A meta-analytic investigation of the psychometric properties of the Health Anxiety Questionnaire and the Health Anxiety Inventory

Nicholas James Stenning

Submitted in accordance with the requirements for the degree of
Doctor of Clinical Psychology (D. Clin. Psychol.)
The University of Leeds
School of Medicine
Academic Unit of Psychiatry and Behavioural Sciences

August 2017

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

The right of Nicholas James Stenning to be identified as Author of this work has been asserted by him in accordance with the Copyright, Designs and Patents Act 1988.

Acknowledgements

Abstract

Introduction: valid, norm-referenced questionnaires are essential for research and clinical practice. After initial scale validation efforts, the validity and reliability of questionnaires are frequently not re-assessed. The Short Health Anxiety Inventory (SHAI) and the Health Anxiety Questionnaire (HAQ) were subject to norm-generation and a validation exercise using meta-analysis. This information informed a commentary on the cognitive-behavioural conceptualisation of heath anxiety.

Method: a reverse citation search from the publication date of each paper to April 2016 was conducted using Medline, PychInfo and Web of Science. Data was extracted regarding mean score, standard deviation, reliability coefficients and correlations with other measures from all studies utilising either scale. Data was included from a total of 137 study arms.

Results: population norms for both scales in various populations were generated via the meta-analysis of mean scores. The SHAI was found to have good average reliability (α =0.870, computed from 54 studies) and correlations with other psychometric support its construct validity. There was insufficient data to assess the reliability and validity of the HAQ using meta-analysis.

Discussion: under the current analysis, the SHAI was confirmed to have good construct validity and reliability. Correlations between the SHAI and other measures provided evidence that supported the cognitive-behavioural conceptualisation of health anxiety. There was no evidence that the HAQ lacked construct validity, however this could not be confirmed with the available data.

Table of Contents

List of Tables

List of Figures

1.1. Introduction

Health anxiety will affect everyone at some point in their life. For some, however this anxiety can be debilitating and will severely interfere with daily life (Salkovskis & Warwick, 1986; Tang et al., 2009). In contrast to non-anxious controls, those with more severe health anxiety or hypochondriasis will not be reassured by a doctor following a benign diagnosis of a bodily state that is causing worry (Lucock, Morley, White, & Peake, 1997). It is this group that are more likely to be seen in a mental health clinic and require input from a clinical psychologist (Taylor & Asmundson, 2004). Those with health anxiety use healthcare services significantly more than the rest of the population (Barsky, Ettner, Horsky, & Bates, 2001), presumably because they do not feel reassured by doctors that they are in fact healthy.

The prevalence of hypochondriasis or of health anxiety is difficult to assess, this is in part due to changes in definition over time. Noyes (2001) points out that under DSM-III (American Psychiatric Association, 1986) criteria a person experiencing panic attacks would be excluded from receiving a diagnosis of hypochondriasis, under DSM-IV (American Psychiatric Association, 1994) this exclusion criteria was extended to anyone with an anxiety disorder or depression. This increasingly stricter criteria means fewer individuals will meet diagnostic level so will not be counted in prevalence studies. More recently the diagnosis of hypochondriasis has been removed from DSM-V (American Psychiatric Association, 2013) and replaced with somatoform disorder and illness anxiety disorder. Accurately establishing prevalence is further complicated by whether one defines hypochondriasis as a categorical diagnosis, or an extreme of a dimensional health anxiety construct (Ferguson, 2009). The fact that hypochondriasis or severe health anxiety can only be established in the absence of an existing medical condition also contributes to the difficult in estimating its prevalence (Noyes, 2001). In a narrative review Noyes (2001) reported that the point prevalence estimates of hpochondraisis in community populations ranged from 4-25%, although the highest end of this estimate was derived from an older study with methodological problems. A systematic review of studies reporting on the prevalence of hypochondriasis in community samples reported rates of 4.5-10% (Creed & Barsky, 2004). When only including studies using DSM or ICD-10 criteria, prevalence drops to 0.2-7% (median 4.2%). The findings of this systematic review are more robust than other prevalence studies as it only included papers that assessed hypochondriasis on the basis of clinical cut-off or diagnostic interview.

Norm-referenced psychometric measures can complement a clinician's structured assessment of health anxiety by offering a quantitative assessment of an individual's symptoms. Use of norm-referenced questionnaires also enable a clinician to

examine whether a client's score differs in comparison to a normative sample, provides an indication of the magnitude of this disparity and whether this difference is in the direction and to the extent that is expected (Sperlinger, 2002). Such measures are also essential in quantitatively evaluating the success of treatment in clinical practice and in formal research (Barkham, Hardy, & Mellor-Clark, 2010). Use of a norm-referenced questionnaire has the advantage of allowing a clinician to determine whether an intervention has reduced an individual's score below an established clinical cut-off (clinically significant change; Jacobson, Follette, & Revenstorf, 1984) and indicate whether this change is sufficient in magnitude to be attributable to the intervention rather than measurement error (reliable change; Jacobson & Truax, 1991).

In order for a measure to be effective it must demonstrate good *construct validity*, that is measure what it purports to measure (Strauss & Smith, 2009) and be *reliable,* or consistently assess a construct across differing testing situations and populations (Cronbach, 1947). Finally, so that the effectiveness of clinical interventions may be effectively evaluated, a measure must also be sensitive to changes in a construct (Vermeersch, Lambert, & Burlingame, 2000). To assess these qualities, a scale will undergo a validation process. This usually involves administering a pool of potential scale items to a large sample, participants scores are then subjected to a factor analysis and redundant items are deleted. The resulting scale then undergoes an assessment of construct validity and reliability checks. All of these steps are essential in developing psychometrically sound measures (Rust & Golombok, 2009). However, I argue this one-off validation process is insufficient and it is possible that a measure may be employed in research and have only undergone validation procedures in a single sample. This is problematic, because a scale cannot be said to have inherent construct validity and reliability (Vacha-Haase & Thompson, 2011). In fact, every application of a scale provides an opportunity to better understand its psychometric properties and the construct it purports to measure (Strauss & Smith, 2009).

This thesis utilises published data to better understand the psychometric properties of two measures of health anxiety: The Health Anxiety Questionnaire (HAQ; Lucock & Morley, 1996) and the Health Anxiety Inventory (SHAI; Salkovskis, Rimes, Warwick, & Clark, 2002). I aim to develop normative data and investigate variations in the reliability and construct validity of both scales. I will also use this information as a means to review the construct of health anxiety from the perspective of the cognitive-behavioural model of health anxiety (Salkovskis & Warwick, 1986). To achieve these aims, I systematically searched the research literature for every published application of the HAQ and SHAI. I then extracted data relating to each scale from these papers and then performed a meta-analysis on this information.

In this first chapter, I provide an overview of the literature pertaining to health anxiety, including a discussion of whether it may be best described as a categorical or

continuous variable. I also systematically review the evidence for the cognitive-behavioural model of health anxiety because the HAQ and SHAI are both based on this conceptualisation of health anxiety. This is followed by a summary of the scale validity literature as applied to scale development, with a focus on construct validity. Finally, I discuss how the scale validity literature has been applied to measuring health anxiety, this includes a brief review of health anxiety measures and a rationale for investigating the HAQ and SHAI. Chapter two outlines the method I have used including literature searching, data extraction, quality inclusion criteria and means of analysis. Chapter three contains the results of the meta-analysis of mean scores and reliability coefficients of the HAQ and SHAI. Chapter four consists of a review and meta-analysis of correlation coefficients between the HAQ/SHAI and other measures. This chapter also includes an evaluation of the variation in reliability coefficients using meta-regression Finally, chapter five will comprise my discussion of these results and places them in the context of the wider literature.

1.2. What is health anxiety?

As previously discussed, the affective experience of anxiety about one's health is something everyone will experience. Most people can reassure themselves they are well or a visit to the doctor will provide sufficient reassurance that there is nothing to fear and their anxiety will dissipate (Salkovskis & Warwick, 2001). However, for a significant minority, this reassurance only provides temporary relief and very soon their worries return (Speckens, Spinhoven, Van Hemert, & Bolk, 2000). It is this group who are particularly troubled by health anxiety and are most likely to require support from a mental health professional (Taylor & Asmundson, 2004).

Before proceeding, it is important to distinguish between hypochondriasis and health anxiety. Confusingly, the terms have been used interchangeably in the literature, but also may refer to different entities (Ferguson, 2009). Hypochondriasis has been considered both as a discrete medical syndrome characterised by a fear of having contracted an illness (DSM-IV; American Psychiatric Association, 1994) and as a dimensional construct at the extreme end of normal health anxiety (Warwick & Salkovskis, 1990). In DSM-V hypochondriasis has been removed and replaced by somatic symptom disorder and illness anxiety disorder (American Psychiatric Association, 2013). Ferguson (2009), points out that understanding the latent structure of both health anxiety and hypochondriasis is important as this will impact on the selection of participants for research, the design of studies and also on the content of clinical interventions. If hypochondriasis is a categorical entity, this would imply a single causal factor or multiple factors in combination that lead to a unique grouping of individuals. If this scenario exists, then research will need to be designed to incorporate individuals who fall within this group and clinical practice would be best suited to

follow a diagnostic approach. However, if hypochondriasis is dimensional, then this implies multiple causal factors and research should focus on identifying and separating these variables. Additionally, study samples should be selected to include all degrees of hypochondriasis rather than those achieving 'caseness' and dimensional measures should be used to assess the degree of health anxiety (Ferguson, 2009).

The latent structure of hypochondriasis and health anxiety is particularly relevant to this thesis because, as indicated above it will affect the way I analyse the data collected from the literature. If both lie on a continuum, treating health anxiety as a continuous variable and seeking correlations with other measures and normative scores in different population would be appropriate. If categorical then I will need to pay extra attention to the description of samples and treat hypochondriasis as a categorical variable. It is also important as I aim to provide a commentary on the cognitive-behavioural construct of health anxiety as assessed by both scales.

A taxometric analysis seeks to determine whether the latent structure of a construct is best described as categorical or dimensional (Ruscio and Ruscio, 2004). Three taxometric analyses have been conducted on samples with health anxiety. The first Ferguson (2009) used the Whitely Index (WI; Pilowsky, 1967) to measure health anxiety in 711 working-age adults who did not have a physical illness. This was to ensure that those who were anxious about having an actual illness rather than being generally anxious about their health were excluded. Ferguson achieved this by asking whether participants were receiving any medical treatment. This may be problematic as individuals may be ill (e.g. ongoing asthma) and not be receiving medical input. The study may also have excluded individuals who are anxious about a medical condition that they are not being treated for. Ferguson (2009) reported that a dimensional latent structure of health anxiety was best supported by the data. Ferguson suggested this procedure be repeated in a medically unwell sample to determine if the latent structure differs in the population. However, this study has been criticised for having too small a sample to detect whether there are two distinct groups; one being health anxious, the other not. This is because health anxiety is thought to have a median point prevalence of 4.2% (Creed & Barsky, 2004), at this incidence rate, a representative sample of 711 would contain only 31 individuals with clinically significant health anxiety (Asmundson, Taylor, Carleton, Weeks, & Hadjstavropoulos, 2012).

A second analysis was conducted by Longley and colleagues (Longley et al., 2010). They argued that Ferguson's (2009) analysis was limited because the sample were middle-old aged so were more likely to be physically unwell or have past experience of severe illness. They also criticised the use of six items of the Whitely index as being too-short a measure to accurately reflect the full range of health anxiety (responses on three of the original nine items were not used in the taxometric analysis, because their validity indicators were too low). Longley et al (2010) analysed data from

1083 undergraduates who completed the Illness Attitudes Scales (IAS; Kellner, 1987), the Multidimensional Inventory of Hypochondriacal Traits (MIHT; Longley et al., 2005) and the Whitely Index (Pilowsky, 1967). They claim that their analysis is superior as it focusses on younger people and uses a range of measures, so will better capture the hypochondriasis construct. They replicated Ferguson's findings, concluding that hypochondriasis is likely to be a dimensional construct.

More recently, Asmundson, Taylor, Carleton, Weeks and Hadjstavropoulos (2012) analysed IAS scores from 1768 university students using factor mixture modelling. This involves performing a factor analysis on the data and determining whether any factors are most closely related with one sub-group compared to another. The authors concluded that two distinct groups could be found in their sample. The two factors that distinguished the groups were those assessing focus on bodily states and the extent that these states interfered with day-to-day life. Surprisingly, the authors report that 81.4% of the sample fell in an 'anxious group' and the remaining 18.6% in a 'non-anxious' group that reported universally low scores across all factors. This anxious group was larger than expected: the authors had predicted an 'anxious' group should be proportionate to the prevalence of health anxiety which would equate to 7-80 individuals. This number is based on the point prevalence rates of 0.4-4.5% reported in Asmundson et al., (2012), the upper estimate of 80 corresponds to the 4.2% median prevalence reported by Creed and Barsky (2004). In their study, Asmundson et al., (2012) argue that for the large 'anxious' sample, health anxiety exists on a continuum while the smaller 'non-anxious' sample is a distinct group that does not experience any health anxiety. This final study has the advantage of a large enough sample size to detect a smaller health-anxious subsample, although I argue the sample is still likely to be insufficient to detect a separate group based on a lower prevalence rate of 0.4%.

In summary, the evidence for the latent structure of health anxiety is mixed, with two studies reporting a dimensional structure and one a taxonic structure. Importantly, all studies were consistent with the view that for those with health anxiety, this is likely to be a dimensional construct.

Another difficulty with research regarding health anxiety is that its definition has altered over time (Starcevic & Noyes, 2014) and it shares some characteristics with other disorders, notably panic disorder, obsessive compulsive disorder (OCD) and generalised anxiety (Starcevic, 2014). I will now discuss the recent changes in the definition of health anxiety. This is followed by a discussion of mental health conditions that are related to health anxiety and somatic concerns.

In terms of defining hypochondriasis, DSM-IV emphasised a fear of illness due to a preoccupation with, and misinterpretation of bodily symptoms. Key to this definition is a persistence of such fears despite reassurance from medical personnel (American Psychiatric Association, 1994). The ICD-10 (World Health Organisation,

1992) criteria are similar to that of DSM-IV, however differ in that they do not require medical investigation to be unable to account for somatic symptoms (Pannekoek & Stein, 2014).

This picture is complicated by the fact that in DSM-V hypochondriasis has been removed and replaced by two diagnoses; illness anxiety disorder and somatic symptom disorder. Illness anxiety disorder emphasises the presence of anxieties about health and concerns about having or catching an illness, in the *absence* of somatic symptoms. The criteria also include avoidance behaviours (e.g. avoiding health-related information) or behaviours related to health (e.g. monitoring one's internal states for symptoms of illness). In contrast, somatic symptom disorder relates to when individuals experiences "…somatic symptoms that are distressing or result in significant disruption in daily life" and "Excessive thoughts, feelings, or behaviours related to the somatic symptoms or associated health concerns: (p.331 American Psychiatric Association, 2013). Crucially it is possible to achieve diagnostic criteria for somatic symptom disorder *without* experiencing any anxiety about one's health. In contrast, the cognitive behavioural model (see next section), which influenced the development of both the HAQ and SHAI, conceptualises health anxiety as an extreme form of worry about one's health. This means that some individuals who may meet diagnostic criteria for somatic symptom disorder would not be considered to have health anxiety according to the cognitive model.

Another important criterion for the definition of health anxiety is lack of responsiveness to reassurance from medical professionals. As I will outline in the next section, according to the cognitive model, lack of effective reassurance is a maintaining factor for health anxiety. Other approaches such as the interpersonal model view lack of responsiveness to reassurance as a causative factor and relates this to an individual's attachment style and the means that they seek help from others (Stuart, 2104).

In this thesis, I will adopt a definition of health anxiety that is consistent with the cognitive behavioural model, this is because both the HAQ and SHAI aim to measure a health anxiety construct that is founded on this model. Accordingly, the definition of health anxiety that I will use is "an extreme and persistent form of worry about one's health that continues despite reassurance from a medical professional". This definition is consistent with the DSM-IV hypochondriasis, and DSM-V illness anxiety disorder, however some individual's meeting criteria for DSM-V somatic symptom disorder would not experience health anxiety by this definition.

In terms of relationship with other mental health problems, health anxiety shares some similarities with other mental health conditions that have a somatic component (Starcevic, 2014). In panic disorder an individual may focus on a bodily state and interpret it as evidence they have a severe illness and that they are about to die as a result. The immediacy of the prospect of death is a difference between panic disorder

and health anxiety: an individual with panic disorder will feel they are going to die from an illness almost immediately, whereas someone with health anxiety will believe they will die of an illness at some time in the future. Other differences include a focus on the symptoms of the panic attack itself and thoughts that occur during an attack, whereas in health anxiety the focus is likely to be on a broader range of bodily states and will be present more frequently (Deacon & Abramowitz, 2008).

OCD is another condition with some overlap with health anxiety (Ania Greeven, Spinhoven, & van Balkom, 2009); the intrusive thoughts characteristic of OCD are similar to the thoughts about health and body experienced by those with health anxiety. Similarly, continually checking one's body for signs of illness may also be thought of as a kind of compulsion. Similarly, the role of reassurance in (only temporarily) reducing uncertainty about a health problem is similar to the temporary relief felt by an individual with OCD when they complete a ritual behaviour (Starcevic, 2014). What distinguishes health anxiety is poorer insight and more strongly held beliefs that are related to the self, for example having an illness because they are 'an unwell person'. Those with OCD have greater insight into their condition and their OCD symptoms are more likely to be experienced as unrelated to the self; for example continually checking on the safety of a loved one despite being aware that this is irrational and that the individual is safe (Starcevic, 2014). An empirical evaluation of the differences between those with hypochondriasis, OCD and healthy controls reported those with hypochondriasis have significantly more worries about illness than those with OCD. Those with hypochondriasis have less obsessive-compulsive symptoms than those with OCD, but have significantly more OCD symptoms than healthy controls (Anja Greeven, van Balkom, van Rood, van Oppen, & Spinhoven, 2006).

As conceptualised by the cognitive model, anxiety and worry are central features of health anxiety (Salkovskis & Warwick, 2001). This has some similarities to generalised anxiety disorder which is characterised by worry across multiple life domains, however health anxiety differs in terms of worries being specifically related to one's health. Illness phobia is another related form of anxiety and concerns fear that one may catch an infectious disease; this may be characterised by avoiding situations where one may catch an illness, a lack of somatic symptoms and a fear of becoming unwell in the future. In contrast, those with health anxiety are more likely to have a somatic symptom and be fearful they are already unwell (Noyes, 2001b).

Finally, there is some relationship between health anxiety and depression, for example those experiencing depression may report some signs of health anxiety such as overly ruminating about a bodily state or symptom. Starcevic (2014) points out that usually when a person is depressed any health anxiety symptoms are secondary and may be better explained by depression. In clinical practice, Starcevic suggests collecting information around a person's experience and to determine the temporal relationship

between the two, or determine if health anxiety is only felt when the person is experiencing depression. Finally, another possibility is that depression may be an emotional reaction to the experience of health anxiety; again this may be determined by careful assessment and examining the temporal relationship between the two (Starcevic, 2014).

In summary, the evidence for the underlying structure of hypochondriasis and health anxiety is mixed and the definition of hypochondriasis has altered over time, including its deletion from DSM-V. Because of this I will use the term hypochondriasis when referring to the clinical syndrome, as defined using DSM-III/IV criteria, whereas I will use the term health anxiety to describe the cognitive and affective experience of anxiety about health.

1.3. Theories of health anxiety

I now outline and evaluate two contemporary theories of health anxiety. I focus on cognitive-behavioural formulation of health anxiety as both the HAQ and SHAI are based on this approach. The majority of this section comprises of a systematic review of the literature pertaining to the model originally conceived by Salkovskis and Warwick (1986). I will also briefly discuss the somatosensory amplification theory (Barsky, 2001), which is a related but distinct account of health anxiety.

A secondary aim of this thesis is to employ scores relating to the HAQ and SHAI to investigate the construct validity of this conceptualisation of health anxiety. I will also include a brief overview of the process of Somatosensory Amplification, a related, but distinct theory of health anxiety.

1.3.1. The cognitive model of health anxiety

The cognitive-behavioural of health anxiety was initially proposed by Salkovskis and Warwick (1986) who reported successfully treating health anxiety in a single-case series. This was expanded in later papers and book chapters (Salkovskis & Warwick, 1990; 2001). The cognitive model draws on Beck's conceptualisation of anxiety being a result of a perception of being under threat (Beck & Emery, 2005). Central to this is the often catastrophic misinterpretation of novel or unusual bodily sensations as being indicative of a health problem. The model suggests that individuals form dysfunctional or inaccurate beliefs or assumptions about their health, generally through their early life experiences. These beliefs are then activated by a critical incident which then leads to negative automatic thoughts related to health and illness.

An example may be a middle-aged man who thinks he is having a stroke because he notices a tingling in his left hand (misinterpretation of bodily states). He may hold the belief that 'all men will die of a stroke in middle age' because two family members had strokes at similar age that he is now (dysfunctional belief about health).

He was more aware of this because he had reached the same age and also because a colleague had recently had a stroke (critical incident). This led to him having the thought that this tingling was due to a stroke rather than part of normal sensory variation (cognitive misattribution). See figure 1 for overview of the cognitive model of health anxiety.
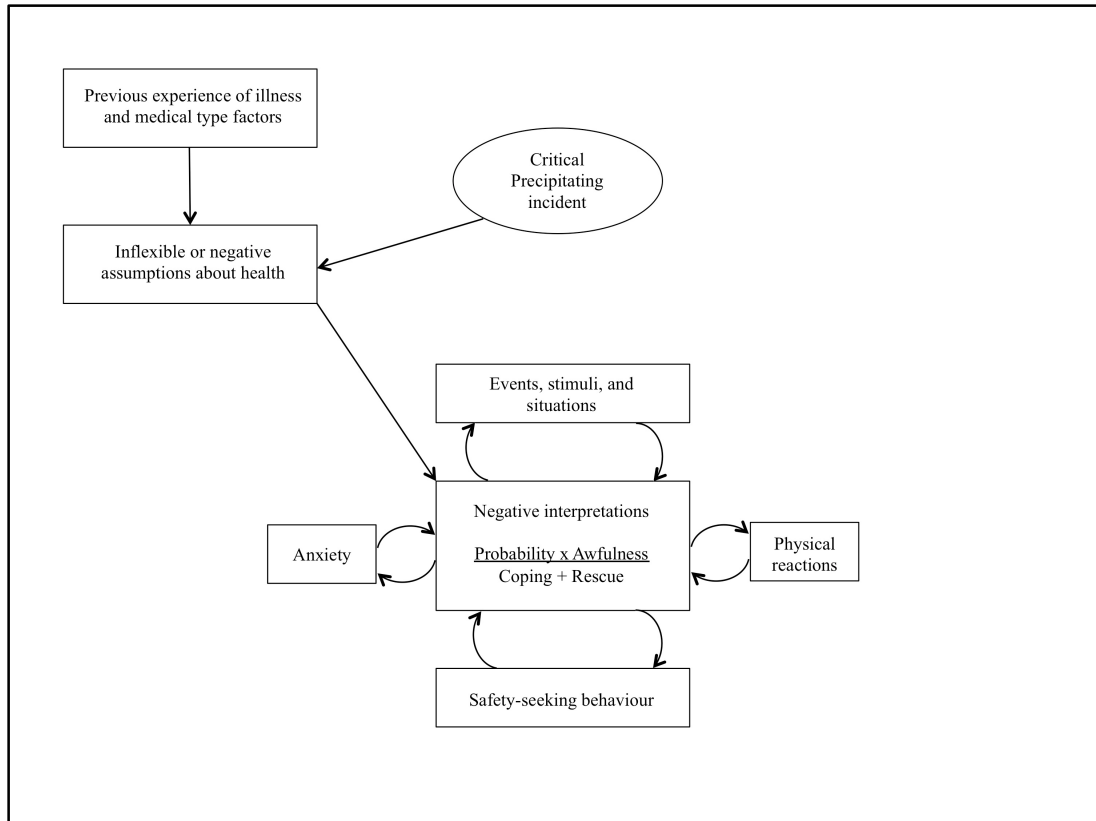
Figure 1. The cognitive model of health anxiety (Salkovskis & Warwick, 2001).

Dysfunctional beliefs about health are then maintained by increased attention being paid towards bodily states or selectively attending towards information that confirms the belief that they are unwell. 'Safety seeking' behaviours may also occur; a person may seek out information that they are unwell by touching the relevant area of their body or by focussing their attention toward unusual bodily states. Another example of this is frequently seeking out reassurance from medical professionals that they are in fact well.

In the example of the man who was fearful of having a stroke, he may then keep looking out for signs he is having a stroke such as his fingers tingling again (selective attention). He may touch or rub his hand, which itself may lead to his fingers tingling more, in turn leading to an increase in anxiety. He may visit his doctor (safety seeking behaviour) and be reassured that he is not ill. Unfortunately, the relief felt by the man will only be temporary, as reassurance will only help a person realise they are not ill at that time. As soon as he notices another bodily state that concerns him, he will then become anxious again and then require additional reassurance. As an alternative to providing reassurance, therapy founded on the cognitive model focusses on a person's

appraisal and misinterpretation of benign bodily states (Taylor & Asmundson, 2004). A therapist will help a client develop an alternative theory of the meaning of unusual or novel bodily states (e.g. that tingling in the fingers is a result of normal variation rather than a sign of illness).

1.3.2. Evidence for the cognitive model

I now review the research investigating each component of the cognitive model, beginning with a discussion of the evidence for differential health related beliefs being held by those with health anxiety. I performed a systematic review of the literature and I included articles if they evaluated a component of the cognitive behavioural model as described by Salkovskis and Warwick (2001). The following search terms were entered into Web of Science; 'hypochond* OR health anxiety' the results were then combined with output from the term 'cognitive'. I also searched all articles citing the two original papers by Salkovskis and Warwick (1986; 1990). Reference lists of included papers were also searched for relevant papers. Search dates ranged from 1986-2016.

1.3.2.1. Health related beliefs

According to the cognitive model, individuals with health anxiety are hypothesised to possess dysfunctional beliefs about illness and wellness. The content of the beliefs may be similar to those held by non-health anxious individuals, but are thought to be less-flexible and more extreme than the wider population.

A study published soon after the initial cognitive model papers, investigated the core beliefs held by 10 people undergoing Cognitive Behavioural Therapy (CBT) for hypochondriasis (Wells & Hackmann, 1993). These were accessed by the treating therapist who asked participants about mental images they may experience that are related to their health. The therapist then used the 'downward arrow' approach used in CBT (Taylor & Asmundson, 2004) to elicit core beliefs about health. Participants reported specific beliefs about illness (e.g. 'worrying about my health will keep me safe', 'if I'm ill people will abandon me') and death (e.g. 'death means this will go on forever', 'death means I'll be weak and punished forever'). Despite this study having a small sample and lacking control participants, it provides some indication of the beliefs held by those with health anxiety. Although this study may be criticised for asking only about images related to health and death: it may be that other images are important to the subjective experience of health anxiety, but these may not have been uncovered because these were not asked about.

When presented with health-related vignettes (e.g. accidently swallowing contaminated water and being advised to seek medical attention), those with hypochondriasis are more likely to believe this will result in a negative health outcome (e.g. becoming unwell) than healthy controls. This effect was limited to health-related vignettes and was not found in non-health related vignettes (e.g. arriving home to find

the house had been burgled). Participants were then given a further piece of information designed to either provide reassurance or increase anxiety. The authors found both the hypochondriasis group and controls equally responsive to reassurance (Haenen, De Jong, Schmidt, Stevens, & Visser, 2000). This is in contrast to the finding that those with health anxiety are less responsive to reassurance (Lucock et al., 1997), although reassurance in the first study may have only been effective because the vignettes concerned hypothetical situations rather than being about the participant themselves. This effect was replicated using a questionnaire study of 158 students. Those scoring more highly on measures of health anxiety were more likely to predict a worse outcome from being unwell than those with lower health anxiety, although this effect only held for more serious illnesses, rather than minor concerns (Marcus, 1999).

Another study asked 133 undergraduate students to rate out of 100 whether twenty different descriptions of bodily symptoms were an indication of illness. Participants were also asked to rate whether the symptom was likely to be a sign of severe or minor illness. Those scoring more highly on the Illness Attitude Scales were more likely to rate a symptom as being a sign of illness and as having a catastrophic outcome (Marcus & Church, 2003). This relationship held even when accounting for negative affect and worry, so suggests these health beliefs are unique to health anxiety rather than being accounted for by another generic cognitive process (e.g. trait anxiety).

Two studies conducted with undergraduates (Fulton, Marcus, & Merkey, 2011) employed the Irrational Health Beliefs Scale (IHBS; Christensen, Moran, & Wiebe, 1999) a measure containing 20 vignettes that assesses biases in an individual's interpretation of health-related information. They reported that those scoring more highly on health anxiety measures were more likely to report irrational beliefs about health, in both samples this relationship was mediated by trait anxiety. The authors point out that on the basis of their study irrational health beliefs cannot be said to be a causal factor in health anxiety. It may be that the relationship could be the reverse or that irrational health beliefs are more likely to be held by those with a particular cognitive style or those who score more highly on measures of trait anxiety.

Compared to healthy controls, individuals with hypochondriasis are more likely to believe a bodily symptom is indicative of a serious illness, to a lesser extent this effect was found in an anxiety disorder sample (Florian Weck, Neng, Richtberg, & Stangier, 2012a). This may be because of some conceptual overlap between health anxiety and generalised anxiety. Additionally, this was assessed using a series of vignettes so refer to a hypothetical situation, it may be that this effect would not be replicated if the symptom was one actually experienced by participants.

There is also evidence that individuals with hypochondriasis have a more rigid and narrow view of being in good health (Weck, Neng, Richtberg, & Stangier, 2012b). Those with hypochondriasis are more likely to hold beliefs that good health involves

being without any signs of illness and are more likely to believe that an ambiguous symptom is a sign of disease (Barsky, Coeytaux, Sarnie, & Cleary, 1993).

Overall, there is convincing evidence that those with severe health anxiety or hypochondriasis hold irrational beliefs about their health. Additionally, this effect appears to be restricted to beliefs about health and illness rather than more general negative or dysfunctional beliefs. There is also emerging evidence that those with health anxiety have a narrower view about what constitutes good health.

1.3.2.2. Critical incidents/triggering events

Another component of the cognitive model of health anxiety is the hypothesis that a particular event (e.g. a colleague becoming unwell) can trigger latent dysfunctional beliefs or assumptions about health. Three studies reported experimental evidence relating to the triggering of health anxiety (Kaur, Butow, & Sharpe, 2013; Lecci & Cohen, 2002; Marcus, 1999).

In the first study, Marcus (1999) asked 2,117 undergraduates to complete a task that involved making sentences out of groups of words that were presented in nonsensical order. Health-anxiety was triggered using illness-related words, participants then completed the State scale of the STAI (Spielberger, Gorsuch, & Lushene, 1970). Marcus reported that threat related words did not increase state anxiety in those scoring high in health anxiety, however it did increase state anxiety in those scoring lower on health anxiety.

Another study conducted with undergraduate participants (Lecci & Cohen, 2002) triggered health anxiety by taking a medical history and then blood pressure readings. Participants were informed they had a 'dangerously high blood pressure' and then completed a Stroop task, control participants were not given this information. Those in the experimental condition showed greater interference on a Stroop task, this effect was specific to negative health related words and did not occur in response to non-health related words.

A third paper (Kaur et al., 2013) randomised undergraduate participants to receive a health anxiety induction using the same procedure as Lecci and Cohen (2002) but found no effect on attentional processing as assessed by a dot-probe task and an emotional Stroop task.

There is mixed evidence from studies attempting to trigger health anxiety in laboratory studies, with studies using the same method reporting conflicting results.

