 95(4): 509-521.

Andrews, B., Hejdenberg, J. &, Wilding, J. (2006). Student anxiety and depression: Comparison of questionnaire and interview assessments. *Journal of Affective Disorders*, 95:29-34.

Angst, J., Gamma, A., Bienvenu, O., Eaton, W., Ajdacic, V., Eich, D. &, Rossler, W. (2006). Varying temporal criteria for generalized anxiety disorder. *Psychological Medicine*, 36:1283-1292.

Aniebue, P. &, Onyema, G. (2008). Prevalence of depressive symptoms among Nigerian medical undergraduates. *Tropical Doctor*. 38 (3): 157-158.

Antunez, Z. &, Vinet, E. (2012). Depression anxiety stress scales (DASS - 21): Validation of the abbreviated version in Chilean university students. *Terapia Psicológica,* 30*(3)*, 49-55.

Aragao, J., Freire, M., Nolasco Farias, L., Diniz, S., Sant'anna Aragao, F., Sant'anna Aragao, I., Lima, T. &, Reis F. (2018). Prevalence of depressive symptoms among medical students taught using problem-based learning versus traditional methods. *International Journal of Psychiatry in Clinical Practice*. 22 (2): 123-128.

Arbona, C., Burridge, A. &, Olvera, N. (2017). The Center for Epidemiological Studies Depression Scale (CES-D): Measurement equivalence across gender groups in Hispanic college students. Journal of Affective Disorders. 219: 112-118).

Aronin, S. &, Smith, M. (2016). One in four students suffer from mental health problems. Retrieved from https://yougov.co.uk/topics/lifestyle/articles-reports/2016/08/09/quarter-britains-students-are-afflicted-mental-hea

Arslan, G., Ayranci, U., Unsal, A. &, Arslantas, D. (2009). Prevalence of depression, its correlates among students, and its effect on health-related quality of life in a Turkish university. *Upsala Journal of Medical Sciences*. 114 (3): 170-177.

Ashton, C. &, Kamali, F. (1995). Personality, lifestyles, alcohol and drug consumption in a sample of British medical students. *Medical Education* 29(3):187-92.

Aydemir, O., Guvenir, T., Kuey, L. &, Kultur S. (1997). The validity and reliability of Turkish version of Hospital Anxiety and Depression Scale. *Journal of Turkish Psychiatry* 8:280–287.

Badpa, S., Hosseinian, T. &, Zarin, F. (2015). Investigating the prevalence of depression and its factors among university students. *Research Journal of Pharmaceutical, Biological and Chemical Sciences.* 6 (2): 25-31.

Baigent, M., Abi-Abdallah, C., & Achal, R. (2015). Prevalence of depression, anxiety and burnout in graduate entry medical students and its relationship to the academic year. *Australian and New Zealand Journal of Psychiatry. Conference: Royal Australian and New Zealand College of Psychiatrists*, 49: 102.

Baldassin, S., Alves, T., Andrade, A. &, Martins, L. (2008). The characteristics of depressive symptoms in medical students during medical education and training: a cross-sectional study. *BMC Medical Education*, 8, 60.

Balsamo, M. &, Saggino, A. (2014). Determining a diagnostic cut-off on the Teate Depression Inventory. *Neuropsychiatric Disease and Treatment*, 10: 987–995.

Barreta Do Carno, M., dos Santos, L., Feitosa, C., Fiaccone, R., Silva, N., dos Santos, D., Barreto, M. &, Amorim, L. (2018). Screening for common mental disorders using the SRQ-20 in Brazil. *Revista Brasileira de Psiquiatria*.40:115–122.

Barton, B. (2012). Prevalence and correlates of depression and anxiety disorders in U.S. graduate students. *Dissertation Abstracts International: Section B: The Sciences and Engineering.* Vol.73(1-B): 214.

Basnet, B., Jaiswal, M., Adhikari, B. &, Shyangwa, P. (2012). "Depression among undergraduate medical students." *Kathmandu University Medical Journal* 10(39): 56-59.

Bassols, A., Okabayashi, L., da Silva, A., Carneiro, B., Feijo, F., Guimaraes, G., Cortes, G., Rohde, L. &, Eizirik, C. (2014). First- and last-year medical students: Is there a difference in the prevalence and intensity of anxiety and depressive symptoms? *Revista Brasileira de Psiquiatria*. 36 (3): 233-240.

Bayati, A., Mohammad Beigi, A. &, Mohammad Salehi, N. (2010). Depression prevalence and related factors in Iranian students. *Pakistan Journal of Biological Sciences*. 12 (20): 1371-1375.

Bayati, A., Mohammad Beigi, A.&, Mohammad Salehi, N. (2009). Depression prevalence and related factors in Iranian students. Pakistan Journal of Biological Sciences. 12 (20),1371-1375.

Baykan, Z., Nacar, M. &, Cetinkaya, F. (2012). Depression, anxiety, and stress among last-year students at Erciyes University Medical School. Academic Psychiatry. 36(1):64-65.

Bayram, N. &, Bilgel, N. (2008). The prevalence and socio-demographic correlations of depression, anxiety and stress among a group of university students. *Social Psychiatry and Psychiatric Epidemiology*. 43 (8): 667-672.

BBC news (2018). Bristol university suicides spark mental health alerts. 24th September 2018. Retrieved from https://www.bbc.co.uk/news/uk-england-bristol-45624695.

Beesdo-Baum, K., Wittchen, H., Craske, M., Hoyer, J., Jurkschat, S., El Hadad, A., Knappe, S. (2014). Psychometric properties of the 1-week dimensional anxiety scales for DSM-5 in a non-clinical student sample. *European Neuropsychopharmacology. Conference*, Conference Publication, 24 (2): S579-S580).

Beiter, R., Nash, R., McCrady, M., Rhoades, D., Linscomb, M., Clarahan, M. &, Sammut, S. (2015). The prevalence and correlates of depression, anxiety, and stress in a sample of college students. *Journal of Affective Disorders*. 173: 90-96.

Bellini, L., Baime, M. &, Shea, J. (2002). Variation of Mood and Empathy during Internship. *Journal of the American Medical Association*, 287(23), 3143-3146.

Bentley, K., Gallagher, M., Carl, J., Barlow D. (2014). Development and validation of the overall depression severity and impairment scale. *Psychological Assessment*. 26 (3): 815-830.

Ben-Zeev, D., Scherer, E., Wang, R., Xie, H. &, Campbell, A. Next-generation psychiatric assessment: Using smartphone sensors to monitor behavior and mental health. *Psychiatric Rehabilitation Journal*. Vol.38(3), 218-226.

Beyond Blue. (2013). National Mental Health Survey of Doctors and Medical Students: Monash University, Australia.

Bhat, S., Cherian, A., Thomas, C., Bhat, A., Kumar, S., Arahantabailu, P. &, Math S. (2015). A study on prevalence of depressive symptoms among young adults in a semi Urban city of India. *Indian Journal of Psychiatry. Conference*: 57 (5 SUPPL. 1) S93.

Bilgel, N. &, Bayram, N. (2010). Turkish version of the Depression Anxiety Stress Scale (DASS-42): Psychometric properties. *Archives of Neuropsychiatry*. Vol.47(2): 118-126.

Bitsika, V., Sharpley, C. &, Hartley, L. (2010). The specific environmental antecedents of anxiety and depression among Australian university students: Development of the Effects of University Study on Lifestyle Questionnaire. *The International Journal of Educational and Psychological Assessment*. 4, 2-18.

Bjelland, I., Dahl, A., Haug, T. &, Neckelmann, D. (2002). The validity of the Hospital Anxiety and Depression Scale. An updated literature review. *Journal of Psychosomatic Research,* **52:** 69-77

Blanco, C., Okuda, M., Wright, C. (2008) Mental health of college students and their non-college-attending peers: results from the National Epidemiologic Study on Alcohol and Related Conditions. *Archives of General Psychiatry*, 65, 1429–1437.

BMA (2019). Supporting the mental health of doctors and medical students. British Medical Association, retrieved from https://www.bma.org.uk/collective-voice/policy-and-research/education-training-and-workforce/supporting-the-mental-health-of-doctors-in-the-workforce.

BMA. (2018). Medical students must be given better mental health support to prepare them for emotional toll of career in the NHS, say leading doctors. Retrieved from:

https://www.bma.org.uk/news/media-centre/press-releases/2018/june/medical-students-must-be-given-better-mental-health-support

Bredemeier, K., Spielberg, J., Silton, R., Berenbaum, H., Heller, W. &, Miller, G. (2010). Screening for Depressive Disorders Using the Mood and Anxiety Symptoms Questionnaire Anhedonic Depression Scale. Psychological Assessment, 22(3), 702-710.

Breenman, S., Cotton, S., Fielding, S. &, Jones, G. (2015). Normative data for the Hopsital Anxiety and Depression Scale. *Quality of Life Research*, 24: 391-398.

Brennan, C., Worral-Davies, A., McMillan, D., Gilbody, S. &, House, A. (2010). The Hospital Anxiety and Depression Scale: A diagnostic meta-analysis of case-finding ability. *Journal of Psychosomatic Research,* 69, 371-378.

Brown, P. (2016). The invisible problem? Improving students' mental health. *Higher Education Policy Institute*. Retrieved from https://www.hepi.ac.uk/wp-content/uploads/2016/09/STRICTLY-EMBARGOED-UNTIL-22-SEPT-Hepi-Report-88-FINAL.pdf

Bukhari, S. &, Khanam S. (2015). Prevalence of depression in university students belonging to different socioeconomic status. *Journal of Postgraduate Medical Institute*. 29 (3): 156-159.

Bunevicius, A., Katkute, A. &, Bunevicius, R. (2008). Symptoms of anxiety and depression in medical students and in humanities students: relationship with big-five personality dimensions and vulnerability to stress. *International Journal of Social Psychiatry*, 54(6): 494–501.

Buote, V., Pancer, S., Pratt, M. &, Adams, G. (2007). The importance of friends, friendship and adjustment among 1st year university students. *Journal of Adolescent Research* 22(6):665-689.

Burns, A., Ofer, S., Curry, P., Sexton, E. &, Doyle, F. (2014). Revisiting the dimensionality of the Hospital Anxiety and Depression Scale in an international sample of patients with ischaemic heart disease. *Journal of Psychosomatic Research*, 77, 116–121.

Caci, H., Bayle, F., Mattei, V., Dossios, C., Robert, P. &, Boyer, P. (2003). How does the Hospital and Anxiety and Depression Scale measure anxiety and depression in healthy subjects? *Psychiatric Research*, 118:89-99.

Cairney, J., Veldhuizen, S., Wade, T., Kurdyak, P. &, Streiner, D. (2007). Evaluation of 2 measures of psychological distress as screeners for depression in the general population. *Canadian Journal of Psychiatry*. 52 (2) (pp 111-120.

Caldwell, P., Hamilton, S., Tan, A. &, Craig, J. (2010). Strategies for Increasing Recruitment to Randomised Controlled Trials: Systematic Review. *PLoS Med* 7(11): e1000368.

Calvo, F., Karras, B., Phillips, R., Kimball, A. &, Wolf, F. (2003). Diagnoses, Syndromes and Diseases: A knowledge representation problem. AMIA, 2003:802.

Cameron, I., Crawford, J., Lawton, K. &, Reid, I. (2008). Psychometric comparison of PHQ-9 and HADS for measuring depression severity in primary care. *British Journal of General Practice*, 58(546), 32-6.

Camp-Arias et al 2007 Campo-Arias, A., Diaz-Martinez, L., Rueda-Jaimes, G., del Pilar Cadena-Afanador, L.,Hernandez, N. (2007). Psychometric properties of the CES-D Scale among Colombian adults from the general population. *Revista Colombiana de Psiquiatria*. Vol.36(4) ;664-674.

Campbell, M., Maynard, D., Roberti, J. &, Emmanuel M. (2012). A comparison of the psychometric strengths of the public-domain zung self-rating depression scale with the proprietary beck depression inventory-II in barbados. *West Indian Medical Journal.* 61 (5) (pp 483-488), 2012.

Campos, R., Goncalves, B. &, Thalbourne, M. (2009). Preliminary psychometric data for a Portuguese scale to assess history of depressive symptomatology with a college student sample. *Psychological Reports*. 104 (3): 1015-1018.

Carleton, R., Thibodeau, M., Teale, M., Welch, P., Abrams, M., Robinson, T. &, Asmundson, G. (2013). The Center for Epidemiologic Studies Depression Scale: A Review with a Theoretical and Empirical Examination of Item Content and Factor Structure. *Plos One*, 8(3): e58067.

Casey, D., Thomas, S., Hocking, D. &, Kemp-Casey, A. (2016). Graduate-entry medical students: older and wiser but not less distressed. *Australasia Psychiatry*.24(1):88-92

Chan, R., Xu, T., Huang, J., Wang, Y., Zhao, Q., Shum, D, O'Gorman, J. &, Potangaroa, R. (2012). Extending the utility of the Depression Anxiety Stress scale by examining its psychometric properties in Chinese settings. *Psychiatry Research.* 200 (2-3): 879-883.

Chandavarkar, U., Azzam, A. &, Mathews, C. (2007). Anxiety symptoms and perceived performance in medical students. *Depression and Anxiety*, 24(2): 103-111.

Chandler, R., Wang, P., Ketter, T.&, Goodwin G. (2008). A new US-UK diagnostic project: Mood elevation and depression in first-year undergraduates at Oxford and Stanford universities. *Acta Psychiatrica Scandinavica*. 118 (1): 81-85.

Chang, E., Eddins-Folensbee, F. &, Coverdale, J. (2012). Survey of the prevalence of burnout, stress, depression, and the use of supports by medical students at one school. Academic psychiatry 36 (3): 177-182.

Chang, S., Abdin, E., Shafie, S., Sambasivam, R., Vaingankar, J., Ma, S., Chong, S. &, Subramaniam, M. (2019). Prevalence and correlates of generalized anxiety disorder in Singapore: Results from the second Singapore Mental Health Study. *Journal of Anxiety Disorders.* 66: 102106.

Chen, C., Chen, Y., Sung, H., Hsieh, T., Lee, M. &, Chang, C. (2015). The prevalence and related factors of depressive symptoms among junior college nursing students: a cross-sectional study. *Journal of psychiatric and mental health nursing*. 22 (8): 590-598.

Chen, L., Wang, L., Qiu, X., Yang, X., Qiao, Z., Yang Y. &, Liang, Y. (2013). Depression among Chinese University Students: Prevalence and Socio-Demographic Correlates. *PLoS ONE*. 8 (3): e58379.

Chew-Graham, C. A., Rogers, A., & Yassin, N. (2003). 'I wouldn't want it on my CV or their records': medical students' experiences of help-seeking for mental health problems. *Medical Education, 37*, 873-880.

Chow, W., Schmidtke, J., Loerbroks, A., Muth, T. &, Angerer, P. (2018). The Relationship between Personality Traits with Depressive Symptoms and Suicidal Ideation among Medical Students: A Cross-Sectional Study at One Medical School in Germany. *International Journal of Environmental Research in Public Health*. 15(7): 1462.

Cinarbas, D. (2014). Cross-cultural metric equivalence of somatization, depression, and anxiety scales across Turkish and U.S.A. University students. Dusunen Adam. 27 (2): 94-107.

Cochrane Library (2013). About Cochrane Reviews. Retrieved from https://www.cochranelibrary.com/about/about-cochrane-reviews.

Coentre, R., Faravelli, C. &, Figueira, M. (2016). Assessment of depression and suicidal behaviour among medical students in Portugal*. International journal of medical educatio*n. 7: 354-363.

Cohen, D. Rhydderch, M., Reading, P. &, Williams, S. (2015). Doctors' health: obstacles and enablers to returning to work. *Occupational Medicine*, 65(6): 459-465.

Connor, K. &, Davidson, J. (2003). Development of a new resilience scale: The Connor-Davidson Resilience Scale (CD-RISC). *Depression and Anxiety*, 18: 76-82.

Cooke, R., Barkham, M. &, Audin, K. (2004) Student debt and its relation to mental health. *Journal of Further and Higher Education*, 28, 53–66.

Cosco, T., Doyle, F., Ward, M. &, McGee, H. (2012). Latent structure of the Hospital Anxiety and Depression Scale: A 10-year systematic review. *Journal of Psychosomatic Research*, 72, 180-184.

Coutinho, F., Leitao, C. &, Oliveira, V. (2012). Prevalence of depressive symptoms in Health Science College students of the University of Beira interior. *European Journal of Epidemiology. Conference: IEA-EEF* 2012: 27: S155.

Cox, J., Holden, J.& Sagovsky, R. (1987). Detection of Postnatal Depression: Development of the 10-item Edinburgh Postnatal Depression Scale. British Journal of Psychiatry, 150(6), 782-786.

Coyne, J. &, van Sonderen, E. (2012). No further research needed: Abandoning the Hospital and Anxiety Depression Scale (HADS). Journal of Psychosomatic Research. Vol.72(3), 173-174.

Crawford, J., Cayley, C., Lovibond, P., Wilcon, P. &, Hartley, C. (2011). Percentile Norms and Accompanying Interval Estimates from an Australian General Adult Population Sample for Self-Report Mood Scales (BAI, BDI, CRSD, CES-D, DASS, DASS-21, STAI-X, STAI-Y, SRDS, and SRAS). Australian Psychological Society, 46(1), 3-14.

Cusin, C., Yang, H., Yeung, A. &, Fava, M. (2009). Rating Scales for Depression. Baer, L., Blais, M. (eds). *Handbook of Clinical Rating Scales and Assessments in Psychiatry and Mental Health.* (pp. 7-35). doi:10.1007/978-1-59745-387-5_2.

Cvetkovski, S., Reavley, N. &, Jorm, A. (2012). The prevalence and correlates of psychological distress in Australian tertiary students compared to their community peers. *Australia and New Zealand Journal of Psychiatry*, 46(5):457-67.

Dahlin, M., Joneborg, N., & Runeson, B. (2005). Stress and depression among medical students: a cross-sectional study. *Medical Education, 39*, 594-604.

Davis, A. (2018). Personality, lifestyles, alcohol and drug consumption in a sample of British medical students. Retrieved from https://www.standard.co.uk/news/education/scale-of-student-anxiety-and-distress-is-worry-for-us-all-says-universities-chief-a3976616.html

De Lim, O., Crippa, J. &, Loureiro, S. (2011). Further psychometric study of the Beck Anxiety Inventory including factorial analysis and social anxiety disorder screening. International *Journal of Psychiatry in Clinical Practice*. 15 (4): 255-262.