1.3.2.3. Core cognitions

The cognitive model of health anxiety proposes four core cognitions thought to be involved in the development and maintenance of health anxiety. These relate to the

perceived: likelihood of illness; cost and awfulness of illness; means of coping with illness; and inadequacy of medical services (Salkovskis & Warwick, 2001).

The Health Cognitions Questionnaire (HCQ; Hadjistavropoulos et al., 2012) was designed to directly measure these core cognitions. In the scale validation paper, high HCQ scores were associated with increased health anxiety and a lack of response to medical reassurance in both a medically well community sample of 273 and another sample of 208 who were physically unwell. In the healthy sample, only 'perceived difficulty coping with illness' and the 'likelihood of illness' predicted health anxiety (medical services inadequacy and awfulness of illness did not). In the medical sample perceived 'medical services inadequacy', 'likelihood of illness' and 'awfulness of illness' predicted increased health anxiety (difficulty coping did not). This indicates there may be some differences in the experience of health anxiety in those who are well, who may think more about the challenges and likelihood of being ill, compared to those who are unwell whose thoughts may be more focussed on the experience of being ill. The HCQ-perceived inadequacy of medical services subscale predicted lack of response to reassurance (assessed by the Reassurance Questionnaire; Speckens et al., 2000) even when the SHAI-14 was included as a predictor. This is a potentially helpful aspect of the HCQ, as it may assist clinicians in determining those who will not be reassured by a medical professional. A potential difficulty of this scale is that it focusses on a hypothesised group of core-cognitions and was developed from a pool of items that aimed to directly assess these hypothesised cognitions. There has been no other investigation as to whether there are other cognitions that relate specifically to health anxiety, nor whether they are the most important; it may be that other cognitions are important to the health anxiety construct or may be more important in predicting response to reassurance. It is not possible to determine this, simply because the focus of the HCQ is the core cognitions described above and items concerning other cognitions have not been asked about. For example, it is notable that the HCQ and the SHAI do not feature questions about death, despite this historically being a feature of hypochondriasis (Kellner, 1987) and the presence of death-related mental images in the study conducted by Wells and Hackman (1993). This is particularly relevant as fear of death has been closely linked to hypochondriasis (Noyes, Stuart, Longley, Langbehn, & Happel, 2002).

Indirect evidence for the importance of the core-cognitions described by Salkovskis and Warwick (1985) comes from the success of a cognitive reattribution intervention in reducing health anxiety (Kerstner et al., 2015). In this study participants were trained to alter their attribution of a bodily state from an illness appraisal to finding an alternative external explanation. A recent meta-analysis of 13 RCTs reported CBT for health anxiety was superior to no-treatment controls, with an average pre/post treatment effect size of 0.95 (Olatunji et al., 2014). CBT interventions for health anxiety

also include other non-cognitive components (Taylor & Asmundson, 2004), therefore it is not possible to determine whether cognitive or non-cognitive aspects of the intervention reduced health anxiety.

In summary, there is some evidence for an association between health anxiety and the core cognitions included in the cognitive model of health anxiety, however these results warrant replication. It is also not possible to determine whether additional cognitions specific to health anxiety are of importance as the current research has focussed only on those outlined by (Warwick & Salkovskis, 1990). The role of cognitions in health anxiety is further supported by evidence from successful treatment of health anxiety with CBT. However, there was only one study that specifically focussed on cognitions, in the other treatment studies non-cognitive treatment variables may also have contributed to the reduction in health anxiety.

1.2.2.4. Maintaining factors

From a cognitive behavioural perspective, various factors are thought to be responsible for the maintenance of health anxiety (Salkovskis & Warwick, 2001). I now review potential maintaining factors for the maintenance of health anxiety.

Cognitive rumination: A study of 198 undergraduate students found a relationship between increased health anxiety and increased cognitive rumination, the authors argue the cognitive model of health anxiety should be extended to include this variable (Marcus, Hughes, & Arnau, 2008). A later study with a medically-healthy community sample of 410 also found increased rumination was associated with increased health anxiety and remained a unique predictor even when negative affect was included in the analysis (Fergus, 2013b).

Mental imagery; In the study reviewed above by Wells and Hackman (1993), those being treated for health anxiety report specific mental images associated with health. For all 10 participants, these involved images of being dead or of others being happy that they are dead. A qualitative study concerning mental imagery in 55 individuals with a DSM-IV diagnosis of hypochondriasis reported 72% of participants experienced distressing imagery about being ill, dying or of how their death may affect their family or friends (Muse, McManus, Hackmann, & Williams, 2010).

Memory bias: an experimental word recognition and recall task was carried out with 28 participants with hypochondriasis or somatoform disorder and 14 healthy controls. Words categorised as either positive, negative, neutral or pain related, were displayed to participants. Compared to controls those with hypochondriasis were more likely to rate a word as being negative or related to pain and also demonstrated better recall of these words than controls. (Pauli & Alpers, 2002).

In a delayed recall task, those with hypochondriasis could recall more health related words than controls, although because the delay in this study was only seven

minutes, it is unclear whether this bias is longer lasting (Brown, Kosslyn, Delamater, Fama, & Barsky, 1999). Specificity of recognition memory of health-related threat words was not replicated in an experimental study using the emotional Stroop. Individuals with hypochondriasis had equally good recognition memory of words relating to health-threat, panic and bodily symptoms, this effect reduced following cognitive-behavioural therapy (Gropalis, Bleichhardt, Hiller, & Witthoft, 2013).

In an undergraduate sample, those with increased levels of health anxiety showed better recognition memory for health-related words, but did not recall more health-related words (Ferguson, Moghaddam, & Bibby, 2007). The delay before completing the recall task was shorter than in the study reported by Brown, Kosslyn, Delamater, Fama and Barsky (1999). Ferguson and colleagues argue that in their study there may have been insufficient time for participants to encode the health-related words in their study.

Attentional bias; has been assessed using three methods, the dot-probe, the emotional Stroop task and visual search tasks.

As assessed by the Stroop task, higher levels of health anxiety related to greater interference in threat related words (Lecci & Cohen, 2002; Lecci & Cohen, 2007). Similarly, those scoring higher on the Illness Attitude Scales are more likely to attend to health related words (Owens, Asmundson, Hadjistavropoulos, & Owens, 2004). Another study also reported that compared to controls, those higher in health anxiety react more quickly to health related words than non-health words (Ferguson et al., 2007). Two additional studies conducted with undergraduates reported that, those higher in health anxiety showed a greater tendency to attend towards health-threat words and then to generic-threat related words (Karademas, Christopoulou, Dimostheni, & Pavlu, 2008; Owens et al., 2004). A final study found those high in health anxiety attended more to words about bodily symptoms, but not to illness related or neutral words (Witthöft et al., 2013). This was an fMRI study with a sample of only 24, so may not have had sufficient statistical power to detect smaller effects on attention bias.

All the previously reviewed studies using the Stroop task were carried out with undergraduate students and excluded those with hypochondriasis. This is potentially problematic as it may be that attentional biases differ in those with higher levels of health anxiety. Overcoming this shortcoming, again using the emotional Stroop task, Gropalis et al., (2013) found those with a DSM-IV diagnosis of hypochondriasis show an attentional bias towards threating words related to health and also to words relating to panic.

The dot-probe task has also been employed as a measure of attention bias. In a sample of 95 undergraduates, Increased health anxiety was not related to attentional bias towards health-related information in a dot-probe task (Jacoby, Wheaton, &

Abramowitz, 2016). This lack of effect was despite a large enough sample size to detect a medium effect size, the authors suggest this may be because the health-related threat words may not have personal relevance to participants or because the dot-probe task has poor reliability. In contrast, an earlier study of 83 undergraduates did find a bias towards health-related information in a dot-probe task (Jasper & Witthoft, 2011). Similarly, following a mood induction that aimed to increase illness worry, a third study using the dot-probe found no relationship between increased health anxiety and attention toward health-threat (Kaur et al., 2013). One study (Lee, Watson, & Frey-Law, 2013) personalised the dot-probe task to include health-threat words of personal relevance to individual participants. They found that rather than attending more rapidly to these words, participants were slower to remove their attention.

Two studies assessed for attentional bias via use of a visual search task. Participants meeting diagnostic criteria for hypochondriasis and controls were shown 'perceptually degraded' health-related and neutral words on a computer screen and read the word out loud. Those with hypochondriasis perceived less health-related words than controls (Brown et al., 1999). A study in undergraduate populations, found all participants (both low and high in health anxiety) selectively attended more towards health-related words (Shields & Murphy, 2011).

Indirect evidence for the cognitive model of health anxiety is found in studies that report a reduction in somatically focused attention following an attention training intervention (Weck, Neng, & Stangier, 2013). A later study provided an attention training intervention to students high in health anxiety (Schwind, Gropalis, Witthoft, & Weck, 2016). There were three groups in this study, the first received an attentional training intervention designed to help participants divert their attention away from bodily states that concern them. The second group were taught to train their attention onto their body (predicted to increase somatically focussed attention) and another group that received no intervention. Contrary to their predictions, only those who were taught the body-focussed attention training showed a reduction in health anxiety. Similarly, attentional biases assessed by the Stroop task have been shown to reduce after completing cognitive-behavioural therapy Gropalis et al., (2013).

Overall, the evidence for attentional biases in those experiencing health anxiety is mixed. Studies employing the emotional Stroop have reported a specific attentional bias and interference when health related words are presented, this effect has been found in both student and health-anxious populations. The only study that did not find an effect may have had too small a sample for differences between those with health anxiety and controls to be detected (Witthöft et al., 2013). Studies using the dot-probe task have been less likely to detect an attentional bias, although when stimulus words are personalised to the participants an effect is found. Visual search tasks have also reported contradictory findings, with those with hypochondriasis seeing less health-

threat words than controls, whereas a study conducted with students found all participants attended towards health-related words. Finally, there is evidence for a reduction in attentional biases after successfully completing psychological therapy.

1.3.2. Somatosensory amplification

Distinct, but complementary to the cognitive theory of health anxiety is the suggested process of somatosensory amplification (Barsky, 2001). A defining feature of health anxiety is the misattribution of benign bodily sensations as an indication of a health problem. Somatosensory amplification seeks to explain this as either being due to an increased sensitivity to one's sensory information or due to a reporting style that overemphasises the severity of this information (Barsky, 2001).

Barsky (2001) points out that an individual's reports of their sensory information cannot be separated from their interpretation of their senses: these reports are almost certainly influenced by a person's belief structures, attentional biases or cognitive misattributions.

I suggest that although somatosensory amplification is a different theory to the cognitive behavioural model, it is in fact another perspective on the same information. The emphasis is placed on the senses and misinterpretations of sensory information: our thoughts, feelings and behaviours modulate this information. In the cognitive model, thoughts are primary and distort the information from the senses via attentional biases.

In his narrative review Barsky (2001) concluded that individuals with heath anxiety tend to report an increased sensitivity to sensory stimulation (e.g. painful stimuli). However, on more objective measures (e.g. awareness of own heartbeat) they perform identically to controls without health anxiety. A meta-analysis conducted by Marcus et al. (2007) reported on an additional four studies that replicated these findings. Both papers point out that there have been inconsistent findings in the literature. Additionally, some of the included studies do not report co-morbidity with other mental health problems, which is problematic as somatosensory amplification has been reported in anxiety, depression and somatising disorder. Overall, I argue this evidence suggests that people with health anxiety experience sensory information related to their bodies as being heightened compared to those without health anxiety. This is most likely due to their appraisal of this information rather than actual sensory differences.

1.4. Classical test theory, validity and reliability

I now provide a brief overview of validity and reliability as applied to psychological research. This is relevant because in chapter 3 I will provide a meta-analytic review of the published reliability coefficients for the HAQ and SHAI, I will also conduct a reliability generalisability analysis (see chapter 2 for method). In chapter

4, I will provide a narrative review of information abstracted from these methods relating to the construct validity of health anxiety.

1.4.1. Validity

Validity may be simply defined as 'measuring what you intend to measure'. However, it is often wrongly thought of as a stable characteristic of a measure that one may continually assess and refine until a test's validity is determined. This view is incomplete, as validity may be more accurately seen as a property of both tests and theories. Validity is also not a static property, rather it is an ongoing process: Information about the reliability and validity of a test can be used to aid development of theory, likewise theory generation should inform the construction of measures (Strauss & Smith, 2009).

In order to conduct research on a particular psychological concept (e.g. depression) or process (e.g. attention bias), it is necessary to have a valid means of measuring it (e.g. Beck depression inventory, emotional Stroop task). Early in the history of psychological research neither robust, valid theories nor robust valid measures existed. This lack of knowledge base meant that researchers would typically rely on observation and description of poorly understood concepts and attempt to measure them with the tools available to them. An initial advance came in the form of criterion validity, or the ability of a test to predict an individual's membership of a certain class or category. For example, a person can be said to be depressed, establishing this could be achieved by direct observation (e.g. sleeping less, slower movements, more tearful) over the course of a number of weeks. The person could then be interviewed to understand their experiences (they may report feeling unhappy, lethargic and lacking in motivation). A scale could then be designed that predicted whether a person would fall into a 'depressed' category. As criterion validity continued to be employed, the problems inherent with it became apparent. The first is that the test is only as good as the criterion selected, which may itself not be valid. A second difficulty is that criterion validity does not facilitate theory building. Both difficulties are problematic, but understandable given the lack of robust psychological theory from which criteria may be developed (Strauss and Smith, 2009).

Construct validity is a term used to describe an overarching concept of validity, that subsumes the other types of validity. This approach argues for the legitimacy of hypothetical constructs that are not directly measurable, but are inferred by direct observation, self-report or scores on psychometric tests (Cronbach and Meehl, 1955). A test can be said to have good construct validity if it accurately and reliably assesses a hypothetical latent construct and any changes in that construct. For example, the existence of a latent construct of depression could be inferred using the criteria described in the previous paragraph and a scale developed to assess it. An example of this is the Beck Depression Inventory (Beck, Steer, & Brown, 1996) which has

undergone validation efforts and has been shown to reliably and accurately reflect a latent construct of depression.

These sources of information can then be placed under further analysis to determine whether they are coherently and meaningfully related. For example, it would be expected that in depression, increased low mood would be related to more negative thinking. If these variables were highly correlated they could be considered to have *convergent validity*, or likely to be two distinct, but related indicators of the same construct. Similarly, depression should not be strongly associated with anxiety as it is considered to be a distinct construct. A measure or test that can distinguish between two separate constructs is said to have *discriminant validity* so can distinguish between two related but distinct constructs.

One means of investigating construct validity is by use of the multi-trait multimethod matrix (MTMM; Campbell and Fiske, 1959). This approach involves assessing the convergent and discriminant validity of various measurements of the construct of interest. This information is then summed in a correlational matrix and can be reviewed qualitatively to determine whether the pattern of results is as predicted and consistent with the current proposed theoretical construct (Campbell and Fisk, 1959). It is no longer necessary to evaluate MTMM data qualitatively as advances in statistical procedures and computation modelling such as Confirmatory Factor Analysis and Structural Equation Modelling allow for quantitative testing of latent variables and constructs (Strauss and Smith, 2009).

As previously described, in line with falsificationist theories of science (Popper, 1972) a measure can never be said to possess construct validity, rather it is property that should be continually evaluated. Additionally, every test of the construct validity of a measure also provides an opportunity to better understand the construct that the measure purports to assess (Strauss and Smith, 2009).

1.4.2. Reliability

Reliability refers to the consistency of a test or instrument. A measure is said to have good reliability if test scores remain constant under the same conditions. It is important to note that reliability is not a fixed property of a test, rather it refers to a test's performance on a particular occasion with a particular sample (Wilkinson, 1999). There are several methods of evaluating reliability, but only *test-retest reliability* and *internal consistency* are relevant to this thesis. This is because these are the forms of reliability that are used when assessing the quality of self-report measures. Parallel forms reliability is also sometimes employed when initially generating a scale, but did not feature in the development of either of the scales investigated in this thesis. *Test-retest* reliability assesses the stability of test scores over time, by being administered to the same participants with a delay between administration. The internal consistency of a

test, assesses whether the items contained in an instrument are related to one another. This may be assessed in two ways: *Split-half reliability* is assessed by dividing an instrument into two halves. To avoid order effects, this is usually achieved by alternating items, with even numbers going into one half, odd numbers to the other. Each half of the scale is treated as a separate measure and then they are correlated with one-another. A test is considered to have good split-half reliability if the correlation coefficient is sufficiently high. *Coefficient alpha* (Cronbach, 1951) is an internal consistency coefficient that is generated by correlating all items in a test with one-another. A systematic evaluation of 696 (out of a total 2078) scale-development papers reported test-retest reliability (19.0% of papers) and the alpha coefficient (66.5% of papers) were the most frequently employed reliability coefficients used when evaluating questionnaires (Hogan, Benjamin, & Brezinski, 2000).

Reliability coefficients may provide additional sources of information about a construct under investigation. Classical test theory (Lord & Novick, 1968) regards an individual's test score to be the sum of their 'true score' and their 'error score' and may be expressed as the following formula:

Observed score = true score + error score

An individual's 'true score' represents their actual score on a particular test. However, an individual's observed score will also contain additional sources of variation, for example testing conditions or individual characteristics such as tiredness. These are summed and termed the 'error score'. The various sources of variance summed in an individual's 'error score' may be used to investigate whether there is systematic variance in test scores across participants (e.g. if reliability is reduced when a scale is used with particular samples). This may indicate specific variables that consistently impact on reliability estimates.

Reliability generalizability analysis is a form of meta-analysis that investigates the variation in reliability scores across studies (Vacha-Haase, 1998). This analysis allows one to calculate an average reliability for a test or an average for the application of a test within particular populations. It also allows one to statistically investigate potential sources of variation in reliability coefficients. For example, if a particular test of depression were administered in a range of studies and its reliability estimate was consistently poorer in older people then this indicates that the test may not adequately capture the experience of depression in older adults. On the other hand, it may have implications for the theory of depression as the issue may lie in the theory rather the test.

More recently it has been argued that a study's mean scores and variances can be incorporated into a reliability generalisability analysis. This may then be used to determine moderators of variability in test scores (Botella, Suero and Gambara, 2010).

Reliability generalisability is relevant to this project as it will provide insight into the performance of the HAQ and SHAI with various populations. In addition, information about consistent sources of variance in reliability coefficients may add additional information to the theory of health anxiety. I return to the topic of reliability generalisability in the method section.

To comment on the reliability of a measure, it is important to briefly review what is considered 'acceptable' reliability. This varies according to which source text is consulted. Cicchetti (1994) suggests the following qualitative descriptions for reliability coefficients (including split-half reliability); $\alpha<0.70$ 'unacceptable', $\alpha=0.70-.79$ 'fair', $\alpha=0.80-0.89$ 'good', $\alpha>0.90$ 'excellent'.

Similar ratings specific to coefficient alpha have been advocated by George and Mallery (2003), who suggest a lower alpha coefficient is acceptable; $\alpha<0.5$ 'unacceptable', $\alpha=0.5-0.6$ 'poor', $\alpha=0.6-0.7$ 'questionable', $\alpha=0.7-0.8$ 'acceptable', $\alpha=0.8-0.9$ 'good', $\alpha>0.9$ 'excellent'. An alpha coefficient over 0.95 is considered undesirable as it indicates there may be items in the scale that are measuring the same concept, so should be deleted (Streiner, 2003).

Interestingly, the general 'rule of thumb' often published in text books is that reliability coefficients should be at least 0.7 or higher. This has been attributed to a text by Nunnally (1978) and has been described as a methodological 'urban legend' (Lance, Butts, & Michels, 2006). In the original text, Nunnally actually states that acceptable reliability for a test depends on its purpose and on the context in which it is to be used. He suggests that for basic correlational research an alpha of .70 is acceptable, whereas if a test is to be used in experimental studies an alpha of 0.80 is more appropriate. If a policy or other decision is being made on the basis of scores derived from a measure, then a minimum alpha coefficient of 0.90-0.95 should be selected (Nunnally, 1978).

It is also important to note that the alpha coefficient may be inflated by larger samples and a larger number of test items. This has led several authors to suggest that psychometricians move to an alternative such as the Omega coefficient (Dunn, Baguley, & Brunsden, 2014; Peters, 2014). Cronbach has argued that the alpha coefficient is likely to be a flawed measure of reliability and should be used as part of a selection of tools and observations that may be used in concert to evaluate reliability (Cronbach & Shavelson, 2004).

1.5 Measures of health anxiety

I now introduce the HAQ and SHAI, a description of the aims for their development and then provide a rationale for this thesis.

1.5.1. The Health Anxiety Questionnaire (HAQ)

Is grounded in a cognitive-behavioural model of health anxiety, has 21 items and is based on the Illness Attitude Scale (IAS; Kellner, Abbott, Winslow, & Pathak, 1987). In the original scale development study, the HAQ demonstrated good test-retest reliability (r = 0.94) and internal consistency (alpha = 0.92). The HAQ also displayed good discriminant validity and could differentiate between those with health anxiety and those with other forms of anxiety. Norms have been published for the general population, nursing students, outpatients and users of a psychology service (Lucock & Morley, 1996).

The HAQ conceptualises health anxiety from a cognitive behavioural perspective and was developed using items from the Illness Attitude Scales (Kellner, 1987) and in consultation with individuals experiencing health anxiety. The HAQ uses words such as 'worry' to describe cognitions (e.g. item 1. "Do you ever worry about your health?") and includes affective states such as 'afraid' (e.g. item 15. "Do you ever feel afraid you may have cancer"). Reassurance seeking is also assessed (e.g. item 5. "Do you ever examine your body to find whether there is something wrong?") as is the extent that bodily symptoms interfere with daily functioning (e.g. item 21. Do your bodily symptoms stop you from enjoying yourself?). Because of a focus on the cognitive behavioural perspective and because the measure was developed under DSM-IV/ICD-10, an individual with DSM-IV hypochondriasis (or severe health anxiety according to the cognitive model) will likely score very highly on this scale. However, someone with a DSM-V diagnosis of illness anxiety disorder may have a lower score as the items concerning somatic symptoms may not apply to them. Likewise, an individual with DSM-V somatic symptom disorder may also score less highly as they may not experience any anxiety about their health, so the items about health worries and fear may not apply to their experience. Related to this is that fact that all items in the HAQ refer to generic worries, fears, body vigilance and somatic symptoms rather than concerning a specific symptom. This means that if an individual with somatic symptom disorder had a concern about a particular bodily state then this would be detected by generic items rather than an item tailored to their experience. See appendix A for the HAQ.

1.5.2. The Short Health Anxiety Inventory (SHAI)

Overview

The SHAI has a long-form 64 item inventory (Salkovskis et al., 2002), which has been shortened to an 18-item version. The 18-item version has four questions concerning an individual's perceptions of the negative consequences of becoming unwell. These items were not designed to measure health anxiety, but are often included in total scores of health anxiety (see Alberts, Hadjistavropoulos, Jones, & Sharpe, 2013

for review). Therefore, some studies provide a total score using either a 14- or 18-item version. Salkovskis et al., (2002) describe aiming to produce a scale that could differentiate between people who have health anxiety from those who have a genuine illness, but are not overly anxious about it. Previous measures can be criticised for having items that are applicable to people with a physical illness, which led to elevated health anxiety scores in groups evaluated with such scales (see below). According to the original validation paper, the SHAI has good discriminative validity and could differentiate between those with health anxiety, social anxiety, panic disorder and non-anxious controls. The short version of the SHAI has good internal consistency (alpha = 0.89).

Similarly to the HAQ, the SHAI was deliberately designed to reflect the cognitive behavioural model of health anxiety, it also uses items that reflect a generic concern about health, including fear of illness (item 5d. "I am always afraid that I have a mental illness") and mental imagery (item 6d. "I constantly have images of myself being ill"). It also assesses somatosensory amplification/preoccupation with bodily states (item 3d. I am constantly aware of bodily sensations or changes; item 10d. If I have a bodily sensation or change I must know what it means). The SHAI also assesses whether a respondent typically is reassured by a visit to a physician (e.g. item 8a. "I am lastingly relieved if my doctor tells me there is nothing wrong"). Similarly to the HAQ, all the question items refer to generic concerns so will not assess the presence (or change) in a specific somatic symptom. Again, those with a DSM-IV diagnosis of hypochondriasis are likely to score more highly than those with a DSM-V diagnosis of illness anxiety disorder (if they do not experience somatic symptoms) or somatic symptom disorder (if they do not experience anxiety). As with the HAQ, this means that if working clinically with individuals with the latter diagnoses care should be taken to capture their experiences that are highly relevant to their diagnosis (e.g. a specific somatic concern). This could be achieved by use of an idiographic measure such as a Personal Questionnaire (Shapiro, 1961) which may be helpful complements to standardised measures such as the HAQ and SHAI (Barkham et al., 2010).

Interestingly, in contrast to the HAQ the SHAI does not contain items about fear of death, which is surprising as fear of death is often thought to be a feature of hypochondriasis (Kellner, 1987) and the two have shown to be related (Noyes et al., 2002; Sirri & Fava, 2014). Although this may be because a specific fear of death does not feature in the cognitive model of health anxiety (Salkovskis & Warwick, 2001).

An existing review and meta-analysis of the SHAI has been published (Alberts et al., 2013). However, it is limited to providing only a narrative review of sample means, standard deviations, reliability coefficients and scale factor structure. I intend to investigate all these variables using meta-analysis, so will significantly extend their

analysis. Alberts et al., (2013) used meta-analysis only to aggregate information concerning the correlations between the SHAI and other psychometric measures.

Other SHAI versions

Factor analysis of the SHAI-14 and -18 has indicated item number 13 does not load onto either of the two factors thought to underlie the scale (Abramowitz, Deacon, & Valentiner, 2007). This has led to some researchers deleting this item and creating a SHAI-13 and -17. The studies using these versions of the scale are detailed in the results section.

1.5.3. Other measures of health anxiety

In this section I will briefly review information regarding the factor structure, reliability and validity of three other measures of health anxiety; the Whitely Index (WI; Pilowsky, 1967), Illness Attitude Scales (IAS; Kellner, 1987) and the more recent Multidimensional Inventory of Hypochondriacal Traits (MIHT; Longley, Watson, & Noyes, 2005).

1.5.3.1. The Whitely Index (WI)

The WI is an older measure of health anxiety developed by Pilowsky (1967), the orignial version of the scale included 14-items, respondents select whether the each statement applies to them in a yes/no format. More recently, the WI was altered to a 5-point likert scale format (Welch, Carleton, & Asmundson, 2009) scored from 0 (no) – 4 (often).

There have been 14 factor analytic studies of the Whitley Index. With six studies supporting a single factor solution (Pilowsky, 1967; Veddegjaerde, Sivertsen, Wilhelmsen, & Skogen, 2014), four finding a 2-factor and finally four studies finding a 3-factor solution (Conradt, Cavanagh, Franklin, & Rief, 2006; Hiller, Rief, & Fichter, 2002; Hinz, Rief, & Brahler, 2003). These have been reviewed by Veddegjaerde, Sivertsen, Wilhelmsen, & Skogen, (2014). The WI has shown good reliability, with alpha coefficients over 0.80 (Speckens, Van Hemert, Spinhoven, & Bolk, 1996).

1.5.3.2. The Illness Attitude Scales (IAS)

Is a 29-item measure, originally designed to include nine subscales (Kellner, 1987). It has been criticised for lacking a stable factor structure, as re-validation attempts have reported three, four or five factors, each corresponding to a different sub-scale (Longley, Meyers, Maxwell, & Letizia, 2014). In a factor analytic study Ferguson and Daniel (1996) report sub-scale alpha coefficients as low as 0.29, with the best performing at 0.75. This level of reliability is insufficient for applied research (Nunnally, 1978).

1.5.3.3. The Multidimensional Inventory of Hypochondriacal Traits (MIHT).

Is a 31-item scale with four subscales, each assessing a distinct dimension of hypochondriasis, these are; affective, behaviour, cognitive and perceptual (referring to attention to bodily states). The scale validation paper reported four factors, each loading on to the four sub-scales (Longley et al., 2005). This factor structure has since been independently replicated (Stewart, Sherry, Watt, Grant, & Hadjistavropoulos, 2008). The sub-scales have reliability coefficients ranging from 0.80 to 0.88.

It would be equally plausible to carry out this study using either the WI or the IAS. However, I elected to assess the properties of the HAQ and SHAI because they are both founded on the cognitive-behavioural conceptualisation of health anxiety. This means they have been designed to assess the same latent construct, so the properties of both may be compared with one another to determine if one or the other is psychometrically superior. The MIHT also emerged from the cognitive-behavioural literature (although is thought to assess dimensions of hypochondriasis), however there was insufficient data available to include the MIHT in this analysis (seven studies meeting inclusion criteria). I now outline the study undertaken for this thesis.

1.6 The current study

Valid and reliable measures are required for research and clinical purposes (Barkham et al., 2010). Scales will typically undergo a validation process where the reliability and construct validity are assessed (Rust & Golombok, 2009). I argue that it is important to re-investigate the reliability and validity of a scale again after it has been published. Validation samples may not have been entirely representative of the populations with which the measure is later used. For example, because of their convenience, university students often participate in scale validation studies, but may be dissimilar in composition to clinical samples. This is especially problematic in health anxiety research as younger people generally have fewer health concerns than older people (Beard et al., 2015) and having a physical health problem is related to increased health anxiety (Noyes, 2001a). Another potential difficulty with only a one-off validation study is that the results reported may have been discovered by chance, although this problem is less likely as scales are usually validated over more than one study.

There are other advantages to reinvestigating psychometric measures. Meta-analysis of the mean and standard deviation of health anxiety scores in various populations will allow the generation of normative data for use in future research or clinical practice. This would be particularly helpful if norms could be generated for populations that were not included in the initial scale validation procedure. It is likely that the mean scores in various populations will be heterogeneous in nature. This may

be due to differences in 'true' scores in the respective populations, may be influenced by differences in sampling or due to errors in measurement. I will also attempt to explore variations in mean score by use of meta-regression to determine if these variables systematically impact in HAQ/SHAI scores.

Meta-analysis of reliability coefficients may offer two helpful sources of data. The first is to provide information about the variation in reliability estimates in different populations. This may be helpful when selecting the most appropriate scale for use with different populations. A second source of useful information may be obtained by using meta-analysis to investigate study-level characteristics that may lead to systematic variations in reliability coefficients. This has been named a reliability generalisability analysis (Vacha-Haase, 1998).

I also argue that re-examining the properties of psychometric measures using meta-analysis has other possibilities for furthering research. As outlined in the 'validity' section above, validity is both a property of tests and of psychological theory. Information about the validity of tests can inform theory generation and vice versa (Strauss & Smith, 2009). Therefore, I suggest that meta-analysis of the published data derived from the HAQ and SHAI provides an opportunity to investigate both the construct validity of both scales and an opportunity to investigate the cognitive-behavioural conceptualisation of health anxiety.

In a validation study, the construct validity of a scale is generally assessed by examining the direction and strength of correlations between it and other measures that have known psychometric properties (Hinkin, Tracey, & Enz, 1997). It is assumed that a scale will correlate most strongly with other measures of the same construct (convergent validity), less strongly with related constructs and not correlate with measures of unrelated constructs (discriminant validity). For the HAQ and SHAI to have good construct validity they should correlate most strongly with other measures of health anxiety, then measures of generalised anxiety, followed by measures of obsessive compulsive disorder (OCD) and should correlate least strongly with measures of depression. I predict that this relationship will be found in this meta-analysis. I selected these measures because there is a theoretical rationale for the described expected pattern of relationships and because my scoping exercise indicated these measures would be available for analysis.

Indirect evidence relating to the construct validity of both scales may also be found by examining their correlations with measures of precursors of anxiety, for example the Intolerance of Uncertainty Scale (IUS; Freeston, Rheaume, Letarte, & Ladoucer, 1994), the Body Vigilance Scale (BVS; Schmidt, 1997) and the Anxiety Sensitivity Index (ASI-3; Taylor & Cox, 1998a). Examining these correlations can also provide evidence regarding the construct of health anxiety (Marcus, Gurley, Marchi, & Bauer, 2007).

Anxiety sensitivity refers to a fear that signs of anxiety will have a negative real-life impact (Taylor & Cox, 1998b). It is based on Reiss's (1991) expectancy model and is thought to be precursor or risk factor for developing anxiety. The initial 16-item version of the ASI was designed to assess anxiety sensitivity (Peterson & Reiss, 1987) although has been criticised as having insufficient items that are phrased in too general terms to detect the factors thought to underlie anxiety sensitivity. This led to the development of the longer, 36-item ASI-3R (Taylor & Cox, 1998a). Factor analysis of this scale revealed an overall anxiety sensitivity factor and the following lower order factors; fear of respiratory symptoms, fear of losing cognitive control, fear of appearing anxious in public and fear of cardiac symptoms. A clinical example of this would be someone fearing that a rapid heartbeat will lead directly to having a panic attack, or that increased breathing will lead to being rejected or humiliated by peers.

Intolerance of uncertainty (IU) refers to the inability to tolerate the uncomfortable feeling that may arise when one is in an uncertain situation and is thought to be a precursor to generalised anxiety disorder (Dugas, Gagnon, Ladoucer, & Freeston, 1998). More recent research has indicated IU may be involved across other emotional disorders including OCD, social anxiety, post-traumatic stress disorder and panic disorder (Boswell, Johanna Thompson-Hollands). The same study found that during CBT treatment, intolerance of uncertainty reduced alongside symptoms of emotional disorder, the authors argue IU may be a helpful component of trans-diagnostic therapy. The 27-item Intolerance of Uncertainty Scale (Freeston et al., 1994) was designed to assess IU and provides a total IU score.

Body vigilance refers to when an individual consciously focusses their attention on somatic sensations and is assessed by the Body Vigilance Scale (Schmidt, Lerew, & Trakowski, 1997). A recent factor analysis of the BVS in a non-clinical and anxiety disorder sample revealed a unitary factor (Olatunji, Deacon, Abramowitz, & Valentiner, 2007).

Correlations with the ASI-3, BVS and IUS were included in an exploratory analysis, so no prediction is made regarding the strength or the direction of correlations with these measures.

1.7 Aims

With the previous rationale in mind, this study aims to achieve the following with both the HAQ and SHAI:

1) Generate normative scores in various populations by extraction and meta-analysis of published mean scores and SDs;
2) Investigate variations in mean scores via use of meta-regression (supplementary analysis);

3) Calculate an average reliability for each scale and where possible for different populations;

4) Investigate sources of systematic variation in reliability via reliability generalisability analysis;

5) Examine the construct validity of both scales by review and meta-analysis of correlations with other measures;

6) Review the cognitive-behavioural conceptualisation of health anxiety by examining correlations between both scales and other measures.

2. Method

My choice of method for this thesis is meta-analysis. I begin with a general discussion of the aims and range of the applications of meta-analysis, including a rationale for why meta-analysis is the appropriate method to meet the research aims outlined in the preceding section. Next, I will consider the various criticisms levelled at meta-analysis and where possible how I overcome or ameliorated their impact.

In the 'meta-analysis procedure' section, I summarise the steps I took in conducting this analysis including my search strategy, inclusion/exclusion criteria and a discussion of risk of bias. I will argue that in this study, assessing risk of bias is not an appropriate strategy. Instead I will directly investigate the impact of various study level characteristics on variations in mean scores and reliability coefficients. Some of these variables are the same as those contained in a risk of bias scale. However, my use of this data differs from the usual application of these variables. This discussion is placed in the 'investigating heterogeneity in mean scores' section.

Finally, in the 'data analysis' section, I outline each component of the analysis conducted on both the SHAI and the HAQ. In summary, these are: (a) developing normative scores for various populations by meta-analysis of published mean scores and standard deviations; (b) reviewing the variation in alpha coefficients in various populations and use of reliability generalisability analysis to reveal potential sources of this variation; (c) review the construct validity of both scales via systematic review and meta-analysis of their correlations with other measures. This final section includes a discussion of the impact of these findings on the cognitive-behavioural conceptualisation of health anxiety as a latent construct.

In summary, the first half of the method section discusses and critically evaluates meta-analysis as a research tool. I then discuss the assumptions that underlie all my analyses such as employing a random-effects model and assessing heterogeneity. The closing half of this method section outlines the detail of the steps undertaken for each type of analysis.

2.1. What is meta-analysis?

The term meta-analyses was coined by Glass (1976), in an early analysis of psychotherapy outcome studies. Since then, meta-analysis has evolved into a collection of statistical procedures used to aggregate and analyse data from published research (Borenstein, Hedges, Higgins, & Rothstein, 2009). The use of meta-analysis as a research tool is increasing. For instance, one study searched Pubmed for all studies 'tagged' as a meta-analysis and reported 334 were published in 1991; this increased to 9,135 published in 2014 (Ioannidis, 2016). Pubmed tags, a form of article keyword are unlikely to be a wholly reliable source of information about the content of a study.

Nevertheless, these figures provide some indication of the increasing popularity of meta-analyses. Two other more methodologically rigorous studies estimated the number of meta-analyses published in 2004 to be over 1,300 (Moher, Tetzlaff, Tricco, Sampson, & Altman, 2007) this increased to 5,000 published in 2014 (Page et al., 2016). Both studies conducted a literature search for one month using Medline (November 2004 and February 2014 respectively) and used these figures to estimate the total likely to be published that year. Again, these studies are estimates, but still point to a large increase in the use of meta-analysis.

A meta-analysis will typically take the following steps (Borenstein, Hedges, Higgins, & Rothstein, 2009):

1. Define research question
2. Conduct literature search
3. Define inclusion/exclusion criteria
4. Extract data
5. Calculate effect sizes for each study
6. Conduct meta-analysis – by calculating a weighted effect size from those available in each study, then provide an overall effect size.

My research questions have been outlined in the introduction and I will detail my methods for each of the above points in the 'meta-analysis procedure' section. Before doing this, I outline an important methodological issue regarding choice of a random- or fixed-effect model. This is followed by a discussion of some of the criticisms levelled at meta-analysis and my efforts to overcome or ameliorate their impact.

2.2. Why employ meta-analysis?

This project aims to develop normative data for the SHAI and the HAQ and investigate the construct validity of both scales. Meta-analysis was selected because it allows data to be collected from all published applications of the HAQ and SHAI. This will allow normative data to be collected from multiple populations (e.g. those with health anxiety, an anxiety disorder and those attending medical outpatients). This permits a far broader range of data to be gathered amongst a greater number of populations than would be possible if collecting raw data. Similarly, collecting reliability coefficients from multiple applications of a test allows a far more comprehensive assessment of its reliability than would be possible via a one-off application with a single population of participants. This approach has the advantage of being able to assess the impact of study-level variables (e.g. language of test, nationality of participants) that may explain variations in HAQ/SHAI mean scores and reliability coefficients that would be impossible without collecting a large amount of raw data amongst multiple populations and languages.

2.2.1. Fixed vs random effects models

An important decision when conducting a meta-analysis is whether to adopt a fixed or random effects model, each is founded in different assumptions about the populations that are under analysis. A fixed-effect model assumes the effect sizes within the sampled population do not vary in magnitude, so will be homogenous (Hunter & Schmidt, 2014). In contrast, a random effects model assumes the effect sizes within the sampled population have a component of random variance, so will be heterogeneous (Hedges & Vevea, 1998).

The choice of whether to adopt a random or fixed effect model depends on which assumption is most appropriate to the study population of interest (Borenstein et al., 2009). This decision is also important as incorrectly applying a fixed-effect model to a population containing varying (heterogeneous) effect sizes can increase the type 1 error rate from the usual 5% up to 80% (see Field & Gillett, 2010). It has been argued that in real-world research concerning the populations of interest to psychology researchers that the average effect sizes are likely to be heterogeneous (Field, 2003).

Rather than selecting between a fixed or random effects model prior to analysis, Hedges and Vevea (1998) argue the choice of model should be determined statistically, by assessing the homogeneity of effect sizes: If the effect sizes are homogenous then a fixed-effect model is most appropriate, if significantly heterogeneous then a random effects model should be employed. If a random effects model is adopted because of significant heterogeneity of effect sizes it may be described as a 'conditional random effects model'. This term has been suggested to make it clear that this decision has been made statistically (Hedges & Vevea, 1998).

A second consideration when choosing between a fixed or random effects model concerns the generalisability of results. If a researcher intends to generalise their findings beyond the strictly defined range of studies included in the analysis then a random effects model must be used. If this is not required then a fixed effects model is more appropriate (Borenstein et al., 2009).

I employed a random effects model because there was significant heterogeneity in effect sizes (see results section) and as the goal of this thesis is to generate population norms for the SHAI and HAQ it is necessary to generalise my findings beyond the included studies. I also conducted a meta-analytic technique known as reliability generalisability analysis (Vacha-Haase, 1998), again assuming random effects. I discuss this in detail in the 'meta-analysis procedure' section.

2.3. Criticisms of meta-analysis

In this section, I discuss some of the general criticisms of meta-analysis and how they may be managed. These criticisms mostly apply to more 'traditional' applications

of meta-analysis such as evaluating pre/post treatment effect sizes or odds ratios in RCTs. In this study, I applied meta-analytic techniques to developing population norms, employed meta-regression to investigate the reliability and performed a meta-analysis of correlation coefficients to investigate the construct validity of the SHAI and the HAQ. I did not evaluate treatment effects, this means that although the usual criticism of meta-analysis are applicable to this study, they required accounting for in a different way. My attempts to do so are detailed in the meta-analysis procedure section. Because of my bespoke application of meta-analysis, I needed to adapt my response to these criticisms to the demands of this study, therefore some of my responses to these criticisms are accounted for in more depth in the meta-analysis procedure section.

2.3.1. The 'file drawer problem'

Journal editors are more likely to publish statistically significant or novel findings (Rosenthal 1979), meaning that non-significant findings are less likely to appear in journals and remain 'in the file drawer'. The usual level of acceptable probability in psychology research is 0.05, which means 5% of significant results are false-positives (type I error). In an extreme example, it may be that of 100 studies investigating a relationship between depression and shoe size, 5 found a significant result and were subsequently published. However, the remaining 95 did not find a significant result so were not published. This may easily lead to the conclusion that there is a relationship between depression and shoe size, even though all the published results are due to chance. Rosenthal (1979) offered a 'fail safe n' calculation that enables a researcher to calculate the number of studies reporting non-significant results required to reject a significant finding. Publication bias is important to consider when accepting the conclusions of published research and can also be an unwanted source of bias when conducting a meta-analysis. It should be noted that Rosenthal's calculation has been criticised for overestimating the number of null-findings necessary to reject a significant result (see Scargle, 2000 for an alternative formula).

The concept of publication bias has been extended to include effect sizes (Hedges, 1992). There is evidence for the existence of publication bias in the medical literature; a meta-analysis of studies investigating publication bias reported that significant effect sizes are more than twice as likely to be published than non-significant (odds ratio (OR) 2.40, 95% CI [1.18, 4.88]; Song et al., 2009). This replicated an earlier study reporting similar findings (OR 2.54, 95% CI [1.44, 4.47]; Dickersin, 1997). This realisation led to the development of a 'fail safe n' calculation for effect sizes in meta-analyses (Orwin, 1983). Similarly, to Rosenthal's (1979) formula discussed above, it allows a researcher to calculate the number of studies reporting zero effect that would be required to reject a positive finding. It is beyond the scope of this thesis to evaluate different means of calculating a failsafe n for effect sizes, but it should be noted that alternative methods

exists, for example employing a maximum likelihood model (Iyengar & Greenhouse, 1988).

Each analysis I conducted required distinct approaches to evaluate the potential effect or extent of publication bias. Therefore, I have discussed these in more detail in the corresponding 'meta-analysis procedure' sections.

2.3.3. Biased study inclusion and data analysis

When reporting the results of a systematic review or meta-analysis it is essential to report the risk of bias in both the included studies and resulting data analysis (Borenstein et al., 2009). This enables the reader to critically evaluate whether the authors' conclusions are warranted or if their analysis is flawed. For example, non-disclosure of the funding organisation is a potential risk of bias as the authors may have a conflict of interest, so selectively report their findings.

Despite this, an analysis of 300 systematic reviews and meta-analyses that were published in November 2004 reported widespread omissions of data essential in evaluating risk of bias (Moher et al., 2007). For example, only 23% assessed publication bias and 33% did not provide a report of how risk of bias was assessed.

Because of these shortcomings, the Preferred Reporting Items of Systematic Reviews and Meta-Analyses (PRISMA) guidelines were developed to ensure systematic reviews and meta-analyses are reported correctly (Moher, Liberati, Tetzlaff, & Altman, 2009). These emphasise the need for clear description of the method used when conducting a meta-analysis or systematic review, and includes a review of risk of bias (see PRISMA checklist items 12, 15, 19 and 22). I have followed the PRISMA guidelines throughout this thesis and have aimed to describe my study inclusion criteria, method and analysis in sufficient detail that they may be replicated.

Bias in included studies is important to consider because meta-analysis is a statistical analysis based on previously published research so cannot overcome the input of poor quality data (Eysenck, 1978). Of course the same can be said about any research that has not been conducted in a scientifically rigorous way (Greenland, 1994). This difficulty is generally overcome by having robust inclusion and exclusion criteria and transparent reporting of the data-analytic strategy.

The inclusion of biased or flawed studies is of relevance to meta-analyses examining treatment effects, where the input of effect-sizes from poorly conducted trials (e.g. researchers not blinded to condition, intervention characteristics not adequately described) may adversely affect the analysis. However, it has less relevance for the study in this thesis as I will not be investigating treatment outcome. Again, I argue that even a poorly conducted trial will be unlikely to affect an individual's baseline health anxiety scores. However, there are some scenarios that may potentially affect my study more dramatically than a meta-analysis of treatment effects; for instance, if the sample

description is poorly described, or if study conditions artificially (and temporarily) alter an individual's health anxiety score. I discuss these examples and others in detail along with a rationale for how I assessed study quality in the 'risk of bias' section.

2.3.4. Comparing 'apples and oranges'

A final criticism is that it is inappropriate for meta-analysts to compare different kinds of studies that employ varying measures or include differing participants or 'apples and oranges' (Sharpe, 1997). In contrast, others argue the ability of meta-analysis to aggregate effect sizes from disparate analyses is a strength (Borenstein et al., 2009). Regardless, I argue this criticism has little impact on this analysis as I will only be comparing samples on the same measure. This is consistent with Sharpe's (1997) recommendation that when the scope and research question of a meta-analysis is sufficiently narrow then the 'apples and oranges' criticism is of less relevance. It is possible to determine whether a study is comparing studies that are too dissimilar be examining the homogeneity of effect sizes (Borenstein et al., 2009). The effect of moderator variables (such as quality of study or sampling strategy) on homogeneity can also shed light on the differences between studies that may influence variability in effect sizes.

2.4. Alternative study designs

I considered alternative designs to the study I eventually carried, I now discuss these designs and my reasons for selecting meta-analysis of published data.

2.4.1. Alternative design 1: collect new data

An alternative design would be to collect new raw data from various populations and develop norms based on this information. This approach would have the advantage of conducting a confirmatory factor analysis and the collection of detailed demographic information that (with a sufficient sample size) would allow precise norms based on age, gender and ethnicity to be produced. However, scale validation papers typically have samples in the hundreds (Anthoine, Moret, Regnault, Sébille, & Hardouin, 2014) and 300 participants are recommended as the minimum required for exploratory factor analysis (Tabachnick & Fidell, 2012).

Collecting new data would involve gathering sufficient information to develop norms for several populations (e.g. students, community controls and hypochondriasis) each requiring a corresponding programme of data collection. It would be important to collect data amongst multiple populations rather than focussing on undergraduates. This is because undergraduate students are a limited sample as they are better educated, younger and at a much earlier life stage than those in the wider community. They also differ in terms of personality traits (as assessed by the Big-5) and their values as compared to older peers. The magnitude of these differences also vary according to

nationality (Hanel et al., 2016). Crucially for health anxiety research, young people are also likely to have stronger immune systems (Roberts-Thomson, Young, Whittingham, & Mackay, 1974), be healthier and have had fewer illness than older populations (Beard et al., 2015). This means their experience of illness and ill health is different to the general public as health anxiety is more common after being physically ill and also is more likely to appear when in one's 30s or 40s (Starcevic & Lipsitt 2001). I suggest this is highly problematic as much health anxiety research is based on studies conducted with undergraduates and there is evidence this group significantly differs from the wider population (Henrich, Heine, & Norenzayan, 2010).

Another difficulty with this approach is that it is incompatible with my goal of investigating the construct validity of each scale. In the study design I eventually selected, this was achieved by examining the correlations between the SHAI/HAQ and psychometric measures of other constructs. Completing this with a new sample would require each participant to complete a large number of questionnaires. Although this would be advantageous as a more complete picture of the nomothetic span of the construct validity of each scale could be obtained, a disadvantage is the extra time required to gather information for multiple populations. As such it may not be achievable given the constraints of completing this thesis. In addition, there may be ethical difficulties in asking participants to complete questionnaires when this data is already available in the literature. Overall, I argue collecting new raw data would be unachievable for a project of this nature and would only allow for a limited range of population norms to be developed.

2.4.2. Alternative design 2: contact authors for existing raw data

A second alternative design would be to contact authors of primary studies and requesting access to their data. This would enable me to calculate normative data with greater precision than meta-analysis of mean scores. Another advantage of this approach is that detailed demographic data may be available for analysis, which would allow for further differentiation of the variability of health anxiety scores across groups, for example dividing a sample by age and gender. This approach may have been an equally helpful means of developing normative data and investigating the construct validity of both scales. I contacted authors via email requesting access to their data, but these attempts were unsuccessful. Furthermore, this method is limited as it relies on authors sharing their data, which although a condition of publication in many journals, few authors do so in practice (Savage, Vickers, Kats, & Molenaar, 2009; Wicherts, Borsboom, Kats, & Molenaar, 2006). This means that the amount of data available is likely to be narrower than that in the published literature, so will limit the range of normative data that can be produced. Therefore, I elected to continue this project by meta-analysis of published data only.

2.5. Meta-analysis procedure

I now outline each step taken in conducting this meta-analysis, beginning with my search strategy.

2.5.1. Search strategy

Medline, PsycINFO and Web of Science were used to search for articles citing the original scale validation study of the SHAI (Salkovskis et al., 2002) and the HAQ (Lucock & Morley, 1996). These databases were selected as they are the major subject databases for clinical psychology and medical research. Other studies may exist in the literature, though this is unlikely given the scope of the databases searched. The output of each database search were collated and duplicate articles were removed. Reference lists of included papers were searched for other citations not discovered in initial literature search. The literature search was restricted by date: The start date for the HAQ literature search was 1996, for the SHAI it was 2002 (the dates when both scales were first published) The closing date for published literature to be used in the meta-analysis was 31 April 2016.

2.5.2. Inclusion and exclusion criteria

Studies were included if they were published in a peer reviewed journal and reported on any of the following; sample mean, sample standard deviation, reliability coefficients and correlations with other measures. Table 1 below states reason for selection of each variable.

When the HAQ/SHAI were utilised in a RCT or at multiple time-points, only baseline data was extracted to provide pre-intervention scores and norms for untreated populations. Follow up data was not included in the meta-analysis as this thesis does not aim to evaluate treatment approaches for health anxiety.

Table 1. Variables extracted for meta-analysis and rationale for inclusion

Variable	Reason for inclusion
Mean score	• Generate normative data for different populations. • Predictor variable in reliability generalizability analysis.
Standard deviation	• Generate normative data for different populations. • Predictor variable in reliability generalizability analysis.
Alpha coefficients	• Generate normative reliability data for different populations.

	• Dependent variable in reliability generalizability analysis.
	• Narrative review of internal consistency estimates in different populations.
Test re-test reliability coefficients	• Narrative review of test re-test reliability.
Correlations with other measures	• Review the construct validity of both scales.
Demographics (age, gender, nationality)	• Predictor variables in exploring heterogeneity in mean scores.
	• Predictor variables in reliability generalizability analysis.
Other study level variables (sample composition, measurement context)	• Predictor variables in exploring heterogeneity in mean scores.

Studies were excluded if: (a) Papers cite either measure but do not report any numerical data (e.g. narrative reviews, editorials); (b) Where multiple studies report on the same sample, data was only extracted once, although if additional data were available in a second study this data was combined with data from the first so that each sample was only reported once. For example, Norr and colleagues published three separate papers, each reporting different information relating to the SHAI (used as a secondary measure) but collected all the data from the same student sample (Norr, Oglesby, et al., 2015; Norr, Albanese, Oglesby, Allan, & Schmidt, 2014; Norr, Allan, Boffa, Raines, & Schmidt, 2015). Data from these three studies were combined into one data-point to avoid the same sample being counted more than once; (c) When extracting data it emerged that 4 studies selected participants based on SHAI cut-off scores (Abraham et al., 2013; Brady & Lohr, 2014; Lee, Goetz, Turkel, & Siwiec, 2015; Roberts, Hart, & Eastwood, 2010; Wattar et al., 2005). In these studies participants were selected for being very high, or very low in health anxiety so will not represent the true range of scores found in that population. Therefore, these studies were excluded from the meta-analysis as their inclusion would bias the normative data produced. For comprehensiveness their mean scores, standard deviations and reliability coefficients are reported in appendix B along with data derived from studies that were included in the analysis.

See the PRISMA diagram in figure 2 for a summary of the quantity of included and excluded studies.
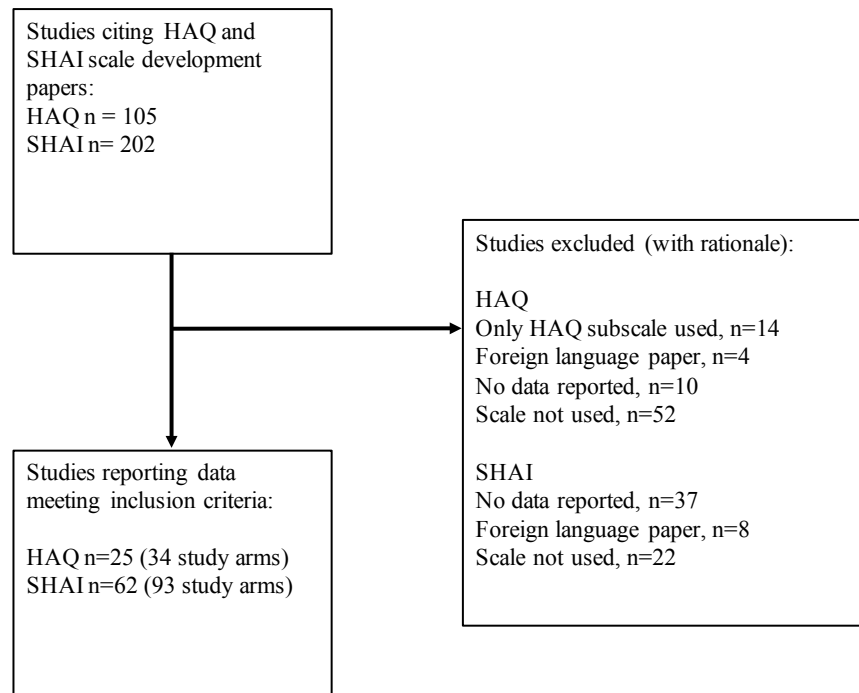
Figure 2. PRISMA diagram of included studies.

2.5.3. Assessing risk of bias (RoB)

As stated previously, it is important to assess RoB as studies with flawed design may produce inaccurate results which in turn will affect the accuracy of the conclusions drawn from a meta-analysis (Gluud, 2006). There has been much consideration of this topic amongst medical researchers using RCTs to investigate the efficacy of treatments (Wood et al., 2008). The effects of poorly conducted trials on a meta-analysis can be striking; for example a review of 250 RCTs reported that including trials with high RoB can increase an odds ratio by 41% (Schulz, Chalmers, Hayes, & Altman, 1995). Government policy such as NICE guidelines are developed based on results from clinical trials, systematic reviews and meta-analysis. Therefore, it is important that these are not based on flawed results, especially as analysis of flawed trials could lead to a treatment being erroneously recommended as effective. The realisation of the need to systematically review RoB in RCTs led to the Cochrane Collaboration development of guidelines in achieving this (Higgins et al., 2011). This guidance focussed on evaluating allocation procedures, blinding of participants and researchers to treatment arm and participation attrition. Of course, not all of these criteria can be achieved in trials of psychotherapy as it impossible for participants and trial therapist to be blind to the intervention being used. The ROBINS-I toolkit is a similar set of guidelines used to evaluate non-randomised intervention studies (Sterne et al., 2016). Clearly, evaluating risk of bias in research is of great importance when carrying out a meta-analysis of treatment studies.

As previously mentioned, I am not researching treatment effects and have only analysed baseline data from RCTs, in fact the vast majority of the studies included in this analysis were from cross-sectional research. Additionally, I did not extract post-treatment data, so all studies in this analysis could be considered cross-sectional. The National Institute of Health (NIH; 2014) have published guidelines for evaluating RoB in for cross-sectional and observational studies. These guidelines assess whether a study was designed and conducted in robust way, has a sufficient sample size with minimal attrition and whether the report has sufficient information to allow replication.

The overarching aim of assessing RoB in a meta-analysis of treatment effects is to uncover the true effect of an intervention rather than any spurious effects that result from a poorly designed study or measurement error. Studies that are deemed to be at greater risk of bias are often excluded from an analysis. Initially I planned to develop a bespoke RoB scale, calculate a score for each study and input these scores as a moderator in a regression analysis to determine if studies with a greater RoB lead to systematic variation in mean scores or reliability coefficients. I generated this scale from the relevant items included in the Cochrane Collaboration guidelines, ROBINS-I toolkit and NIH guidelines. After constructing a bespoke scale, I realised the items included in this scale may be potential sources of heterogeneity in mean scores. Therefore, rather than creating a RoB scale, collating the data and generating an overall rating I have entered various RoB items that were initially planned to comprise the scale as predictors of mean scores. These predictors along with a rationale for inclusion are detailed in the next section along with a description of each step taken in the analysis.

As with the Cochrane Collaboration and ROBINS-I toolkit, the NIH RoB guidelines were designed to evaluate medical studies, so also includes information about exposure to a particular intervention, for example diabetes medication. This means that none of the reviewed guidelines could be used in unadulterated form in this analysis, instead I assessed RoB using my own items. This has the disadvantage of not being standardised and has not been agreed with other researchers, unlike the above toolkits which were developed by several researchers in collaboration. However, I am applying meta-analysis to a research problem that does not concern treatment effects, so required an alternative means of assessing RoB. I am unaware of any other study that attempts to generate population norms using meta-analysis, so there are no established guidelines or precedent to follow when conducting this study.

The rationale for assessing RoB in a meta-analysis is to describe and (via exclusion of problematic studies) remove heterogeneity in calculated effect sizes. During the analysis phase of this project, I realised I could investigate potential sources of heterogeneity in mean scores by utilising this RoB information. Therefore, I will describe this component of my analysis as investigating heterogeneity in mean scores.

As there is no precedent for this approach, I attempted to follow the scientific reasoning fundamental to the Cochrane Collaboration, ROBIN-I toolkit and the NIH RoB guidelines and applied it to this study. I now describe which variables I selected to investigate as potential sources of heterogeneity in means scores.

2.5.4. Investigating heterogeneity in mean scores

In this section I discuss the study level variables that I selected to investigate their impact on mean health anxiety score. These variables fall into three domains; sample composition, recruitment strategy and testing conditions, I now describe the rationale for selecting each variable.

2.5.4.1. Exploring heterogeneity domain 1: Sample composition

For the normative data generated by this analysis to be helpful in research or clinical practice it will need to refer to specific populations, for example students or those with a chronic medical condition. Therefore, the composition of the sample may impact on mean scores. If a study sample is less homogenous, for example including both students and those with clinical levels of health anxiety, then any normative data developed may only be accurately applied to samples of the same population. I aimed to establish risk of sampling bias by inclusion of the following items, all coded yes/no:

1. Homogeneous sample (defined as homogenous if containing only one diagnostic group);
2. Representative of population of interest (rated as 'yes' if the sample demographics and description matches with the study design);
3. Well described sample (met this criteria if demographics and diagnostic criteria were reported);
4. Medically unwell included.

The medically unwell participants item was included as being ill can increase an individual's health anxiety (Katon, Lin, & Kroenke, 2007). Therefore, it may be inappropriate to aggregate studies that exclude the medically unwell with those that include them as health anxiety is likely to be higher in the latter. By collecting study-level data about the inclusion of the physically unwell, I can determine whether this variable does have an impact on mean score. All of these items are designed to assess whether the sample in a study represents the target population of interest.

2.5.4.2. Exploring heterogeneity domain 2: Recruitment strategy

A related domain concerns bias in recruitment, which may then influence the composition of the final sample. An example of a biased recruitment strategy would be use of 'snowball' sampling, where existing participants recommend friends to participate in the study; a less biased strategy would be asking consecutive referrals at

an anxiety disorder clinic. Recruitment strategy was grouped according to the following categories:

1. Random sampling;
2. Opportunity samples;
3. Snowball sampling;
4. Not reported.

2.5.4.3. Exploring heterogeneity domain 3: Testing conditions

The conditions that a person is assessed under may affect their test scores, therefore for accurate comparison between groups it is important that these are standardised (Wechsler, 2008). This was assessed by the following items:

1. Standardised testing conditions?
2. Test administered correctly?

Another consideration is the context within which the assessment of health anxiety occurred. If a health anxious person is assessed in a hospital setting then I argue that this may temporarily increase their anxiety and perhaps elevate their scores. Similarly, if measurement of health anxiety occurred at a time point in treatment or diagnosis likely to temporarily increase health anxiety, such as immediately prior to receiving information about a medical condition then this was recorded. I aimed to capture this information with the following items:

3. Measurement context likely to elevate anxiety?
4. Measurement point likely to elevate anxiety?

2.6. Data analytic strategy

There are four overarching aims, or domains of this thesis regarding the creating normative data and investigating the psychometric properties of the SHAI and the HAQ. Table 2 contains a summary of each aims and associated analysis.

Table 2. Research questions and choice of analysis for each thesis domain.

Research domain	Research question(s)	Analysis employed
Creating normative data	1. What are normative mean scores, SDs and reliability coefficients for various populations?	1. Mean scores, SDs and reliability coefficients entered as effect sizes and meta-analysed.
	2. Can the variability in mean scores be explained by demographic or study variables?	2. Meta regression with mean scores as the dependent variable, demographic and study variables as predictors.
Investigation of reliability data	1. Is there systematic variation in reliability coefficients?	Reliability generalisability meta-analysis.
	2. Can this variation be accounted for by demographic or study variables?	
Scale construct validity	1. Do the SHAI and HAQ have adequate construct validity?	1. Meta-analysis of correlations with other measures.
		2. Qualitative review of correlation coefficients.

In order to answer the research questions in table 2, I conducted three distinct meta-analyses, each with different aims, methodology and analysis. As previously stated, a random effects model (Hunter & Schmidt, 2000) was assumed in all analyses. I now describe each step taken to analyse the data for each study domain. This is reported in line with PRISMA guidelines so this analysis may be replicated and critically evaluated.

2.6.1. Meta-analysis 1: Calculating population norms

Population norms were developed by extracting the mean score, standard deviation and reliability coefficients for each population reported in the included studies. Mean scores and standard deviations were extracted in order to determine average health anxiety scores in different populations. Similarly, where possible reliability coefficients were extracted to provide an average reliability for each subpopulation.