Deepak, P., Usmani, U., Washdev, W., Mirza, D., Naz, F., Das, K. &, Rehman, R. (2017). Prevalence of depression and anxiety among undergraduate medical students in a government medical college of Karachi. *Journal of Postgraduate Medical Institute*. 31 (3): 281-285.

DeJesus, R., Vickers, K., Melin, G. &, Williams, M. (2007). A System-Based Approach to Depression Management in Primary Care Using the Patient Health Questionnaire-9. *Mayo Clinical Procedings*. 2007;82(11) : 1395-1402.

Demirci, K., Akgönül, M. &, Akpinar, A. (2015). Relationship of smartphone use severity with sleep quality, depression, and anxiety in university students. *Journal of Behavioral Addictions* 4, (2), 85-92.

Demyttenaere, K., Mortier, P., Kiekens, G. &, Bruffaerts, R. (2017). Is there enough "interest in and pleasure in" the concept of depression? The development of the Leuven Affect and Pleasure Scale (LAPS). *CNS Spectrums*: 1-10.

Dere, J., Watters, C., Yu, S., Michael Bagby, R., Ryder, A. &, Harkness, K. (2015). Cross-cultural examination of measurement invariance of the beck depression inventory-II. *Psychological Assessment*. 27 (1): 68-81.

Dhariwal, R., Verma, K., Goyal, S., Baniya, G. &, Dewanand, A. (2018). Prevalence of anxiety symptoms among newly joined medical students: A cross sectional study from northwest rajasthan. *Indian Journal of Psychiatry*. Conference, 60 :S104-S105.

Dimopoulou, M., Giannopoulou, D., Bartsokas, C., Savoglidou, I., Argyropoulos, K. &, Jelastopulu, E. (2013). Prevalence of anxiety and depression in undergraduate students: A cross-sectional study at the University of Patras, Greece. *European Journal of Epidemiology*. Conference, 28: S118-S119.

Djukanovic, I., Carlsson, J. &, Arestedt, K. (2017). Is the Hospital Anxiety and Depression Scale (HADS) a valid measure in a general population 65-80 years old? A psychometric evaluation study. *Health and Quality of Life Outcomes*. 15 (1) 193.

Downs, N., Feng, W., Kirby, B., McGuire, T., Moutier, C., Norcross, W., Norman, M., Young, I. &, Zisook, S. (2014). Listening to depression and suicide risk in medical students: the Healer Education Assessment and Referral (HEAR) Program. Academic psychiatry. *The journal of the American Association of Directors of Psychiatric Residency Training and the Association for Academic Psychiatry*. 38 (5): 547-553.

Du, N., Yu, K., Ye, Y. &, Chen, S. (2017). Validity study of Patient Health Questionnaire-9 items for Internet screening in depression among Chinese university students. *Asia-Pacific Psychiatry*. 9 (3) e12266.

Dunn, L., Iglewicz, A. &, Moutier, C.(2008). A Conceptual Model of Medical Student WellBeing: Promoting Resilience and Preventing Burnout. *Academic Psychiatry*, 32(1): 44-53.

Dunstan, D., Scott, N. &, Todd, A. (2017). Screening for anxiety and depression: Reassessing the utility of the Zung scales. *BMC Psychiatry*. 17 (1): 329.

Dyrbye, L., & Shanafelt, T. (2011). Medical Student Distress: A Call to Action. *Academic Medicine, 86*(7), 801-803.

Dyrbye, L., Power, D., Massie, F., Eacker, A., Harper, W. &, Thomas, M, (2010). Factors associated with resilience to and recovery from burnout: a prospective, multi-institutional study of US medical students. *Medical Education*. 2010;44(10):1016–26.

Dyrbye, L., Satele, D., Sloan, J. &, Shanafelt, T. (2013). Utility of a Brief Screening Tool to Identify Physicians in Distress. *Journal of General Internal Medicine*, 28(3), 421-427.

Dyrbye, L., Thomas, M. &, Shanafelt, T. (2005). Medical Student Distress: Causes, Consequences, and Proposed Solutions. *Mayo Clinic Proceedings*, 80(12),1613-1622.

Dyrbye, L., Thomas, M. R., & Shanafelt, T. D. (2006). Systematic review of depression, anxiety, and other indicators of psychological distress among U.S. and Canadian medical students. *Academic Medicine, 81*, 354-373.

Dzubur, A., Abdulahovic, D., Kurspahic-Mujcic, A., Dzubur, A., Loga-Zec, S. &, Skrijelj, V. (2018). Depressive Symptoms Among Sarajevo University Students: Prevalence and Socio-Demographic Correlations. *Acta Medica Academica*. 47(2):155-164.

Eack, S., Greeno, C. &, Lee, B. (2006). Limitations of the Patient Health Questionnaire in Identifying Anxiety and Depression in Community Mental Health: Many Cases Are Undetected. *Research on Social Work Practice*, 16(6), 625-631.

Ediz, B., Ozcakir, A. &, Bilgel, N. (2017). Depression and anxiety among medical students: Examining scores of the beck depression and anxiety inventory and the depression anxiety and stress scale with student characteristics. *Cogent Psychology*. Vol.4(1), 1283829.

Edwards, M., Cheavens, J., Heiy, J. &, Cukrowicz, K. (2010). A Reexamination of the Factor Structure of the Center for Epidemiologic Studies Depression Scale: Is a One-Factor Model Plausible? *Psychological Assessment*, 22(3), 711-715.

Eisenberg, D., Gollust, S., Golberstein, E. &, Hefner, J. (2007). Prevalence and Correlates of Depression, Anxiety, and Suicidality Among University Students. *American Journal of Orthopsychiatry*. 77 (4): 534-542.

Eisenberg, D., Gollust, S., Golberstein, E. &, Hefner, J. (2010). Prevalence and Correlates of Depression, Anxiety, and Suicidality Among University Students. *American Journal of Orthopsychiatry, 77(4): 534-542.*

Eley, D., Leung, J., Hong, B., Cloninger, K. &, Clininger, C. (2016). Identifying the Dominant Personality Profiles in Medical Students: Implications for Their Well-Being and Resilience. PLoS ONE 11(8): e0160028.

El-Gilany, A., Amr, M. &, Hammad, S. (2008). Perceived stress among male medical students in Eygypt and Saudi Arabia: effect of sociodemographic factors. *Annals of Saudi Medicine*, 28(6): 442-448

Elhai, J., Contractor, A., Tamburrino, M., Fine, T., Prescott, M., Shirlet, E., Chan, P., Slembarski R., Liberzon, I., Galea, S. &, Calabrese, J. (2012). The factor structure of major depression symptoms: A test of four competing models using the Patient Health Questionnaire-9. *Psychiatry Research*, 199:169-173.

Ercolani, M., Vohra, R., Carmichael, F., Mangat, K. &, Alderson, D. (2015). The lifetime cost to English students of borrowing to invest in a medical degree: a gender comparison using data from the Office for National Statistics. *Medical education and training* 4 (5).

Facundes, V. &, Ludermir, A. (2005). Common mental disorders among health care students. *Rev Bras Psiquiatr*. 2005;27(3):194-200.

Farahangiz S, Mohebpour F, Salehi A. Assessment of mental health among Iranian medical students: a cross-sectional study. International Journal of Health Science, 10(1):49–55.

Fawzy, M. &, Hamed, S. (2017). Prevalence of psychological stress, depression and anxiety among medical students in Egypt. *Psychiatry Research*. 255: 186-194.

Fernandes, M., Vieira, F., Silva, J., Avelino, F. &, Santos, J. (2018). Prevalence of anxious and depressive symptoms in college students of a public institution. Brazilian Journal of Nursing. 71(5): 2169-2175.

Fineberg, N., Haddad, P., Carpenter, L., Gannon, B., Sharpe, R., Young, A., Joyce, E., Rowe, J., Wellsted, D., Nutt, D. and Sahakian, B. (2013). The size, burden and cost of disorders of the brain in the UK. *Journal of Psychopharmacology*, 27(9), pp.761-770.

Firth-Cozens, J. (1987). The stresses of medical training. In R. Payne & J. Firth-Cozens (Eds.), *Stress in health professionals* (pp. 3-22). Chichester: John Wiley & Sons Ltd.

Firth-Cozens, J., Caceres, V. &, Firth, R. (1999). Speciality choice, stress and personality: their relationships over time. Hospital Medicine, 60(10), 751–755.

Fitzpatrick, O., Biesma, R., Conroy, R. &, McGarvey, A. (2019). Prevalence and relationship between burnout and depression in our future doctors: A cross-sectional study in a cohort of preclinical and clinical medical students in Ireland. *BMJ* Open. 9 (4): e023297.

Fong, T. &, Ho, R. (2013). Factor analyses of the Hospital Anxiety and Depression Scale: a Bayesian structural equation modeling approach. *Quality of Life Research*, 22:2857-2863.

Fong, T. &, Ho, R. (2014). Testing gender invariance of the Hospital Anxiety and Depression Scale using the classical approach and Bayesian approach. *Quality of Life Research*, 23: 1421-1426.

Fong, T., Chan, C., Ho, R., Chan, J., Chan, C. &, Ng, S. (2016). Dimensionality of the Center for Epidemiologic Studies Depression Scale: An exploratory bi-factor analytic study. *Quality of Life Research*, 25: 731-737.

Forkmann, T., Boecker, M., Wirtz, M., Glaesmer, H., Brahler, E., Norra, C.&, Gauggel, S. (2010). Validation of the Rasch-based Depression Screening in a large scale German general population sample. *Health and Quality of Life Outcomes*. 8:105.

Franco-Diaz, K., Fernandez-Nino, J. &, Astudillo-Garcia, C. (2018). Prevalence of depressive symptoms and factorial invariance of the Center for Epidemiologic Studies (CES-D) Depression Scale in a group of Mexican indigenous population. *Biomedica : revista del Instituto Nacional de Salud.* 38: 127-140.

Gabilondo, A., Vilagut, G., Pinto-Meza, A., Haro. &, Alonso, J. (2012). Comorbidity of major depressive episode and chronic physical conditions in Spain, a country with low prevalence of depression. *General Hospital Psychiatry.* Vol.34(5): 510-517.

Gale, L. (2015). Anxiety and depression assessment: Using the Depression Anxiety Stress Scales. Social Wok Practice and Skill. December 11.

Galvan-Molina, J., Jimenez-Capdeville, M., Hernandez-Mata, J. &, Arellano-Cano, J. (2017). Psychopathology screening in medical school students. Gaceta Medica de Mexico. 153 (1): 5-87).

Gan, G &, Hue, Y. (2019). Anxiety depression and quality of life of medical students in Malaysia. *Medical Journal of Malaysia,* 74 (1).

Garb, H. (1999). Call for a moratorium on the use of the Rorschach Inkblot test in clinical and forensic settings. *Assessment, 6(4), 313-318.*

Garlow, S., Rosenberg, J., Moore, J., Haas, A., Koestner, B., Hendin, H. &, Nemeroff, C. (2008). Depression, desperation, and suicidal ideation in college students: Results from the American Foundation for Suicide Prevention College Screening Project at Emory University. *Depression and Anxiety*. 25 (6): 482-488.

Geethanjali, B. &, Adalarasu, K. (2014). Assessment of depression, anxiety and stress among female students. *International Journal of Biomedical Engineering and Technology*. 16 (2): 182-193.

Gelaye, B., Tadesse, M., Lohsoonthorn, V.,…&, Williams, M. (2014). Psychometric properties and factor structure of the General Health Questionnaire as a screening tool for anxiety and depressive symptoms in a multi-national study of young adults. Journal of Affective Disorders, 187: 197-202.

George, L. &, Kumar, R. (2008). Diagnostic efficiency of new Rorschach Depression Index (DEPI). *Journal of Projective Psychology & Mental Health*. Vol.15(2): 118-127.

Ghaedi, L., Kosnin, A. &, Mohd, B. (2014). Prevalence of depression among undergraduate students: Gender and age differences*. International Journal of Psychological Research*. Vol.7(2): 38-50.

Ghalechi, E., Kazemi, R. (2010). The evaluation of anxiety prevalence between students of first and last semesters of Ardabil University. *European Psychiatry. Conference*: 25 (SUPPL. 1).

Gilbody, S., Sheldon, T. &, Wessely, S. (2006). Should we screen for depression? 12. *BMJ*, 332 : 1027 – 30

Gill, S., Butterworth, P., Rodgers, B. &, Mackinnon, A. (2007). Validity of the mental health component scale of the 12-item short-form health survey (MCS-12) as measure of common mental disorders in the general population. Psychiatry Research, 152: 63-71.

Gloria, A., Castellanos, J., Kanagui-Munoz, M. &, Rico, M. (2012). Assessing Latina/o undergraduates' depressive symptomatology: Comparisons of the Beck Depression Inventory-II, the Center For Epidemiological Studies-Depression Scale, and the Self-Report Depression Scale. *Hispanic Journal of Behavioral Sciences.* Vol.34(1): 160-181.

GMC (2013). Factors that impact on medical student wellbeing. Retrieved from https://www.gmc-uk.org/about/what-we-do-and-why/data-and-research/research-and-insight-archive/factors-that-impact-on-medical-student-wellbeing.

GMC (2015). Supporting students with mental health conditions. *General Medical Council*, London.

GMC. (2016). Professional behaviour and fitness to practise: guidance for medical schools and their students. *General Medical Council*, London.

Goebert, D., Thompson, D., Takeshita, J., Beach, C., Bryson, P., Ephgrave, K., Kent, A., Kunkel, M., Schechter, J. &, Tate, J. (2009). Depressive symptoms in medical students and residents: a multischool study. *Academic Medicine*, 84(2):236-41.

Goldberg, S., Niemierko, A. &, Turchin, (2015). A. Analysis of data errors in clinical research databases. (2008). AMIA Annual Symposium Proceedings, 6:242-6.

Gomez, R. &, Mclaren, S. (2015). The Center for Epidemiologic Studies Depression Scale: Support for a bifactor model with a dominant general factor and a specific factor for positive affect. *Assessmen*t, 22(3): 351-360.

Gonzalez, D. &, Jenkins, S. (2014). Cross-measure equivalence and communicability in the assessment of depression: a focus on factor-based scales. *Assessment*. 21 (6): 731-741.

Granillo, T. (2012). Structure and function of the Patient Health Questionnaire-9 among Latina and non-Latina white female college students. *Journal of the society for social work and research*, 3(2), 80-93.

Greeson, J., Toohey, M. &, Pearce, M. (2015). An adapted, four-week mind-body skills group for medical students: reducing stress, increasing mindfulness, and enhancing self-care. *Explore (NY)*, 11(3):186-192.

Gress-Smith, J., Roubinov, D., Andreotti, C., Compas, B. &, Luecken, L. (2015). Prevalence, severity and risk factors for depressive symptoms and insomnia in college undergraduates. *Stress and Health*. 31 (1): 63-70.

Gulec, M., Gulec, H., Simsek, G., Turhan, M. &, Sunbul, E. (2012). Psychometric properties of the Turkish version of the Patient Health Questionnaire-Somatic, Anxiety, and Depressive Symptoms. *Comprehensive Psychiatry*. Vol.53(5), 623-629.

Guthrie, E., Black, D., Bagalkote, H., Shaw, C., Campbell M. &, Creed. F. (1998). Psychological stress and burnout in medical students: a five-year prospective longitudinal study. Journal of the Royal Society in Medicine, 91: 237-243.

Haaland, A., Bonga, M. &, Cohen D. (2015). Fuelling resilience through reflective learning. *EAPH* Barcelona.

Habibi, M., Khawaja, N., Moradi, S., Dehghani, M. &, Fadaei, Z. (2014). University Student Depression Inventory: Measurement model and psychometric properties. *Australian Journal of Psychology*. Vol.66(3), 149-157.

Haglund, M., Rot, M., Cooper, N., Nestadt, P.., Muller, D. &, Southwick, S. (2009). Resilience in the third year of medical school. *Academic medicine*, 84(2): 258-268.

Hajebi, A., Motevalian, S., Rahimi-Movaghar, A., Sharifi, V., Amin-Esmaeili, M., Radgoodarzi, R. &, Hefazi, M. (2018). Major anxiety disorders in Iran: Prevalence, sociodemographic correlates and service utilization. *BMC Psychiatry*. 18 (1): 261.

Hajian-Tilaki, H. (2017). The choice of methods in determining the optimal cut-off value for quantitative diagnostic test evaluation. *Statistical Methods in Medical Research*, 27(8), 2374-2383.

Halligan, S., Altman, D. &, Mallett, S. (2015). Disadvantages of using the area under the receiver operating characteristic curve to assess imaging tests: A discussion and proposal for an alternative approach. European Radiology, 25(4), 932-939.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery and Psychiatry*, 23, 56-62.

Hanley, J. &, McNeil, B. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology, 143(1): 29-36.

Hannah, M., Battey, D. &, Benzeval, M. (2013). Common mental disorders and mortality in the West of Scotland Twenty-07 Study: comparing the General Health

Questionnaire and the Hospital Anxiety and Depression Scale. *Journal o*f *Epidemiology* and *Community Health*, 67(7):558-63.

Hasegawa, A., Koda, M., Hattori, Y., Kondo, T. &, Kawaguchi, J. (2013). Longitudinal predictions of the Brooding and Reflection subscales of the Japanese Ruminative Responses Scale for depression. *Psychological Reports*. 113(2):566-85,

Hazibzadeh, F., Habibzadeh, P. &, Yadollahie, M. (2016). On determining the most appropriate test cut-off value: the case of tests with continuous results. Biochemical Medicine, 26(3), 297-307.

Henning, K., Ey, S. &, Shaw, D. (2002). Perfectionism, the impostor phenomenon and psychological adjustment in medical, dental, nursing and pharmacy students. Medical Education, 32(5): 456-464.

Higgins, J. P. T., & Green, S. (2011). Cochrane handbook for systematic reviews of interventions version 5.1.0 [updated March 2011]: The Cochrane Collaboration.

Hillis, J., Perry, W., Carroll, E., Hibble, B., Davies, M. &, Yousef, J. (2010). Painting the picture: Australasian medical student views on wellbeing teaching and support services. *Medical Journal of Australia,* , 192(4): 188-190.

Hinz, A. &, Brahler, E. (2011). Normative values for the Hospital Anxiety and Depression Scale (HADS) in the general German population. *Journal of Psychosomatic Research*. Vol.71(2), 74-78.

Hinz, A., Finck, C., Gomez, Y., Daig, I., Glaesmer, H. &, Singer, S. (2014). Anxiety and depression in the general population in Colombia: reference values of the Hospital Anxiety and Depression Scale (HADS). *Social psychiatry and psychiatric epidemiology*. 49 (1): 41-49.

Hinz, A., Klein, A.M., Brahler, E., Glaesmer, H., Luck, T., Riedel-Heller, S.G., Wirkner, K. &, Hilbert, A. (2017). Psychometric evaluation of the Generalized Anxiety Disorder Screener GAD-7, based on a large German general population sample. *Journal of Affective Disorders*. 210. 338-344.

Hirst,T. (2007). What's the point of university? Retrieved from https://www.theguardian.com/commentisfree/2007/jul/27/whatsthepointofuniversity

Holm, M., Tyssen, R., Stordal, K. &, Haver, B. (2010). Self-development groups reduce medical school stress: a controlled intervention study. BMC Medical Education, 10(23).

Hong, S. &, Min, S. (2007). Mixed Rasch modeling of the Self-Rating Depression Scale: Incorporating latent class and Rasch rating scale models. *Educational and Psychological Measurement*. Vol.67(2): 280-299.

Honney, K., Buszewicz, M., Coppola, W. &, Griffin, M. (2010). Comparison of levels of depression in medical and non-medical students**.** *Clinical Teacher*. 2010 Sep;7(3):180-4.

Hooper, l., Qui, L., Crusto, C. &, Huffman, L. (2012). Scalar Equivalence in Self-Rated Depressive Symptomatology as Measured by the Beck Depression Inventory-II: Do Racial and Gender Differences in College Students Exist. Psychology 3, 762-774.

Hope, V. &, Henderson, M. (2014). Medical student depression, anxiety and distress outside North America: A systematic review. *Medical Education*, 48: 963-97.

Houpy, J., Lee, W., Woodruff, J. &, Pincavage, A. (2017). Medical student resilience and stressful clinical events during clinical training. Medical Education Online, 22(1) doi: 10.1080/10872981.2017.1320187.

Huang, F., Chung, H., Kroenke, K., Delucchi, K. &, Spitzer, R., (2007) Using the Patient Health Questionnaire-9 to Measure Depression among Racially and Ethnically Diverse Primary Care Patients. Journal of *General Internal Medicine* 21(6):547-52.

Hunt, K. &, Gable K. (2013). Prevalence of depressive symptoms and obsessive-compulsive personality traits among pharmacy students. *Currents in Pharmacy Teaching and Learning*. 5 (6): 541-545.

Ibrahim ,A., Kelly, S., Adams, C. &, Glazebrook, C. (2013). A systematic review of studies of depression prevalence in university students. *Journal of Psychiatric Research*, 47(3), 391-400.

Ibrahim, A., Kelly, S. &, Glazebrook, C. (2012). Reliability of a shortened version of the Zagazig Depression Scale and prevalence of depression in an Egyptian university student sample. Comprehensive Psychiatry. 53(5):638-47, 2012.

Ibrahim, M. &, Abdelreheem, M. (2015). Prevalence of anxiety and depression among medical and pharmaceutical students in Alexandria University. *Alexandria Journal of Medicine*. 51 (2): 167-173.

Iqbal, S., Gupta, S. &, Venkatarao, E. (2015). Stress, anxiety and depression among medical undergraduate students and their socio-demographic correlates. Indian Journal of Medical Research, 141(3):354-7.

Irving, G., Neves, A., Dambha-Miller, H., Oishi, A., Tagashira, H., Verho, A. &, Holden, J. (2017). International variations in primary care physician consultation time: a systematic review of 67 countries. BMJ Open;7:e017902.

Ito, M., Bentley, K., Oe, Y., Nakajima, S., Fujisato, H., Kato, N., Miyamae, M., Kanie, A., Horikoshi, M. &, Barlow, D, (2015). Assessing depression related severity and functional impairment: The overall depression severity and impairment scale (ODSIS). PLos One, 10(4): e0122969.

Jackson-Koku, G. (2016). Questionnaire review: Beck Depression Inventory. *Occupational Medicine*, 66: 174-175.

Jadoon, N., Yaqoob, R., Raza, A., Shehzad, M. &, Choudhry, S. (2010). Anxiety and depression among medical students: A cross-sectional study. *Journal of Pakistan Medical Association*, 60(8): 699-702.

Janca, A., Ustun, T. &, & Sartorius, N. (1994). New Versions of World Health Organization instruments for the assessment of mental disorders. *Acta Psychiatrica Scandinavica* 90:73-83.

Jessop, D. C., Herberts, C. &, Solomon, L. (2005) The impact of financial circumstances on student health. *British Journal of Health Psychology*, 10, 421–439.

Johansson, R., Carlbring, P., Heedman, A., Paxling, B. &, Andersson, G. (2013). Depression, anxiety and their comorbidity in the Swedish general population: Point prevalence and the effect on health-related quality of life. *Peer J.*: e98

Johnson, C. S., McLeod, P. J., Sharpe, D., &, Johnston, E. M. (2008). Differences among core dimensions of the Centre for Epidemiological Studies Depression (CES-D) scale across age and gender groups. *Canadian Journal of Community Mental Health, 27*(1), 79-91.

Johnson, J. &, Crenna-Jennings, W. (2018). Prevalence of mental health issues within the student-aged population. https://epi.org.uk/publications-and-research/prevalence-of-mental-health-issues-within-the-student-aged-population/

Jones, R. &, Finlay, F. (2013). Medical students 'experiences and perception of support following the death of a patient in the UK, and while overseas during their elective period. BMJ: Postgraduate Medical journal, 90 (1060).

Julian, L. (2011). *Arthritis Care Research*, 63**;** 0-11.

Kadri, N., Agoub, M., El Gnaoui, S., Berrada, S. &, Moussaoui, D. (2007). Prevalence of anxiety disorders: A population-based epidemiological study in metropolitan area of Casablanca, Morocco. *Annals of General Psychiatry*. 6 : 6.

Kandell, J. (2009). Internet addiction on campus: The vulnerability of college students. CyberPsychology and Behaviour, 1(1).

Kanter, J., Rusch, L., Busch, A. &, Sedivy, S. (2009). Validation of the behavioural activation for depression scale (BADS) in a community sample with elevated depressive symptoms. *Journal of Psychopathological Behaviour Assessment*, 31L 36

Karaoglu, N. &, Seker, M. (2010). Anxiety and depression in medical students related to desire for and expectations from a medical career. T*he West Indian Medical Journal*. 59 (2): 196-202.

Karaoglu, N. &, Seker, M. (2011). Anxiety and depression levels of preclinic years' medical students and probable effective factors. *Preventive Medicine Bulletin*. 10 (3): 303-312.

Kash K.M, et al. Stress and Burnout in Oncology. Oncology 2000;14(11):1621-33.

Kashner, M., Rush, J., Suris, A., Biggs, M., Gajewski, V., Hooker, D,m Shoaf, T. &, Altshuler, K. (2003). Impact of structured clinical interviews on physicians' practices in community mental health settings. *Psychiatric Services*, 54(5):712-718.

Kaya, M., Genc, M., Kaya, B. &, Pehilvan, E. (2007). Prevalence of depressive symptoms, ways of coping, and related factors among medical school and health services higher education students. Turk Psikiyatri Derg, 18(2):137-46.

Kebede, M., Anbessie, B. &, Ayano, G. (2019). Prevalence and predictors of depression and anxiety among medical students in Addis Ababa, Ethiopia. *International Journal of Mental Health Systems.* 13 (1): 30.

Kendrick, T., Dowrick, C., McBride, A., Howe, A., Clarke, P. &, Maisey S. (2009). Management of depression in UK general practice in relation to scores on depression severity questionnaires: analysis of medical record data. *BMJ*;338:b750.

Kenton, W. (2018). Human Development Index. Retrieved from https://www.investopedia.com/terms/h/human-development-index-hdi.asp

Kerr, L. &, Kerr, L. (2001). Screening tools for depression in primary care. Western Journal of Medicine, 175(5): 349–352.

Kessler, R. C., Angermeyer, M., Anthony, J. C., De Graaf, R., Demyttenaere, K., Gasque, I.,. (2007). Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Survey Initiative. *World Psychiatry, 6*, 168–76.

Kessler, R., Brandenburg, N., Lane, M., Roy-Byrne, P., Stang, P., Stein, D. &, Wittchen, H. (2005). Rethinking the duration requirement for generalized anxiety disorder: evidence from the National Comorbidity Survey Replication. Psychological Medicine, 35(7):1073-82.

Khaled, S. (2019). Prevalence and potential determinants of subthreshold and major depression in the general population of Qatar. *Journal of Affective Disorders*. 252: 382-393.

Khoury, B. Langer, E. &, Pagnini. (2014). The DSM: Mindful science or mindless power? A critical review. Frontiers in Psychology, 5: 602.

Khubchandani, J., Brey, R., Koteck,i J., Kleinfelder, J. &, Anderson, J. (2016). The Psychometric Properties of PHQ-4 Depression and Anxiety Screening Scale Among College Students. *Archives of psychiatric nursing*. 30 (4): 457-462.

Kiely, K.M. &, Butterworth, P. (2015). Validation of four measures of mental health against depression and generalized anxiety in a community based sample. *Psychiatry Research*. 225 (3): 291-29.

Kiessling, C., Schubert, B., Schefner, D. &, Burger, W. (2004). First year medical students' perceptions of stress and support: a comparison between reformed and traditional track curricula. Medical Education, 38: 504-509.

Kim, S., Park, S., Choo, I., Kim, S. (2015). Prevalence and factors related to depressive symptoms among Korean university students. *European Neuropsychopharmacology. Conference:* 25 (SUPPL. 2): S357-S358.

Kivrak, Y., Kokacya, M., Sevim, E. &, Copoglu U. (2016). Depressive symptom prevalence and risk factors in eastern Turkish university students. *Journal of Clinical and Analytical Medicine*. 7 (4).

Kjaergaard, M., Arfwedson Wang, C., Waterloo, K. &, Jorde, R. (2013). A study of the psychometric properties of the Beck Depression Inventory-II, the Montgomery and Asberg Depression Rating Scale, and the Hospital Anxiety and Depression Scale in a sample from a healthy population. *Scandinavian journal of psychology*. 55 (1): 83-89.

Knappe, S., Klotsche,, J., Heyde, F., Hiob, S., Siegert, J., Hoyer, J., Strobel, A., Lebeau, R.T., Craske, M.G., Wittchen, H. &, Beesdo-Baum, K. (2014). Test-retest reliability and sensitivity to change of the dimensional anxiety scales for DSM-5. *CNS Spectrums*. 19 (3): 256-267.

Knipe, D., Maughan, C., Gilbert, J., Dymock, D., Moran, P. &, Gunnell, D. (2018). Mental health in medical, dentistry and veterinary students: cross-sectional inline survey. *BJPsych*, 4: 441-446.

Kobak, K., Williams, J. &, Engelhardt, N. (2008). A comparison of face to face and remote assessment of inter-rater reliability on the Hamilton Depression Rating Scale via videoconferencing. *Psychiatry Research,* 158: 99-103.

Kocalevant, R., Hinz, A. &, Brahler, E. (2013). Standardization of the depression screener Patient Health Questionnaire (PHQ-9) in the general population. *General Hospital Psychiatry*, 35: 551-555.

Korszun, A., Dinos, S., Ahmed, K. &, Bhui, K. (2012). Medical student attitudes about mental illness: does medical-school education reduce stigma? *Academic Psychiatry*, 36(3):197-204.

Kötter, T., Tautphäus, Y., Scherer, M. &, Voltmer, E. (2014). Health-promoting factors in medical students and students of science, technology, engineering, and mathematics: design and baseline results of a comparative longitudinal study. BMC Medical Education 14(134): 134.

Kroenke, K., Spitzer, R. &, Williams, J. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*,16(9):606-13.

Kroenke, K., Spitzer, R., Williams, J. &, Lowe, B. (2010). The Patient Health Questionnaire somatic, anxiety, and depressive symptom scales: A systematic review. *General Hospital Psychiatry*. Vol.32(4): 345-359.

Kroenke, K., Strine, T., Spitzer, R., Williams, J., Berry, J., Mokdad, A. (2009). The PHQ-8 as a measure of current depression in the general population. *Journal of Affective Disorders*. 114(1-3):163-73.

Krohne, H., Schmukle, S., Spaderna, H. &, Spielberger, C. (2002) The State-Trait Depression Scales: An International Comparison, Anxiety,Stress & Coping, 15:2, 105-122.

Kulsoom, B. &, Afsar, N. (2015). Stress, anxiety, and depression among medical students in a multiethnic setting. Neuropsychiatr Dis Treat. 2015 Jul 16;11:1713-22.

Kumar, G., Jain, A. &, Hegde S. (2012). Prevalence of depression and its associated factors using Beck Depression Inventory among students of a medical college in Karnataka. *Indian Journal of Psychiatry*. 54 (3): 223-226.

Kung, S., Alarcon, R., Williams, M., Poppe, K., Moore, M. &, Frye, M. (2013). Comparing the Beck Depression Inventory-II (BDI-II) and Patient Health Questionnaire (PHQ-9) depression measures in an integrated mood disorders practice. Journal of Affective Disorders, 145:341-343.

Langvik, E., Hjemdal, O. &, Nordahl, H. (2016). Personality traits, gender differences and symptoms of anhedonia: What does the Hospital Anxiety and Depression Scale (HADS) measure in nonclinical settings? *Scandinavian journal of psychology*. 57 (2): 144-151.

Leahy, C., Peterson,R., Wilson, I., Newbury, J., Tonkin, A. &, Turnbull, D. (2010). Distress levels and self-reported treatment rates for medicine, law, psychology and mechanical engineering tertiary students: cross-sectional study. Australian and New Zealand Journal of Psychiatry, 44:608–615**.**

Lee, A., Tsand, A., Ruscio, A., Haro, J., Stein, D., Alonso, J., Angermeyer, M. &, Kessler, R. (2008). Implications of modifying the duration requirement of generalised anxiety disorder in developed and developing countries. *Psychological Medicine*, 39(7): 1163-1176.

Lee, J. &, Graham, A. (2001). Students' perception of medical school stress and their evaluation of a wellness elective. Medical Education35:652–9.

Lee, S., Ma, Y. &, Tsang, A. (2011). Psychometric properties of the Chinese 15-item Patient Health Questionnaire in the general population of Hong Kong. *Journal of Psychosomatic Research.* Vol.71(2): 69-73.

Lei, X., Xiao, L., Liu, Y., Li, Y. (2016). Prevalence of Depression among Chinese University Students: A Meta-Analysis. *PLoS ONE*. 11 (4): e0153454.

Leray, E., Camara, A., Drapier, D., Riou, F., Bougeant, N., Pelissolo, A., Lloyd, K., Bellamy, V., Roelandt, J. &, Millet, B. (2011). Prevalence, characteristics and comorbidities of anxiety disorders in France: Results from the " Mental Health in General Population" Survey (MHGP). *European Psychiatry*. 26 (6): 339-345.

Lessin, A. &, Pardo, N. (2017). The impact of perfectionism on anxiety and depression. Jounral of Psychology and Cognition, 2(1), 78-82.

Levine, S. (2013). Evaluating the seven-item Center for Epidemiologic Studies Depression Scale short-form: a longitudinal US community study. *Social Psychiatry and Psychiatric Epidemiology,* 48:1519–1526

Levis, B., Yan, X., He, C., Sun, Y., Benedetti, A. &, Thombs, B. (2019). Comparison of depression prevalence estimates in meta-analyses based on screening tools and rating scales versus diagnostic interviews: a meta-research review. BMC Medicine, 17: 65.

Leyfer, O., Ruberg, J. &, Woodruff-Borden, J. (2006). Examination of the utility of the Beck Anxiety Inventory and its factors as a screener for anxiety disorders. *Journal of Anxiety Disorders*, 20(4): 444-58.

Li M., Ding, C., Kanter, J., Zeng, Q. &, Yang, D. (2014). Further Evaluation of the Psychometric Properties of the Behavioral Activation for Depression Scale. *International Journal of Psychology and Psychological Therapy, 14*, 1, 45-57,

Liang, W., Chen, Z., Zhang, H., Zhang, J., Weng, X., Yang, S., Zhang, L., Shen, L. &, Zhang, Y. (2013). 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*. 5 (4): 268-275.

Lim, Y., Kim, Y., Lee, E. &, Kwon, S. (2018). The Penn State Worry Questionnaire: Psychometric properties of the Korean version. *Depression and Anxiety*, 25:E97-E103.

Liu, S., Yeh, Z., Chun-Huang., JuSUn, F., Tjung, J., ChingHwang, L., Shih, Y. &, Yeh, A. (2011). Validation of Patient Health Questionnaire for depression screening among primary care patients in Taiwan. *Comprehensive Psychiatry*, 52(1), 96-101.

Loan, F. (2011). Internet use by the college students across disciplines: a study. *Annals of Library and Information Studies*, 58: 118-127.

Loerch, B., Szegedi, A., Kohnen, R. &, Benkert, O. (2000). The primary care evaluation of mental disorders (PRIME-MD), German version: a comparison with the CIDI. *Journal of Psychiatric Research,*34(3):211-20.

Lotfi, M., Aminian, A., Ghomizadea, A. &, Zarea, S. (2010). Prevalence of depression amongst students of Shaheed Sadoughi University of Medical Sciences, Yazd, Iran. *Iranian Journal of Psychiatry and Behavioral Sciences*. 4 (2): 51-55.

Low, C., Chan, Y., Sung, S. &, Fung, D. (2012). The validation of PHQ-9 and QIDS-SR as screening tools for depression in primary care clinic in Singapore. *Proceedings of Singapore Healthcare Conference.*

Lowe, B., Decker, O., Muller, S., Brahler, E., Schellberg, D., Herzog, W. &, Herzberg, P. (2008). Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. *Medical care*. 46 (3): 266-274.