Because this study relied on previously published data, normative data could only be generated from that already available. Consequently, studies reporting data from distinct, but similar populations were grouped together for analysis. The descriptions of the sample types grouped to develop normative data are summarised in table 3 and apply to those derived from both the SHAI and HAQ.

Table 3. Description of samples included in the meta-analysis to generate normative data.

Normative data group	Included sample types	Selection criteria in primary studies
Anxiety disorder	Generalised anxiety disorder, panic, obsessive compulsive disorder, social phobia, specific phobia.	DSM-IV Diagnostic interview.
Cardiac	Cardiac syndrome X, coronary heart disease, angina, myocardial infarction, stress cardiomyopathy.	Clinician diagnosis.
Clinic attenders	Those attending an amniotic fluid test, a breast cancer diagnostic test, general practice or attending a cardiac, gastroscopy or neurology clinic.	Attendance at clinic.
Gastro-enterology patients	Active Crohn's Disease, inactive Crohn's disease, active ulcerative colitis, inactive ulcerative colitis, non-specific gastro symptoms.	Clinician diagnosis.
Health anxiety*	Health anxiety, hypochondriasis.	DSM-IV, Mini International Diagnostic Interview.
Healthy controls	Heterogeneous, includes the physically well and physically unwell.	No consistent definition of 'healthy'.

| Pain | Chronic pain (lasting>6 months), acute pain (from tissue damage). Exclusion criteria; psychiatric disorders, fibromyalgia, arthritis. | Physician diagnosis. |
| Students | Attenders at university | N/A |

*Note Barsky and Ahern (2004b) defined health anxiety using WI and Somatic Symptom Inventory cut-off scores, only half of their sample met DSM-IV diagnostic criteria. Data from this study was not included in the meta-analysis.

If multiple populations were reported in the same study, for example those experiencing hypochondriasis compared to healthy controls, then each sample in the study was reported separately. It is likely that health anxiety scores will vary across different populations (e.g. those attending medical outpatients vs. anxiety disorders). Therefore, this step was necessary so that a comparison of the typical level of health anxiety in each population could be made.

If the SHAI or HAQ were employed in a treatment trial or other longitudinal follow-up study then data was extracted from baseline scores only. This is because this study aims to determine the typical health anxiety in various populations that have not been treated.

In their earlier meta-analysis of the SHAI, Alberts et al., (2013) noted that some authors incorrectly scored the SHAI from 1-4, rather than 0-3 this lead to greatly inflated scores in some studies. Prior to analysis, I corrected these studies by subtracting 14 points from the mean score if using the SHAI-14 and 18 points if using the SHAI-18. These corrected scores were employed in the meta-analysis, both corrected and raw mean scores for each study are presented in appendix c.

In 10 of the studies included in this thesis, participants from the same population were divided into two or more groups for analysis. In these studies, HAQ/SHAI information was reported for each sub-group but study-level data such as demographic and reliability information was reported for the entire sample. This lack of information pertaining to each sub-sample is problematic as it prevents further analysis of heterogeneity in mean score and reliability as the two sub-samples may differ in some way. To overcome this difficulty, I combined the sub-samples to create an overall sample. For example, in an RCT of mindfulness based cognitive therapy (McManus, Surawy, Muse, Vazquez-Montes, & Williams, 2012), a total of 72 participants were randomised to either mindfulness based cognitive therapy (n=36) or unrestricted services (n=36). In this analysis, the health anxiety mean scores and standard deviations were combined prior to meta-analysis to create a single sample of 72 with an overall mean and standard deviation.

I combined the scores in each study by multiplying each variable of interest by the sub-sample size to create a weighted score. The weighted score for each sample was summed and then divided by the overall sample size. This is outlined in the equation below, where x denotes the variable of interest (mean, standard deviation or reliability coefficient) in each study and n the size of the corresponding sub-sample:

$$\bar{x} = \frac{(x_1 + n_1) + (x_2 + n_2) \dots (x_i + n_i)}{n_1 + n_2 \dots n_i}$$

To prepare the data for meta-analysis, I grouped each sub-sample with others derived from the same population. Using the SHAI-14 as an example, there were 16 studies reporting on undergraduate psychology students and 7 studies reporting on populations with hypochondriasis.

Next, I calculated a weight for each variable of interest (mean, standard deviation and reliability coefficient). This step is required so that larger samples have more influence in the analysis and avoids studies reporting on smaller samples from overly influencing the meta-analysis. There are numerous methods for calculating the weight of a variable, the method I used involves weighting by the inverse of the variance of each variable. I selected this method because it has the advantage of accounting for both sample size and variability of scores, with a larger sample size and smaller variability being allocated a larger weight. It is the most frequently employed means of weighting variables in meta-analysis (Rodriguez & Maeda, 2006) and is also recommended for use in reliability generalisability analysis (Botella, Suero, & Gambara, 2010) which I outline in the next section.

Table 4 outlines the equations I used to generate a weight for the mean, standard deviation and reliability coefficients. Note, se = standard error, sd = standard deviation, v_i = variance of the ith group, n_i = sample size of the ith group, j_i = the number of items in the scale under analysis and finally w_i = weight. The equations for weighting the standard deviation and alpha coefficients were taken from Rodriguez and Maeda (2006) and Botella et al. (2010) respectively. Data analysis was conducted in SPSS using the macro designed by Wilson (2010).

Table 4. Calculations of weights for means, standard deviations and alpha coefficients.
(Table adapted from Morley, 2016).

Meta-analysis variable	Standard error/variance calculation	Weight
Mean	$se = \dfrac{sd}{\sqrt{n}}$	$w_i = \dfrac{1}{se^2}$
Standard deviation	$v_i = \dfrac{2 * sd^4}{(n-1)}$	$w_i = \dfrac{1}{v_i}$
Transformed alpha coefficient	$v_i = \dfrac{18 * j_i * (n_i - 1) * (1 - \alpha)^{2/3}}{(j_i - 1) * (9 * n_i - 11)^2}$	$w_i = \dfrac{1}{v_i}$

2.6.2. Meta-analysis 2: Reliability generalisability analysis

Reliability generalisability (RG) analysis is a meta-analytic procedure developed to empirically assess the variability in reliability coefficients of a test across different populations and studies (Vacha-Haase, 1998). It allows the computation of the average reliability of a test across studies, the extent of this variability and an indication of potential sources of this variance. In the first paper utilising RG analysis, Vacha-Haase (1998) meta-analysed reliability coefficients of the Bem Sex Role Inventory (BSRI; Bem, 1974) using the ordinary least squares method and reported the average reliability across included studies with corresponding confidence intervals. The second component of the analysis involved the statistical evaluation of potential sources of variability in reliability coefficients, this was achieved by running a meta-regression with untransformed reliability coefficients as the dependent variable and various study-level variables (e.g. sample size, language of questionnaire) as predictors. These predictors were added to determine which explained a significant proportion of the variance in reliability coefficients.

An example of a reliability generalisability analysis of the State-Trait Anxiety Inventory reported that STAI had poorer reliability when conducted in studies with younger people under the age of 16 (Barnes, Harp, & Woo, 2002). The authors also reported that the average re-test reliability for the state form of the scale was 0.7, this is a problematically low level of reliability. This example demonstrates the information about the reliability of a test that can be inferred from reliability generalisability analysis.

Since Vacha-Haase's initial paper, RG analysis has developed into a sub-field of meta-analysis in its own right with a diversity of analytical procedures being employed (Vacha-Haase & Thompson, 2011). RG studies differ according to the means of weighting reliability coefficients, whether transformed or untransformed coefficients are

analysed and the choice of meta-analytic model, e.g. ordinary least squares, fixed- or random- effects model (Sánchez-Meca, López-López, & López-Pina, 2013).

Botella, Suero, and Gambara (2010) extended the work of Vacha-Haase (1998) and argued that the sources of variation in scale reliability coefficients can be better assessed when considered in relation to the means and standard deviation in the same study. They argue that this allows one to determine whether variability in reliability is due to biases in sampling, measurement error or variation in correlations between scale items. The distribution of alpha coefficients is highly skewed, violating an assumption of meta-regression, and therefore must be transformed prior to analysis. Several methods of adjusting alpha coefficients have been suggested (see Sánchez-Meca et al., 2013), I used the following formula recommended by Botella et al., (2010) where T_i refers to the transformed version of the alpha coefficient, α_i:

$$T_i = (1 - \alpha_i)^{1/3}$$

After meta-analysis, the computed alpha coefficient was back-transformed using this formula:

$$\alpha_i = |1 - T_i^3|$$

When conducting the RG analysis I followed the procedure outlined by Botella et al., (2010). They advocate weighting means, standard SDs and alpha coefficients by the inverse of the variance of each. This was achieved using the same equations as those used in the meta-analysis to develop normative data and are outlined in table 4.

After calculating weights for the mean score, SD and reliability I then computed a meta-regression analysis with alpha coefficients as the dependent variable and study-level variables (e.g. language of scale) entered as predictors. The study variables inputted as predictors are outlines in the results section.

2.6.2.1. Limitations of RG analysis

A difficulty with RG analysis is that authors often do not report reliability information: a review of 12,994 papers reported that 54.6% did not mention reliability and a further 15.7% reported previously published reliability coefficients from other studies or manuals (Vacha-Haase & Thompson, 2011). Both of these situations are problematic for different reasons. I will now outline these reasons and how I attempted to overcome these difficulties. It has been suggested a lack of reporting of reliability coefficients is because authors erroneously believe reliability is a fixed property of tests rather than property of a specific test in a specific population at a specific time (Wilkinson, 1999).

2.6.2.1.1. Limitation 1: Lack of reported reliability coefficients

A lack of reported reliability coefficients is problematic for the same reasons as the file drawer problem discussed in the 'criticisms of meta-analysis' section. If a study

is less likely to be published for having non-significant results, it also follows that studies reporting lower reliability coefficients may also face the same difficulty. It is possible to envision a hypothetical scenario where only studies reporting acceptable reliability are published and that there are many other unpublished studies all reporting coefficients of 0. If this is the case then a measure that is unreliable, or has poor reliability in certain populations may have been incorrectly used in research or clinical practice. Although such an extreme is unlikely, it is plausible that a measure may be less reliable when used outside of the rarefied clinical or student samples often recruited to validate measures.

To overcome this difficulty, Howell & Shields, (2008) offer two calculations: the first allows one to estimate the average reliability in an overall population comprised of studies that do and do not report a reliability coefficient. The second equation is analogous to the failsafe-n calculation used in treatment studies. This calculation generates an estimation of the number of studies that would need to be published with reliability below a certain threshold (selected by the researcher) for the overall average reliability to drop below an unacceptable level.

The first calculation provides an estimate of the average reliability in a population of studies. This population will include papers that have published a reliability coefficient and those that do not. Because the average reliability in the non-reporting sample is not known, Howell and Shields (2008) recommend assuming the average reliability in this sample is 0.8 standard deviations below the average reliability generated from the reporting studies. This figure is suggested as a difference of 0.8 standard deviations corresponds to a large effect size (Cohen, 1988).

This information is inputted into the below equation, where N_{RG} is the number of studies reporting an alpha coefficient and $\alpha_{RG\ sample}$ is the average reliability of this sample. The number of studies not reporting an alpha coefficient is $N_{NR\ sample}$ and the estimated reliability (0.8 SD below the reporting sample) is $\alpha_{NR\ sample.}$

$$\alpha_{LB\ estimate} = \frac{\left(N_{RG} \times \alpha_{RG\ sample}\right) + \left(N_{NR\ sample} \times \alpha_{NR\ sample}\right)}{N_{RG\ sample} + N_{NR\ sample}}$$

The second equation provides the failsafe-N of the number of unpublished studies with poor reliability that would be required for the average reliability of a test to become too low. In this equation, a threshold for the lowest acceptable reliability of a scale is selected by the researcher according to the intended use of a scale. If a greater degree of precision is required or if the consequences of a false-positive (type one error) finding are important then the chosen alpha threshold should be higher. An estimated alpha value for unpublished studies must also be selected, again Howell and Shields, (2008) recommend this is set at 0.8 standard deviations below the value selected for the

alpha threshold. This difference is suggested as it corresponds to a large effect size. In the equation below, the alpha threshold is denoted $\alpha_{Threshold}$, and the file drawer alpha $\alpha_{File\ drawer}$.

$$Failsafe\ N = N_{RG\ sample} \times (\frac{\alpha_{RG\ sample} - \alpha_{Threshold}}{\alpha_{Threshold} - \alpha_{File\ drawer}})$$

I present the results of the alpha lower-bound estimate and the failsafe-n calculation for the SHAI in the results section.

2.6.2.1.2. Limitation 2: Reliability induction

Another difficulty associated with RG analysis occurs if researchers report reliability figures abstracted from test manuals as if they refer to their own sample; this has been termed reliability induction (Vacha-Haase & Thompson, 2011). As stated previously, reliability is not a static property of tests, rather it refers only to a specific application of a test. The induction of published reliability coefficients is problematic as they refer to previous applications so will not refer to the sample in question. This is problematic as it is possible that the test was insufficiently reliable, thus flawing any analysis based on the test.

Reliability induction cannot be assessed numerically as it is an issue with incorrect reporting rather than incorrect analysis. I overcame this potential difficulty by paying close attention to the wording of results sections, for example if an author wrote 'the reliability in the current sample was...' then it was assumed that coefficient alpha had been calculated for the study sample rather than inducted. More ambiguous wording included no mention of where the alpha coefficient was derived from. In such cases I cross-referenced the reported value with other studies published by the same research group to determine if the same value was reported on multiple occasions. There was no evidence of the second issue in the studies included in this analysis.

2.6.3. Meta-analysis 3: Assessing scale construct validity

As outlined in the introduction, when a scale has good construct validity it can measure a latent construct. To assess the construct validity of the HAQ/SHAI I searched published articles for correlation coefficients between both questionnaires and other measures. I then used meta-analysis to combine the correlations derived from measures assessing the same construct to provide a grand weighted correlation coefficient expressing the direction and strength of the relationship between the HAQ/SHAI and the construct of interested. For example, there were 6 correlation coefficients expressing the relationship between the SHAI-14 and three other measure of health anxiety (3 correlations with the WI, two with the IAS and one with the MIHT). These were

combined via meta-analysis to provide an overall correlation between the SHAI and other measures of health anxiety.

If there were more than three correlation coefficients available for a single measure (e.g. between the SHAI and Body Vigilance Scale) then these were combined using meta-analysis to provide an overall coefficient for this measure alone. I took this step to determine if there were differences in the strength of the relationship between the SHAI and other measures that assess the same construct in different ways. For example, the relationship between the SHAI and Beck Anxiety Inventory compared to the relationship between the SHAI and Trait version of the STAI.

I then reviewed the relationship between the HAQ/SHAI and other measures to determine whether these were consistent with my predictions. This provided an indication of the convergent and discriminant validity of the HAQ and SHAI (see results section). I now outline the steps taken to perform the meta-analysis of correlation coefficients.

The steps taken when performing a meta-analysis of correlation coefficients is the same as any other form of meta-analysis: Compose the research question; define inclusion/exclusion criteria; evaluate risk of bias; weight the coefficients; and finally perform the analysis. The first three steps have already been described in the previous section, so I will now discuss my method of weighting the coefficients and then my approach to analysis.

As discussed in the 'why a employ a meta-analysis' section, a random effects model is most appropriate to this analysis as the findings are intended to be generalizable beyond the scope of the included studies. Field (2005) conducted a Monte Carlo simulation of two meta-analytic methods of analysing correlation coefficients; the Hunter and Schmidt (2004) and the Hedges and Vevea (1998) approaches. Both methods assume a random effects model, but differ regarding their use of transformed or untransformed correlation coefficients and method of weighting studies. Whilst a lengthy comparison of the differences between these two approaches is beyond the scope of this thesis, of relevance is Field's (2005) conclusion that the Hunter and Schmidt approach produced less biased correlation estimates than that of Hedges and Vevea. Therefore, the approach advocated by Hunter and Schmidt was employed in this analysis.

It is important to note that Hunter and Schmidt (2004) offer a range of techniques to adjust effect sizes according to various sources of error. To be consistent with Field's (2005) simulation, this analysis employs the fundamental version of Hunter and Schmidt's (2004) analysis.

The first step in the Hunter and Schmidt (2004) random effects model meta-analysis is to weight the raw correlation coefficients by multiplying each coefficient r_i by the

size of the sample n_i it is derived from. The sum of these coefficients is divided by the sum of the sample, yielding an average coefficient $\bar{r}$ that has been weighted by sample size. This is outlined in the following equation:

$$\bar{r} = \frac{(n_1 * r_1) + (n_2 * r_2) + \dots (n_i * r_i)}{(n_1 + n_2 + \dots n_i)}$$

Information extracted from the sample size and the variance of correlation coefficients may then be employed to calculate credibility intervals (analogous to confidence intervals) around the grand correlation coefficient. Description and evaluation of the steps taken to calculate these credibility intervals is beyond the scope of this thesis, but can be found in Borenstein et al. (2009).

As reviewed in the 'Criticisms of meta-analysis' section, publication bias can lead to an inflated estimation of effect size (Rosenthal, 1979). The presence of publication bias in correlational studies has been investigated empirically: the authors concluded there is little evidence for this in the literature (Dalton, Aguinis, Dalton, Bosco, & Pierce, 2012). Therefore, I have not assessed for publication bias in this section of my analysis.

This meta-analysis of correlation coefficients was calculated in Microsoft Excel using the Hunter and Schmidt (2004) method described in Borenstein et al., (2009).

2.7 Method summary

In summary, three domains of analysis were completed, these are as follows:

1. Develop normative data for the HAQ/SHAI by meta-analysis of mean scores, standard deviations and reliability coefficients. Investigate heterogeneity in mean scores by use of meta-regression with study-level variables inputted as predictor variables;
2. Conduct a reliability generalisability analysis to report on average reliability for the HAQ and SHAI and an investigation of study characteristics that moderate reliability;
3. Investigate the construct validity of the HAQ and SHAI by qualitative review and meta-analysis of correlation coefficients expressing the relationship between each measure and other standardised measures.

3. Analysis

3.1 Structure of analysis

This analysis is divided into four sections, in the first I describe the results of the meta-analysis of mean scores and standard deviations (SDs) for each scale according to sample population (e.g. students, health anxiety). There was significant heterogeneity in mean score for each population, I then report on my attempts to assess the sources of this variation using meta-regression.

The second section comprises the results of the reliability generalizability (RG) analysis, including an overview of the range and average reliability across various populations for each scale. There was significant heterogeneity of alpha coefficients, so this variation was investigated via use of meta-regression.

Part three contains a review and meta-analysis of the size and direction of correlations between the health anxiety measures and other psychometrics and includes a commentary of the construct validity of both scales. Finally, I conclude with a summary section and report on the normative data for various population as derived from the previously described analyses.

As detailed in the method section, a random effects model was selected so that my results can be generalised beyond the current studies. This decision was also appropriate from a statistical perspective as the majority of computed effect sizes were heterogeneous (Hedges & Vevea, 1998).

3.2 Population means and normative data

In total 137 studies (or study arms reporting on two or more populations) reporting on the HAQ or a version of the SHAI were included in analysis. This yields an overall sample of 31,347 participants.

The following sections are divided according to scale type (HAQ, SHAI-14, SHAI-18). Within each subsection, results are reported by sample type. Where there are three or more studies reporting on the same sample, I computed a meta-analysis to generate normative data. As the focus of this thesis is to generate mean score forest plots are provided for all the mean scores included in the analysis. I begin with the SHAI as there was more data available than the HAQ.

3.2.1. SHAI-14

A total of 37 studies provided a mean health anxiety score or standard deviation for the SHAI-14. These have been grouped into the following categories: healthy

community controls, undergraduate students, health anxiety, pain, and vomiting phobia. Finally, a mixed group contains all the studies that could not be readily combined with others as they report on distinct populations, this data is included in the forests plot in figure 4 and raw data for all studies is listed in appendix c. I performed a meta-analysis on all groups except the mixed sample, the results of which are displayed in table 5.

The magnitude of the average mean SHAI-14 score for each group included in the meta-analysis is presented in figure 3. The healthy control grand mean was lowest at 12.56 and the highest mean score was in health anxious populations. Of note is the mean score of 17.16 calculated in student populations. This is higher than the mean score generated in both healthy controls and populations experiencing pain (14.17). This result is unexpected as students are generally in better health and have experienced less illness, so it follows they should have less reason to be anxious about their health. This mean score is higher than the mean score in both control samples (9.4) and anxiety disorder (14.9) populations reported in the SHAI validation paper (Salkovskis et al., 2002).

Sample	Mean SHAI-14 score	k	N
Healthy controls	12.56	16	3,045
Students	17.16	6	1,307
Hypochondriasis	32.13	7*	663
Pain	14.17	3	259

Note: k=number studies, N=number of participants *Three studies excluded from meta-analysis as SD not reported
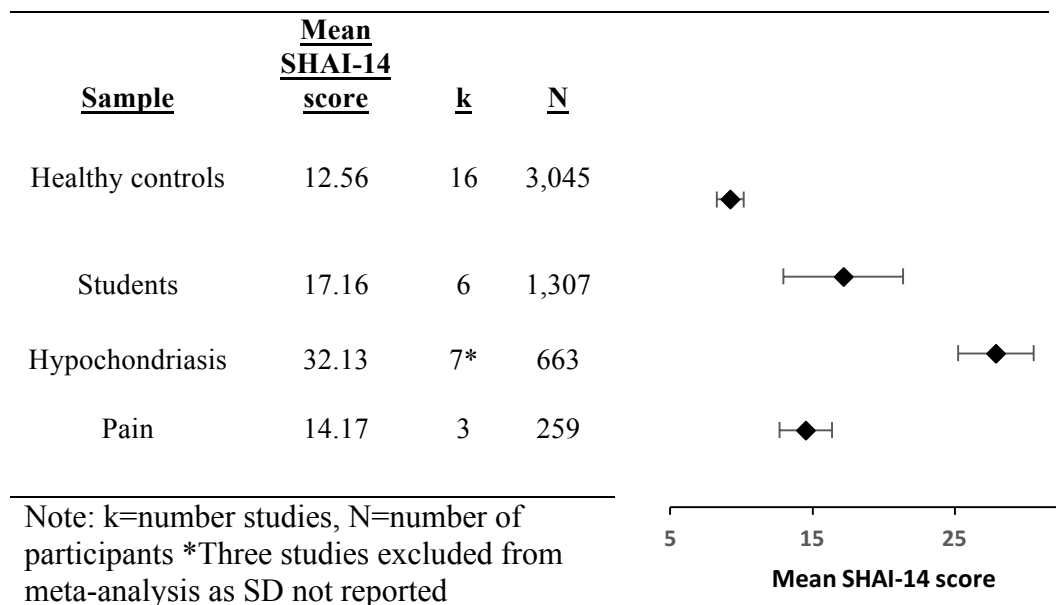
Figure 3 Grand mean SHAI-14 scores and sample size for healthy control, student, hypochondriasis and pain populations. Error bars represent 95% confidence intervals.

It is notable that, as assessed by the Q statistic (see table 5) there was significant heterogeneity in mean scores in all populations except the pain group. This means that a significant proportion of the heterogeneity in means score is explained by non-random factors. The significance of the Q statistic also affirms that the application of a random effects model was appropriate in this analysis. Examination of the I^2 values reveals that in all samples except the pain group, over 95% of the variance in mean scores is explained by non-random factors.

The meta-analysed SDs for all population were similar with range 4.88-7.07 with the lowest dispersion in health anxiety populations and the largest distribution in pain populations. The I^2 values reveal over 94% of the heterogeneity in SDs was explained by non-random factors in healthy, student and pain populations. Only 18% of the variance in health anxiety SDs was explained by non-random factors, although this may be due to the smaller range of studies (k=3) included in the analysis.

Table 5. SHAI-14 meta-analytic results of mean score, standard deviation, Q statistic, degrees of freedom (df) and I^2 value for healthy, student, health anxiety and pain populations.

Population	Variable	Value [95%CI]	Q	df	P	I^2
Healthy	Mean	9.21 [8.26, 10.16]	285.73	13	<0.0001	95.45
	SD	5.36 [4.71, 6.01]	321.76	13	<0.0001	95.96
Student	Mean	17.16 [12.94, 21.37]	662.44	5	p<0.0001	99.25
	SD	5.78 [5.03, 6.53]	86.45	5	p<0.0001	94.22
Health anxiety	Mean	27.90 [25.23-30.56]	108.96	3	p<0.0001	97.25
	SD	4.88 [4.56, 5.20]	3.69	3	0.297	18.70
Pain	Mean	14.52 [12.66-16.38]	5.36	2	0.069	62.69
	SD	7.07 [4.43, 9.71]	1708.19	4	p<0.0001	99.77

The heterogeneity detected in mean scores detected by the Q statistic is apparent when viewed in the forest plot in figure 4. For ease of visual analysis, these have been divided into different populations to reflect those reported in the literature. The 'mixed medically unwell' section contains a diverse range of samples that could not be readily grouped to provide mean scores as they were too heterogeneous in nature. The variation in mean scores in student populations is striking, with two studies both reporting mean scores within the range usually consistent with hypochondriasis (study 1 and 2; Karademas, Christopoulou, Dimostheni, & Pavlu, 2008). I explore this heterogeneity via use of meta-regression in the supplementary analysis section as it was not a part of my original aims.

Please note that three studies by Barsky and colleagues reported a mean item score for the SHAI (Barsky & Ahern, 2004; Barsky, Ahern, Bauer, Nolido, & Orav, 2013; Lovas & Barsky, 2010), yielding a possible score range of 0-3 or 1-4 according to scoring method. The first study (Barsky & Ahern, 2004a) reported a mean item-score of

2.69 and employed the SHAI-14, scoring responses from 1-4. To include this data in the meta-analysis, I multiplied the reported score by 14 to give a mean SHAI-14 score of 37.71. The second study (Lovas & Barsky, 2010) reported a mean item score of 2.07, employed the SHAI-14, scoring from 0-3. Again, I multiplied this score by 14 to give a mean SHAI-14 score of 28.98. There was insufficient data available to calculate the standard deviation so this data was not included in the meta-analysis. The third study (Barsky et al., 2013) did not report the SHAI version or scoring used so it was not possible to calculate an overall mean score. I contacted the corresponding author on the paper to request this data but did not receive a response. Because of the lack of this data, this study was eliminated from the analysis. The mean scores from these papers are displayed along with other SHAI-14 mean scores in the forest plot in figure 4.
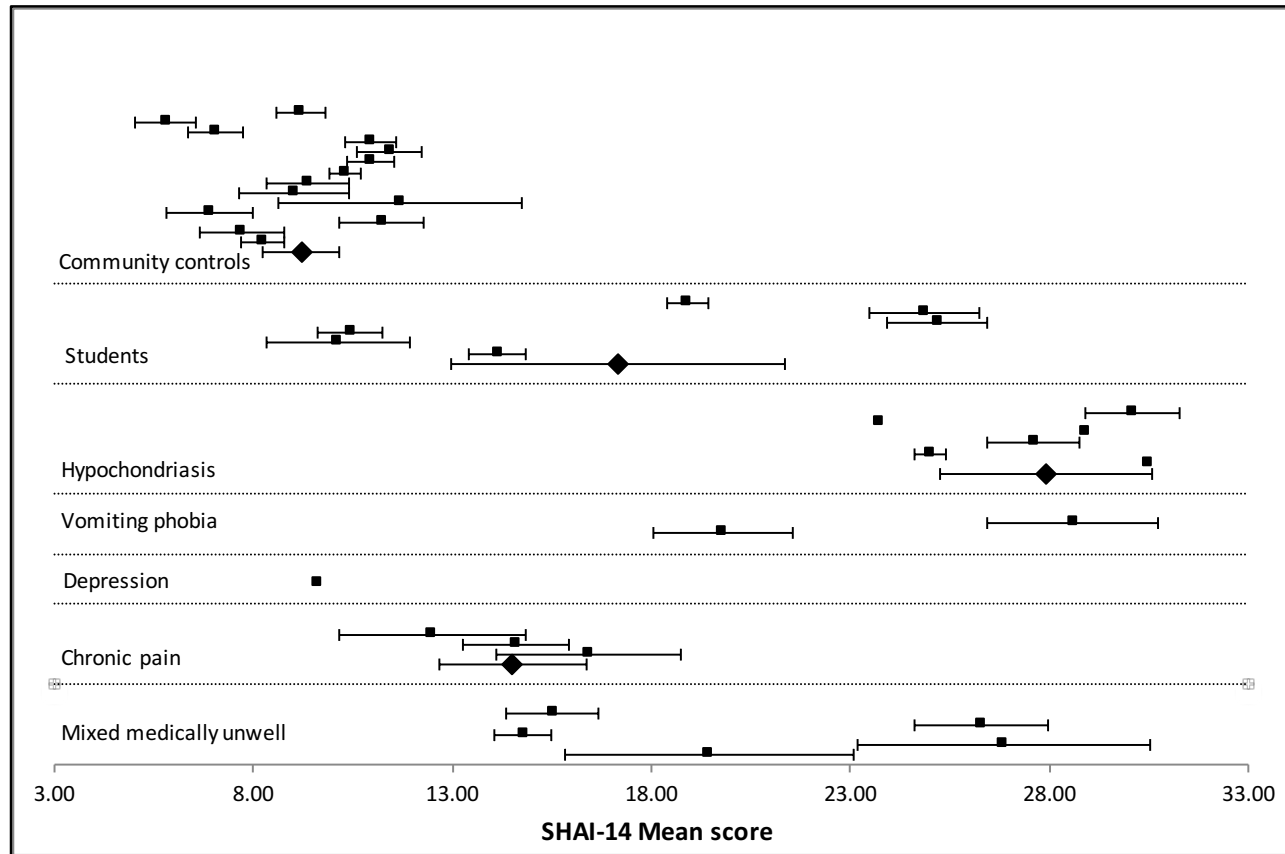
Figure 4. Forest plot of mean SHAI-14 scores by sub-population. Square markers indicate individual studies, diamond markers the meta-analysed grand means score, error bars indicate 95% confidence intervals. Note, confidence intervals could not be calculated in 4 studies as SD unreported.

3.2.2. SHAI-18

A total of 56 studies provided a mean health anxiety score or standard deviation for the SHAI-18. Similarly to the previous section, these have been grouped into sub-populations, with the following categories available for analysis; healthy community controls, undergraduate students, health anxiety, pain, and anxiety disorders. A meta-analysis was performed on all these groups, the results of the meta-analysis of mean scores are displayed in the forest plot in figure 5. All SHAI-18 mean scores, SDs and reliability coefficients are reported in appendix c.