Lowe, B., Wahl, I., Rose, M., Spitzer, C., Glaesmer, H., Wingenfeld, K., Schneider, A. &, Brahler, E. (2010). A 4-item measure of depression and anxiety: Validation and standardization of the Patient Health Questionnaire-4 (PHQ-4) in the general population. *Journal of Affective Disorders.* 122 (1-2): 86-95.

Lu, W., Bian, Q., Song, Y., Ren, J., Xu, X. &, Zhao, M. (2015). Prevalence and related risk factors of anxiety and depression among Chinese college freshmen. *Journal of Huazhong University of Science and Technology*. 35 (6): 815-822.

Ludwig, A., Burton, W., Weingarten, J., Milan, F., Myers, D. &,, Kligler, B. Depression and stress amongst undergraduate medical students. *BMC Medical Education,* 27;15:141.

Lundin, A. Hallgren, M. &, Forsell, Y. (2015). The validity of the symptom checklist depression and anxiety subscales: A general population study in Sweden. Journal of Affective Disorders, 183: 247-52.

Lundin, A., Hallgren, M., Theobald, H., Hellgren, C. &, Torgen, M. (2016). Validity of the 12-item version of the General Health Questionnaire in detecting depression in the general population. *Public Health*. 136: 66-74.

Lupo, M. &, Strous, R. (2011). Religiosity, anxiety and depression among Israeli medical students. *Israel Medical Association Journal,* 3(10):613-8.

Lyoo, Y., Ju, S., Kim, E., Kim, J. &, Lee, J. (2014). The patient health questionnaire-15 and its abbreviated version as screening tools for depression in Korean college and graduate students. *Comprehensive Psychiatry*. 55 (3): 743-748.

Lyrakos, G.N., Arvaniti, C., Smyrnioti, M. &, Kostopanagiotou, G. (2011). Translation and validation study of the depression anxiety stress scale in the greek general population and in a psychiatric patient's sample. European Psychiatry. *Conference: 19th European Congress of Psychiatry*, 26 (SUPPL. 1)

Macaskill, A. (2012). The mental health of university students in the United Kingdom. *British Journal of Guidance and Counselling*, 41 (4), 426-441.

Madhusudan, M., Aravind, M., Gopi, A., Ajanya, V., Pradeep, V., Fabeena Mangaladan, U. &, Sakeer Hussain, O. (2018). Prevalence and factors associated with depression among the students of a medical college in Kerala*. Indian Journal of Public Health Research and Development*. 9 (12): 59-64.

Mahajan RP. (2010). Critical incident reporting and learning. *British Journal of Anaesthesia;* 105*:* 69–75.

Maideen, S., Sidik, S., Rampal, L. &, Mukhtar, F. (2015). Prevalence, associated factors and predictors of anxiety: A community survey in Selangor, Malaysia. *BMC Psychiatry*. 15 (1): 262.

Makhubela, M. &, Mashegoane, S. (2016). Validation of the Beck Depression inventory-ii in South Africa: Factorial validity and longitudinal measurement invariance in university students. *South African Journal of Psychology*. Vol.46(2): 203-217.

Malik, S. (2000). Students, tutors and relationships: the ingredients of a successful student support scheme. Medical Education, 34(8):635-41.

Mancevska, S., Bozinovska, L., Teece, J. Pluncevik-Gilgoroska J. &, Sivevska-Smilevska, E. (2008). Depression, anxiety and substance use in medical students in the Republic of Macedonia. *Bratisl Lek Listy*, 109(12): 568-572.

Manchevska, S. &, Pluncevic-Gligoroska, J. (2014). The prevalence of high anxiety and substance use in university students in the Republic of Macedonia. *Prilozi,* 35 (2): 67-74.

Manea , L., Gilbody, S. &, McMillan, D. (2015). A diagnostic meta-analysis of the Patient Health Questionnaire-9 (PHQ-9) algorithm scoring method as a screen for depression. *General Hospital Psychiatry,* 37(1): 67-75.

Mann, R., Cheung, J., Lalomiteanu, A., Stoduto, G., Chan, V., Wickens, C., Ala-Ieppilampi, K., Goldbloom, D. &, Rehm, J. (2011). Estimating prevalence of anxiety and mood disorder in survey data using the GHQ12: Exploration of threshold values. *European Journal of Psychiatry*. 25 (2): 81-91.

Manos, R., Kanter, J. &, Luo, W. (2011). The behavioural activation for depression scale – short form: Development and validation. *Behaviour Therapy*, 42: 726-739.

Martin, A., Rief, W., Klaiberg, A. &, Braehler, E. (2006). Validity of the Brief Patient Health Questionnaire Mood Scale (PHQ-9) in the general population. General Hospital Psychiatry.28(1):71-7.

Martin-Merino, E., Ruigomez, A., Wallander, M., Johansson, S. and GarciaRodriguez, L. (2009). Prevalence, incidence, morbidity and treatment patterns in a cohort of patients diagnosed with anxiety in UK primary care. *Family Practice*, 27(1), pp.9-16.

Marwat, M. (2013). Prevalence of depression and the use of antidepressants among third year medical students of khyber medical college, Peshawar. *Journal of Postgraduate Medical Institute*. 27 (1): 26-28.

Maske, U., Buttery, A., Beesdo-Baum, K., Riedel-Heller, S., Hapke, U. &, Busch M. (2016). Prevalence and correlates of DSM-IV-TR major depressive disorder, self-reported diagnosed depression and current depressive symptoms among adults in Germany. *Journal of Affective Disorders*. 190: 167-177.

Matsudaira, T., Igarashi, H., Kikuchi, H., Kano, R., Mitoma, H., Ohuchi, K. &, Kitamura, T. (2009). Factor structure of the Hospital Anxiety and Depression Scale in Japanese psychiatric outpatient and student populations. *Health and Quality of Life Outcomes*. 7.

Maurer, D. (2012). Screening for depression. *American Family Physician*. 85(2):139-44.

Maxim, D., Neibo, R. &, Utell. M. (2014). Screening tests: A review with examples. *Inhalation Toxicology*, 26(13), 811-828.

Mayer, F., Santoa, I., Silveira, P., Lopes, M., De Souza, A., Campos, E., de Abreu, B., Hoffman, I., Magalhaes, C., Lima, M., Almeida, R., Spinardi, M. &, Tempski, P. (2016). Factors associated to depression and anxiety in medical students: a multicenter study *BMC Medical Education***,** 16: 282.

McManus S, Bebbington P, Jenkins R, Brugha T. (eds.) (2016) *Mental health and wellbeing in England: Adult Psychiatric Morbidity Survey 2014*. Leeds: NHS Digital.

Meit, S., Borges, N. &, Early, L. (2007) Personality Profiles of Incoming Male and Female Medical Students: Results of a Multi-Site 9-Year Study, *Medical Education Online*, 12:1, 4462.

Mellal, A., Albluwe, T. &, Al-Ashkar, D. (2014). The prevalence of depressive symptoms and its socioeconomic determinants among university students in Al Ain, UAE. *International Journal of Pharmacy and Pharmaceutical Sciences*. 6 (5): 309-312.

Melo-Carrillo, A., Van Oudenhove, L. &, Lopez-Avila, A. (2012). Depressive symptoms among Mexican medical students: High prevalence and the effect of a group psychoeducation intervention. *Journal of Affective Disorders*. 136 (3):1098-1103.

Mian, S., Pugliese, M., Su, B., Newson, R., Tyrer, P. &, Soljiak M. (2014). Health survey for England 2014 depression analysis and local prevalence model. Imperial College London. Retrieved from https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=2ahUKEwi7zvSogo7nAhVUtXEKHTuKDn8QFjAAegQIARAB&url=https%3A%2F%2Ffingertips.phe.org.uk%2Fdocuments%2FDepression%2520-model-2016-Technical-Document-v1.1%2520(1).docx&usg=AOvVaw0u34PaQG7yzR-_c6irzSCb

Michell, A &, Coyne, J. (2007). DO ultra-short screening instruments accurately detect depression in primary care? *British Journal of General Practice*, 57(535): 144-151.

Midtgaard, M., Ekeberg, O., Vaglum, P. &m Tyssen, R. (2008). Mental health treatment needs for medical students: a national longitudinal study. *European Psychiatry*, 23(7), 505-511.

Mikolajczyk, R., Maxwell, A., El Ansari, W., Naydenova, V., Stock, C., Ilieva, S., Dudziak, U. &, Nagyova, I. (2008). Prevalence of depressive symptoms in university students from Germany, Denmark, Poland and Bulgaria. *Social Psychiatry and Psychiatric Epidemiology*. 43 (2): 105-112.

Miletic, V., Lukovic, J., Ratkovic, N., Aleksic, D. &, Ggurevic, A. (2015). Demographic risk factors for suicide and depression among Serbian medical school students. *Social Psychiatry and Psychiatric Epidemiology*, 50(4), 633-638.

Mitchel, C., McMillan, B. &, Hagan, T. (2017). Mental health help-seeking behaviours in young adults. *British Journal of General Practice*, 67 (654): 8-9

Mitchell, A., Meader, N. &, Symonds, P. (2010). Diagnostic Validity of the Hopsital Anxiety and Depression Scale in cancer and palliatvve settings: A meta-analysis. Journal of Affective Disorders, 126(3), 335-348.

Mitsui, N., Asakura, S., Shimizu, Y., Fujii, Y., Kako, Y., Tanaka, T., Oba, K., Inoue, T. &, Kusumi, I. (2013). Temperament and character profiles of Japanese university students with depressive episodes and ideas of suicide or self-harm: A PHQ-9 screening study. Comprehensive Psychiatry. 54 (8): 1215-1221.

Moffat, K., McConnachie, A., Ross, S. &, Morrion, J. (2004). First year medical student stress and coping in a problem-based learning medical curriculum. Medical Education, 38(5), 482-491.

Mokros, Ł., Witusik, A., Michalska, J., Łężak, W., Panek, M., Nowakowska-Domagała, K. (2017). Sleep quality, chronotype, temperament and bipolar features as predictors of depressive symptoms among medical students. Chronobiology International, 34(6):708-720.

Mondal, A. &, Kumar, M. (2017). A study on Rorschach Depression Index in patients suffering from depression. *Eastern Journal of Psychiatry*, 18(2).

Monrouxe, L. Rees, C., Dennis, I., (2015). Professionalism dilemmas, moral distress and the healthcare student: insights from two online UK-wide questionnaire studies. BMJ Open;5:e007518. doi: 10.1136/bmjopen-2014-007518

Moreno, M., Jelenchick, L., Grant, A., Pumper, M. &, Richardson, L. (2011). Depression disclosures on facebook: Associations with self-report and perceptions of peers. *Journal of Adolescent Health. Conference. Conference Publication*. 48(2): S81.

Moullec , G., Plourde, A. &, Lavoie, K. (2014). Beck Depression Inventory II: determination and comparison of its diagnostic accuracy in cardiac outpatients. *European Journal of Preventative Cardiology*, 22(5):665-72.

Mousa, O.Y., Dhamoon, M.S., Lander, S., and Dhamoon, A.S. (2016). The MD blues: under-recognized depression and anxiety in medical trainees. *PLoS One*, 11: e0156554.

Mundia, L. (2010). The prevalence of depression, anxiety and Stress in Brunei preservice student teachers. *Internet Journal of Mental Health*. Vol.6(2): 1-9.

Muntingh, A., Van der Feltz-Cornelis, C., van Marwijk, H., Spinhoven, P., Penninx, B. &, van Balkom, A.(2011). Is the Beck Anxiety Inventory a good tool to assess the severity of anxiety? A primary care study in the Netherlands Study of Depression and Anxiety (NESDA). BMC *Family Practice*, 4;12:66.

Mykletun, A., Stordal, E. &, Dahl, A. (2001). Hospital Anxiety and Depression scale: Factor structure, item analyses and internal consistency in a large population. *British Journal of Psychiatry,* 179, 540-544.

Nahas, A., Elkalmi, R., Al-Shami, A. &, Elsayed, T. (2019). Prevalence of depression among health sciences students: Findings from a public university in Malaysia. *Journal of Pharmacy and Bioallied Sciences.* 11 (2): 170-175.

Naja, W., Kansooun, A. &, Haddad, R. (2016). Prevalence of Depression in Medical Students at the Lebanese University and Exploring its Correlation With Facebook Relevance: A Questionnaire Study. JMIR Research Protocol. 2016 May 31;5(2):e96

Najafipour, S., Najafipour, F. &, Yaktatalab, S. (2012). The prevalence depression and relationships with academic failure on students of Jahrom university medical science. *European Psychiatry. Conference: 20th European Congress of Psychiatry:* 27.

National Collaborating Centre for Mental Health. The NICE guideline on the management and treatment of depression in adults (updated edition). National Institute for Health and Clinical Excellence, 2010.

National Collaborating Centre for Mental Health. Depression in adults: recognition and management. National Institute for Health and Clinical Excellence, 2009.

Navarro, J., del Moral, R., Alonso, M., Loste, P., Garcia-Campayo, J., Lahoz-Beltra, R. &, Marijuan, P. (2014). Validation of laughter for diagnosis and evaluation of depression. *Journal of Affective Disorders*. 160:43-9.

Neves, J. &, Hillman, N. (2016). HEPI / HEA Student Academic Experience Survey. Retrieved from https://www.hepi.ac.uk/wp-content/uploads/2016/09/STRICTLY-EMBARGOED-UNTIL-22-SEPT-Hepi-Report-88-FINAL.pdf

Newbury Birch, D., Lowry, R. &, Kamali, F. (2002). The changing patterns of drinking, illicit drug use, stress, anxiety and depression in dental students in a UK dental school: a longitudinal study. British Dental Journal, 192(11), 646-649.

Newbury-Birch, D., Walshaw, D. &, Kamali, F. Drink and drugs: from medical students to doctors. *Drug Alcohol Depend*, 64(3):265-270.

Ngasa, S., Sama, C., Dzekem, B., Nforchu, K., Tindong, M., Aroke, D. &, Dimala C. (2017). Prevalence and factors associated with depression among medical students in Cameroon: A cross-sectional study. *BMC Psychiatry*. 17 (1): 216.

NICE, (2009). Depression in adults: Recognition and managements. National Institute for Health and Care Excellence. Leicester.

NICE. (2011). Depression in adults. National Institute for Health and Care Excellence. Retrieved from https://www.nice.org.uk/guidance/qs8/chapter/Quality-statement-1-Assessment.

NICE (2011). Generalised Anxiety Disorder in Adults: Management in Primary, Secondary and Community Care. National Institute for Health and Care Excellence Clinical Guidelines, No. 113; Leicester.

NICE (2019). The NICE guideline on the treatment and management of depression in adults. National Institute for Health and Care Excellence. National Clinical Practice Guideline 90**.** Leicester.

Niemi, P. &, Vainiomaki, P. (2006). Medical students' distress--quality, continuity and gender differences during a six-year medical programme. *Medical Teacher*, 28(2):136-41.

Nienhuis, F., Van de Willige, G., Rijnders, C., De Jonge, P. &, Wiersma, D. (2010). Validity of a short clinical interview for psychiatric diagnosis: the mini-SCAN. British *Journal of Psychiatry.* 2010 Jan;196(1):64-8.

Nieuwenhuijsen, K., de Boer, A, Verbeek, J., Blonk, R. &, van Dijk, F.(2003). The Depression Anxiety Stress Scales (DASS): detecting anxiety disorder and depression in employees absent from work because of mental health problems. *Occupational and Environmental Medicine,*60 Suppl 1:i77-82.

Nilges, P &, Essau, C. (2015). Depression, anxiety and stress scales: DASS--A screening procedure not only for pain patients. *Schmerz*, 29(6):649-57.

NIMH, (2018). Depression. National Institute of Mental Health Retrieved from https://www.nimh.nih.gov/health/topics/depression/index.shtml

Norman, S., Campbell-Sills, L., Hitchcock, C., Sullivan, S., Rochlin, A., Wilkins, K. &, Stein, M. (2011). Psychometrics of a brief measure of anxiety to detect severity and impairment: the Overall Anxiety Severity and Impairment Scale (OASIS). *Journal of Psychiatric Research*. 45(2):262-8.

Norton, P. &, Robinson, C. (2010). Development and evaluation of the anxiety disorder diagnostic questionnaire. *Cognitive Behaviour Therapy*. 39 (2): 137-149.

Norton, S., Cosco T., Doyle. F., Done, J. &, Sacker, A. (2013). The Hospital Anxiety and Depression Scale: a meta confirmatory factor analysis. *Journal of Psychosomatic Research*, 74: 74-81.

Nuevo, R., Lehtinen, V., Reyna-Liberato, P., & Ayuso-Mateos, J. (2009). Usefulness of the Beck Depression Inventory as a screening method for depression among the general population of Finland. *Scandinavian Journal of Public Health*. Vol.37(1), 28-34.

Office of National Statistics (2016). Student suicides in those aged 18 years and above, by sex and usual place of residence indicator, deaths registered in England and Wales between 2001 and 2015. Retrieved from https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/adhocs/005991studentsuicidesinthoseaged18yearsandabovebysexandusualplaceofresidenceindicatordeathsregisteredinenglandandwalesbetween2001and2015

Oh, D, Kim, S, Lee, H., Seo, J., Choi, B. &, Nam, J. (2013). Prevalence and correlates of depressive symptoms in Korean adults: Results of a 2009 Korean Community Health Survey. *Journal of Korean Medical Science*. 28 (1): 128-135.

Ohaeri, J., Awadalla A. &, Gado, O. (2010). Relationship between symptoms of major depressive and generalized anxiety disorders in an Arab population using diagnostic criteria-based instruments. *Medical Science Monitor*. 16(12):PH103-15.

Olson, T.R., Presniak, M.D. &, MacGregor, M.W. (2010). Reevaluating positive affect in the center for epidemiologic studies-depression scale. *Psychiatry Research*. 178 (3): 545-549.

Olsson, I., Mykletun, A. &, Dahl, A. (2015). The hospital anxiety and depression rating scale: A cross-sectional study of psychometrics and case finding abilities in general practice. *BMC Psychiatry,* 5(46):

Opoliner, A., Blacker, D., Fitzmaurice, G. &, Becker, A. (2014). Challenges in assessing depressive symptoms in Fiji: A psychometric evaluation of the CES-D. *International Journal of Social Psychiatry*. Vol.60(4): 367-376.

Oppong Asante, K. &, Andoh-Arthur, J. (2015). Prevalence and determinants of depressive symptoms among university students in Ghana. *Journal of Affective Disorders*. 171: 161-166.

Osman, A., Wong, J.L., Bagge, C.L., Freedenthal, S., Gutierrez, P.M. &, Lozano G. (2012). The Depression Anxiety Stress Scales-21 (DASS-21): Further Examination of Dimensions, Scale Reliability, and Correlates. *Journal of Clinical Psychology*. 68 (12): 1322-1338.

Othieno, C., Okoth, R., Peltzer, K., Pengpid, S. &, Malla, L. (2014). Depression among university students in Kenya: prevalence and sociodemographic correlates. *Journal of Affective Disorders*. 165:120-5.

Oxford Dictionary of English (2010). Oxford University Press. Oxford.

Ozen, N., Ercan, I., Irgil, E. &, Sigirli, D. (2010). Anxiety prevalence and affecting factors among university students. *Asia-Pacific journal of public health / Asia-Pacific Academic Consortium for Public Health*. 22 (1): 127-133.

Pan, X., Wen, Y., Zhao, Y., Hu, J., Li, S., Zhang, S., Li, X., Chang, H., Xue, Q., Zhao, Z., Gu, Y., Li, C., Zhang, Y., Sun, X., Yang, C. &, Fu, C. (2016). Prevalence of depressive symptoms and its correlates among medical students in China: a national survey in 33 universities. *Psychology, health & medicine*. 21 (7): 882-889.

Parikh, R., Mathai, A., Parikh, S., Chandra Sekha, G. &, Thomas, R. (2008). Understanding and using sensitivity, specificity and predictive values. *Indian Journal of Ophthalmology*, 56(1):45-50.

Parikh, R., Sonawalla, S., Sharma-Pandey, S., Farabaugh, A., Yeung, A., Safren S. &, Fava, M. (2009). A cross-cultural, comparative study of the prevalence and symptom patterns of depression among college students. *European Neuropsychopharmacology. Conference*: 19: S378.

Patel, V., Fisher, A., Hetrick, S. &, McGorry, P. (2007). Mental health of young people: a global public health challenge. *Lancet*, 369:1302-13.

Perarson Clinical (2019). Beck depression Inventory. Retrieved from http://www.pearsonclinical.co.uk (forms BDI)

Pez, O., Gilbert, F., Bitfoi, A., Carta, M.G., Jordanova, V., Garcia-Mahia, C., Mateos-Alvarez, R., Prince, M., Tudorache, B., Blatier. C. &, Kovess-Masfety, V. (2010). Validity across translations of short survey psychiatric diagnostic instruments: CIDI-SF and CIS-R versus SCID-I/NP in four European countries. *Social psychiatry and psychiatric epidemiology*. 45 (12): 1149-1159.

Pfalzgraf, A. &, Tipton, D. (2013). Self-reported depression, anxiety, stress, and related coping strategies among pharmacy students. Journal of Pharmacy Practice. *Conference: 16th Annual Meeting of the College of Psychiatric and Neurologic Pharmacists. Conference Publication:*26 (3) (pp 335-336), 2013.

Pickard, M., Bates, L., Dorian, M., Greig, H. &, Saint, D. (2000). Alcohol and drug use in second year medical students at the University of Leeds. *Medical Education*, 34(2):148-50.

Primack, B., Shensa, A., Escoba-Viera, C., Barrett, E., Sidani, J., Colditz, J. &, James, A. (2017). Use of multiple social media platforms and symptoms of depression and anxiety: A nationally representative study among US young adults. Computers in *Human Behaviour*, 69:1-9.

Prinz, P., Hertrich, K., Hirschfelder, U. &, de Zwaan, M. (2012). Burnout, depression and depersonalisation – Psychological factors and coping strategies in dental and medical students. *GMS Zeitschrift für Medizinische Ausbildung* 2012, Vol. 29(1)

Puthran, R., Zhang, M., Tam, W. &, Ho, R. (2016). Prevalence of depression amongst medical students: a meta-analysis. *Medical Education*, 50(4), 456-68.

Quince, T., Wood, D., Parker, R. &, Benson J., (2012). Prevalence and persistence of depression among undergraduate medical students: A longitudinal study at one UK medical school. *BMJ Open*. 2 (4): e001519.

Rab, F., Mamdou, R. &, Nasir, S. (2008). Rates of depression and anxiety among female medical students in Pakistan. *La Revue de Santé de la Méditerranée orientale*, 14(1), 126-133.

Radloff, l. (1977). The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Applied Psychological Measurement*. 1 (3), 385-401.

Raes, F., Hoes, D., Van Gucht, D., Kanter, J. &, Hermans, D. (2010). The Dutch version of the Behavioral Activation for Depression Scale (BADS): psychometric properties and factor structure. *Journal of Behavioural Therapy and Experimental Psychiatry*, 41(3):246-50.

Rami Reddy, K., Rani, P. &, Sandeep, P. (2016). Prevalence of depression and its associated factors among medical students: Andhra medical college, visakhapatnam. Indian Journal of Psychiatry. Conference: 68th Annual National *Conference of the Indian Psychiatric Society, ANCIPS* 2016. India. 58 S99.

Rancans, E., Vrublevska, J., Kivite, A., Ivanovs, R., Logins, R. &, Berze, L. (2016). Prevalence of depression in primary care settings in latvia-the results of the national research program biomedicine. *European Neuropsychopharmacology. Conference*: 26: S478-S479.

Rancans, E., Vrublevska, J., Snikere, S., Koroleva, I. &, Trapencieris, M. (2014). The point prevalence of depression and associated sociodemographic correlates in the general population of Latvia. *Journal of Affective Disorders*. 156: 104-110.

RCP (2011). Mental health of students in higher education. *Royal College of Psychiatrists.* London.

Reed, D., Shanafelt, T. &, Satele, D. (2011). Relationship of pass/fail grading and curriculum structure with well-being among preclinical medical students: a multi-institutional study. *Academic Medicine*.; 86(11):1367-1373.

Reisbig A, Hafen M, White M, & Rush B. (2007). Improving response rates: Introducing an anonymous longitudinal survey research protocol for veterinary medical students. *Journal of Veterinary Medical Education*, 2007: 34(2), 194-201.

Reynolds-Haertle, R., McBride, R. (1992). Single vs. double data entry in CAST. *Controlled Clin Trials,* 13:487-94.

Rich, A., Viney, R., Needleman, S., Griffin, A &, Woolf, K. (2016). You can't be a person and a doctor': the work–life balance of doctors in training—a qualitative study. BMJ Open, 6(12).

Richards, D. &, Sanabria, A.S. (2014). Point-prevalence of depression and associated risk factors. *Journal of Psychology: Interdisciplinary and Applied*. 148 (3): 305-326.

Richardson, T. &, Roberts, E. (2015). The impact of tuition fees amount on mental health over time in British students. *Journal of Public Health*, 37(3): 412–418.

Risal, A., Manandhar, K., Linde, M., Koju, R., Steiner, T.J. &, Holen, A. (2015). Reliability and validity of a Nepali-language version of the hospital anxiety and depression scale (HADS). *Kathmandu University Medical Journal.* 13 (50): 115-124.

Ristic-Ignjatovic, D., Hinic, D., Jakovljevic, M., Fountoulakis, K., Siepera, M. &, Rancic, N. (2013). A ten-year study of depressive symptoms in Serbian medical students. Acta Clinica Croatica. 52(2):157-63.

Rizvi, S., Quilty, L., Sproule B., Cyriac, A., Bagby, M. &, Kennedy, S. (2015). Development and validation of the Dimensional Anhedonia Rating Scale (DARS) in a community sample and individuals with major depression. *Psychiatry Research*, 229:109-119.

Roberts, R., Golding, J., Towell, T. &, Weinreb, I . (1999). The Effects of Economic Circumstances on British Students' Mental and Physical Health. Journal of American College Health, 48(3), 103-109.

Roberts, S., Glod, C., Kim, R. &, Hounchell, J. (2010). Relationships between aggression, depression, and alcohol, tobacco: Implications for healthcare providers in student health. *Journal of the American Academy of Nurse Practitioners,* 22(7):369-75.

Robinson, C.M., Klenck, S.C. &, Norton, P.J. (2010). PsychometriC Properties Of The Generalized Anxiety Disorder Questionnaire for DSM-IV among four racial groups. *Cognitive Behaviour Therapy*. 39 (4): 251-261.

Robinson, P. &, Lowe, J. (2015). Literature reviews vs systematic reviews. Australian and New Zealand Journal of Public Health, 39(2), 103.

Robotham, D. &, Julian, C. (2006). Stress and the higher education student: a critical review of the literature. *Journal of Further and Higher Education*, 30(2), 107-117.

Roh, M., Jeon, H., Kim, H., Han, S. &, Hahm, B. (2010). The prevalence and impact of depression among medical students: A nationwide cross-sectional study in South Korea. *Academic Medicine.* 85 (8): 1384-1390.

Rosal, M., Ockene, I., Ockene, J., Barrett, S., Ma, Y. &, Hebert, J. (1997). A longitudinal study of students' depression at one medical school. *Academic Medicine*, 72(6), 542-546.

Rosenberg, D. (2018). 1 in 5 college students have anxiety or depression, new report reveals. Retrieved from https://www.dailymail.co.uk/health/article-5373403/1-5-college-students-anxiety-depression.html

Rosenstrom, T., Elovainio, M., Jokela, M., Pirkola, S., Koskinen, S., Lindfors, O. &, Keltikangas-Jarvinen, L. (2015). Concordance between Composite International Diagnostic Interview and self-reports of depressive symptoms: a re-analysis. *International journal of methods in psychiatric research.* 24 (3): 213-225.

Roshandel, A., Safavi, M., Rafienejad-Zahra Kheiri, A. &, Mohamadi, A. (2010). Prevalence of substance use and depression among the university students in Iran. *European Psychiatry. Conference*: 25 (SUPPL. 1).

Ross, S. Cleland, J. &, Macleod, M. (2006). Stress, debt and undergraduate medical student performance. Medical Education, 40(6): 584-589.

Rotenstein, L. S., Ramos, M. A., Torre, M., Segal, J. B., Peluso, M. J., Guille, C., . . . Mata, D. A. (2016). Prevalence of depression, depressive symptoms, and suicidal ideation among medical students: a systematic review and meta-analysis. *Journal of the American Medical Association, 316*(21), 2214-2236.

Royal College of Psychiatrists (2003) *The Mental Health of Students in Higher Education* (Council Report CR112). Royal College of Psychiatrists.

Ruiz, F., Garcia M., Suarez Falcon, J. &, Odriozola Gonzalez, P. (2017). The hierarchical factor structure of the Spanish version of Depression Anxiety and Stress Scale -21. *International Journal of Psychology & Psychological Therapy*. Vol.17(1): 97-105.

Sahoo, S. &, Khess, C. (2010). Prevalence of depression, anxiety, and stress among young male adults in India: A dimensional and categorical diagnoses-based study. *Journal of Nervous and Mental Disease.* 198 (12): 901-904.

Saltani, M., Smooth, S., Beck, E. &, Johnson, M. (2015). Universal depression screening, diagnosis, management and outcomes at a student-run free clinic. Academic Psychiatry, 39: 259-266.

Samaranayake, C. &, Fernando, A. (2011). Satisfaction with life and depression among medical students in Auckland, New Zealand. New Zealand Medical Journal, 124(1341): 12-17.

Sasaki, J., Sakamoto, S., Moriwaki, A., Inoue, K. &, Ugajin, K. (2013). The recognized benefits of negative thinking/affect in depression and anxiety: Developing a scale. *Japanese Psychological Research*. Vol.55(3): 203-215.

Schneider S.E, et al. (1993). Depression and Anxiety in Medical, Surgical, and Paediatric Interns. *Psychological Reports*; 72: 1145-6

Schulte-Van Maaren, Y., Giltay, E., Van Hemert, A., Zitman, F., De Waal, M. &, Carlier, I. (2013). Reference values for anxiety questionnaires: The leiden routine outcome monitoring study. *Journal of Affective Disorders.* 150 (3): 1008-1018.

Schwenk, T., Davis, L. &, Wimsatt, L. (2010). Depression, stigma and suicidal ideation in medical student. *Journal of the America Medical Association*, 304(11), 1181-1190.

Sepehrmanesh, Z. &, Moraveji, A. (2015). Prevalence of depression among interns of kashan university of medical science in 2013. *European Psychiatry. Conference: 23rd European Congress of Psychiatry, EPA* 2015. 30: 718.

Serra, R., e Dinato, S. &, Caseiro, M. (2015). Prevalence of depressive and anxiety symptoms in medical students in the city of Santos. *Jornal Brasileiro de Psiquiatria*. 64 (3): 213-220.

Shafiullah, K., Khan, S., Nisar, S., Iftikhar-ud-Din. &, Sadaf, S. (2016). Prevalence of depression among students at Bacha Khan Medical College, Mardan. *Journal of Medical Sciences (Peshawar).* 24 (4) (pp 202-205

Shean, G. &, Baldwin, G. (2012). The latent structure of the center for epidemiological studies- depression scale. *Journal Psychopathological Behaviour Assessment*, 34:502-509.

Shi, M., Liu, L., Wang, Z. &, Wang, L. (2016). Prevalence of depressive symptoms and its correlations with positive psychological variables among Chinese medical students: An exploratory cross-sectional study. *BMC Psychiatry*. 16 (1):3.

Shin, C., Kim, Y., Park, S., Yoon, S., Ko, Y., Kim, Y., Kim, S., Jeon, S., & Han, C. (2017). Prevalence and Associated Factors of Depression in General Population of Korea: Results from the Korea National Health and Nutrition Examination Survey, 2014. *Journal of Korean medical science*. 32 (11): 1861-1869.

Shudo, Y., Yamamoto, T. &, Sakai, M. (2017). Longitudinal predictions of depression symptoms using the activation and avoidance subscales of the Japanese behavioral activation for depression scale-short form. *Psychological Reports*. 120(1), 130-140.

Sidana, S., Kishore, J., Ghosh, V., Gulati, D., Jiloha, R. C., & Anand, T. (2012). Prevalence of depression in students of a medical college in New Delhi: a cross-sectional study. *Australasian Medical Journal, 5*(5).

Siddaway, A., Wood, A. &, Taylor, P. (2017). The Center for Epidemiologic Studies-Depression (CES-D) scale measures a continuum from well-being to depression: Testing two key predictions of positive clinical psychology. *Journal of Affective Disorders*, 213: 180-186.

Silva, V., Costa, P., Pereira, I., Faria, R., Salgueira, A., Costa, M., Sousa, N., Cerqueira, J &, Morgado, P. (2017). Depression in medical students: insights from a longitudinal study. *Medical Education* 17:184.

Simic-Vukomanovic, I., Mihajlovic, G., Kocic, S., Djonovic, N., Bankovic, D., Vukomanovic, V. &, Djukic-Dejanovic, S. (2016). The prevalence and socioeconomic correlates of depressive and anxiety symptoms in a group of 1,940 Serbian university students. *Vojnosanitetski Pregled*. 73 (2):169-177.

Singer, S., Kuhnt, S., Gotze, H., Hinz, A., Liebmann, A., Kraus, O., Lehmann, A. &, Schwarz, R. (2009). Hospital anxiety and depression scale cut-off scores for cancer patients in acute care. British Journal of Cancer, 100:908-912.

Skorikov, V. &, Vandervoort, D. (2003). Relationships between the underlying constructs of the beck depression inventory and the center for epidemiological studies depression scale. Educational and Psychological Measurement, , 63(2): 319-335.

Slavin, S., Schindler, D. &, Chibnall, J. (2014). Medical Student Mental Health 3.0: Improving Student Wellness Through Curricular Changes. Academic Medicine, 89(4): 573-577.

Smarr, K. &, Keefer, A. (2011). Measures of Depression and Depressive Symptoms. Arthritis Care and Research, 63(s11): S454-S466.

Smith, C., Peterson, D., Degenhardt, B. &, Johnson, J. (2007). Depression, anxiety and perceived hassles among entering medical students. *Psychology Health and Medicine,* 12(1), 31-39.

Sobowale, K., Zhou, A., Fan, J., Liu,N. &, Sherer, R. (2014). Depression and suicidal ideation in medical students in China: a call for wellness curricula. International Journal of Medical Education, 5:31-36.

Sokratous, S., Merkouris, A., Middleton, N. &, Karanikola, M. (2014). The prevalence and socio-demographic correlates of depressive symptoms among Cypriot university students: A cross-sectional descriptive co-relational study. *BMC Psychiatry*. 14 (1): 235.

Soltani, M., Smith, S., Beck, E. &, Johnson, M. (2015). Universal depression screening, diagnosis, management and outcomes at a student-run free clinic. *Academic Psychiatry,* 39: 259-266.

Song, Y., Huang, Y., Lui, D. &, Kwan, J. (2008). Depression in college: Depressive symptoms and personality factors in Beijing and Hong Kong college freshmen. Comprehensive Psychiatry, 49(5): 496-502.

Spencer, E., Brassey, J. &, Mahtani, K. (2017). Catalogue of Bias Collaboration: Recall bias**.** https://www.catalogueofbiases.org/biases/recall-bias

Spielberger, C. (2010). Test Anxiety Inventory. In Craighead, W &, Nemeroff, C. The Corsini Encylopedia of Psychology. John Wiley and Sons: New York.

Spitzer, R., Kroenke, K., Williams, J. &, Lowe, B. (2006). A brief measure for assessing generalised anxiety disorder: the GAD-&. *Archives of Internal Medicine*, 166(10): 1092-1097.

Spitzer, R., Kroenke, K., Williams, J. &, Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. Archives Internal Medicine, 166, 1092-1097.

Sreeramareddy, C., Shankar, P., Binu, V., Mukhopadhyay, C., Ray, B. &, Menezes, R. (2007). Psychological morbidity, sources of stress and coping strategies among undergraduate medical students of Nepal. *BMC Medical Education*, 7(26):7-26.

Stallman, H. (2010). Psychological distress in university students: A comparison with general population data. *Australia Psychologist* 45,4, 2010.