Sample	Mean SHAI score	k	n
Healthy controls	12.56	7	768
Students	12.35	17	7985
Hypochondriasis	32.13	5	335
Anxiety disorders	16.28	3	522
Pain	14.17	5	554

Note: k=number of studies, N=total number of participants.
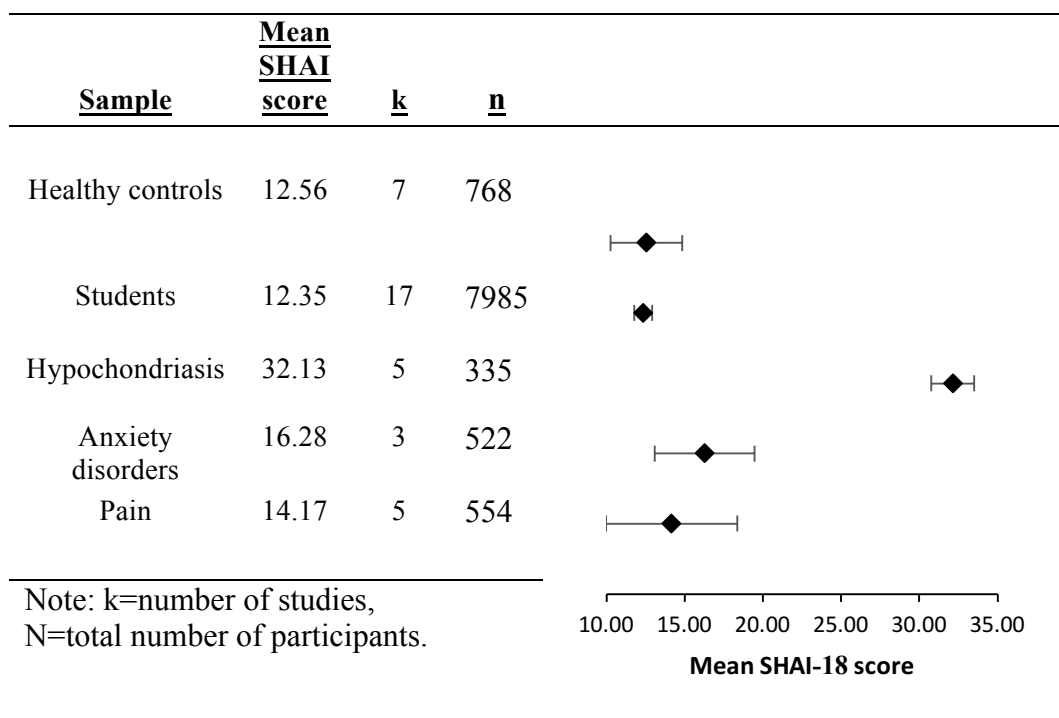
Figure 5. Grand mean SHAI-18 scores, number of studies and sample size for healthy control, student, hypochondriasis, anxiety disorder and pain populations. Error bars represent 95% confidence intervals

Data from one study (Kate Muse, McManus, Leung, Meghreblian, & Williams, 2012) reports an overall mean score and SD from a sample consisting of both students and a group of treatment-seeking individuals with a diagnosis of hypochondriasis. Raw data from this study is reported in appendix c, but was not included in the meta-analysis that generated population norms because the overall sample is likely to be highly heterogeneous. The hypochondriasis sub-sample was reported separately in another study (McManus et al., 2012) and does features in this meta-analysis.

There were 2 studies investigating emetophobia (fear of vomiting), but the total number were insufficient to perform a meta-analysis. A further 8 studies

reporting mean scores for various samples that could not be combined into a coherent population group, so again were not subjected to meta-analysis. All mean scores, along with meta-analysed grand mean scores for each population group are presented in the forest plot in figure 6.

The lowest SHAI-18 scores were in the student and healthy control populations, which is as one would expect, those with an anxiety disorder scored more highly and the highest scores were in populations with hypochondriasis. All meta-analysed scores were similar to that reported in the SHAI validation paper (Salkovskis et al., 2002), I compare my findings with those from the validation paper in the results summary section.

Table 6. SHAI-18 meta-analytic results of mean score, SD, Q statistic, degrees of freedom (df) and I^2 value for healthy, student, health anxiety, anxiety disorder and pain populations.

Population	Variable	Value [95%CI]	Q	df	P	I^2
Healthy	Mean	12.56 [10.26,14.86]	189.37	6	<0.0001	96.83
	SD	6.03 [5.37, 6.68]	125.57	6	<0.0001	95.22
Student	Mean	12.35 [11.78, 12.93]	151.23	14	<0.0001	90.74
	SD	6.51 [6.11, 6.91]	119.78	14	<0.0001	88.31
Health anxiety	Mean	32.13 [30.77, 33.49]	2.918	2	0.2324	31.46
	SD	8.56 [7.47, 9.66]	267.70	2	<0.0001	99.25
Anxiety disorders	Mean	16.28 [13.07, 19.48]	23.13	6	0.0008	74.06
	SD	12.04 [10.16, 13.92]	78649.62	6	<0.0001	99.99
Pain	Mean	14.17 [9.97, 18.36]	145.86	4	<0.0001	97.25
	SD	7.07 [4.43, 9.71]	1708.19	4	<0.0001	

As assessed by the Q statistic (see table 5), there was significant heterogeneity in mean scores in all populations except the health anxiety group. Therefore, a significant proportion of heterogeneity in means score is explained by non-random factors. In addition, the significance of the Q statistic also affirms that the application of a random effects model was appropriate in this analysis. Examining the I^2 values reveals that, excepting the health anxiety group, over 90% of the variance in mean scores is explained by non-random factors. The meta-analysis of studies reporting on health anxiety populations only included three studies so any heterogeneity in mean scores may not be detected, particularly as the Q statistic is not sensitive to heterogeneity in smaller

samples (Bowden, Tierney, Copas, & Burdett, 2011). This heterogeneity is apparent when viewing the forest plot in figure 6, which presents all extracted mean scores by population. I explore sources of this heterogeneity in the supplementary analysis section.

Figure 6. Forest plot of mean SHAI-18 scores by sub-population. Square markers indicate individual studies, diamond markers the meta-analysed grand means score, error bars indicate 95% confidence intervals. Note, confidence intervals could not be calculated in 5 studies as SD unreported.

The meta-analysed SDs were all broader than those calculated from the SHAI-18, this may be due to the addition of 4-item 'negative consequences of illness' sub-scale which does not measure health anxiety. As the two parts of the scale measure different constructs, this may explain the greater dispersion in scores. The smallest SDs were derived from healthy, student and pain populations, indicating that for these groups the dispersion of scores in published studies was smaller. The largest SD was found in anxiety disorder populations.

Examination of the I^2 value reveals that for all samples over 88% of the variance in SDs is due to non-random factors.

3.2.3. Other SHAI versions

As outlined in the introduction, researchers have advocated the deletion of item 13 because it did not load onto either of two factors uncovered in a factor analysis (Wheaton, Berman, Franklin, & Abramowitz, 2010). Because of this, two studies (Fergus & Russell, 2016; Norr, Albanese, Oglesby, Allan, & Schmidt, 2015) reported scores for a 13-item and a 17-item version. Neither study recorded how the SHAI was scored (either 0-3 or 1-4), so could not be included in the meta-analysis. I attempted to contact the authors of both these studies and did not receive any response. The mean scores, SD and reliability coefficients reported for these studies are reported in appendix c.

3.2.4. HAQ

A total of 34 studies reported an HAQ mean score or standard deviation. The lowest grand mean score was reported in healthy controls (12.29) and the highest was in gastroenterology patients (15.59). There was less dispersion in grand mean scores compared to the SHAI-14 and SHAI-18. This may be because the SHAI has been more frequently employed in mental health samples that include individuals with anxiety disorders or hypochondriasis (which are more likely to score more highly on health anxiety measures). The 95% confidence intervals around each meta-analysed mean are also all approximately broader by two points for the HAQ compared to the SHAI, meaning the dispersion of mean scores within each group is greater. This is likely to be because the studies I have combined in the HAQ analysis contain more heterogeneous populations than in the SHAI analysis. For example the cardiac group in the HAQ analysis includes populations with angina, cardiac syndrome x, coronary heart disease, myocardial infarction and stress cardiomyopathy. See table x for the results of the meta-analysis of HAQ mean scores. The SDs were similar in magnitude across all populations, ranging from 7.78-9.80, this indicates the dispersion in scores in each population is similar in size.

Table 7. Results of Meta-analysis of HAQ mean score, SD, Q statistic, degrees of freedom (df) and I^2 value for healthy, cardiac patients, gastro patients and medical clinic attenders.

Population	Variable	Value [95%CI]	Q	df	P	I^2
Healthy	Mean	12.29 [8.78, 15.80]	113.36	3	<0.0001	97.35
	SD	8.39 [7.04, 9.74]	285.13	3	<0.0001	98.95
Cardiac	Mean	13.51 [9.57, 17.47]	14.01	2	0.0009	78.59
	SD	8.06 [2.78, 13.35]	6693.53	2	<0.0001	99.97
Clinic attenders	Mean	15.36 [13.86, 16.86]	19.36	4	0.0007	79.34
	SD	9.80 [8.36, 11.24]	4341.80	6	<0.0001	99.86
Gastro patients	Mean	15.59 [10.79, 20.38]	275.32	4	<0.0001	98.18
	SD	7.78 [6.07, 9.50]	1158.98	4	<0.0001	99.65

The grand mean scores for each population are presented in the forest plot in figure 7 and demonstrates that the grand mean health anxiety score in each population was remarkably similar, with only a 5 point difference between the lowest (healthy controls) and highest (cardiac patients). According to classical test theory (Lord & Novick, 1968) this lack of variation may originate in three sources either alone or in combination: 1) the true scores are similar in each population; 2) there are systematic non-random variables causing scores to appear similar; 3) the similarity is due to measurement error. I will attempt to investigate potential sources of variation in the next section.

Sample	Mean HAQ score	k	N					
Healthy controls	12.29	6	1,169					
Cardiac patients	17.16	7	353					
Clinic attenders	16.36	9	1,057					
Gastro patients	15.59	5	407					

Note: k=number of studies, N=total number of participants.
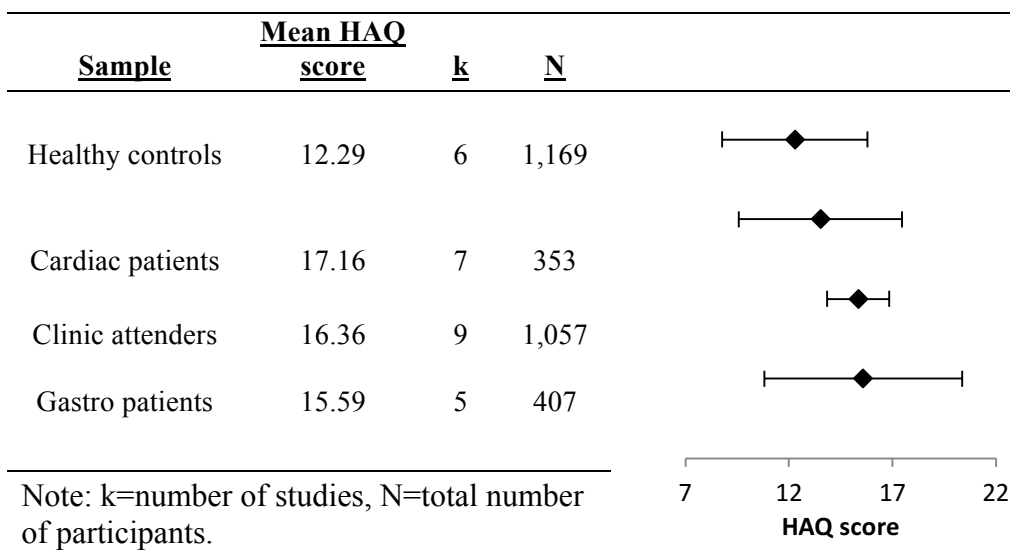
Figure 7. Grand mean HAQ scores, number of studies and sample size for healthy control, cardiac patients, medical clinic attenders and gastro patients. Error bars represent 95% confidence intervals.

Examination of the 95% confidence intervals in figure 7 indicates the variation in mean scores were narrowest for the clinic attenders group, this may be due to the larger number of studies (9) included in this analysis compared to the other groups or due to a narrow range of mean scores generated from studies in this population. The broadest 95% confidence intervals were in the gastro patients group (also see table 6).

To better view the variation in mean scores within each population, I have plotted the mean score (and 95% confidence intervals) derived from every included study in the forest plot in figure 8. This reveals a similar range of scores across each population, for example the lowest reported score for healthy control participants was 8 in a British female-only sample (Asbury, Creed, & Collins, 2004). The highest was over double at 16.83 (Meana & Lykins, 2009) in a USA female only sample. It is remarkable that two very similar populations, albeit of different nationalities, using the same measure can have such a difference in mean score, that as assessed by the Q statistic, is statistically significant.
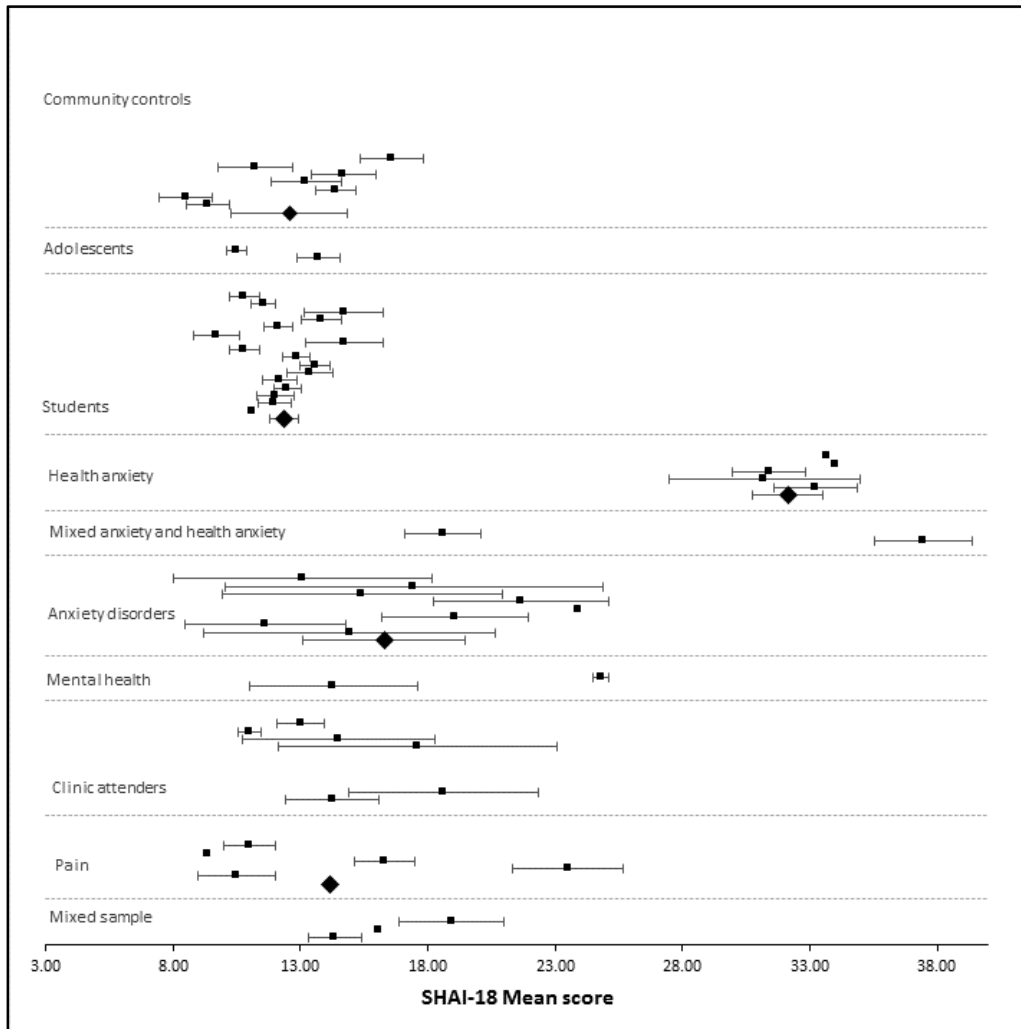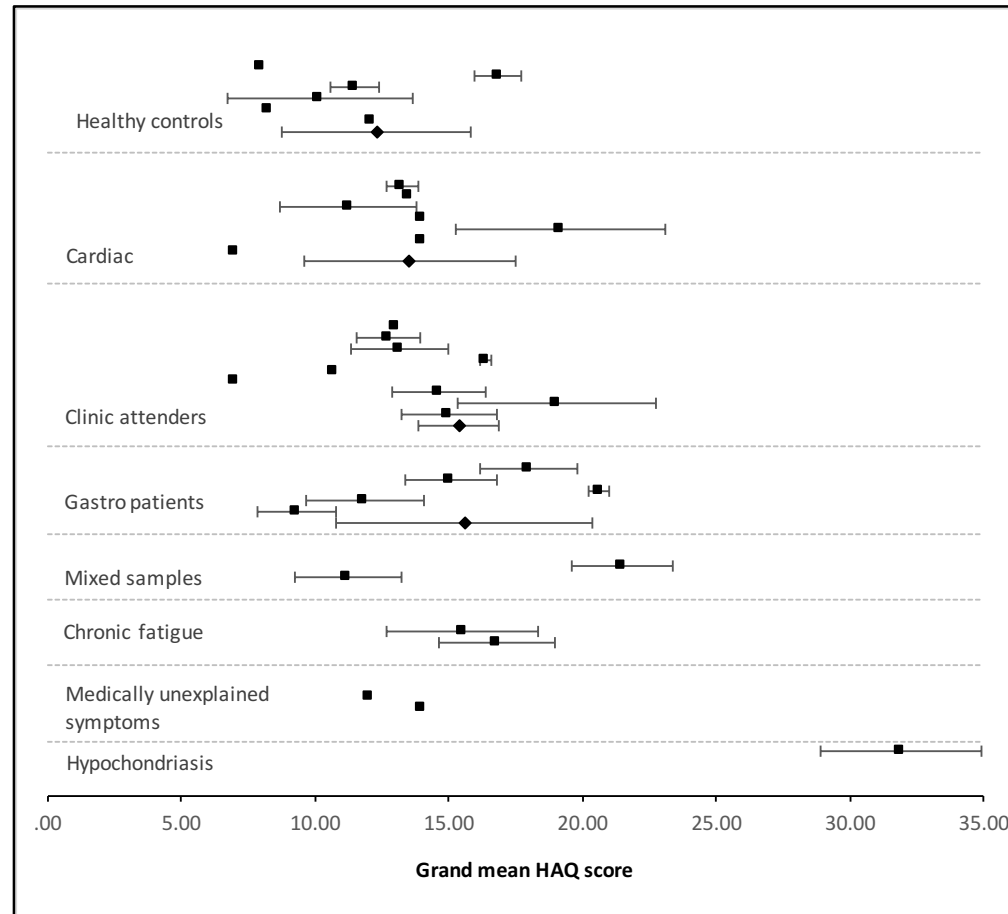
Figure 8. Forest plot of mean HAQ scores by sub-population. Square markers indicate individual studies, diamond markers the meta-analysed grand mean score. Error bars indicate 95% confidence intervals. Note, confidence intervals could not be calculated in 11 studies as SD unreported.

Examination of scores in the other sub-populations revealed a similar range of mean scores. Studies reporting on cardiac patients had a lowest mean score of 7 in individuals seeking a cardiac screen in New Zealand (Zarifeh, Mulder, Kerr, Chan, & Bridgman, 2012) and a highest score of 19.15 in a UK study of cardiac rehabilitation (Asbury et al., 2012). The lowest score in clinic attenders was 7 (Aggarwal et al., 2014), the highest was 19 (Jackson, Kincey, Fiddler, Creed, & Tomenson, 2004), the variation in this population was more expected as it contained studies reporting on more dissimilar populations (e.g. attenders at neurology clinics and attenders in a general medicine clinic). Finally, in the gastro patients the lowest score was 9.29 (Graff et al., 2006) and the highest 20.60 (Esfandyari & Harewood, 2007). A single study reported on hypochondriacal populations reported a mean score of 31.90 (Noyes et al., 2002). This score is consistent with a scale designed to measure health anxiety. It is the variation in mean scores described in the preceding paragraphs that I aim to investigate in the following section using meta-regression.

3.2.5. Supplementary analysis: Exploring heterogeneity in mean scores

As stated in the method section, I initially intended to assess risk of bias in this study via use of a bespoke scale and determine if the overall score on this scale relates to mean scores. This scale included items that assessed study-level variables that may be responsible for heterogeneity (or bias) in mean scores. After collecting this data, I realised that similar to a reliability generalisability analysis, I could input these study-level variables into a regression equation as predictors to investigate their potential impact on heterogeneity in mean scores. In this section I describe the results of this exploratory analysis.

Prior to analysis, nine study arms were removed this was because the authors did not report which version of the SHAI they employed, so could not be entered as predictor variables. These were derived from nine papers (Farmer, Doll, Levy, & Salkovskis, 2003; Fergus & Russell, 2016; Hollo, Kothy, Garas, Geczy, & Vargha, 2012; Hollo, Kothy, Geczy, & Vargha, 2010; Inamura et al., 2015; Kirby & Yardley, 2009; Norr, Albanese, Oglesby, Allan, & Schmidt, 2015; Aaron M. Norr, Oglesby, et al., 2015; Norr, Allan, et al., 2015) although the three studies by Norr and colleagues reported on the same sample.

3.2.5.1. Heterogeneity domain 1: Sample composition

As detailed in the method section, the composition of a sample included in a study may be a potential source of variation in SHAI mean scores. I investigated this by generating a dichotomous variable (0=no, 1=yes) for each of the following variables: homogenous sample; sample representing the population of interest; whether the sample was well described; whether the physically unwell were included. These variable were inputted as predictors in a meta-regression with mean SHAI score as the dependent variable.

The results of this regression are displayed in table 8. None of the independent variables significantly predicted SHAI mean score, the overall model explained a non-significant ($F(4,96)=1.03$, $p=0.532$) 0.05% of the between studies variance. Inspection of the I^2 value reveals 99.49% of the between studies variance is due to non-random sources of heterogeneity.

Table 8. Results of meta regression of sample composition on SHAI mean score.

Predictor	Coefficient	SE	t	p
Homogenous	0.278	1.08	0.26	0.797
Population of interest	6.45	4.74	1.36	0.177
Well described	-5.97	3.91	-1.52	0.131
Unwell included	1.32	1.60	0.83	0.41q
Constant	13.27	4.90	2.71	0.008

3.2.5.2. Heterogeneity domain 2: Sampling strategy

A second domain of potential sources of heterogeneity is sampling strategy, studies were rated according to which sampling strategy was used to recruit participants; random, part-random, or snowball sampling. If sampling strategy was not reported then this was also noted. Sampling strategy was inputted as a categorical predictor variable with four levels into a meta-regression with mean SHAI score as the dependent variable.

The results presented in table 9 reveal that sampling strategy did not predict mean score. Examining the adjusted R^2 value reveals this model explained a non-significant ($F(4,95)=1.93$, $p=0.111$), 4.16% of the between studies variance. The I^2 value indicated that 99.48% of between studies variance is due to non-random heterogeneity.

Table 9. Results of meta regression of sample composition on SHAI mean score.

Predictor	Coefficient	SE	t	p
Random sampling	-0.203	4.90	-0.04	0.867
Part non-random	-2.34	4.93	-0.47	0.636
Snowball sampling	-4.41	6.12	-0.72	0.473
Not reported	2.96	5.07	0.58	0.561
Constant	16.36	4.78	3.42	0.001

3.2.5.3. Heterogeneity domain 3: Administration conditions

The final domain of potential sources of variation I investigated was whether the context that the SHAI was administered impacted on mean score. Again, I created a dummy variable (0=no, 1=yes) and rated studies according to whether the administration conditions of a test (e.g. in a hospital vs classroom) would be likely to temporarily increase health anxiety. I also rated studies as to whether the time point that the SHAI was used may be a time of increased health anxiety (e.g. immediately prior to receiving a medical test result).

The results are presented in table 10, neither variable predicted variation in SHAI mean score. The regression model explained a non-significant ($F_{(2,94)}=1.41$, $p=0.249$), 1.01% of between studies variance was explained by this model. A remaining 99.45% of the between studies variance is due to unexplained non-random factors.

Table 10. Results of meta regression of sample composition on SHAI mean score.

Predictor	Coefficient	SE	t	p
Measurement context	1.90	5.30	0.35	0.720
Measurement point	-0.361	5.17	-0.07	0.944
Constant	14.81	0.948	15.62	<0.001

Finally, again using a dichotomous dummy variable with 'no' coded as 0 and 'yes' as 1, I entered whether the SHAI was administered correctly and in a standardised manner. The Adjusted R^2 value demonstrated that 1.50% of the variance in SHAI mean score was explained by this regression model and was not statistically significant ($F_{(2,97)}=0.34$, $p=0.706$. The results of this meta-regression are presented in table 11.

Table 11. Results of meta regression of sample composition on SHAI mean score.

Predictor	Coefficient	SE	t	p
Standardised administration	-2.18	2.85	-0.77	0.446
Correct administration	1.80	2.85	0.63	0.529
Constant	16.21	0.975	16.63	<0.001

4. Construct validity of the HAQ and SHAI

In this section I review the range and strength of correlation coefficients between the SHAI/HAQ and other psychological measures. As detailed in the method section, I have focussed on other measures of health anxiety, anxiety, depression, OCD and scales assessing precursors to anxiety. There were 15 correlation coefficients available reporting a relationship between the HAQ and other psychometric measures. These correlations were derived from four studies and represent a relationship with 15 other scales. This means there were insufficient correlation coefficients to run a meta-analysis. Because of a lack of data this section only comprises my findings relating to variables correlated with the SHAI.

4.1 SHAI – construct validation

There were 98 studies reporting at least one correlation coefficient between the SHAI and another psychometric measure. As detailed in the method, I extracted correlation coefficients expressing the relationship between the SHAI and measures of health anxiety, anxiety, depression and OCD.

The results of the meta-analysis of correlation coefficients is depicted in the forest plot in figure 9. It was not possible to perform a meta-analysis on individual measures of health anxiety as each was reported in three or less studies, therefore I combined all the available coefficients to provide a single grand correlation coefficient.
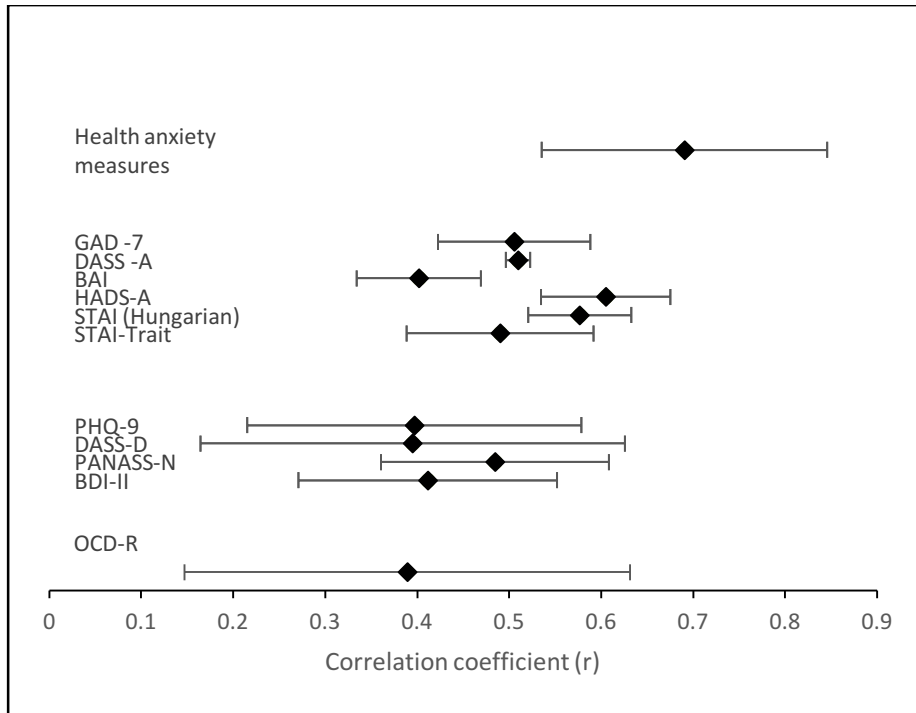
Figure 9. Forest plot of meta-analysis of correlation coefficients between the SHAI
    and measures of health anxiety, anxiety, depression and OCD.

The pattern of correlations was broadly as expected, with the SHAI correlating most strongly with other measures of health anxiety, followed by the anxiety measures. It is notable that the grand correlation between the OCD-R and SHAI was lower than the depression measures, which was opposite to the predicted pattern. This data is presented in table 12 along with the number of studies and total sample size.

Table 12. Results of meta-analysis of correlations between the SHAI and measures of
    anxiety, depression and OCD.

Scale	Total n	No. studies	Meta-analyzed r [95%CI]
Health anxiety (multiple)	1481	6	0.690 [0.535, 0.846]
GAD-7	440	3	0.505 [0.422, 0.588]
DASS-Anxiety	633	4	0.510 [0.496, 0.523]
BAI	1732	6	0.402 [0.334, 0.469]
HADS-Anxiety	554	2	0.605 [0.535, 0.675]
STAI-Hungarian	234	2	0.577 [0.521, 0.633]
STAI - Trait	1329	3	0.490 [0.388, 0.591]
PHQ-9	440	3	0.397 [0.216, 0.579]
DASS-Depression	633	4	0.395 [0.164, 0.626]
PANASS-Negative	2517	6	0.485 [0.361, 0.609]
BDI-II	593	2	0.412 [0.271, 0.552]
OCD-R	702	4	0.389 [0.147, 0.631]

The 95% credibility intervals of the health anxiety measures, indicates with 95% probability that the true correlation falls between 0.535 and 0.846. This is due to the variability in the raw correlation coefficients, as reported in table 13. The highest correlations were with the WI (r=0.71-0.84), followed by the IAS (r=0.63-0.73) and the lowest correlations was with the MIHT (r=0.61). These differences may be due to the differences in conceptualisation of health anxiety inherent in each scale, for example the MIHT considers health anxiety to have an interpersonal element, whereas the SHAI focusses on a cognitive-behavioural model. Alternatively, it may be that either the SHAI or another scale (or both) does not adequately assess health anxiety. As there were only six correlation coefficients available for analysis it is impossible to determine whether these relationships are consistent. Consequently, further analysis with more studies is warranted. If a consistent pattern emerges, for example correlations with the MIHT being consistently lower, this indicates the two scales may be measuring different constructs.

Table 13. Correlations between the SHAI, Whitely Index (WI), Multidimensional Inventory of Hypochondriacal Traits (MIHT) and Illness Attitude Scales (IAS).

Population	Author (date)	n	WI	MIHT	IAS
Community	Fergus (2013b), (Fergus, 2014a)	410	.80	.61	-
Pregnant women	Kowalyk, Hadjistavropoulos, & Jones, 2009)	45	-	-	.73
Students	Abramowitz et al. (2007)	442	-	-	.63
Hypochondriasis	McManus et al. (2012)	74	.84	-	-
Arm pain	Vranceanu, Safren, Cowan, & Ring (2010)	100	.71	-	-

4.2 SHAI – correlations with precursors to anxiety

As an exploratory analysis, correlations between the SHAI and measures of precursor or risk-factor variables for anxiety were collected and meta-analysed.

These measures were the Anxiety Sensitivity Index-3 (ASI-3) the Body Vigilance Scale (BVS) and the Intolerance of Uncertainty Scale (IUS). The results of this meta-analysis are presented in the forest plot in figure 10.
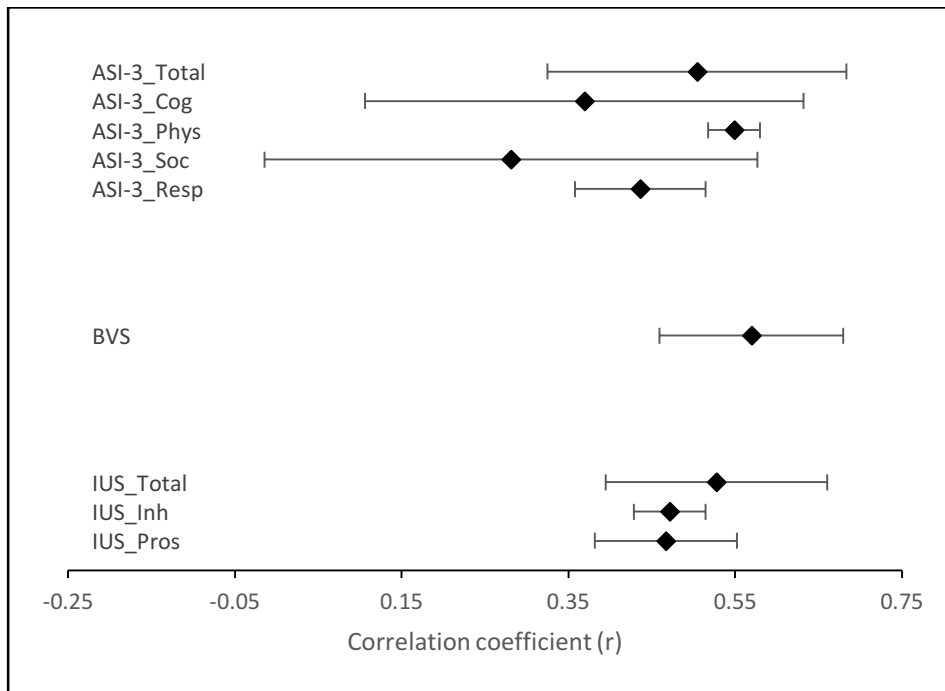
Figure 10. Forest plot of results of meta-analysis of correlations between the SHAI and ASI-3, BVS and IUS.