Stansfeld, S., Clark, C., Bebbingston, P., King, M., Jenkins, R. &, Hinchliffe, S. (2016). Chapter 2: Common mental disorders. In S. McManus, P. Bebbington, R. Jenkins, & T. Brugha (Eds.), *Mental health and wellbeing in England: Adult Psychiatric Morbidity Survey* 2014. Leeds: NHS Digital.

Stecker, T. (2004). Well-being in an academic environment. *Medical Education*, 38: 465-4778.

Stern, A. (2014). Questionnaire review: The hospital Anxiety and Depression Scale. *Occupational Medicine*, 64: 393-394.

Straat, J., Van der Ark., A. &, Sijtsma, K. (2013). Methodological artifacts in dimensionality assessment of the Hospital Anxiety and Depression Scale (HADS). *Journal of Psychosomatic Research,* 74: 116-121.

Student BMJ (2014). Medical students with mental health problems do not feel adequately supported. Retrieved from: www.bmj.com/company/wp-content/uploads/2014/07/student-bmj-survey.pdf.

Stulz, N., & Crits-Christoph, P. (2010). Distinguishing anxiety and depression in self-report: Purification of the Beck Anxiety Inventory and Beck Depression Inventory- II. *Journal of Clinical Psychology*, 66(9), 927–940.

Sugawara, N., Yasui-Furukori, N., Sasaki, G., Umeda, T., Takahashi, I., Danjo, K., Matsuzaka, M., Kaneko, S. &, Nakaji, S. (2011). Assessment of the Center for Epidemiological Studies Depression Scale factor structure among middle-aged workers in Japan. *Psychiatry and Clinical Neurosciences.* 65 (1): 109-111.

Sujatha Baskaran, B., Manoj Abraham, M. &, Britesaghayarayna, A. (2018). Prevalence rate of generalized anxiety disorder among young adults between age 18-21 years. *Biomedicine (India).* 38 (4): 561-565.

Sunderland, M., Mahoney, A. &, Andrews, G. (2012). Investigating the factor structure of the Kessler Psychological Distress Scale in community and clinical samples of the Australian population. *Journal of Psychopathology and Behavioural Assessment*, 34(2): 253-259.

Sunderland, M., Slade, T., Stewart, G. &, Andrews, G. (2011). Estimating the prevalence of DSM-IV mental illness in the Australian general population using the Kessler Psychological Distress Scale. *Australian and New Zealand Journal of Psychiatry*. 45 (10): 880-889.

Tabuse, H., Kalali, A., Azuma, H., Ozaki, N., Iwata, N., Naitoh, H., Higuchi, T., Kanba, S., Shioe, K., Akechi, T. &, Furukawa, T. (2007). The new GRID Hamilton Rating Scale for Depression demonstrates excellent inter-rater reliability for inexperienced and experienced raters before and after training. *Psychiatry Research*. 153(1):61-7.

Takahashi, O. &, Ohde, S. (2016). Somatic symptoms in patients with undiagnosed depression: Prevalence and health care utilization in Japan. *Value in Health. Conference: ISPOR 19th Annual European Congress. Austria*. 19 (7): A840.

Talkovsky, A. &, Norton, P. (2015). The Mood and Anxiety Symptom Questionnaire across four ethnoracial groups in an undergraduate sample. *American Journal of Orthopsychiatry*. 85(5):431-40.

Talwar, P., Othman, M., Othman, A., Mustaffa, M. &, Mughal, Y. (2017). Socio-demographic determinants and prevalence of depression, anxiety, and stress among Malaysian university students. *Journal of the Indian Academy of Applied Psychology*. Vol.43(2): 296-304.

Talwar, P., Othman, M., Othman, A., Mustaffa, M. &, Mughal, Y. (2017). Socio-demographic determinants and prevalence of depression anxiety, and stress among Malaysian university students. *Journal of the Indian Academy of Applied Psychology.* Vol.43(2): 296-304.

Teismann, T., Ertle, A., Furka, N., Willutzki, U. &, Hoyer, J. (2016). The German Version of the Behavioral Activation for Depression Scale (BADS): A Psychometric and Clinical Investigation. *Clinical Psychology & Psychotherapy*. 23(3):217-25.

Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., Parkinson, J., Secker, J., Stewart-Brown, S., (2007). The Warwick–Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. *Health and Quality of Life Outcomes* 5, 63.

Thibodeau, M. &, Asmundson, G. (2014). The PHQ-9 assesses depression similarly in men and women from the general population. *Personality and Individual Differences*, 56: 149-153.

Thiboutot, C., Nichol,s S., Morris, J. &, Piper, B. (2016). Prevalence of potentially undiagnosed anxiety among pharmacy, nursing, and physical therapy students. *Journal of Pharmacy Practice. Conference*: 19th Annual Meeting of the College of Psychiatric and Neurologic Pharmacists, CPNP. United States. 29 (3): 320-321.

Thombs, B. &, Zieggelstein, R. (2014). Does depression screening improve depression outcomes in primary care? *British Medical Journal*, 4: 348.

Thombs, B., Arthurs, E., El-Baalbaki, G., Ziegelstein, R. &, Steele, R. (2011). Risk of bias from inclusion of patients who already have diagnosis of or are undergoing treatment for depression in diagnostic accuracy studies of screening tools for depression: systematic review. *British Medical Journal, 343.*

Thombs, B., Benedetti, A,m Kloda, L., Levis, B., Azar, M., Riehm, K., Saadat, N., Cuijpers, P,m Gilbody, S,m Loannidism J., McMillan, D., Patten, S., Shrier I., Steele, R.,

Ziegelstein, R., Loiselle, C., Henry, M., Ismail, Z., Mitchell, N. &, Tonelli, M. (2016). Diagnostic accuracy of the Depression subscale of the Hospital Anxiety and Depression Scale (HADS-D) for detecting major depression: protocol for a systematic review and individual patient data meta-analyses. BMJ Open, 6(4).

Thompson, D., Goebert, D. &, Takeshita, J. (2010). A program for reducing depressive symptoms and suicidal ideation in medical students. Academic Medicine, 85(10):1635-9.

Tjia J, Givens JL, Shea JL. Factors associated with undertreatment of medical student depression. *Journal American College Health*. 2005;53(5):219-26.

Trainor, K., Mallett, J. &, Rushe, T. (2013). Age related differences in mental health scale scores and depression diagnosis: Adult responses to the CIDI-SF and MHI-5. J*ournal of Affective Disorders*, 151: 639-645.

Trevethan, R. (2017). Sensitivity, Specificity, and Predictive Values: Foundations, Pliabilities, and Pitfalls in Research and Practice. *Front Public Health*, 5: 307.

Twenge, J. (2009). Generational changes and their impact in the classroom: teaching Generation Me. *Medical Education*, 43(5): 398-405.

Tyrala, K., Seweryn, M., Bonk, M., Bulska, W., Orszulak, K., Bratek, A. &, Krysta, K. (2015). Evaluation of the utility of Liebowitz Social Anxiety Scale and Barratt Impulsiveness Scale in the diagnosis of social anxiety, impulsivity and depression. Psychiatria Danubina. 27: S223-S226.

Tyssen, R., Dolatowski, F. C., Rovik, J. O., Thorkildsen, R. F., Ekeberg, O., Hem, E., Gude, T., Gronvold, N. &, Vaglum, P. (2007). Personality traits and types predict medical school stress: a six-year longitudinal and nationwide study. *Medical Education, 41*(781-787).

Uehara, T., Takeuchi, K., Kubota, F., Oshima, K. &, Ishikawa, O. (2010). Annual transition of major depressive episode in university students using a structured self-rating questionnaire. *Asia-Pacific Psychiatry*. 2 (2): 99-104.

Uglesic, B., Lasic, D., Zuljan-Cvitanovic, M., Bukovic, D., Karelovic, D., Delic-Brkljacic, D., Bukovic, N. &, Radan, M. (2014). Prevalence of depressive symptoms among college students and the influence of sport activity. *Collegium Antropologicum.* 38(1):235-9.

Uher, R., Perlis, R., Placentino, A., Dernovšek, M., Henigsberg, N., Mors, O., Maier, W., McGuffin, P. &, Farmer, A. (2012). Self-report and clinician-rated measures of depression severity: can one replace the other? *Depression and Anxiety*. 2012 Dec;29(12):1043-9.

Umegaki, Y. &, Todo, N. (2017). Psychometric properties of the Japanese CES-D, SDS, and PHQ-9 depression scales in university students. *Psychological Assessment*. 29 (3): 354-359.

Urban, E. (2018). Depression history and memory bias for specific daily emotions. PLoS One, 13(9), e0203574.

Van Ameringen, M., Mancini, C., Simpson, W. &, Patterson, B. (2009). Internet-based screening of anxiety disorders. *European Neuropsychopharmacology. Conference: 22 ECNP Congress. Istanbul Turkey. Conference Publication*: 19 (3):S588-S589.

Van Dam, N. &, Earleywine, M. (2011). Validation of the Center for Epidemiologic Studies Depression Scale-Revised (CESD-R): Pragmatic depression assessment in the general population. *Psychiatry Research*. Vol.186(1): 128-132.

Van hemert, D., van de vijer, F. &, Poortinga, Y. (2002). The beck depression inventory as a measure of subjective well-being: a cross-national study. *Journal of Happiness Studies* **3:** 257–286.

Vankar, J., Prabhakaran, A. &, Sharma, H. (2014). Depression and Stigma in Medical Students at a Private Medical College. *Indian Journal Psychological Medicine*, 36(3): 246–254.

Vasegh, S. &, Baradaran, N. (2014). Using the persian-language version of the beck depression inventory-II (BDI-II-persian) for the screening of depression in students. *Journal of Nervous and Mental Disease*. 202 (10): 738-744.

Venkatesh, P., Badesaab, B., Kumar, A. &, Patil, R. (2014). A study on the prevalence of depression among adolescent medical students, SS institute of medical sciences & research centre, Davangere, Karnataka. *Indian Journal of Public Health Research and Development*. 5 (4): 244-247.

Vera-Villarroel, P., Buela-Casal, G., Zych, I., Cordova-Rubio, N., Celis-Atenas, K., Zepeda, L. &, Spielberger, C. (2010). Chilean experimental version of the state-trait depression questionnaire (ST-DEP): State subscale (S-DEP). *Psychological Reports.* 106 (1): 65-77.

Verger, P., Guagliardo, V., Gilbert, F., Rouillon, F. &, Kovess-Masfety, V. (2010). Psychiatric disorders in students in six French universities: 12-month prevalence, comorbidity, impairment and help-seeking. *Social psychiatry and psychiatric epidemiology.* 45 (2): 189-199.

Verma, S. (2011). Prevalence of depression among 1st year medical students and it's reflection on their academic performance. *Indian Journal of Physiology and Pharmacology. Conference: 57th Annual Conference of Association of Physiologists and Pharmacologists of India*. (5 SUPPL. 1): 250-251.

Verma, S.K., Mishra, A., Kumar, S. &, Dixit, V. (2013). Could academic performance be a tool for screening depression: A cross-sectional study among 1st year medical students? *International Journal of Pharma and Bio Sciences*. 4 (3): B749-B756.

Videnova, V., Manuseva, N., Gjorgjevikj, D. &, Isjanovski, V. (2016). Screening for depression in medical students and sports academy students. *European Neuropsychopharmacology. Conference*: 26 (2): S389.

Vilagut G., Forero C., Barbaglia G. &, Alonso, J. (2016). Screening for Depression in the General Population with the Center for Epidemiologic Studies Depression (CES-D). *PLoS ONE*. 11(5):e015543.

Vilagut, G., Forero, C., Pinto-Meza, A., Haro, J., de Graaf, R., Bruffaerts, R., Kovess, V., de Girolamo, G., Matschinger, H., Ferrer, M.&, Alonso, J. (2013). The mental component of the Short-Form 12 Health Survey (SF-12) as a measure of depressive disorders in the general population: Results with three alternative scoring methods. *Value in Health*. Vol.16(4): 564-573.

Vivek, U., Ajeesh, P. Pabhakaran, A. &, Kumar, V. (2011). Prevalence of depressive symptoms among junior residents of medical college trivandrum- A cross sectional study. Indian Journal of Psychiatry. Conference Publication: 53 (5 SUPPL. 1): S108).

Voltmer, E., Kotter, T. &, Spahn, C. (2012). Perceived medical school stress and the development of behavior and experience patterns in German medical students. Medical teache, 34(10): 840-847.

Vrublevska, J., Rancans, E., Trapencieris, M., Snikere, S. Wahl. (2013). 12-month and point prevalence of depression and related characteristics in Latvian population assessed by the Mini International Neuropsychiatric Interview (MINI) in 2012. *European Neuropsychopharmacology. Conference: 26th European College of Neuropsychopharmacology, ECNP Congress*. Barcelona Spain. 23: S335).

Vrublevska, J., Trapencieris, M., Snikere, S., Grinberga, D., Velika, B., Pudule, I. &, Rancans, E. (2017). The 12-month prevalence of depression and health care utilization in the general population of Latvia. *Journal of Affective Disorders*. 210: 204-210.

Wang, Y. &, Gorenstein, C. (2013). Psychometric properties of the Beck Depression Inventory-II: a comprehensive review. *Revista Brasileira de Psiquiatria*. 2013;35:416–431.

Wardenaar, K., Wanders, R., Jeronimus, B. &, dr Jonge, P. (2017). The Psychometric Properties of an Internet-Administered Version of the Depression Anxiety and Stress

Scales (DASS) in a Sample of Dutch Adults. Journal of Psychopathological Behaviour Assessment, 40(2): 318-333.

Watson, D., O'Hara, M. &, Simms, L. (2007). Development and Validation of the Inventory of Depression and Anxiety Symptoms (IDAS). *Psychological Assessment*, 19:3: 253-268.

Webler, R., Gallagher, B., McClintock, M. &, Perera, T. (2016). Timing of depressive symptom response to rTMS: An analysis of psychometric scales. *Brain Stimulation. Conference: 2016 Clinical TMS Society Annual Meeting*. 9 (5): e8.

Weir, K.F. &, Jose, P.E. (2007). Can child self-report measures of depression and anxiety be used in college samples? *Psychological Reports*. 100 (3 I): 827-837.

Whisman, M.A., Judd, C.M., Whiteford, N.T. &, Gelhorn, H.L. (2013). Measurement Invariance of the Beck Depression Inventory-Second Edition (BDI-II) Across Gender, Race, and Ethnicity in College Students. *Assessmen*t. 20 (4): 419-428.

Whyte, R., Quince, T., Benson, J., Wood, D. &, Barclay, S. (2013). Medical students' experience of personal loss: incidence and implications. Medical Education 13 (36).

Willacy, H. (2019). Screening for Depression in Primary Care. Retrieved from https://patient.info/doctor/screening-for-depression-in-primary-care

Williamson, J. (2015). Measurement Invariance of the Beck Depression Inventory-Second Edition (BDI-II) Across Gender, Race, and Ethnicity in College Students. *Assessmen*t. 20 (4): 419-428.

Wiltink, J., Beutel, M., Till, Y., Ojeda, F., Wild, P., Munzel, T., Blankenberg, S. &, Michal, M. (2011). Prevalence of distress, Comorbid conditions and well being in the general population. *Journal of Affective Disorders*. 130 (3): 429-437.

Wing, J., Babor, T., Brugha, T., Burke, J., Cooper, J., Giel, R., Jablenski, A., Regier, D. &, Sartorius, N. (1990). Schedules for Clinical Assessment in Neuropsychiatry. *JAMA*, 47: 6

Wood, A., Taylor, P. &, Joseph, S. (2010). Does the CES-D measure a continuum from depression to happiness? Comparing substantive and artifactual models. Psychiatry Research, 177:120-123.

World Health Organisation, (1992). The ICD-10 Classification of Mental and Behavioural Disorders. (2010). *World Health Organization*.

World Health Organisation. International statistical classification of diseases and related health problems.10th revision, edition.

Wouters, E., Booysen, F., Ponnet, K. &, Baron Van Loon, F. (2012). Wording Effects and the Factor Structure of the Hospital Anxiety & Depression Scale in HIV/AIDS Patients on Antiretroviral Treatment in South Africa. *PLoS ONE* 7(4): e34881.

Wright, B. &, Mynett, J. (2019). Training medical students to manage difficult circumstances – a curriculum for resilience and resourcefulness? *Medical Education*, 19:280.

Wu, Q., Erbas, Y., Brose, A., Kuppens, P. &, Janssen, R. (2016). The factor structure, predictors, and percentile norms of the Center for Epidemiologic Studies Depression (CES-D) Scale in the Dutch-speaking adult population of Belgium. *Psychologica Belgica*. Vol.56(1): 1-12.

Xu, Y., Chi. X., Chen, S., Qi, J., Zhang. P. &, Yang, Y. (2014). Prevalence and correlates of depression among college nursing students in China. *Nursing Education Today,* 34(6): 7-12.

Yap, J. (2018). The declining state of student mental health in universities and what can be done? Retrieved from https://www.mentalhealth.org.uk/blog/declining-state-student-mental-health-universities-and-what-can-be-done

Yin, H., Xu, G., Tian, H., Yang, G., Wardenaar, K. &, Schoevers, R. (2018). The prevalence, age-of-onset and the correlates of DSM-IV psychiatric disorders in the Tianjin Mental Health Survey (TJMHS). *Psychological medicine*. 48 (3): 473-487.

Yiu, V. (2005). Supporting the well-being of medical students. *CMAJ*, 172 (7).

Yoon, S., Lee, Y., Han, C., Pae, C., Yoon, H., Pakar, A., Steffens, D. &, Kim, Y. (2014). Usefulness of the Patient Health Questionnaire-9 for Korean medical students. Academmic Psychiatry, 38(6): 6661-667.

Yu, S., Lin, Y. &, Hsu, W. (2013). Applying structural equation modeling to report psychometric properties of Chinese version 10-item CES-D depression scale. *Quality & Quantity: International Journal of Methodology*. Vol.47(3): 1511-1518.

Yu, X., Stewart, S.M., Wong, P.T.K. &, Lam, T.H. (2011). Screening for depression with the Patient Health Questionnaire-2 (PHQ-2) among the general population in Hong Kong. *Journal of Affective Disorders*. 134 (1-3): 444-447.

Yu, X., Tam, W.W.S., Wong, P.T.K., Lam, T.H. &, Stewart, S.M. (2012). The Patient Health Questionnaire-9 for measuring depressive symptoms among the general population in Hong Kong. *Comprehensive Psychiatry*. 53 (1): 95-102.