The SHAI correlated moderately with all measures of precursors to anxiety, with the lowest meta-analysed correlation coefficient being with the ASI-3_social subscale (r=0.281) and the largest with the BVS (r=0.569). Overall, this provides indirect evidence that the SHAI has good construct validity as one would expect the SHAI to have only moderate correlations with the above variables. If the correlations were larger it would imply they are measuring similar or the same constructs; if they were nearer zero then it implies the constructs are unrelated. The latter scenario would be a problematic finding as there is sound theoretical (and experimental) evidence that that health anxiety should be related to anxiety sensitivity, body vigilance and intolerance of uncertainty. It is possible to present this claim as all the above scales have undergone validation programmes so their psychometric properties are robust. If the measures had not undergone validation then they may be potentially psychometrically unstable and these correlations may be due to measurement error.

Examination of the correlations between the SHAI and ASI-3 total score and ASI-3 subscales indicates some variation in the strength of relationship across studies. This was greatest in the ASI-3 total, ASI-3_cognitive and ASI-3_Social subscales (see table 14). In contrast, the variation in correlation between the ASI-

3_Phys and ASI-3_Respiritory displayed little variation. This variability also provides some support for the construct validity of the SHAI as one would expect fear of physical and respiratory symptoms of anxiety to be consistently correlated with health anxiety. This is because catastrophic misinterpretation of somatic states are inherent components of the cognitive-behavioural health anxiety construct.

The BVS was also moderately correlated with the SHAI and showed moderate variation in strength of correlation coefficient across studies. This variation was broader than might be expected as body vigilance is thought to be an important variable in the maintenance of health anxiety so I would imagine this relationship would be more consistent.

Finally, there was less variation in the correlations between the SHAI and two IUS subscales, although this data was generated from only two studies, conducted by the same research group and within the same population (undergraduates). Therefore large heterogeneity is unlikely in these circumstances.

The results of the meta-analysis of the correlations between the SHAI and ASI-3, BVS and IUS along with the total number of studies and total sample size are presented in numerical form in table 14.

Table 14. Meta-analysis results of correlation between the SHAI and ASI-3 subscales, BVS and IUS subscales.

Scale	Total n	No. studies	Meta-analyzed r [95%CI]
ASI-3 Total	3314	10	0.504 [0.325, 0.683]
ASI-3 Cognitive	3988	10	0.369 [0.106, 0.632]
ASI-3 Physical	3988	10	0.549 [0.517, 0.580]
ASI-3 Social	3988	10	0.281 [-0.014, 0.577]
ASI-3 Respiratory	1021	4	0.436 [0.357, 0.514]
Body Vigilance Scale	1958	6	0.569 [0.459, 0.680]
IUS Total	1758	7	0.527 [0.394, 0.660]
IUS Inhibitory	1086	2	0.471 [0.428, 0.514]
IUS Prospective	1086	2	0.467 [0.381, 0.552]

4.3 Reliability generalisability analysis

I now provide the results of my investigation into the alpha coefficients reported for both scales. I begin by reviewing the quantity and range of published alpha coefficients. This is followed by the results of the meta-analysis where I report on the average reliability reported for various populations. Finally, I describe the

results of the meta-regression of SHAI reliability coefficients which revealed several potential moderating variables that may influence reliability. There was insufficient data to perform a meta-regression on HAQ alpha coefficients.

4.3.1. Overview of reliability coefficients

A summary of the availability and range of published alpha coefficients is reported in table 15. There were more alpha coefficients reported for the SHAI (n=54, range 0.74-0.97) than the HAQ (n=4, range 0.79-0.95). As only four alpha coefficients were published for the HAQ it was not possible to perform a meta-analysis. Therefore, I could not produce normative reliability data or investigate moderators of reliability coefficients. See appendix c for a summary of all alpha coefficients extracted.

Next I calculated the lower bound estimate for alpha coefficients and a failsafe-n for the SHAI. The (unweighted) average alpha derived from this sample was 0.875 (SD=0.052) and was reported in 54 studies. A lower bound estimate of the level of reliability in these studies is 0.8 SD (0.041) below the average of 0.875, yielding an estimate of the average alpha=0.834 in the non-reporting studies. There were 83 studies that did not publish an alpha coefficient. Using this information, a lower bound estimate of alpha can be calculated as follows:

$$\alpha_{LB \ estimate} = \frac{(54 \times 0.875) + (83 \times 0.834)}{54 + 83} = 0.850$$

This means that the overall average reliability in both reporting and non-reporting studies is an acceptable 0.850, although this is a lower average reliability than the 0.91 coefficient reported in the SHAI validation paper (Salkovskis et al., 2002).

The threshold I selected as an acceptable level for the SHAI is 0.80, this is because the SHAI is used in treatment trials so requires a higher level of reliability than the 0.7 level that would be required for exploratory or correlational research (Nunnally, 1978). Subtracting a further 0.8 SD from the alpha lower-bound threshold gives an estimate of 0.758 this may be combined with the average unweighted reliability (0.875) and inputted into the equation described in the method to provide a failsafe-n of the number of unpublished studies required to reduce the average alpha coefficient to an unacceptable level. This is as follows:

$$Fail \ safe \ N = \ 54 \times \left(\frac{0.875 - \ 0.80}{0.80 - \ 0.758}\right) = 97.99$$

Therefore 100 studies would need to be published with an alpha coefficient less than 0.758 for the mean population alpha to be lower than 0.8. This is a large number of studies, especially given that only one study reported an alpha coefficient lower than this value (α=0.74; Zhang, Liu, Li, Mao, & Yuan, 2015). Therefore, I argue that the

results of this meta-analysis of SHAI alpha coefficients is unlikely to be affected by publication bias.

Table 15. Summary of availability of SHAI and HAQ alpha coefficients.

Scale	Total studies utilising scale	Alpha reported	Percentage reporting	Alpha min	Alpha max
SHAI-14	44	20	45.5	0.81	0.97
SHAI-18	78	33	42.3	0.74	0.96
SHAI-13	1	1	100	0.90	0.90
SHAI-17	1	0	0	-	-
SHAI-unknown	8	0	0	-	-
HAQ	41	4	9.8	0.79	0.95

Next, I computed the meta-analysis of reliability coefficients to generate an average reliability, including 95% confidence intervals for the SHAI-14, SHAI-18. I also calculated an overall average reliability generated from all the available alpha coefficients from all versions of the scale (SHAI-13, 14,17 and 18). The results are presented in table 16.

The range of the three average alpha coefficients is small, with a difference of only 0.003 between the smallest and the largest, indicating that the SHAI has good average reliability in the studies included in this analysis. It is notable that these coefficients are lower that the $\alpha=0.91$ reported in the original validation paper (Salkovskis et al., 2002).

Table 16. Results of meta-analysis of SHAI alpha coefficients.

Scale type	Alpha [95%CI]	V	Q(df)	P	I^2
SHAI-14	0.873 [0.857,0.889]	0.001174	6122.82(17)	<0.0001	99.72
SHAI-18	0.870 [0.857,0.883]	0.001481	21660.24(31)	<0.0001	99.86
SHAI-All	0.871 [0.861,0.882]	0.001446	30149.15(49)	<0.0001	99.84

Note, v = random effects component, Q=homogeneity statistic, df=degrees of freedom, I^2=proportion of variance explained by true heterogeneity rather than measurement error.

Examination of the meta-analysis Q statistics (see table 17) indicates significant heterogeneity in alpha coefficients. This means there is statistically significant variation in SHAI scores that cannot be explained by random error. The I^2 statistic indicates that for all scale types, over 99% of the heterogeneity in alpha coefficients can be explained by non-random factors. This is important to investigate further as it is possible the SHAI may have systematically poorer reliability under certain conditions or when employed with certain populations.

To investigate the heterogeneity of alpha coefficients I have plotted all available coefficients and the average reliability derived from the meta-analysis in the forest plot depicted in figure 11. Examining this graph reveals both scales have greater reliability in mixed samples comprised of both community controls and people with health anxiety (range 0.84-0.96). In student populations, there was greater dispersion in alpha coefficients (range 0.74-0.91) compared to healthy controls (0.76-0.92) and the medically unwell (0.82-0.90).

Table 17. Meta-analysis of alpha coefficients, for healthy, student, medically unwell and mixed anxiety disorders populations.

Population	Alpha [95%CI]	V	Q(df)	P	I^2
Healthy	0.867 [0.847, 0.885]	0.001022	3235.90 (10)	<0.0001	99.69
Student	0.851 [0.837, 0.865]	0.000913	3225.79 (17)	<0.0001	99.47
Anxiety mixed	0.922 [0.910, 0.934]	0.000332	2234.33 (8)	<0.0001	99.64
Mixed medical	0.857 [0.830, 0.884]	0.001134	370.05 (10)	<0.0001	97.30
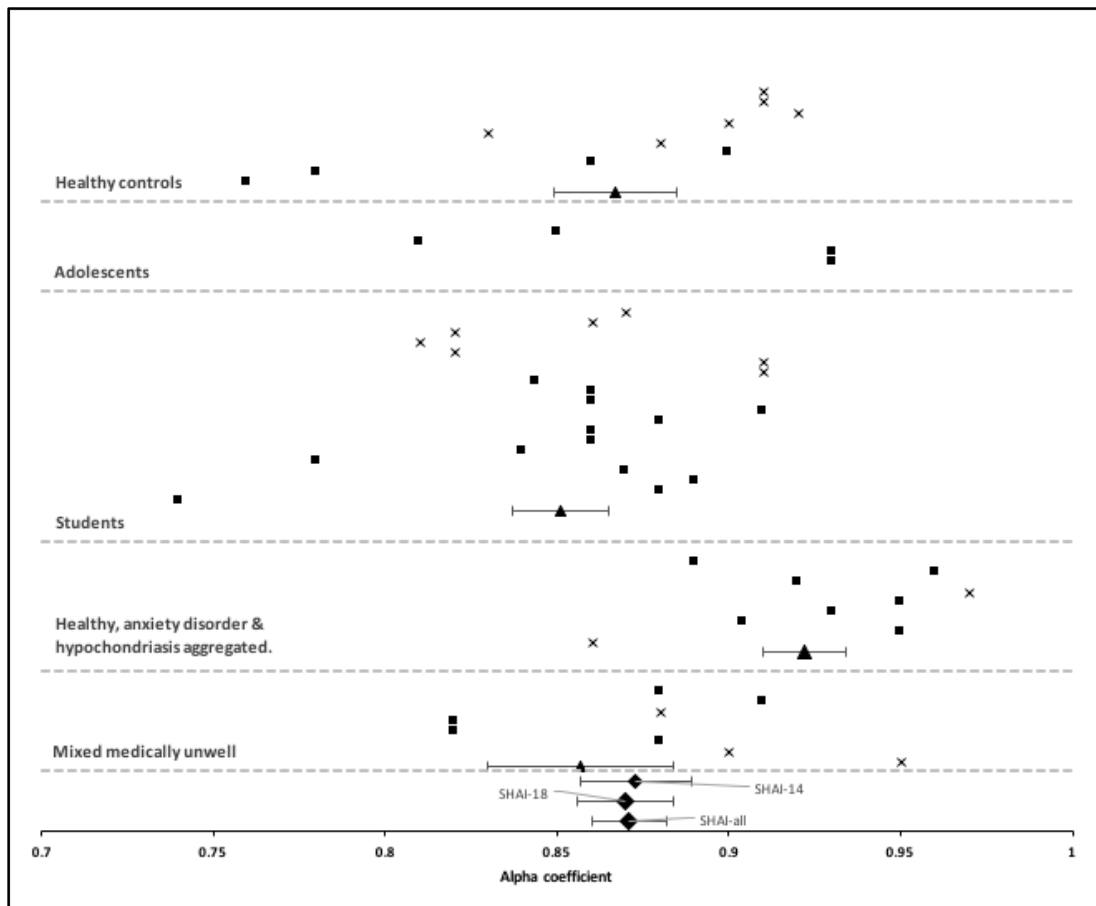
Figure 11. Forest plot of all SHAI alpha coefficients and meta-analysis providing overall mean reliability divided by sample composition. Note, square markers indicate SHAI-18 coefficients, crosses SHAI-14 coefficients, diamonds indicate the results of the meta-analysis of reliability coefficients, triangles indicate meta-analysis of reliability coefficients for each sub-population, error bars indicate 95% confidence intervals.

It is notable that that SHAI-18 has a broader dispersion of alpha coefficients (as denoted by the square markers). To further illustrate this, I have plotted the frequency of magnitude of SHAI alpha coefficients in figure 12. Examining figure 12a, the histogram summarising all published alpha coefficients, reveals a positive skew. As detailed in the method section this was corrected prior to meta-analysis. The distribution of SHAI-18 alpha coefficients (0.74-0.96) is broader than SHAI-14 coefficients (0.81-0.97). This suggests that either the version of the scale may impact on reliability, or potentially that each scale has been systematically employed in different populations, which themselves impact on reliability. Also of note is the criticism of the 4 additional 'negative consequences of illness items in the SHAI-18 as not being a measure of health anxiety (Alberts et al., 2013). Inclusion of these additional items in the SHAI-18 may reduce the scale's reliability as they are measuring a different construct.

Figure 12. a, b & c. Histograms displaying frequency of SHAI-all, SHAI-14 and SHAI-18 alpha coefficients.

Figure 12a. SHAI-all versions

Figure 12b. SHAI-14

Figure 12c. SHAI-18
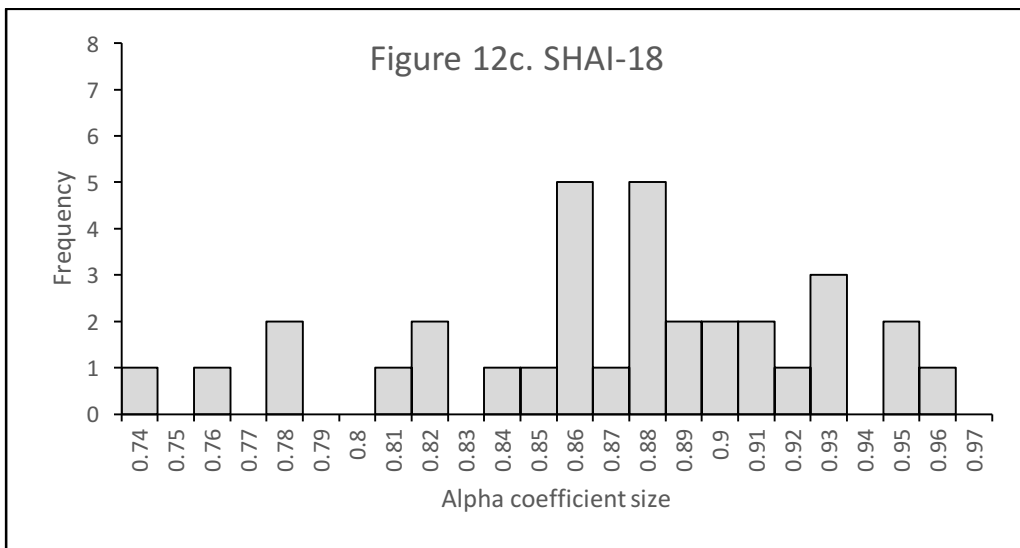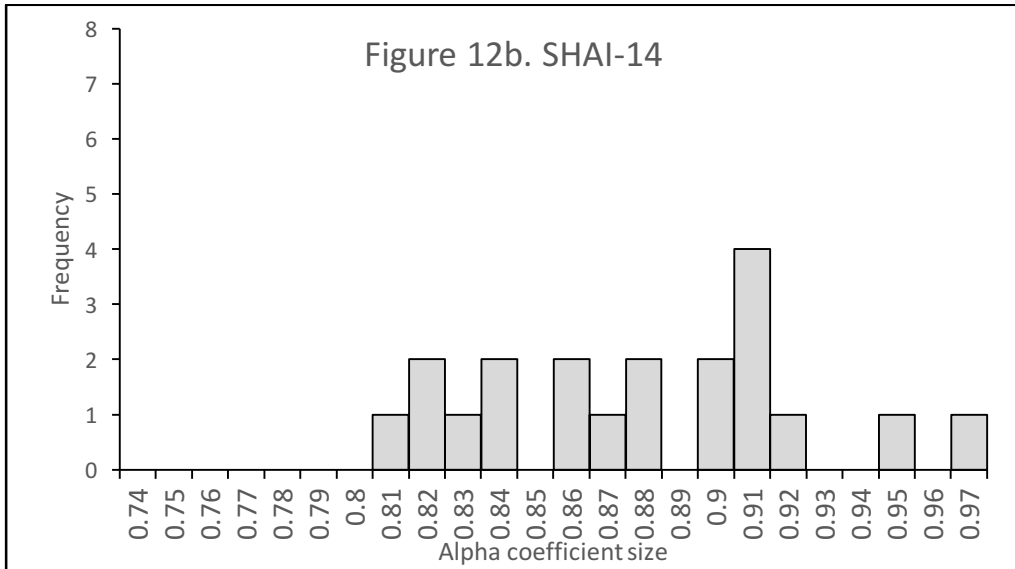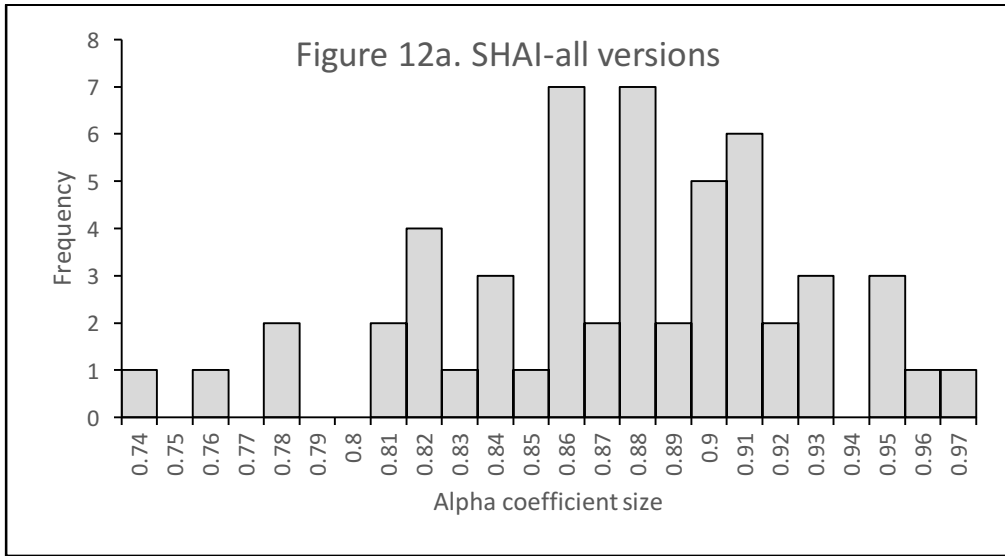
The data presented in figures12a-c indicate there may be differences in SHAI reliability according to scale type and population composition. Therefore, these variables were entered as predictors in a meta-regression of reliability coefficients. The results of this first meta-regression model are displayed in table 18 below.

Table 18. Model 1: Meta-regression of scale type and diagnostic group on coefficient alpha.

Predictor	Coefficient	SE	t	p
Scale type	.0006718	.0061682	0.11	0.914
Diagnosis	-.0008571	.0003246	-2.64	0.011
Constant	.5135555	.0153503	33.46	<0.001

The results of model 1 indicated that diagnostic group significantly predicted alpha, this first result appeared to statistically confirm the observation gleaned from the forest plot in figure 11 that alpha coefficients are generally higher in mixed samples. Scale type (SHAI-14 vs SHAI-18) did not predict variation in the strength of alpha coefficient. Therefore, the variation observed in the histograms reported in figure 12 was not responsible for statistically significant variations in alpha coefficients. It may be that the studies using the SHAI-18 that report lower reliability may have other factors (e.g. language of scale) that account for the lower alpha value. The R^2 value indicates this model accounts for a non-significant ($F(3,43)=2.62$, $p=0.062$), 9.84% of the between studies variation in alpha coefficients.

Botella and colleagues (Botella et al., 2010) recommend interpreting the variation in alpha coefficients in relationship to the mean score and variance of the included studies. They argue that if alpha varies as a function of mean score, then variability in reliability is related to the true score, whereas if alpha varies as a function of the sample variance then the variability in alpha is due to sampling error. Therefore, I added these as predictors in a second regression model presented in table 19.

Table 19. Model 2: Meta-regression of mean SHAI score and variance on coefficient alpha.

Predictor	Coefficient	SE	t	p
Mean	.0019543	.0014143	1.38	0.176
Variance	-.0016114	.000304	-5.30	<0.001
Constant	.5485561	.0240314	22.83	<0.001

This second model explained a total of 44.31% of the between-studies variance in alpha coefficient size and was statistically significant (F(2,33)=17.53, p<0.0001. Only the variance significantly predicted variations in coefficient alpha, as the variance increased, the alpha coefficient reduced in size, indicating the variation in alpha coefficients may be due to error in measurement rather than variations in alpha true scores.

Finally, as a third, exploratory analysis I also inputted the average age of sample, the language of scale and the nationality of participants in a third regression model. This was because the SHAI may perform differently in different nationalities due to the cross-cultural differences in the conceptualisation of health anxiety. Language of scale was included as a translated scale may affect reliability, again because the health anxiety construct may not be accurately related to non-English speaking languages or cultures. Finally, I inputed average age and gender mix to determine their impact on reliability. The results of this analysis are presented in table 19.

Table 20. Model 3: Meta-regression of sample average age, nationality and scale language on coefficient alpha.

Predictor	Coefficient	SE	t	p
Age	-.0011043	.0007498	-1.47	0.148
Nationality	.000448	.0022986	0.19	0.846
Language	.0065963	.0034817	1.89	0.065
Constant	.5132198	.0240359	21.35	<0.001

The third regression model explained a combined 10.00% of the between-studies variance in alpha coefficients and was not statistically significant (F(3,43)=2.62, p=0.0626). Only scale language emerged as a significant predictor of coefficient alpha, nationality and average age of participants did not predict variations in alpha coefficients.

The final regression model is presented in table 21 where diagnostic group, sample variance and scale language were entered as predictors, this model explained 69.00% of the variance in alpha coefficients and was statistically significant (F(3,32)=24.86, p<0.0001). In this model, diagnostic group was no longer a significant predictor of coefficient alpha, whereas language and variance in mean scores did predict variations in alpha. It may be that the variance explained by diagnostic group is shared with scale language and errors in measurement in these samples rather than due to an inherent property of the sample in these studies. Because diagnostic group no-longer predicted variations in alpha I did not investigate

this further, if diagnostic group had been a significant predictor then I would have investigated which groups did predict variation in coefficient alpha.

Table 21. Model 4: Meta-regression of diagnostic group, variance and scale language on coefficient alpha.

Predictor	Coefficient	SE	t	p
Diagnosis	-.0000827	.0002313	-0.36	0.723
Variance	-.001296	.0002379	-5.45	<0.001
Language	.0117084	.0022283	5.25	<0.001
Constant	.5399614	.0134814	50.05	<0.001

A summary of all 4 regression models, with R^2 values representing the proportion of variation in alpha coefficients along with F probability statistics are presented in table 22.

Table 22. Summary of results of all regression models.

Model	Predictors	τ^2	R^2	F(df)	p
M1	Scale type, diagnosis	.003725	9.84%	3.50(2,48)	0.0382
M2	Mean, variance	.001877	44.31%	17.53(2,33)	<0.0001
M3	Age, nationality, language	.00373	10.00%	2.62(3,43)	0.0626
M4	Diagnosis, variance, language	0.001044	69.00%	24.86(3,32)	<0.0001

4.4. Summary of results

In summary, I have generated means scores for the SHAI-14, -18 and the HAQ for various populations, these may be employed as normative values. I have investigated the construct validity of the SHAI via use of the correlations between the SHAI and other psychometric measures. This is complemented by a reliability generalisability analysis of SHAI alpha coefficients. Unfortunately there was insufficient data for an investigation of the construct validity and reliability of the HAQ. I will now discuss these findings.

5. Discussion

There were four aims of this project, the first was to generate normative data for the HAQ and SHAI for all populations with three or more published mean scores or alpha coefficients. The second aim was to generate an average reliability coefficient for each scale and then to explore variation in alpha coefficients using meta-regression. The third aim involved evaluating the construct validity of both scales by examining their correlations with other psychometric measure.

I will now summarise and discuss my findings regarding each of the above aims. This is followed by a discussion of the strengths and weaknesses of the project, clinical and research implications and finally suggestions for future research. I begin by discussing the population norms I calculated through meta-analysis.

5.1. Population norms

As there was a larger quantity of data available for the SHAI, I begin by discussing the norms generated for this scale, how I explored heterogeneity in mean scores and then move on to discuss norms generated for the HAQ.

5.1.1. SHAI population norms

Using published data, I was able produce norms for the SHAI-14 and SHAI-18 in the following populations; healthy community controls, undergraduate students, health anxiety and pain populations. A norm for anxiety disorder populations was generated for the SHAI-18 only. This study adds to the literature by producing norms for student and pain populations, which did not feature in the original scale validation study. Additionally, there were sufficient alpha coefficients available to generate an average alpha for healthy control and student populations for both the SHAI-14 and -18. This data is presented in table 23.

The meta-analysed grand mean scores and SDs may be used as a normative value for each population. The score for the health anxiety populations for the SHAI-14 and -18 may be used to develop a clinical cut-off score. This score can be used to indicate whether an individual has made clinically significant change. Jacobson et al., (1984) recommend three different means to calculating a cut off for clinical significance. Criterion 'a' is a score that is two standard deviations below the average in a clinical population. The data produced in this analysis can be used to calculate a clinical cut-off score.

Using Jacobson's criterion 'a', the clinical cut-off score for the SHAI-14 is 18.34 and for the SHAI-18 it is 15.35. Reliable change (Jacobson & Truax, 1991) refers to a change in an individual's psychometric scores that are greater than can be attributed to measurement error alone. This is calculated by use of a measure's reliability coefficient and represents a magnitude of change that only 5% of a population would achieve by chance. This yields a reliable change index of 4.82 for the SHAI-14 and 8.46 for the SHAI-18. Clinically significant and reliable change calculations were performed using the Leeds Reliable Change Indicator (Morley & Dowzer, 2014). This information may be used in clinical practice or for research purposes to determine if an individual has made clinically significant or reliable change during treatment. Use of clinically significant and reliable change is of particular use when conducting single case research (Kazdin, 2011).

It is notable that the grand mean scores I calculated for healthy control populations are very similar to those in the original SHAI validation paper (see table 23; Salkovskis et al., 2002). In contrast the grand mean scores in health anxiety populations are lower in my analysis than in the SHAI validation paper. Although they are within one SD of the scores provided by Salkovskis and colleagues. It is difficult to determine why this is the case, it may be that health anxiety was simply higher in the original sample compared to the populations from which I generated norms. I next attempted to discover whether sample composition or other study-level variables were responsible for variations in SHAI mean score, I now discuss my findings regarding this variation.

Table 23. Normative data for the SHAI-14 and SHA-18 and equivalent values reported in Salkovskis et al., (2002).

Population	Variable	Meta-analysis score [95%CI]		Validation paper score	
		SHAI-14	**SHAI-18**	**SHAI-14**	**SHAI-18**
Healthy	Mean	9.21 [8.26, 10.16]	12.56 [10.26,14.86]	9.4	12.2
	SD	5.36 [4.71, 6.01]	6.03 [5.37, 6.68]	5.1	6.2
	Alpha	0.885 [0.863, 0.907]	0.840 [0.767, 0.913]		
Student	Mean	17.16 [12.94, 21.37]	12.35 [11.78, 12.93]	-	-
	SD	5.78 [5.03, 6.53]	6.51 [6.11, 6.91]	-	-
	Alpha	0.844 [0.813, 0.876]	0.853 [0.835, 0.870]	-	-
Health anxiety	Mean	27.90 [25.23-30.56]	32.13 [30.77, 33.49]	30.1	37.9
	SD	4.88 [4.56, 5.20]	8.56 [7.47, 9.66]	5.5	6.8
Anxiety disorders	Mean	-	16.28 [13.07, 19.48]	14.9	18.5
	SD	-	12.04 [10.16, 13.92]	6.2	7.3
Pain	Mean	14.52 [12.66-16.38]	14.17 [9.97, 18.36]	-	-
	SD	7.07 [4.43, 9.71]	7.07 [4.43, 9.71]	-	-

5.1.1.1. Exploring heterogeneity in SHAI mean scores

The majority of the effect sizes I computed using mean health anxiety scores were heterogeneous in nature. The presence of heterogeneity is often seen as a problem with a meta-analysis as it can indicate the included studies are too different to be compared (see the 'comparing apples and oranges' section; Sharpe, 1997). Indeed, this is potentially a fair criticism of my attempts to develop population norms for different populations; the existence of significant heterogeneity in mean health anxiety score means it is difficult to state that my calculated grand mean adequately reflects the population of interest. It may be that statistical outliers or extreme variables may bias these findings. There is a risk that a mean score reported in any study could be an outlier, so could equally be an issue in a scale development study. The advantages of using multiple studies is that the effect of extreme scores should be lessened by the presence of other mean scores. A second advantage of using meta-analysis to develop normative data is that it allows statistical investigation of heterogeneity in mean scores. This would not be possible in studies with only one sample.

In terms of classical test theory (Lord & Novick, 1968), there are three potential reasons for this heterogeneity: 1) the studies I analysed employed an inadequate sampling strategy therefore producing mean scores that do not adequately represent the population of interest; 2) the performance of the SHAI varies when used with different samples; 3) studies reporting on the same kind of participants (e.g. students) were sampled from true populations that were different in composition and had different health anxiety scores.

I attempted to explore various sources of heterogeneity by use of meta-regression, I explored option 1) by entering study-level variables regarding sampling strategy as predictors in a meta-regression but did not find any association between sampling strategy and SHAI mean score. It may be that there was no effect of sampling strategy on mean score or that this effect was too small to be detected. As assessed by the I^2 statistic, the levels of heterogeneity in SHAI mean score were above 80% across all populations (except health anxious populations assessed by the SHAI-18). This is a high level of heterogeneity left unexplained and none of the study level variables I selected could account of a significant proportion of this variation. It is possible that other study level variables exist that may explain this variation. Although meta-analysis will always be restricted to study-level data that has been published.

The second potential source of heterogeneity according to classical test theory (option number 2 above) is that the SHAI varies in accuracy (a component of method variance and error variance; Campbell & Fiske, 1959) when used with different populations. Some indirect evidence for this comes from the finding that the SHAI has lower reliability when used in mixed-medical and student samples. However, the meta-analysed reliability coefficient was above 0.85 in all populations, so is sufficiently high to have sufficient reliability for research purposes, so does not support the notion that this variability is due to error in measurement. I will comment on this again in the construct validity discussion section.

The final option is simply that the levels of health anxiety true scores differed between the various samples across the same populations, this hypothesis has the most support from the data that I collected.

It is unfortunate that the heterogeneity in mean scores could not be explained, this is potentially problematic as it is difficult to determine the accuracy of the normative data that I have produced. On the other hand, identifying that there is significant heterogeneity in SHAI mean score even within groups with health anxiety that were selected by diagnostic interview may be helpful to future researchers: this study demonstrates that SHAI scores can significantly vary even between highly similar samples and this effect may be a limitation of the measure itself. I argue the impact of heterogeneity in mean score may be reduced by use of the 95% confidence intervals giving the range of health anxiety score in each population rather that than the mean score that was calculated by meta-analysis.

5.1.1.2. HAQ population norms

There were less studies reporting data for the HAQ (n=34) compared to the SHAI (n=86). I have produced grand mean scores for healthy control populations, cardiac patients, clinic attenders and gastroenterology patients, these are presented in table 24. There was only one published use of the HAQ with a health anxiety population (Noyes et al., 2002). Therefore, a norm could not be created for this population, nor could I calculate a cut-off for clinically significant change or a reliable change index.

Similarly, to the SHAI, the data in table 24 may be utilised as HAQ normative data for each population. Compared to the scores reported in the HAQ validation paper (Lucock & Morley, 1996), the meta-analysed grand mean score for healthy controls was higher (12.29 vs. 8.62) and the clinic attenders grand

mean score was lower (15.36 vs. 17.35). Although, the differences in scores in both populations was within one SD (of both the original paper and my meta-analysis), so may be within normal variation for the HAQ.

Table 24. Normative data for the HAQ and equivalent values reported in Lucock and Morley (1996).

Population	Variable	HAQ meta-analysis score [95%CI]	HAQ validation score
Healthy controls	Mean	12.29 [8.78, 15.80]	8.62
	SD	8.39 [7.04, 9.74]	7.96
Cardiac	Mean	13.51 [9.57, 17.47]	-
	SD	8.06 [2.78, 13.35]	-
Clinic attenders	Mean	15.36 [13.86, 16.86]	17.35
	SD	9.80 [8.36, 11.24]	11.16
Gastro patients	Mean	15.59 [10.79, 20.38]	-
	SD	7.78 [6.07, 9.50]	-

One potential criticism of my analysis of HAQ scores is that the clinic attenders sample was comprised of several populations (e.g. medical outpatients, attenders at neurology) so may be too heterogeneous in nature to be appropriately combined into a single population norm.

Exploring heterogeneity in mean SHAI scores was a supplementary and exploratory analysis. Because not a single study level variable predicted variation in SHAI mean score, I did not repeat this analysis with the HAQ. This is also because the HAQ and SHAI are both questionnaire measures, so will presumably account for a similar proportion of method-variance. As detailed in the method, there were insufficient HAQ alpha coefficients available for meta-analysis (Lopez-Lopez, Marin-Martinez, Sanchez-Meca, Noortgate, & Viechtbauer, 2014), so an average reliability could not be generated for any population.

5.2 Construct validity

I now consider information relating to the construct validity of the SHAI generated from the meta-analysis of correlation coefficients between the SHAI and other measures.

Overall, the pattern of relationships between the SHAI and measures of health anxiety, anxiety, depression and OCD was broadly as expected. The

SHAI correlated most strongly with other measures of health anxiety, followed by generalised anxiety measures. A finding that was contrary to predictions was that the average correlation with measures of depression were higher than with measures of OCD.

There was significant heterogeneity in all correlation coefficients, unfortunately there was insufficient data to perform a moderator analysis. It is possible that sample type may impact on the strength of these relationships, or that there are multiple populations within each study sample. It is also possible that other study level variables that may explain the heterogeneity in correlation coefficients between the SHAI and other measures.

The correlation coefficients between the SHAI and OCD-R were highly heterogeneous, in fact the credibility interval was broader than the strength of the correlation itself. This is problematic, as such a large interval implies there may be no relationship between these variables. This finding is hard to interpret, but is surprising nevertheless as I predicted there would be some relationship between the two measures.

Another possibility is that the construct of health anxiety is categorical rather than dimensional. Consequently, even in a homogenous sample (e.g. undergraduate students) there would be two populations that differ in their experience of health anxiety. The latent structure of health anxiety is still debated (Asmundson et al., 2012; Ferguson, 2009; Longley et al., 2010), however the most methodologically rigorous taxometric analysis found a categorical latent structure comprising two distinct groups or 'taxons'. One 'non-anxious' group scored universally low on all IAS items, a second 'anxious' had a dimensional latent structure (Asmundson et al., 2012). This non-anxious group comprised approximately 15% of the undergraduate study sample. It may be that a portion of the heterogeneity in correlation coefficients is due to two distinct groups, each with a differring experience of health anxiety being included in study samples. This is of particular relevance as the taxometric analysis conducted by Asmundson and colleagues was conducted with North American undergraduate students, which are the same population that the majority of SHAI studies that reported correlations were carried out with.

Correlations between the SHAI, IUS, ASI and BVS were entered as an exploratory analysis. These variables are all considered to be precursors or risk factors for anxiety (Alberts et al., 2013) and correlated moderately with the SHAI. This provides evidence that these risk factors may also be involved in some way with health anxiety. Interestingly, there was heterogeneity in the

magnitude of correlations between these precursor variables. The narrowest ASI-3R credibility intervals (indicating less heterogeneity) occured in correlations between the SHAI and the 'respiratory' and 'physical' subscales. These subscales assess an individual's fear that a bodily symptom will lead to negative consequences. I argue the finding that most studies reported a similar strength of relationship (so have narrower credibility intervals) with the SHAI lends support to the cognitive model, as in this model a catastropic misinterpretation of bodily states is central to the development of health anxiety.

The correlation coefficients between the SHAI and the measures assessing precursors to anxiety (ASI-3, BVS and IUS) were heterogeneous, with some having large credibility intervals. This variation warrants further investigation. It may be that the relationships between these scales and the SHAI vary as a function of some other variable, for instance the sample composition or mean SHAI score or by some other demographic variable. This could be investigated via a moderation analysis, unfortunately there were insufficient correlation coefficients available for this kind of analysis. Simulation studies have revealed that a minimum of 20 studies are required to have sufficient power to detect variation in heterogeneity (Lopez-Lopez et al., 2014). In the included studies, the ASI-3 was used ten times, the BVS six and the IUS on 7 occasions, so were insufficient in number to perform a moderator analysis of the impact of other variables on heterogeneity. Studies typically utilised more than one of these measures in one analysis (e.g. IUS and BVS), which is problematic for further analysis as their variance will be shared (Borenstein et al., 2009).

The correlations between the ASI-3 cognitive and social subscale and the SHAI were more heterogenous. These variables assess fear of the consequences of the thoughts associated with anxiety and fear of the social consequences of anxiety respectively. Again, I argue this provides suport for the cognive model, as these variables are thought to be more peripheral in the maintenance of health anxiety.

The data extracted regarding correlation coefficients provides some support for the cognitive-behavioural conceptualisation of health anxiety and for the ability of the SHAI to measure this construct. This is because the pattern of correlation coefficients were broadly in the pattern that was expected, demonstrating that SHAI has 'discriminant validity'. That is, for example the SHAI can distinguish between those scoring highly on health anxiety from those scoring more highly on generalised anxiety. Additionally, the pattern of variation in correlation coefficients broadly appeared in a theoretically consistent

way with the correlations between the SHAI and measures of generalised anxiety being strongest. On the basis of these results, health anxiety can be considered to be a construct that is distinct, but related to generalised anxiety, (if the two were the same construct, then correlations between SHAI and anxiety measures would be higher). The fact that there is a moderate relationship between anxiety and health anxiety may provide some support for the notion that health anxiety is an anxiety disorder (Salkovskis & Warwick, 1986), as if there were no relationship between the two then the meta-analysed correlation coefficient would be smaller. There was heterogeneity in the magnitude of the correlation coefficents between the SHAI and anxiety measures. Assessing the relationship between anxiety and health anxiety solely in 'pure' samples (e.g only studies with samples of indiviuals experiencing hypochondraisis) may be another helpful means of assessing whether health anxiety is an anxiety disorder. If the relationship between the two is consistent across studies then may provide support for this hypothesis, if there is variability then this may indicate that anxiety is a less important component of hypochondriasis. Conducting a moderator analysis with diagnosic label inputted as a moderating variable may be an alternative means of achieving this using meta-analysis. This was not possible in the current study due to a lack of available data.

An unexpected finding was that the correlation between the SHAI and measures of depression were greater than those with measures of OCD. This relationship is the opposite to my predictions as some similarities between OCD and health anxiety have been noted, for example compulsive checking or monitoring one's body for signs of illness (Ania Greeven et al., 2009).

Another means of assessing the cognitive behavioural construct of health anxiety may be to examine its relationship to quality of life (QoL). It follows that worry about one's health should lead to increased stress and this will consequently reduce one's QoL. Similarly to the previously described analysis, this could be assessed by examining the correlation coefficients between measures of health anxiety and measures of QoL and clinical stress and if consistent with this prediction, clinical stress should be positively related to health anxiety and QoL negatively related. One study (Porritt, Sufi, Barlow, & Baker, 2014) reported a correlation of r=0.49, p<0.001 between the SHAI and the Index of Clinical Stress (Walmyr Publishing Company, 1992), another study (Olatunji et al., 2009) reported a correlation of r=-0.19,p=n.s. between the SHAI and the EuroQol (The EuroQol Research Group, 1990) a measure of Quality of Life. The moderate correlation between the SHAI and ICS is consistent with the

cognive behavioural conceptualisation of health anxiety, but the non-significant relationship between the SHAI and QoL is not. The lack of research investigating between health anxiety, stress and QoL is a limitation of the literature and could not be adequately assessed using meta-analysis in this study.

A note of caution is warranted about these conclusions, the first is that correlation coefficients only indicate a potential relationship. It may be this is a spurious finding and this relationship is an artifact due to them both co-relating to another unknown variable. A second issue is the potential existence of two taxons within an undergradute sample. This is problematic for my attempts to assess the health anxiety construct using correlations between the SHAI, IUS, ASI and BVS, this is because all of the studies making use of these scales were in student samples. It may be that the precursers to anxiety relate differently to the each of the two taxons. I suggest this is highly likely, as Asmundson et al. (2012) report that a non-anxious sub-sample scored very low on all IAS items. Future research should attempt to separate these groups prior to analysis.

The factor structure of both scales has relevance to a discussion of the cognitive behavioural construct of health anxiety as factor analysis has revealed that each scale may be comprised of multiple underlying factors. In the original scale validation paper, the HAQ was found to have four factors; 'interference with life', 'fear of illness and death', health worry and preoccupation' and 'reassurance seeking-behaviour' (Lucock & Morley, 1996). The presence of multiple latent factors indicate that the cognitive behavioural construct of health anxiety (as assessed by the HAQ) may comprise related, but distinct constructs. Beyond the original validation paper, there have been no attempts to reinvestigate the factor structure of the HAQ and doing so may be a helpful avenue for future research.

The factor structure of the SHAI has been re-evaluated and both a two- and three- factor solution have been reported in the literature (Alberts et al., 2013). The initial validation paper of the SHAI-18 reported a two-factor solution, with one factor comprising the the SHAI-14 items, the second was formed of the 4-item 'negative consequences of illness' subscale (Salkovskis et al., 2002), this structure has been replicated in a study using a student sample (Wheaton et al., 2010). However, two additional studies also employing student samples reported a three-factor solution. In both studies this included the 4-item 'negative consequences of illness' subscale, the remaining 14-items comprising the SHAI-14 were divided into two subscales named 'illness likelihood' and 'body vigilance' (Abramowitz, Deacon, et al., 2007; Olatunji, 2009). A more recent

study conducted in a community and a multiple sclerosis sample found the SHAI-14 was better explained by a two-factor solution, the authors named these 'fear of illness' and 'thought intrusion' (Alberts, Sharpe, Kehler, & Hadjistavropoulos, 2011). The study conducted by Abramowitz, Deacon, et al., (2007) also found that the 'body vigilance' scale alone predicted use of medical services (as assessed by the medical utilisation questionnaire) and that 'body vigilance' and the 'illness likelihood' scales predicted safety behaviours related to health (e.g. telephoning for medical advice). These findings indicate that the cognitive behvioural conceptualisation of health anxiety as measured by the SHAI may be considered to have multiple components rather than a unified and generic health anxiety construct. The fact that each factor may differentially predicit medical use and safety behaviours provides some prelminary evidence that separating body vigilance and perceived likelihood of illness may be helpful for future research and may also indicate separate constructs within the health anxiety domain.

Related to the findings regarding the factor analysis is that the average alpha coefficient reported for all the SHAI versions were above 0.87. This high alpha value indicates that even if the SHAI (and cognitive behavioural construct of health anxiety) may be divided, it is likely that the scale items comprising each component are consistently related to all the other items and therefore each construct is also likely to be related. This could be investigated in the future by reporting correlation coefficients between factors and also by publication of alpha coefficients specific to each factor.

5.2.1 Reliability

In this section I discuss the reliability information gathered about each scale. There was a large difference between the SHAI and HAQ in terms of the number of available reliability coefficients (n=53 and n=4 respectively). I also observed the same pattern in terms of correlations between the SHAI/HAQ and other scales, with fewer coefficients reported by researchers employing the HAQ. I suggest this may be explained by differences in the disciplines of the scholars employing the scales: the HAQ was far more likely to be employed in medical research being conducted by medical personnel, whereas the SHAI was used more frequently by psychologists. I suggest that due to the differences in research training and applications of psychometrics, psychologists are more likely to report additional psychometric information. Because of lack of HAQ data, this section focusses on data derived from use of the SHAI.

The mean SHAI-14 and -18 alpha coefficients were all at 0.870 or above. This indicates that on average, for the samples included in this analysis the SHAI has good reliability. The meta-regression of SHAI reliability coefficients explained 69% of the variance in alpha coefficients, this was explained by a combination of scale language and study variance. This indicates that majority of this variance is due to poor translations of the SHAI, or that the health anxiety construct is not applicable in other cultures. Unfortunately, it is not possible with the data available to determine which of these is that case. It was only possible to say there was an overall effect of scale language on reliability. This is because the majority of foreign language translations of the SHAI were used in only one or two studies, making it difficult to determine if a there is a systematic relationship between one particular translation of the SHAI and reliability. However, further investigation of this effect may be potentially fruitful direction for future research.

5.3. Strengths and limitations

A strength of this study is the breadth of analysis. Population norms were generated for a range of samples, this exceeded the range of populations that were assessed in the original validation papers of both papers. Use of data gleaned from multiple sources regarding the same populations also has the advantage of reducing the impact of erroneous findings or outliers. It is possible that in the original scale validation papers the SHAI and HAQ included samples that were not entirely representative of the wider population. There was some evidence for this in my analysis, as the mean health anxiety scores I generated were lower than in the SHAI validation paper (Salkovskis et al., 2002). There is a risk that samples included in this analysis were not from the same population as those in the Salkovskis paper, for example it may be that some samples were simply not as health anxious. However, this is unlikely as in the included studies, health anxiety was generally assessed using DSM interview, the same criteria used by Salkovskis et al. In addition, I assessed for the impact of sample composition on mean score and found no relationship. The latter analysis would not be possible without collecting a large amount of data from multiple populations.

Another strength of this study is the large total sample size, for example there were nearly 4,000 participants included in the correlations between the ASI-3 subscales and the SHAI, far exceeding the usual size in a scale validation paper. This means that providing included studies were conducted in a

methodologically rigorous way, this offers a more precise estimate of scale mean scores as the impact of outliers within study samples has less impact in larger samples (Cousineau & Chartier, 2010).

A limitation is the use of data derived from mean scores which were then aggregated via meta-analysis. If this data were available in raw format, with participant demographics then more precise norms could be generated and related to age and gender. However, as discussed in the method section, this was not possible as authors did not respond to requests to share their data.

Another limitation of this approach is that there is some heterogeneity in the composition of samples included in each normative value generated by meta-analysis. This means the normative values may have been influenced by a sample containing unusually high numbers of individuals with (or without) health anxiety, so may affect the accuracy of these normative values. I attempted to correct for this in the meta-regression of sample characteristics (whether the sample was well described or homogenous in composition) on SHAI score and found no statistical relationship. Heterogeneity of samples should still be considered a limitation as there may be insufficient data for an effect of sample composition to be observed statistically, but this may still be a potential explanatory factor for the significant levels of heterogeneity found in mean health anxiety score found in each group of studies. This difficulty may be overcome by solely including studies that identify participants using diagnostic interview, this will only be possible when additional data regarding each measure has been published.

It is unfortunate that there was insufficient data to perform an analysis on the construct validity and reliability of the HAQ. The available alpha coefficients did not indicate that the HAQ lacks reliability, but there was no additional information that could confirm the scale has good reliability.

In summary, this meta-analysis allowed for the production of norms from a large number of studies, which can help shed light on the range of health anxiety scores across a broader range of populations than the original validation paper. Employing mean scores from a greater number of applications of the scale in the same population, which has the advantage of reducing the impact of outliers or erroneous findings. A robust scale re-validation attempt such as this adds to the evidence that the SHAI and to a lesser extent the HAQ generally have good reliability and validity in the studies in this analysis.

5.4. Research implications

With the data reviewed in this study, the SHAI can be confirmed as a valid and reliable research instrument. Translated versions should be used with caution as this negatively impacts reliability. A future study may investigate which translations are particularly problematic, this could be achieved by analysis of reliability coefficients published in future research. Another option might be to review translated scales and back-translate them and conduct face validity checks. A more robust procedure would be to assess the construct validity of scales numerically via comparison with other measures that have been validated in the corresponding language.

It was not possible to examine the construct validity of health anxiety within certain groups (e.g. comparing students with clinic attenders), as there were insufficient correlation coefficients available. It may be that health anxiety has a different form or quality at different intensities, there is some evidence for this as reported by Asmundson et al. (2012). This is not a limitation of the design of the present study, as it is due to insufficient data. Nevertheless, this may be a fruitful area for research in the future when more data is available.

Another area for research that may shed light on the cognitive behavioural construct of health anxiety may be investigating its relationship with stress and QoL. Due to a lack of research it was not possible to achieve this in the current study, future research could investigate whether increased health anxiety leads to increased stress and reduced quality of life. It may also be helpful to assess whether this relationship varies according to population, for example it may be that those with medically unexplained symptoms do not experience a high level of anxiety, but nevertheless are preoccupied with a somatic symptom and have a correspondingly lower QoL.

It was not possible to calculate information beyond normative data for the HAQ. It was not possible to provide additional information about the psychometric properties of the HAQ.

5.5. Clinical implications

In terms of the implications for clinical practice, the SHAI can be supported as a measure with good construct validity and reliability. This makes it a suitable measure for use within clinical practice. The normative data that was generated in this analysis may be used as a point of comparison in clinical practice. Additionally, the data provided regarding clinical cut-off scores and a

reliable change index can be employed in clinical practice to determine if a client has made clinically significant and reliable change (Jacobson & Truax, 1991). Similarly, to the research implications, translated versions of the SHAI should be used with caution until their reliability and validity has been better assessed.

5.6. Future research

When more studies have been published, across a range of populations it may be possible to generate more precise normative data and investigate whether there are more fine-grained differences in health anxiety between populations.

The relationship between OCD and health anxiety could be investigated further, there was a large heterogeneity in the calculated grand correlation coefficients. In the current study, it was not possible to determine whether this heterogeneity was due to some conceptual overlap in some people (perhaps with lower levels of health anxiety) and not in others (anxiety/hypochondriasis samples). As previously stated, if more correlations were available a moderator analysis could be conducted to determine if severity of health anxiety (or another variable) moderates this relationship.

Another potential area for research would be evaluating the heterogeneity in HAQ and SHA standard deviations. I did not investigate this in the current study, nevertheless moderators of the standard deviation may provide further information about sources of error of measurement captured by the scales.

Studies conducted by the same research group, using the same methods but in different populations may lead to a better understanding of sources of heterogeneity in means scores. An example of this is in the studies by (Graff et al., 2006) who in the same study, reported different HAQ scores according to whether a patient had active or inactive Crohne's Disease or Ulcerative Colitis. If more studies are published using the same method, then we can be more confident that this difference is due to changes in true scores rather than due to measurement error.

5.7. General discussion of method

There are a number of advantages and disadvantages of applying meta-analysis to better understand psychometric questionnaires. Beginning with an advantage, this application of meta-analysis has the potential to develop

normative data for a wider range of populations than would be possible in a single scale development study. Similarly, this approach allows one to generate a grand mean and standard deviation for a non-clinical control population, this data may then be used to generate criteria for reliable (Jacobson et al., 1984) and clinically significant change (Jacobson & Truax, 1991).

A potential difficulty inherent in this approach is use of published data. It is unknown what the effect of publication bias may be on baseline mean health anxiety scores. I am unaware of a means to asses this statistically as existing techniques such as funnel plots (Egger, Smith, Schneider, & Minder, 1997) are designed for use with pre/post treatment effect sizes or correlational data. Also, as previously discussed there is a risk that the samples' composition in included studies may be poorly described or be biased in some way, this may then lead to a biased mean score being included in the meta-analysis. However, these same criticisms may also be levelled at the norms generated from any scale development study and it is equally plausible that any such paper is at risk of publication bias or of inadvertently employing a non-representative sample. An advantage of using meta-analysis to develop normative data is that a greater number of studies may be included. This may help reduce the impact of using norms generated from a single study that may have a non-representative sample or other confounding variable. On the other hand, there is the potential for normative data to be appear more precise than is warranted given the potential for confounding variables or non-representative sampling. It may be possible to minimise the impact of this by only including studies with well described samples. Another possibility is relying on the 95% confidence intervals, so reporting a range of values as a norm rather than using the grand mean generated from the meta-analysis. This way a lower and upper bound for the range of health anxiety could be calculated for various samples.

Calculation of an average reliability and exploration of the variation in reliability coefficients is another strength of this approach as it has the potential to reveal samples or circumstances that a test should not be employed. Assessing reliability across many studies can provide much additional information about a test than can be gleaned from a single application (Vacha-Haase & Thompson, 2011).

Finally, meta-analysis of correlation coefficients as a means to investigate the convergent and discriminant validity of the SHAI has been previously investigated (Alberts et al., 2013). I followed the same steps of Alberts and colleagues and I am unaware of any formalised guidance for investigating the

construct validity of a scale in this way. The advantages and disadvantages of this approach are similar to those relevant to generating normative data using meta-analysis. The advantages include collecting data from a wider range of sources than a scale development paper and the potential to investigate moderating variables such as sample or study characteristics to determine whether these may impact on the relationship between constructs. A disadvantage of this approach is that it may lead to what Egger and colleagues (Egger, Schneider, & Smith, 1998) refer to as 'spurious precision'. That is, the potential for data from observational studies to appear as if they provide an accurate reflection of the relationship between two or more variables even though confounding variables or sample characteristics may distort this relationship. Egger and colleagues recommend paying attention to moderating variables that explain the heterogeneity in correlation coefficients in different studies. This was not possible in this study due to insufficient data, however when more research is published using the SHAI or HAQ this may be a possible extension to this analysis.

This method could easily be applied to other psychometric measures, for example the PHQ-9 (Kroenke & Spitzer, 2002) a widely used measure of depression. A meta-analysis conducted using the same method employed in this thesis would complement an existing meta-analysis of studies that assesses the diagnostic accuracy of the PHQ-9 (Manea, Gilbody, & McMillan, 2015). This is because the study conducted by Manea and colleagues focussed only on the effectiveness of the PHQ-9 as a tool for diagnosing depression and recommends a clinical cut off score (they report a score of between 8 and 11 will have sufficient sensitivity and specificity for accurate diagnosis). The approach I have used also allows the potential for normative data for different groups to be collected and for a great deal more information about the reliability and validity of a measure to be collected.

5.8. Conclusion

In conclusion, the analysis in this thesis has provided normative data for the HAQ and SHAI. Meta-analysis of correlation coefficients indicated that the SHAI has good construct validity. The average reliability of the SHAI was sufficiently high that the scale can be recommended for use within clinical practice and applied research (Nunnally, 1978). The reliability generalisability analysis also indicated that translated versions of the SHAI may have poorer reliability and future research should investigate which translated versions are

responsible for reducing reliability. There was no evidence for the HAQ to be lacking in reliability. Due to insufficient data, it was not possible to provide any additional information about the construct validity or reliability of the HAQ. When more data regarding both scales is published, future research could repeat and extend the analyses reported in this thesis and provide additional information about the reliability and validity of both measures.

Bibliography

Abraham, A., Kaufmann, C., Redlich, R., Hermann, A., Stark, R., Stevens, S., & Hermann, C. (2013). Self-referential and anxiety-relevant information processing in subclinical social anxiety: an fMRI study. *Brain Imaging and Behavior*, *7*(1), 35–48. http://doi.org/10.1007/s11682-012-9188-x

Abramowitz, J. S., Deacon, B. J., & Valentiner, D. P. (2007). The short health anxiety inventory: Psychometric properties and construct validity in a non-clinical sample. *Cognitive Therapy and Research*, *31*(6), 871–883. http://doi.org/10.1007/s10608-006-9058-1

Abramowitz, J. S., & Moore, E. L. (2007). An experimental analysis of hypochondriasis. *Behaviour Research and Therapy*, *45*(3), 413–424. http://doi.org/10.1016/j.brat.2006.04.005

Abramowitz, J. S., Olatunji, B. O., & Deacon, B. J. (2007). Health anxiety, hypochondriasis, and the anxiety disorders. *Behavior Therapy*, *38*(1), 86–94. http://doi.org/10.1016/j.beth.2006.05.001

Aggarwal, V. R., Macfarlane, G. J., Tajar, A., Mulvey, M. R., Power, A., Ray, D., & McBeth, J. (2014). Functioning of the hypothalamic-pituitary-adrenal and growth hormone axes in frequently unexplained disorders: Results of a population study. *European Journal of Pain*, *18*(3), 447–454. http://doi.org/10.1002/j.1532-2149.2013.00413.x

Alberts, N. M., & Hadjistavropoulos, H. D. (2014). Parental illness, attachment dimensions, and health beliefs: testing the cognitive-behavioural and interpersonal models of health anxiety. *Anxiety Stress and Coping*, *27*(2), 216–228. http://doi.org/10.1080/10615806.2013.835401

Alberts, N. M., Hadjistavropoulos, H. D., Jones, S. L., & Sharpe, D. (2013). The short health anxiety inventory: A systematic review and meta-analysis. *Journal of Anxiety Disorders*, *27*(1), 68–78. http://doi.org/10.1016/j.janxdis.2012.10.009

Alberts, N. M., Sharpe, D., Kehler, M. D., & Hadjistavropoulos, H. D. (2011). Health anxiety: Comparison of the latent structure in medical and non-medical samples. *Journal of Anxiety Disorders*, *25*(4), 612–614. http://doi.org/10.1016/j.janxdis.2011.01.011

American Psychiatric Association. (1986). *Diagnostic and statistical manual of mental disorders* (3rd ed). Washington, DC: Author.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed). Washington, DC: Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed). Arlington, VA: Author.

Annagur, B. B., Avci, A., Demir, K., & Uygur, O. F. (2015). Is there any difference between the early age myocardial infarction and late age myocardial infarction in terms of psychiatric morbidity in patients who have survived acute myocardial infarction? *Comprehensive Psychiatry*, *57*, 10–15. http://doi.org/10.1016/j.comppsych.2014.11.001

Anthoine, E., Moret, L., Regnault, A., Sébille, V., & Hardouin, J.-B. (2014). Sample size used to validate a scale: A review of publications on newly-developed patient reported outcomes measures. *Health and Quality of Life Outcomes*, *12*(2), 176. http://doi.org/10.1186/s12955-014-0176-2

Asbury, E. A., Chandrruangphen, P., & Collins, P. (2006). The importance of

continued exercise participation in quality of life and psychological well-being in previously inactive postmenopausal women: A pilot study. *Menopause*, *13*(4), 561–567. http://doi.org/10.1097/01.gme.0000196812.96128.e8

Asbury, E. A., Creed, F., & Collins, P. (2004). Distinct psychosocial differences between women with coronary heart disease and cardiac syndrome X. *European Heart Journal*, *25*(19), 1695–1701. http://doi.org/10.1016/j.ehj.2004.07.035

Asbury, E. A., Slattery, C., Grant, A., Evans, L., Barbir, M., & Collins, P. (2008). Cardiac rehabilitation for the treatment of women with chest pain and normal coronary arteries. *Menopause: The Journal of the North American Menopause Society*, *15*(3), 454–460. http://doi.org/10.1097/gme.0b013e31815982eb

Asbury, E. A., Webb, C. M., & Collins, P. (2011). Group support to improve psychosocial well-being and primary-care demands among women with cardiac syndrome X. *Climacteric*, *14*(1), 100–104. http://doi.org/10.3109/13697137.2010.499181

Asbury, E. A., Webb, C. M., Probert, H., Wright, C., Barbir, M., Fox, K., & Collins, P. (2012). Cardiac rehabilitation to improve physical functioning in refractory angina: A pilot study. *Cardiology*, *122*(3), 170–177. http://doi.org/10.1159/000339224

Asmundson, G. J. G., Taylor, S., Carleton, R. N., Weeks, J. W., & Hadjstavropoulos, H. D. (2012). Should health anxiety be carved at the joint? A look at the health anxiety construct using factor mixture modeling in a non-clinical sample. *Journal of Anxiety Disorders*, *26*(1), 246–251. http://doi.org/10.1016/j.janxdis.2011.11.009

Aydemir, O., Kirpinar, I., Sati, T., Uykur, B., & Cengisiz, C. (2013). Reliability and Validity of the Turkish Version of the Health Anxiety Inventory. *Archives of Neuropsychiatry*, *50*(4), 325–331. http://doi.org/10.4274/npa.y6383

Bailer, J., Witthoft, M., Wagner, H., Mier, D., Diener, C., & Rist, F. (2014). Childhood maltreatment is associated with depression but not with hypochondriasis in later life. *Journal of Psychosomatic Research*, *77*(2), 104–108. http://doi.org/10.1016/j.jpsychores.2014.06.004

Barkham, M., Hardy, G. E., & Mellor-Clark, J. (2010). *Developing and delivering practice-based evidence: A guide for the psychological therapies*. Chichester: Wiley-Blackwell.

Barnes, L. B., Harp, D., & Woo, S. J. (2002). Reliability generalizability of scores on the Spielberger State-Trait Anxiety Inventory. *Educational and Psychological Measurement*, *62*(4), 603–618.

Barsky, A. J., & Ahern, D. K. (2004a). Cognitive behavior therapy for hypochondriasis: A randomised controlled trial. *JAMA*, *291*(12), 1464–1470. http://doi.org/10.1001/jama.291.12.1464

Barsky, A. J., & Ahern, D. K. (2004b). Cognitive behavior therapy for hypochondriasis - A randomized controlled trial. *Jama-Journal of the American Medical Association*, *291*(12), 1464–1470. http://doi.org/10.1001/jama.291.12.1464

Barsky, A. J., Ahern, D. K., Bauer, M. R., Nolido, N., & Orav, E. J. (2013). A Randomized trial of treatments for high-utilizing somatizing patients. *Journal of General Internal Medicine*, *28*(11), 1396–1404.

http://doi.org/10.1007/s11606-013-2392-6

Barsky, A. J., Coeytaux, R. R., Sarnie, M. K., & Cleary, P. D. (1993). Hypochondriacal patients' beliefs about good health. *American Journal of Psychiatry*, *150*(7), 1085–1089. http://doi.org/10.1176/ajp.150.7.1085

Barsky, A. J., Ettner, S. L., Horsky, J., & Bates, D. W. (2001). Resource utilization of patients with hypochondriacal health anxiety and somatization. *Medical Care*, *39*(7), 705–715. http://doi.org/10.1097/00005650-200107000-00007

Beard, J. R., Officer, A., de Carvalho, I. A., Sadana, R., Pot, A. M., Michel, J.-P., … Chatterji, S. (2015). The world report on ageing and health: A policy framework for healthy ageing. *Lancet*, *6736*(15), 1–10. http://doi.org/10.1016/S0140-6736(15)00516-4

Beck, A. T., & Emery, G. (2005). *Anxiety disorders and phobias: A cognitive perspective*. Cambridge, MA: Basic Books.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *The Beck Depression Inventory-II*. San Antonio, TX: The Psychological Corporation.