Yusoff, M., Abdul Rahim, A., Baba, A., Ismail, S., Mat Pa, M. &, Esa, R. (2013). Prevalence and associated factors of stress, anxiety and depression among prospective medical students. *Asian Journal of Psychiatry*. 6 (2): 128-133.

Yusoff, M., Rahim, A. &, Yaacob, M. (2011). The prevalence of final year medical students with depressive symptoms and its contributing factors. *International Medical Journal*. 18 (4): 305-309.

Zakrewska, J. (2012). Should we still use the hospital anxiety and depression scale. *Pain*, 153(6): 1332.

Zalta, A. &, Chambless, D. (2012). Understanding gender differences in anxiety: The mediating effects of instrumentality and mastery. *Psychology of Women Quarterly*. Vol.36(4): 488-499.

Zhang, YL., Liang, W., Chen, ZM., Zhang, HM., Zhang, JH., Weng, XQ., Yang, SC., Zhang, L., Shen, LJ. &, Zhang, YL. (2013). 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*, 5(4):268-75.

Zhao, F., Yang, X., Pan, X. &, Yang, C. (2018). Prevalence of anxiety symptoms and its correlates among medical students in China: A national survey in 33 universities. *Value in Health. Conference: 23rd Annual Meeting of the International Society for Pharmacoeconomics and Outcomes Research*, ISPOR 2018. United States. 21: S182-S183.

Zigmond, A. &, Snaith, R. (1983). The Hospital Anxiety and Depression Scale. *Acta Psychiatry Scand*, 67:361-70.

Zong, J., Cao, X., Coa, Y., Shi, Y., Wang, Y., Yan, C., Abela, J., Gan, Y., Gong, Q. &, Chan, R. (2010). Coping flexibility in college students with depressive symptoms. *Health and Quality of Life Outcomes*, 8: 66.

Appendix

List of appendices

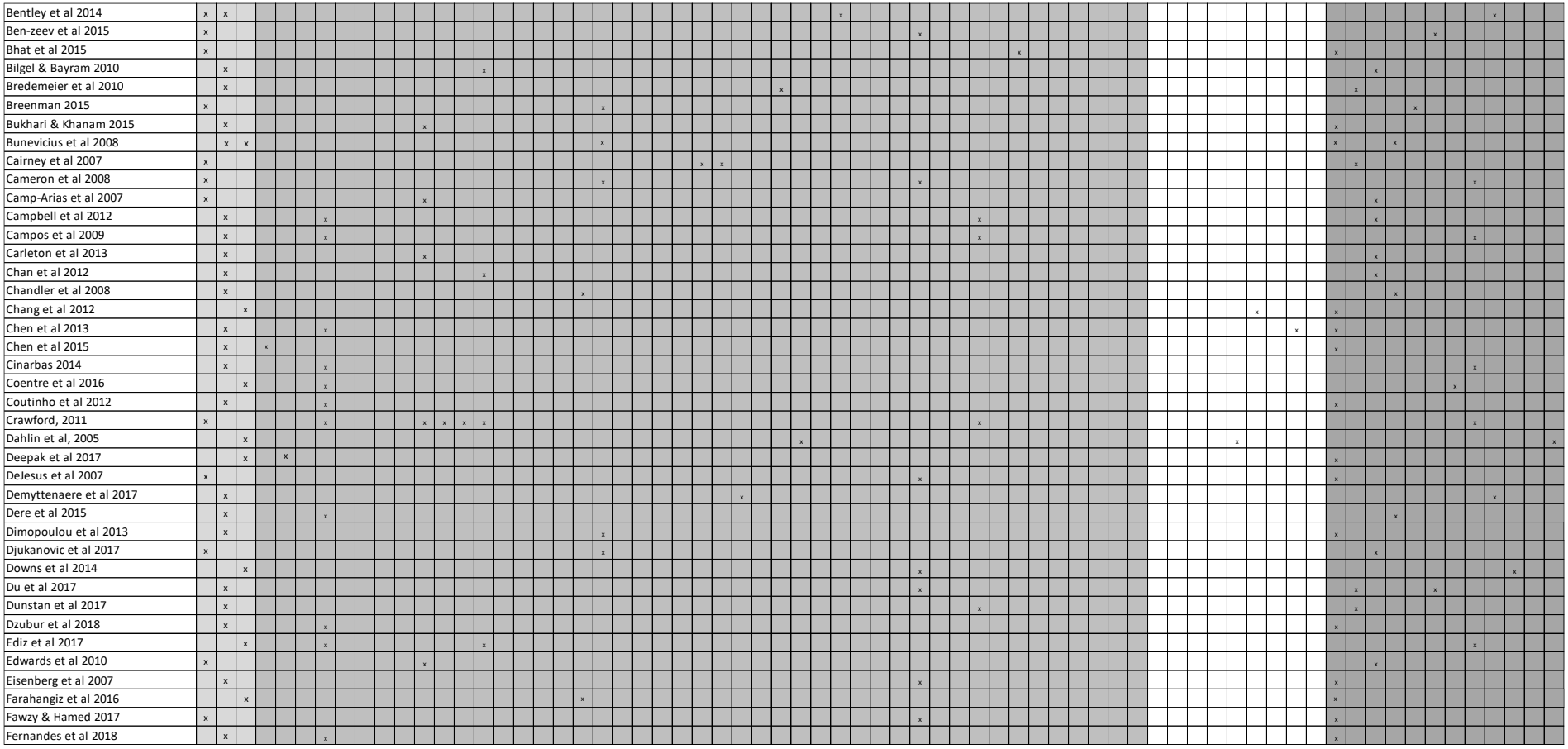
Appendix 1 – Depression reference table

Author	general	students	medical students	ADI	AKUADS	BADS	BDI	BDI- FS	MBDI	BSI	CES-D – short versions	CESD-R	CRSD	DASS-21	DASS-42	DESC	DSM	GDS	GHQ-12	GHQ-28	HADS
Aalto et al 2012	x									x									x		
Abdel Wahed & Hassan 2017			x											x							
Abdulahovic et al 2018			x				x														
Aghakhani et al 2011		x					x														
Ahmad et al 2017		x					x														
Ahmed et al 2009			x				x														
Akbari et al 2014			x																	x	
Akin & Bayram 2007		x													x						
Al Rashed et al 2019	x																				
Al-Alawi et al 2017			x																		
Al-Faris et al 2012			x				x														
AlFaris et al 2016		x	x				x														
Alfonsson et al 2017		x												x							
Al-Ghafri et al 2014			x																		
Alpizar et al 2017		x																			
Alvi et al 2010			x		x																
Amarasuriya et al 2015		x																			
Amir & Gillany 2010			x																		x
Anderson et al 2011	x																				
Aniebue & Onyema 2008			x																		
Aragao et al 2018			x				x														
Arbona et al 2017		x										x									
Arslan et al 2009		x					x														
Badpa et al 2015		x																			
Baigent et al 2015			x																		
Baldassin et al 2008			x				x														
Barreta Do Carno et al 2018	x																				
Barton 2012		x																			
Basnet et al 2012			x																		
Bassols et al 2014			x				x														
Bayati et al 2009		x																		x	
Baykan et al 2012			x												x						
Bayram & Bilgel 2008		x													x						
Beiter et al 2015		x												x							

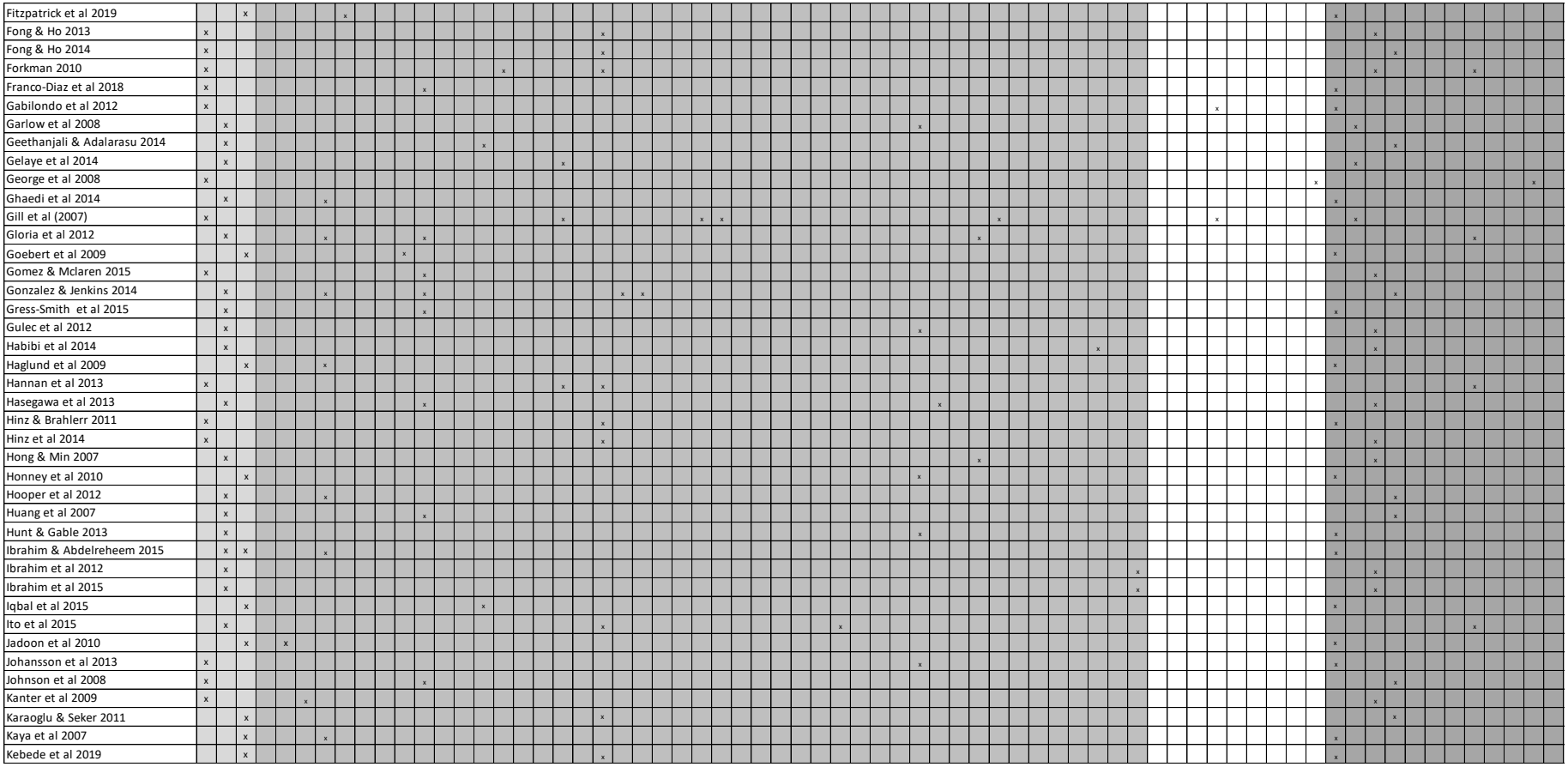
Author	IDAS	IDS	ISR	KADS	K10	K6	LAPS	MACL	MASQ	MDI	MHI-5	ODSIS	PHQ-2	PHQ-4	PHQ-8	PHQ-9	RRS – 10	SCL-90	SDS	SF-12	SRQ	SQ-48	SRE-20	ST-DEP	USDI	WHO-5	ZDS
Aalto et al 2012																											
Abdel Wahed & Hassan 2017																											
Abdulahovic et al 2018																											
Aghakhani et al 2011																											
Ahmad et al 2017																											
Ahmed et al 2009																											
Akbari et al 2014																											
Akin & Bayram 2007																											
Al Rashed et al 2019																x											
Al-Alawi et al 2017																x											
Al-Faris et al 2012																											
AlFaris et al 2016																											
Alfonsson et al 2017																											
Al-Ghafri et al 2014																x											
Alpizar et al 2017															x												
Alvi et al 2010																											
Amarasuriya et al 2015																x											
Amir & Gillany 2010																											
Anderson et al 2011					x	x																					
Aniebue & Onyema 2008																			x								
Aragao et al 2018																											
Arbona et al 2017																											
Arslan et al 2009																											
Badpa et al 2015																											
Baigent et al 2015					x																						
Baldassin et al 2008																											
Barreta Do Carno et al 2018																							x				
Barton 2012																x											
Basnet et al 2012																			x								
Bassols et al 2014																											
Bayati et al 2009																											
Baykan et al 2012																											
Bayram & Bilgel 2008																											
Beiter et al 2015																											

Author	HAM-d/HRSD	MADRS	QIDS	CIDI	MINI	PRIME MD	SCAN	SCID	DEPI	Prevalence	Screening	psychometric properties	Compariosn groups	norms	technology	language	Comaprison tools	tool development	service evaluatio	testing as diagnostic	longitudinal
Aalto et al 2012				x								x									
Abdel Wahed & Hassan 2017										x											
Abdulahovic et al 2018										x											
Aghakhani et al 2011										x											
Ahmad et al 2017										x											
Ahmed et al 2009										x											
Akbari et al 2014										x											
Akin & Bayram 2007												x									
Al Rashed et al 2019										x											
Al-Alawi et al 2017										x											
Al-Faris et al 2012										x											
AlFaris et al 2016										x											
Alfonsson et al 2017												x									
Al-Ghafri et al 2014										x											
Alpizar et al 2017												x									
Alvi et al 2010										x											
Amarasuriya et al 2015										x											
Amir & Gillany 2010										x											
Anderson et al 2011										x											
Aniebue & Onyema 2008										x											
Aragao et al 2018										x											
Arbona et al 2017													x								
Arslan et al 2009										x											
Badpa et al 2015										x											
Baigent et al 2015										x											
Baldassin et al 2008													x								
Barreta Do Carno et al 2018											x										
Barton 2012										x											
Basnet et al 2012										x											
Bassols et al 2014										x											
Bayati et al 2009										x											
Baykan et al 2012										x											
Bayram & Bilgel 2008										x											
Beiter et al 2015										x											

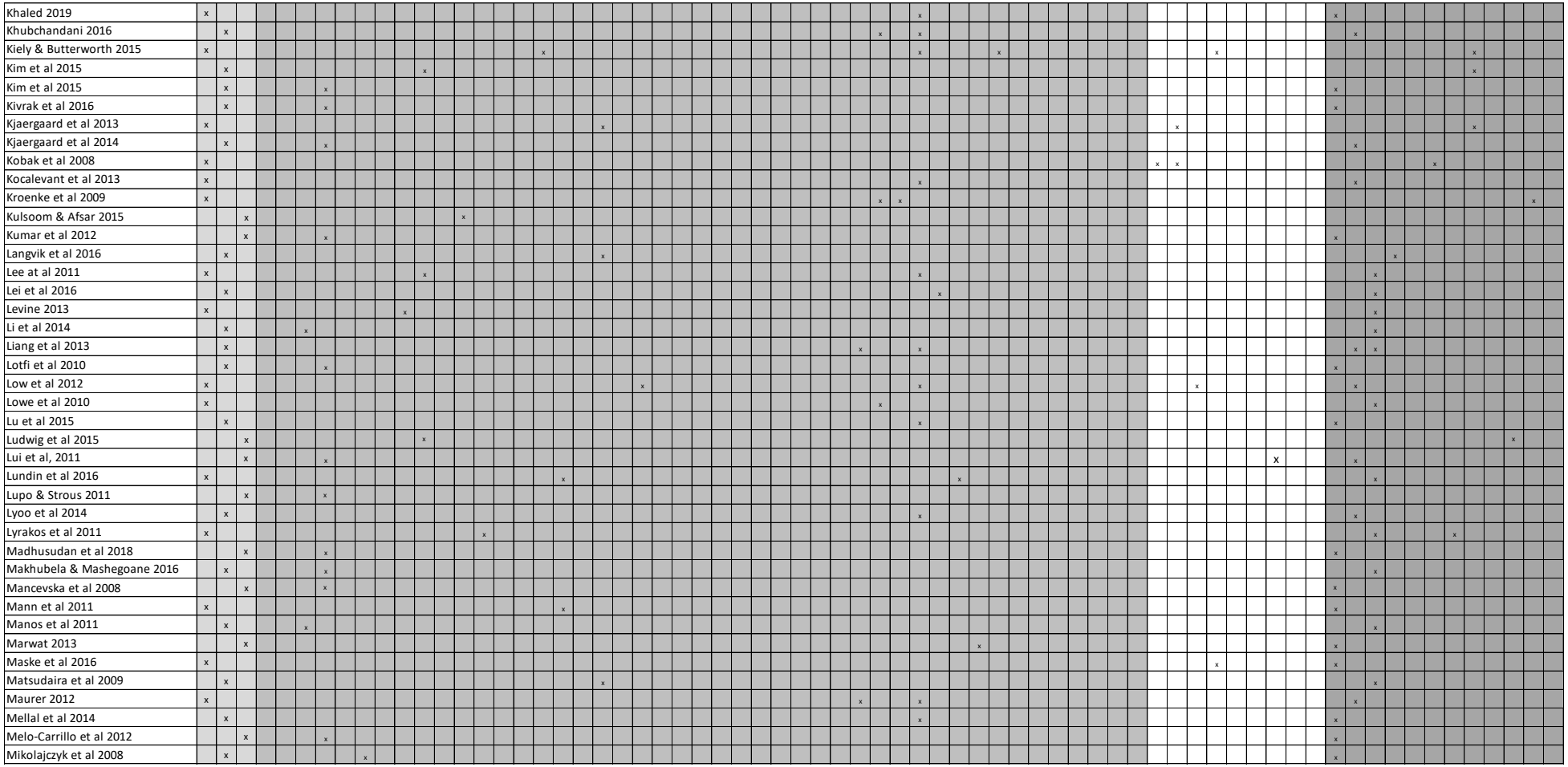
Appendix 1 - continued



Appendix 1 - continued



Appendix 1 - continued



Appendix 1 – continued

Miletic et al 2015			x																																		x																					x											
Mitsui et al 2013		x																																			x																									x								
Moreno et al 2011		x																																			x																											x						
Mousa 2016			x																															x																												x										
Mundia 2010		x													x																																											x											
Nahas et al 2019		x																																			x																					x											
Naja et al 2016			x																																		x																					x											
Najafipour et al 2012			x				x																																																			x											
Navarro et al 2014	x																																																x																x				
Neiuwenhuijsen et al 2003	x														x																																												x									
Ngasa et al 2017			x																																		x																					x											
Nuevo et al 2009	x						x																																																						x								
Oh et al 2013	x											x																																														x											
Olsonet al 2010	x											x																		x																												x									
Opoliner et al 2014		x										x																																																x				x					
Oppong Asante & Andoh-Arthur 2015		x										x																																														x											
Osman et al 2012		x												x																x																														x									
Othieno et al 2014		x										x																																														x											
Pan et al 2016			x				x																																																			x											
Parikh et al 2009		x					x																																																			x											
Prinz et al 2012		x	x																		x																																								x								
Quince et al 2012			x																		x																																					x											
Raes et al 2010	x					x																																																						x									
Rami Reddy et al 2016			x				x																																																			x											
Rancans et al 2014	x																																				x																					x											
Rancans et al 2016	x																																																				x					x											
Richards & Sanabria 2014		x					x																																																			x											
Risal et al 2015	x																				x																																											x					
Ristic-Ignjatovic et al 2013			x				x																																																														x
Rizvi et al 2015			x												x																																											x											
Roberts et al 2010		x					x																																																			x											
Roh et al 2010			x																																																		x					x											
Rosenstrom et al 2015	x						x												x																																	x													x				
Roshandel et al 2010		x					x																																																			x											
Sahoo & Khess 2010		x													x																																											x											
Samaranayake & Fernando 2011		x	x																																		x																									x								
Sasaki et al et al 2013		x										x																																																						x			
Schwenk et al 2010			x																																		x																					x											
Sepehrmanesh & Moraveji 2015		x					x																																																			x											
Serra et al 2015			x				x																																																			x											
Shafiullah et al 2016			x																																																							x											

Appendix 1 - continued

Shean & Baldwin 2012		x					x					x																																													x												
Shi et al 2016			x									x																																													x												
Shin et al 2017	x																																			x																					x												
Shudo et al 2017		x				x																																																																x
Shulte-Van Maaren et a 2013	x						x																x																											x										x									
Sidana et al 2012			x																																	x																					x												
Siddaway et al 2017	x											x																																																									
Simic-Vukomanovic et al 2016		x					x																																																		x												
Sobowale et al 2014			x																																	x																					x												
Sokratous et al 2014		x										x																																													x												
Song et al 2008		x										x																																													x												
Spielberger 2010		x																																											x															x									
Sreeramareddy et al 2007			x															x																																								x												
Straat et al 2013	x																				x																																							x									
Sugawara et al 2011	x											x																																																									
Sunderland et al 2011	x																									x	x																																	x												
Sunderland et al 2012	x																									x																																				x									
Tabuse et al 2007		x																																															x											x									
Takahashi & Ohde 2016	x																																			x																					x												
Talkovsky & Norton 2015		x																											x																																		x								
Talwaret al 2017		x											x																																											x												
Teismann et al 2016		x				x																																																									x					
Thibodeau et al 2014	x																																			x																								x								
Thompson et al 2010		x										x																																																						x		
Trainor et al 2013	x																														x																				x							x											
Tyrala et al 2015			x				x																																																		x			x								
Uehara et al 2010		x														x																																										x												
Uglesic et al 2014		x					x																																																		x												
Umegaki & Todo 2017		x										x																								x			x																					x									
Van Dam & Earleywine, 2011		x										x																																														x									
Vankar er ak 2014			x																																	x																					x												
Vasegh & Baadaran 2014		x					x																																																								x					
Venkatesh et al 2014		|x|																				|x|																																	|x|											
Vera-Villarroel et al 2010		x																																											x																			x					
Verger et al 2010		x																																																		x						x												
Verma 2011			x																																														x										x												
Verma et al 2013			x																																														x											x											
Videmova et al 2016		x	x																																	x																						x											
Vilagut et al 2013	x																																								x																		x											
Vilagut et al 2016	x											x																																														x											
Vivek et al 2011			x				x																																																	x												

Appendix 1 – continued

Voltmer er ak 2012			x																		x																																								x								
Vrublevska et al 2017	x																																																							x					x										
Vrublevska J., Rancans E., Trapencieris M., Snikere S.	x																																																							x					x										
Wahl et al 2013	x						x			x		x									x			x					x			x		x			x												x														x									
Watson et al 2007	x																						x																																													x			
Webler et al 2016		x					x																																										x															x				
Weir & Jose 2007		x					x																																																									x				
Whisman et al 2013		x					x																																																					x							
Wiltink et al 2011	x																																				x																						x												
Wood et al 2010	x											x																																																	x									
Wu et al 2016	x						x																																																						x							
Xu et al 2014		x										x																																													x												
Yin et al 2018	x																			x																																							x												
Ying-Li ey al 2012	x																																			x																									x									
Yoon et al 2014			x																																			x																							x									
Yu 2013		x								x																																																	x									
Yu et al 2012	x																																				x																							x									
Yu et all 2011	x																																	x																									x										
Yusoff et al 2013			x										x																																												x												
Yusoff et al 2011			x				x																																																		x												
Zhang et al 2013		x																																	x			x																							x									
Zong et al 2010		x					x																																																		x												

Appendix 2 – Anxiety reference table

Author	general	students	medical students	ADDQ	AKUADS	BAI	BSI	BSI-18	DASS-21	DASS-42	GAD-Q-IV	GAD-7	HADS	K10	K6	MASQ	OASIS	phq-4 (gad-2)	PSWQ	SAS - Zung	scat	STAI	STAIS	WDQ	GAS	HARS	CIDI	MINI	SCID	cis-r	**Prevalence**	**Screening**	**psychometric**	**Compariosn groups**	**technology**	**language**	**Comaprison tools**	**tool development**	**service evaluatio**	**testing as diagnostic**	**longitudinal**
Abas (2017)		x				x																								x										
Abdel Wahed & Hassan 2017			x						x																					x										
Akin & Cetin(2007)		x								x																						x									
Alfonsson et al (2017)	x								x			x																				x									
Alvi et al 2010			x			x																								x										
Andersen et al (2011)	x													x	x												x				x										
Antunez & Vinet (2012)		x							x																							x									
Barton 2012		x										x																			x										
Bassols 2014			x			x																									x										
Bayani (2010)		x								x																									x						
Baykan et al 2012			x							x																					x										
Beesdo-Baum et al (2014).		x							x																												x				
Beiter et al 2015		x							x																						x										
Bilgel & Bayram 2010		x								x																							x									
Bitsika et al 2010		x																		x																	x				
Bunevicius et al 2008			x										x																		x			x							
Chan et al 2012		x								x																						x									
Chang et al 2019	x																									x					x										
De Lima et al (2011).		x				x																						x				x									
Deepak et al 2017			x		x																									x										
Dhariwal et al 2018			x																		x										x										
Dimopoulou et al 2013		x											x																		x										
Dunstan et al (2017)		x								x										x												x									
Ediz et al 2017			x			x				x																										x					

Appendix 2 – continued

Eisenberg et al 2007		x										x																			x										
Fernandes 2018		x				x																									x										
Galvan-Molina et al 2017			x																			x										x									
Gelaye et al 2014		x																														x									
Ghalechi & Kazemi 2010		x																													x										
Hajebi et al 2018	x																										x				x										
Hinz & Brahlerr (2011)	x												x																		x										
Hinz et al (2017)	x											x																					x								
Ibrahim & Abdelreheem 2015		x	x			x																									x										
Johansson et al 2013	x											x																						x							
Kadri et al 2007	x																											x			x										
Karaloglu & Seker 2010			x										x																					x							
Kebede et al 2019			x										x																		x										
Keily & Butterworth (2015)	x																								x		x										x				
Khubchandani et al (2016)		x																x														x									
Knappe et al (2014)		x									x																						x								
Kroenke et al (2009)	x																	x																						x	
Langvik et al 2016		x											x																					x							
Leray et al 2011	x																											x			x										
Lowe et al (2008)	x											x																					x								
Lowe et al (2010)	x																	x															x								
Lu et al 2015		x										x																			x										
Lyrakos et al (2011)	x									x													x													x					
Maideen et al 2015	x											x																			x										
Manchevska & Pluncevic-Gligoroska 2014		x				x																									x										
Mann et al 2011	x																														x										
Matsudaira et al 2009		x											x																				x								
Mousa et al (2016)			x									x																				x									
Mundia 2010		x								x																					x										
Norman et al (2011)		x						x									x												x									x			
Norton & Robinson (2010)		x		x		x					x								x																			x			

Appendix 2 continued

Ohaeri et al (2010)	x										x																														
Osman et al 2012		x				x			x							x																	x								
Ozen et al 2010		x											x																							x					
Pez et al 2010	x																										x		x	x										x	
Pfalzgraf & Tipton (2013)		x																		x											x										
Risal et al (2015)	x												x																							x					
Robinson et al (2010)		x				x					x								x														x								
Ruiz et al 2017		x							x																								x								
Sahoo & Khess 2010		x							x																						x										
Schulte Van Maaren et al (2013)	x						x												x					x				x					x								
Serra et al 2015			x			x																									x										
Simic-Vukomanovic 2016		x				x																									x										
Sujatha Baskaranet al 2018		x										x																			x										
Sunderland et al 2011	x													x	x																x										
Talkovsky & Norton, 2015		x				x				x						x																		x							
Talwar et al 2017		x							x																						x										
Thiboutot et al 2016		x																		x											x										
Thompson et al 2010	x																									x													x		
Van Ameringen et al (2009)	x											x																							x						
Verger et al 2010		x																									x				x										
Wardenaar et al (2017)	x									x																							x								
Weir et al 2007		x				x																															x				
Williamson 2015		x								x																															
Wiltink et al 2011	x											x																			x										
Yin et al 2018	x																														x										
Yusoff et al 2013			x						x																						x										
Zalta & Chambless (2012)		x							x										x															x							
Zhao et al 2018			x																	x											x										

Appendix 3 HADS completion sheet

HADS questionnaire

In order to match your responses to different parts of this study we would like to use a unique identifier. To generate this please can we ask you for the **first two letters** of the answers to the following three questions?

1. What was the name of your primary (or first) school?

_____

2. What is your mother's first name?

_____

3. Name of town where you were born?

_____

The following questions are about how you are feeling most of the time (whether in medical school or in your own time). Please answer these how they best fit you and how they best reflect how you have felt **over the last week**. Don't spend too long thinking about each question, we would like your immediate response.

1. I feel tense or 'wound up'

Most of the time	
A lot of the time	
From time to time, occasionally	
Not at all	

2. I still enjoy the things I used to enjoy

Definitely as much	
Not quite so much	
Only a little	
Hardly at all	

3. I get a sort of frightened feeling as if something awful is about to happen

Very definitely and quite badly	
Yes, but not too badly	
A little, but it doesn't worry me	
Not at all	

4. I have lost interest in my appearance

Definitely	
I don't take as much care as I should	
I may not take quite as much care	
I take just as much care as ever	

5. Worrying thoughts go through my mind

A great deal of the time	
A lot of the time	
From time to time but not too often	
Only occasionally	

6. I can laugh and see the funny side of things

As much as I always could	
Not quite as much now	
Definitely not so much now	
Not at all	

7. I can sit at ease and feel relaxed

Definitely	
Usually	
Not often	
Not at all	

8. I feel as if I am slowed down

Nearly all of the time	
Very often	
Sometimes	
Not at all	

9. I get a sort of frightened feeling like 'butterflies' in the stomach

Not at all	
Occasionally	
Quite often	
Very often	

10. I look forward with enjoyment to things

As much as I ever did	
Rather less than I used to	
Definitely less than I used to	
Hardly at all	

11. I feel restless as if I have to be on the move

Very much indeed	
Quite a lot	
Not very much	
Not at all	

12. I feel very cheerful

Never	
Not often	
Sometimes	
Most of the time	

13. I get sudden feelings of panic

Very often indeed	
Quite often	
Not very often	
Not at all	

14. I can enjoy a good book or radio or TV programme

Often	
Sometimes	
Not often	
Very seldom	

Appendix 4 – Participant information sheet

PARTICIPANT INFORMATION SHEET

1. Study title

Measuring prevalence of common mental health disorders among medical students

2. Invitation paragraph

You are being invited to take part in the above research study. Before you decide whether to participate, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

Thank you for reading this.

3. What is the purpose of the study?

Providing support for medical students is an area of growing concern in the UK and internationally. Despite existing support, there is clear evidence that we are not meeting students' needs. The General Medical Council (GMC) has recently published a report outlining how medical schools can more effectively support their students with mental health concerns. Gaining a better understanding of the prevalence of common mental health disorders (CMHD) can more effectively support medical students. This will ensure that support is tailored in the right area. The provision of effective support at medical school will improve coping strategies, and the health and wellbeing of both medical students and future doctors.

A number of prevalence studies on psychological distress in medical students have been conducted but these have measured symptoms rather than the presence of a clinical disorder. The estimated prevalence varied widely across these studies, which related to the measures used and data interpretation. As such, the true prevalence of CMHD in medical students is not known. Therefore a more robust study to evaluate the true prevalence of CMHD is required. This study aims to investigate CMHD in medical students in more detail.

4. Why have I been chosen?

Medical students from Cardiff University are being invited to take part in a study looking at medical student wellbeing.

5. Do I have to take part?

It is up to you to decide whether or not to take part in an interview as part of this study. If you wish to take part, page you will be asked to complete

a consent form. If you consent to take part you are still free to withdraw at any time by providing your unique identifier (see question 7).

6. What will happen to me if I take part?

If you would like to take part, an interview will be arranged at a mutually convenient time. The interviews will be held at the student support unit at 53-54 Park Place. Before the interview you will be asked to complete a short questionnaire asking about symptoms relating to depression and anxiety. The interview will take no longer than 30 minutes. The interviews will ask you questions about symptoms relating to depression and anxiety, you are not obliged to answer all questions. Your interview data will be analysed along with your questionnaire data. The interviews will be audio-recorded to allow the researchers to analyse the data as a team.

7. What about confidentiality?

The interview data will be stored anonymously using your unique identifier. In order to match your responses from different parts of the study you will be asked to produce a unique ID. To generate this you will be asked to give the first two letters of the answers to the following three questions:

What was the name of your primary (or first) school?

What is your mother's first name?

Name of town where you were born?

All data will be stored securely on password protected computers. This anonymised data may be retained indefinitely in accordance with the Data Protection Act (1998). This is so that we can refer to the original data if anyone questions any findings in our final reports.

8. What do I have to do?

If you decide to take part in the study, please complete the consent form and send it to the researchers using the address below. You will then be contacted by the researchers to arrange an interview.

Are there any risks?

This study contains minimal risk. We do request you consider all the above information carefully before you decide whether you would like to take part. If you are distressed by any issues raised in this study please contact your student support services, or contact Dr Debbie Cohen at Cardiff University (see contact details below).

9. What will happen to the results of the research study?

This study will serve as a pilot to a further national longitudinal study and will form part of Naomi Marfell's Mphil report. The results of the study will be written up as a report. In addition the results may also be published in a peer review journal and presented at appropriate

conferences. You will not be identified in any publication related to this study; all data will be anonymous.

10. Who is organising and funding the research?

The research is being organised and funded by researchers from the Department of Primary Care and Public Health, Cardiff University.

11. Contact for Further Information

If you have any questions or queries about the project please contact Naomi Marfell at Cardiff University.

Naomi Marfell
53 -54 Park Place
Cardiff
CF10 3AT
E: marfelln@cardiff.ac.uk

If you are concerned or distress about anything concerning this study please contact Dr Debbie Cohen at Cardiff University

Dr Debbie Cohen OBE
Senior Medical Research Fellow and Director of The Individual Support Programme
Centre for Psychosocial Research, Occupational and Physician Health (PROPH)
53 -54 Park Place
Cardiff
CF10 3AT
T: 02920 870686
E: cohenda@cardiff.ac.uk

THANK YOU FOR TAKING THE TIME TO READ THIS INFORMATION SHEET

Appendix 5 – Participant consent form

CONSENT FORM - Interviews

Title of Project: Measuring prevalence of common mental health disorders among medical students

Name of Researcher: Naomi Marfell

Please initial box

1. I confirm that I have read and understand the information sheet and have had the opportunity to ask questions. ☐
2. I understand that my participation is voluntary ☐
3. I agree to be audio-recorded ☐
4. I understand that the information provided by me will be recorded and will be stored anonymously using my unique identifier so that it is impossible to trace this information back to me individually. I understand that, in accordance with the Data Protection Act (1998), this information may be retained indefinitely. ☐
5. I understand that data collected may be presented at conferences and meetings. ☐
6. I understand that in completing this consent page I am giving my permissions for any data collected today to be used for research purposes. ☐
7. I agree to take part in the above study. ☐
8. I agree to give my unique identifier to the researchers to match my data ☐

Name of Participant _____

Date _____

Appendix 6 – Participant debrief sheet

PARTICIPANT DEBRIEF SHEET

Measuring prevalence of common mental health disorders among medical students

Thank you for your participation in this project.
This study willserve as a pilot to a further national longitudinal study and will form part of Naomi Marfell's Mphil report. The results of the study will be written up as a report. In addition, the results may also be published in a peer review journal and presented at appropriate conferences. You will not be identified in any publication related to this study; all data will be anonymous.

If you are concerned about any issues that this study has raised you can contact the Medic Support Unit at School of Medicine, Cardiff University. You can email Medic support at: medicsupport@cardiff.ac.uk or you can find out more information at
http://medicine.cf.ac.uk/medical-education/undergraduate/medic-support-cardiff/student-support-unit/

Alternatively, you can contact your own GP or the student counselling services at Cardiff University if you have any health concerns.

If you have any questions or queries about this project, please contact Naomi Marfell at Cardiff University.
Naomi Marfell
53 -54 Park Place
Cardiff
CF10 3AT
E: marfelln@cardiff.ac.uk

If you feel you need urgent help please contact:
Samaritans (0845 790 9090) http://www.samaritans.org/branches/cardiff-district-samaritans

BMA Doctors-for-Doctors (08459 200 169) http://bma.org.uk/practical-support-at-work/doctors-well-being/about-doctors-for-doctors

Nightline (029 2087 0555)
http://www.cardiffstudents.com/activities/studentled/nightline/