Bem, S. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology*, *42*(42), 155–162. http://doi.org/https://doi.org/10.1037/h0036215

Borenstein, M., Hedges, L. V, Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester: John Wiley & Sons. http://doi.org/10.1002/9780470743386.refs

Boschen, M. J., Veale, D., Ellison, N., & Reddell, T. (2013). The emetophobia questionnaire (EmetQ-13): Psychometric validation of a measure of specific phobia of vomiting (emetophobia). *Journal of Anxiety Disorders*, *27*(7), 670–677. http://doi.org/10.1016/j.janxdis.2013.08.004

Boston, A. F., & Merrick, P. L. (2010). Health anxiety among older people: An exploratory study of health anxiety and safety behaviors in a cohort of older adults in New Zealand. *International Psychogeriatrics*, *22*(4), 549–558. http://doi.org/10.1017/s1041610209991712

Bot, A. G. J., Anderson, J. A., Neuhaus, V., & Ring, D. (2013). Factors associated with survey response in hand surgery research. *Clinical Orthopaedics and Related Research*, *471*(10), 3237–3242. http://doi.org/10.1007/s11999-013-3126-0

Botella, J., Suero, M., & Gambara, H. (2010). Psychometric inferences from a meta-analysis of reliability and internal consistency coefficients. *Psychological Methods*, *15*(4), 386–397. http://doi.org/10.1037/a0019626

Bowden, J., Tierney, J. F., Copas, A. J., & Burdett, S. (2011). Quantifying, displaying and accounting for heterogeneity in the meta-analysis of RCTs using standard and generalised Q statistics. *BMC Medical Research Methodology*, *11*, 41. http://doi.org/10.1186/1471-2288-11-41

Brady, R. E., Cisler, J. M., & Lohr, J. M. (2014). Specific and differential prediction of health anxiety by disgust sensitivity and propensity. *Anxiety Stress and Coping*, *27*(1), 90–99. http://doi.org/10.1080/10615806.2013.772588

Brown, H. D., Kosslyn, S. M., Delamater, B., Fama, J., & Barsky, A. J. (1999). Perceptual and memory biases for health-related information in hypochondriacal individuals. *Journal of Psychosomatic Research*, *47*(1), 67–78. http://doi.org/10.1016/S0022-3999(99)00011-2

Brown, R. J., Skehan, D., Chapman, A., Perry, E. P., McKenzie, K. J., Lloyd, D.

M., … Poliakoff, E. (2012). Physical symptom reporting is associated with a tendency to experience somatosensory distortion. *Psychosomatic Medicine*, *74*(6), 648–655. http://doi.org/10.1097/PSY.0b013e3182595358

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105. http://doi.org/10.1037/h0046016

Christensen, A. J., Moran, P. J., & Wiebe, J. S. (1999). Assessment of irrational health beliefs: Relation to health practices and medical regimen adherence. *Health Psychology*, *18*(2), 169–176. http://doi.org/10.1037/0278-6133.18.3.220

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284–290. http://doi.org/10.1037/1040-3590.6.4.284

Claude, J. A. J., Hadjistavropoulos, H. D., & Friesen, L. (2014). Exploration of health anxiety among individuals with diabetes: Prevalence and implications. *Journal of Health Psychology*, *19*(2), 312–322. http://doi.org/10.1177/1359105312470157

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Lawrence Erlbaum Associates. http://doi.org/10.1234/12345678

Conradt, M., Cavanagh, M., Franklin, J., & Rief, W. (2006). Dimensionality of the Whiteley Index: Assessment of hypochondriasis in an Australian sample of primary care patients. *Journal of Psychosomatic Research*, *60*(2), 137–143. http://doi.org/10.1016/j.jpsychores.2005.07.003

Cousineau, D., & Chartier, S. (2010). Outliers detection and treatment: A review. *International Journal of Psychological Research*, *3*(1), 58–67.

Creed, F., & Barsky, A. (2004). A systematic review of the epidemiology of somatisation disorder and hypochondriasis disorder and hypochondriasis. *Journal of Psychosomatic Research*, *56*, 391–408. http://doi.org/10.1016/S0022-3999(03)00622-6

Cronbach, L. J. (1947). Test "reliability": Its meaning and determination. *Psychometrika*, *12*(1), 1–16. http://doi.org/10.1007/BF02289289

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. http://doi.org/10.1007/BF02310555

Cronbach, L. J., & Shavelson, R. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement*, *64*(3), 391–418. http://doi.org/10.1177/0013164404266386

Dalton, D. R., Aguinis, H., Dalton, C. M., Bosco, F. A., & Pierce, C. A. (2012). Revisiting the file drawer problem in meta-analysis: An assessment of published and nonpublished correlation matrices. *Personnel Psychology*, *65*(2), 221–249. http://doi.org/10.1111/j.1744-6570.2012.01243.x

Deacon, B., & Abramowitz, J. (2006). Anxiety sensitivity and its dimensions across the anxiety disorders. *Journal of Anxiety Disorders*, *20*(7), 837–857. http://doi.org/10.1016/j.janxdis.2006.01.003

Deacon, B., & Abramowitz, J. S. (2008). Is hypochondriasis related to obsessive compulsive-disorder, panic disorder, or both? An empirical evaluation. *Journal of Cognitive Psychotherapy*, *22*(2), 115–127. http://doi.org/10.1891/0889-8391.22.2.115

DeMarinis, V., Barsky, A. J., Antin, J. H., & Chang, G. (2009). Health

psychology and distress after haematopoietic stem cell transplantation. *European Journal of Cancer Care*, *18*(1), 57–63. http://doi.org/10.1111/j.1365-2354.2008.00931.x

Dickersin, K. (1997). How important is publication bias? A synthesis of available data. *AIDS Education and Prevention*, *9*(1 Suppl), 15–21.

Dolin, S. J., Williams, A. C. D., Ashford, N., George, J., Pereira, L., & Perello, A. (2003). Factors affecting medical decision-making in patients with osteoarthritis of the hip: allocation of surgical priority. *Disability and Rehabilitation*, *25*(14), 771–777. http://doi.org/10.1080/0963828031000090551

Dugas, M., Gagnon, F., Ladoucer, R., & Freeston, M. H. (1998). Generalized anxiety disorder: A preliminary test of a conceptual model. *Behaviour Research and Therapy*, *36*, 215–226.

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*(3), 399–412. http://doi.org/10.1111/bjop.12046

Dzhambov, A. M., & Dimitrova, D. D. (2014). Elderly visitors of an urban park, health anxiety and individual awareness of nature experiences. *Urban Forestry & Urban Greening*, *13*(4), 806–813. http://doi.org/10.1016/j.ufug.2014.05.006

Eastin, M. S., & Guinsler, N. M. (2006). Worried and wired: Effects of health anxiety on information-seeking and health care utilization behaviors. *Cyberpsychology & Behavior*, *9*(4), 494–498. http://doi.org/10.1089/cpb.2006.9.494

Egger, M., Schneider, M., & Smith, G. D. (1998). Spurious precision? Meta-analysis of observational studies. *BMJ*, *316*, 140–144.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*, 629–634.

Esfandyari, T., & Harewood, G. C. (2007). Value of a negative colonoscopy in patients with non-specific gastrointestinal symptoms. *Journal of Gastroenterology and Hepatology*, *22*(10), 1609–1614. http://doi.org/10.1111/j.1440-1746.2006.04753.x

Eysenck, H. (1978). An exercise in mega-silliness. *American Psychologist*, *33*(5), 517.

Fan, Q. Q., & Olatunji, B. O. (2013). Individual differences in disgust sensitivity and health-related avoidance: Examination of specific associations. *Personality and Individual Differences*, *55*(5), 454–458. http://doi.org/10.1016/j.paid.2013.04.007

Farmer, A. J., Doll, H., Levy, J. C., & Salkovskis, P. M. (2003). The impact of screening for Type 2 diabetes in siblings of patients with established diabetes. *Diabetic Medicine*, *20*, 996–1004. http://doi.org/DOI: 10.1046/j.1464–5491.2003.01042.x

Fergus, T. A. (2013a). Cyberchondria and intolerance of uncertainty: Examining when individuals experience health anxiety in response to internet searches for medical information. *Cyberpsychology Behavior and Social Networking*, *16*(10), 735–739. http://doi.org/10.1089/cyber.2012.0671

Fergus, T. A. (2013b). Repetitive thought and health anxiety: Tests of specificity. *Journal of Psychopathology and Behavioral Assessment*, *35*(3), 366–374. http://doi.org/10.1007/s10862-013-9340-y

Fergus, T. A. (2014a). Health-related dysfunctional beliefs and health anxiety: Further evidence of cognitive specificity. *Journal of Clinical Psychology*, *70*(3), 248–259.

Fergus, T. A. (2014b). The Cyberchondria Severity Scale (CSS): An examination of structure and relations with health anxiety in a community sample. *Journal of Anxiety Disorders*, *28*(6), 504–510. http://doi.org/10.1016/j.janxdis.2014.05.006

Fergus, T. A., & Russell, L. H. (2016). Does cyberchondria overlap with health anxiety and obsessive-compulsive symptoms? An examination of latent structure and scale interrelations. *Journal of Anxiety Disorders*, *38*, 88–94. http://doi.org/10.1016/j.janxdis.2016.01.009

Fergus, T. A., & Valentiner, D. P. (2011). Intolerance of Uncertainty Moderates the Relationship Between Catastrophic Health Appraisals and Health Anxiety. *Cognitive Therapy and Research*, *35*(6), 560–565. http://doi.org/10.1007/s10608-011-9392-9

Ferguson, E. (2009). A taxometric analysis of health anxiety. *Psychological Medicine*, *39*, 277–285.

Ferguson, E., Moghaddam, N. G., & Bibby, P. A. (2007). Memory bias in health anxiety is related to the emotional valence of health-related words. *Journal of Psychosomatic Research*, *62*(3), 263–274. http://doi.org/10.1016/j.jpsychores.2007.01.015

Fetzner, M. G., Asmundson, G. J. G., Carey, C., Thibodeau, M. A., Brandt, C., Zvolensky, M. J., & Carleton, R. N. (2014). How do elements of a reduced capacity to withstand uncertainty relate to the severity of health anxiety? *Cognitive Behaviour Therapy*, *43*(3), 262–274. http://doi.org/10.1080/16506073.2014.929170

Fiddler, M., Jackson, J., Kapur, N., Wells, A., & Creed, F. (2004). Childhood adversity and frequent medical consultations. *General Hospital Psychiatry*, *26*(5), 367–377. http://doi.org/10.1016/j.genhosppsych.2004.04.001

Field, A. P. (2003). The problem in using fixed-effects models of meta-analysis on real-world data. *Understanding Statistics*. http://doi.org/10.1207/S15328031US0202_02

Field, A. P. (2005). Is the meta-analysis of correlation coefficients accurate when population correlations vary? *Psychological Methods*, *10*(4), 444–467. http://doi.org/10.1037/1082-989X.10.4.444

Freeston, M. H., Rheaume, J., Letarte, H., M, D., & Ladoucer, R. (1994). Why do people worry? *Personality and Individual Differences*, *17*(6), 791–802.

Fulton, J. J., Marcus, D. K., & Merkey, T. (2011). Irrational health beliefs and health anxiety. *Journal of Clinical Psychology*, *67*(6), 527–538. http://doi.org/10.1002/jclp.20769

George, D., & Mallery, P. (2003). *SPSS for windows step by step: A simple guide and reference, 11.0 update*. Boston, MA: Allyn & Bacon. http://doi.org/10.1177/1073191112446654

Gerolimatos, L. A., & Edelstein, B. A. (2012a). Anxiety-related constructs mediate the relation between age and health anxiety. *Aging & Mental Health*, *16*(8), 975–982. http://doi.org/10.1080/13607863.2012.688192

Gerolimatos, L. A., & Edelstein, B. A. (2012b). Predictors of health anxiety among older and young adults. *International Psychogeriatrics*, *24*(12), 1998–2008. http://doi.org/10.1017/s1041610212001329

Glass, G. V. (1976). Primary, secondary, and meta-analysis research.

Educational Researcher, *5*, 3–8.

Gluud, L. L. (2006). Bias in clinical intervention research. *American Journal of Epidemiology*, *163*(6), 493–501. http://doi.org/10.1093/aje/kwj069

Goetz, A. R., Lee, H. J., & Cougle, J. R. (2013). The association between health anxiety and disgust reactions in a contamination-based behavioral approach task. *Anxiety Stress and Coping*, *26*(4), 431–446. http://doi.org/10.1080/10615806.2012.684241

Goodwin, L., Fairclough, S. H., & Poole, H. M. (2013). A cognitive-perceptual model of symptom perception in males and females: The roles of negative affect, selective attention, health anxiety and psychological job demands. *Journal of Health Psychology*, *18*(6), 848–857. http://doi.org/10.1177/1359105312456321

Graff, L. A., Walker, J. R., Lix, L., Clara, I., Rawsthorne, P., Rogala, L., … Bernstein, C. N. (2006). The relationship of inflammatory bowel disease type and activity to psychological functioning and quality of life. *Clinical Gastroenterology and Hepatology*, *4*(12), 1491–1501. http://doi.org/10.1016/j.cgh.2006.09.027

Greenland, S. (1994). Can meta-analysis be salvaged? *American Journal of Epidemiology*, *140*(9), 783–787.

Greeven, A., Spinhoven, P., & van Balkom, A. J. L. M. A. (2009). Hypochondriasis Y-BOCS: A study of the psychometric properties of a clinician-administered semi-structured interview to assess hypochondriacal thoughts and behaviours. *Clinical Psychology & Psychotherapy*, *16*(5), 431–443. http://doi.org/10.1002/cpp.634

Greeven, A., van Balkom, A. J. L. M., van Rood, Y. R., van Oppen, P., & Spinhoven, P. (2006). The boundary between hypochondriasis and obsessive-compulsive disorder: A cross-sectional study from the Netherlands. *The Journal of Clinical Psychiatry*, *67*(11), 1682–1689.

Gropalis, M., Bleichhardt, G., Hiller, W., & Witthoft, M. (2013). Specificity and modifiability of cognitive biases in hypochondriasis. *Journal of Consulting and Clinical Psychology*, *81*(3), 558–565. http://doi.org/10.1037/a0028493

Hadjistavropoulos, H. D., Janzen, J. A., Kehler, M. D., Leclerc, J. A., Sharpe, D., & Bourgault-Fagnou, M. D. (2012). Core cognitions related to health anxiety in self-reported medical and non-medical samples. *Journal of Behavioral Medicine*, *35*(2), 167–178. http://doi.org/10.1007/s10865-011-9339-3

Haenen, M. A., De Jong, P. J., Schmidt, A. J. M., Stevens, S., & Visser, L. (2000). Hypochondriacs' estimation of negative outcomes: Domain-specificity and responsiveness to reassuring and alarming information. *Behaviour Research and Therapy*, *38*(8), 819–833. http://doi.org/10.1016/S0005-7967(99)00128-X

Hanel, P. H. P., Vione, K. C., Simonsohn, U., Nosek, B., Flint, J., & Robinson, E. (2016). Do student samples provide an accurate estimate of the general public? *PLOS ONE*, *11*(12), e0168354. http://doi.org/10.1371/journal.pone.0168354

Hedges, L. (1992). Modeling publication selection effects in meta-analysis on JSTOR. *Statistical Science*, *7*(2), 246–255.

Hedges, L. V, & Vevea, J. L. (1998). Fixed-and random-effects models in meta-analysis. *Psychological Methods*, *3*(4), 486–504. http://doi.org/10.1037/1082-989X.3.4.486

Hedman, E., Lindefors, N., Andersson, G., Andersson, E., Lekander, M., Ruck, C., & Ljotsson, B. (2013). Predictors of outcome in Internet-based cognitive behavior therapy for severe health anxiety. *Behaviour Research and Therapy*, *51*(10), 711–717. http://doi.org/10.1016/j.brat.2013.07.009

Henrich, J., Heine, S., & Norenzayan, A. (2010). The weirdest people in the world. *Behavioral and Brain Sciences*, *33*, 61–135.

Hewison, J., Nixon, J., Fountain, J., Cocks, K., Jones, C., Mason, G., … Thornton, J. (2006). Amniocentesis results: Investigation of anxiety. The ARIA trial. *Health Technology Assessment*, *10*(50).

Higgins, J. P. T., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., … Sterne, J. A. C. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*, *343*(d5928).

Hiller, W., Rief, W., & Fichter, M. M. (2002). Dimensional and categorical approaches to hypochondriasis. *Psychological Medicine*, *32*(4), 707–718. http://doi.org/10.1017/S0033291702005524

Hinkin, T., Tracey, J., & Enz, C. (1997). Scale construction: Developing reliable and valid measurement instruments. *Journal of Hospitality and Tourism Research*, *21*(1), 100–120.

Hinz, A., Rief, W., & Brahler, E. (2003). Hypochondriasis in the general population: Psychometric properties and norm values of the Whiteley Index. *Diagnostica*, *49*(1), 34–42. http://doi.org/10.1026//0012-1924.49.1.34

Hogan, T. P., Benjamin, A., & Brezinski, K. L. (2000). Reliability methods: A note on the frequency of use of various types. *Educational and Psychological Measurement*, *60*(4), 523–531. http://doi.org/10.1177/00131640021970691

Hollo, G., Kothy, P., Garas, A., Geczy, A., & Vargha, P. (2012). Nonpopulation-based glaucoma screening exercise in an osteoporosis patient organization: Can elderly persons with advanced health awareness identify their risk for glaucoma? *Journal of Glaucoma*, *21*(8), 510–515. http://doi.org/10.1097/IJG.0b013e3182208a48

Hollo, G., Kothy, P., Geczy, A., & Vargha, P. (2010). Health anxiety in a non-population-based, pre-publicised glaucoma screening exercise. *Eye*, *24*(4), 699–705. http://doi.org/10.1038/eye.2009.131

Howell, R. T., & Shields, A. L. (2008). The file drawer problem in reliability generalization: A strategy to compute a fail-safe N with reliability coefficients. *Educational and Psychological Measurement*, *68*(1), 120–128. http://doi.org/10.1177/0013164407301528

Hunter, J. E., & Schmidt, F. L. (2000). Fixed effects vs. random effects meta-analysis models: Implications for cumulative research knowledge. *International Journal of Selection and Assessment*, *8*(4), 275–292. http://doi.org/10.1111/1468-2389.00156

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed). London, UK: Sage.

Hunter, J. E., & Schmidt, F. L. (2014). *Methods of meta-analysis: Correcting error and bias in research findings* (3rd ed). London, UK: Sage.

Inamura, K., Shinagawa, S., Nagata, T., Tagai, K., Nukariya, K., & Nakayama, K. (2015). Cognitive dysfunction in patients with late-life somatic symptom disorder: A comparison according to disease severity. *Psychosomatics*, *56*(5), 486–494.

Ioannidis, J. P. (2016). The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. *The Milbank Quarterly*, *94*(3), 485–514. http://doi.org/10.1111/1468-0009.12210

Iyengar, S., & Greenhouse, J. (1988). Selection models and the file drawer problem. *Statistical Science*, *3*(1), 109–117.

Jackson, J., Kincey, J., Fiddler, M., Creed, F., & Tomenson, B. (2004). Differences between out-patients with physical disease and those with medically unexplained symptoms with respect to patient satisfaction, emotional distress and illness perception. *British Journal of Health Psychology*, *9*, 433–446. http://doi.org/10.1348/1359107042304597

Jacobson, N. S., Follette, W. C., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy*, *15*(4), 336–352. http://doi.org/10.1016/S0005-7894(84)80002-7

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*(1), 12–19. http://doi.org/10.1037/0022-006X.59.1.12

Jacoby, R. J., Wheaton, M. G., & Abramowitz, J. S. (2016). Attentional biases in illness anxiety: Null findings from the dot probe paradigm. *Journal of Obsessive-Compulsive and Related Disorders*, *10*, 19–25. http://doi.org/10.1016/j.jocrd.2016.04.007

Jasper, F., & Witthoft, M. (2011). Health anxiety and attentional bias: The time course of vigilance and avoidance in light of pictorial illness information. *Journal of Anxiety Disorders*, *25*(8), 1131–1138. http://doi.org/10.1016/j.janxdis.2011.08.004

Jeffers, A. J., Benotsch, E. G., Green, B. A., Bannerman, D., Darby, M., Kelley, T., & Martin, A. M. (2015). Health anxiety and the non-medical use of prescription drugs in young adults: A cross-sectional study. *Addictive Behaviors*, *50*, 74–77. http://doi.org/10.1016/j.addbeh.2015.06.012

Jones, S. L., Hadjistavropoulos, H. D., & Gullickson, K. (2014). Understanding health anxiety following breast cancer diagnosis. *Psychology Health & Medicine*, *19*(5), 525–535. http://doi.org/10.1080/13548506.2013.845300

Karademas, E. C., Christopoulou, S., Dimostheni, A., & Pavlu, F. (2008). Health anxiety and cognitive interference: Evidence from the application of a modified Stroop task in two studies. *Personality and Individual Differences*, *44*(5), 1138–1150. http://doi.org/10.1016/j.paid.2007.11.007

Katon, W., Lin, E. H. B., & Kroenke, K. (2007). The association of depression and anxiety with medical symptom burden in patients with chronic medical illness. *General Hospital Psychiatry*, *29*(2), 147–155. http://doi.org/10.1016/j.genhosppsych.2006.11.005

Kaur, A., Butow, P. N., & Sharpe, L. (2013). Health threat increases attentional bias for negative stimuli. *Journal of Behavior Therapy and Experimental Psychiatry*, *44*(4), 469–476. http://doi.org/10.1016/j.jbtep.2013.06.003

Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings*. New York, NY: Oxford University Press.

Kehler, M. D., & Hadjistavropoulos, H. D. (2009). Is health anxiety a significant problem for individuals with multiple sclerosis? *Journal of Behavioral Medicine*, *32*(2), 150–161. http://doi.org/10.1007/s10865-008-9186-z

Kellner, R. (1987). *Illness Attitude Scales: Abridged manual*. Albuquerque, NM:

University of New Mexico.

Kellner, R., Abbott, P., Winslow, W. W., & Pathak, D. (1987). Fears, beliefs and attitudes in DSM-III hypochondriasis. *Jounal of Nervous and Mental Disease, 175*, 20–25.

Kerstner, T., Witthoft, M., Mier, D., Diener, C., Rist, F., & Bailer, J. (2015). A diary-based modification of symptom attributions in pathological health anxiety: Effects on symptom report and cognitive biases. *Journal of Consulting and Clinical Psychology, 83*(3), 578–589. http://doi.org/10.1037/a0039056

Kirby, S. E., & Yardley, L. (2009). The contribution of symptoms of posttraumatic stress disorder, health anxiety and intolerance of ncertainty to distress in Meniere's disease. *Journal of Nervous and Mental Disease, 197*(5), 324–329. http://doi.org/10.1097/NMD.0b013e3181a20866

Kobori, O., Okita, M., Shiraishi, T., Hasegawa, T., & Iyo, M. (2014). Health anxiety and healthcare costs in Japanese individuals: An Internet survey. *Health Psychology and Behavioural Medicine, 2*(1), 833–840. http://doi.org/10.1080/21642850.2014.945935

Koteles, F., Freyler, A., Kokonyei, G., & Bardos, G. (2015). Family background of modern health worries, somatosensory amplification, and health anxiety: A questionnaire study. *Journal of Health Psychology, 20*(12), 1549–1557. http://doi.org/10.1177/1359105313516661

Kowalyk, K. M., & Hadjis-Tavropoulos, H. D. (2007). Measuring health anxiety during pregnancy: Examination of the short version of the health anxiety inventory. *Depression and Anxiety, 24*(4), 292–293.

Kowalyk, K. M., Hadjistavropoulos, H. D., & Jones, S. L. (2009). What impact does pregnancy have on anxiety about health? *Journal of Psychosomatic Obstetrics and Gynecology, 30*(4), 223–230. http://doi.org/10.3109/01674820903276453

Kroenke, K., & Spitzer, R. L. (2002). The PHQ-9: A new depression diagnostic and severity measure. *Psychiatric Annals, 32*(9), 509–515. http://doi.org/10.3928/0048-5713-20020901-06

Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria: What did they really say? *Organizational Research Methods, 9*(2), 202–220. http://doi.org/10.1177/1094428105284919

Lecci, L., & Cohen, D. (2007). Altered processing of health threat words as a function of hypochondriacal tendencies and experimentally manipulated control beliefs. *Cognition & Emotion, 21*(1), 211–224. http://doi.org/10.1080/02699930600625248

Lecci, L., & Cohen, D. J. (2002). Perceptual consequences of an illness-concern induction and its relation to hypochondriacal tendencies. *Health Psychology, 21*(2), 147–156. http://doi.org/10.1037//0278-6133.21.2.147

Lee, H. J., Goetz, A. R., Turkel, J. E., & Siwiec, S. G. (2015). Computerized attention retraining for individuals with elevated health anxiety. *Anxiety Stress and Coping, 28*(2), 226–237. http://doi.org/10.1080/10615806.2014.918964

Lee, J. E., Watson, D., & Frey-Law, L. A. (2013). Psychological factors predict local and referred experimental muscle pain: A cluster analysis in healthy adults. *European Journal of Pain, 17*(6), 903–915. http://doi.org/10.1002/j.1532-2149.2012.00249.x

Longley, S. L., Broman-Fulks, J. J., Calamari, J. E., Noyes, R., Wade, M., & Orlando, C. M. (2010). A taxometric study of hypochondriasis symptoms. *Behavior Therapy*, *41*(4), 505–514.

Longley, S. L., Meyers, K., Maxwell, R., & Letizia, B. (2014). Assessment of hypochondriasis and health anxiety. In *Hypochondriasis and health anxiety: A guide for clinicians* (pp. 65–84). New York, NY: Oxford University Press.

Longley, S. L., Watson, D., & Noyes, R. (2005). Assessment of the hypochondriasis domain: The Multidimensional Inventory of Hypochondriacal Traits (MIHT). *Psychological Assessment*, *17*(1), 3–14. http://doi.org/10.1037/1040-3590.17.1.3

Lopez-Lopez, J. A., Marin-Martinez, F., Sanchez-Meca, J., Noortgate, W. Van Den, & Viechtbauer, W. (2014). Estimation of the predictive power of the model in mixed-effects meta-regression: A simulation study. *British Journal of Mathematical and Statistical Psychology*, *67*, 30–48. http://doi.org/10.1111/bmsp.12002

Lord, F. M., & Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.

Lovas, D. A., & Barsky, A. J. (2010). Mindfulness-based cognitive therapy for hypochondriasis, or severe health anxiety: A pilot study. *Journal of Anxiety Disorders*, *24*(8), 931–935. http://doi.org/10.1016/j.janxdis.2010.06.019

Lucock, M., & Morley, S. (1996). The Health Anxiety Questionnaire. *British Journal of Health Psychology*, *1*, 137–150.

Lucock, M. P., Morley, S., White, C., & Peake, M. D. (1997). Responses of consecutive patients to reassurance after gastroscopy: results of self administered questionnaire survey. *British Medical Journal*, *315*(7108), 572–575.

Macatee, R. J., & Cougle, J. R. (2013). The roles of emotional reactivity and tolerance in generalized, social, and health anxiety: A multimethod exploration. *Behavior Therapy*, *44*(1), 39–50.

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. http://doi.org/10.1016/j.genhosppsych.2014.09.009

Marcus, D. K. (1999). The cognitive-behavioral model of hypochondriasis: Misinformation and triggers. *Journal of Psychosomatic Research*, *47*(1), 79–91. http://doi.org/10.1016/S0022-3999(99)00008-2

Marcus, D. K., & Church, S. E. (2003). Are dysfunctional beliefs about illness unique to hypochondriasis? *Journal of Psychosomatic Research*, *54*(6), 543–547. http://doi.org/10.1016/S0022-3999(02)00526-3

Marcus, D. K., Gurley, J. R., Marchi, M. M., & Bauer, C. (2007). Cognitive and perceptual variables in hypochondriasis and health anxiety. *Clinical Psychology Review*, *27*, 127–139.

Marcus, D. K., Hughes, K. T., & Arnau, R. C. (2008). Health anxiety, rumination, and negative affect: A mediational analysis. *Journal of Psychosomatic Research*, *64*(5), 495–501. http://doi.org/10.1016/j.jpsychores.2008.02.004

McManus, F., Surawy, C., Muse, K., Vazquez-Montes, M., & Williams, J. M. G. (2012). A randomized clinical trial of mindfulness-based cognitive therapy versus unrestricted services for health anxiety (hypochondriasis). *Journal of*

Consulting and Clinical Psychology, *80*(5), 817–828.
http://doi.org/10.1037/a0028782

Meana, M., & Lykins, A. (2009). Negative affect and somatically focused anxiety in young women reporting pain with intercourse. *Journal of Sex Research*, *46*(1), 80–88. http://doi.org/10.1080/00224490802624422

Meechan, G. T., Collins, J. P., Moss-Morris, R. E., & Petri, K. J. (2005). Who is not reassured following benign diagnosis of breast symptoms? *Psycho-Oncology*, *14*(3), 239–246. http://doi.org/10.1002/pon.841

Melli, G., Carraresi, C., Poli, A., & Bailey, R. (2016). The role of metacognitive beliefs in health anxiety. *Personality and Individual Differences*, *89*, 80–85. http://doi.org/10.1016/j.paid.2015.10.006

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ*, *339*, b2535.

Moher, D., Tetzlaff, J., Tricco, A. C., Sampson, M., & Altman, D. G. (2007). Epidemiology and reporting characteristics of systematic reviews. *PLoS Medicine*. http://doi.org/10.1371/journal.pmed.0040078

Morales, A., Espada, J. P., Carballo, J. L., Piqueras, J. A., & Orgiles, M. (2015). Short Health Anxiety Inventory: Factor structure and psychometric properties in Spanish adolescents. *Journal of Health Psychology*, *20*(2), 123–131. http://doi.org/10.1177/1359105313500095

Morley, S., & Dowzer, C. (2014). *Manual for the Leeds Reliable Change Indicator: Simple Excel® applications for the analysis of individual patient and group data*. Leeds, UK: University of Leeds.

Muse, K., McManus, F., Hackmann, A., & Williams, M. (2010). Intrusive imagery in severe health anxiety: Prevalence, nature and links with memories and maintenance cycles. *Behaviour Research and Therapy*, *48*(8), 792–798. http://doi.org/10.1016/j.brat.2010.05.008

Muse, K., McManus, F., Leung, C., Meghreblian, B., & Williams, J. M. G. (2012). Cyberchondriasis: Fact or fiction? A preliminary examination of the relationship between health anxiety and searching for health information on the Internet. *Journal of Anxiety Disorders*, *26*(1), 189–196. http://doi.org/10.1016/j.janxdis.2011.11.005

National Institute of Health. (2014). Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies. Retrieved May 14, 2017, from https://www.nhlbi.nih.gov/health-pro/guidelines/in-develop/cardiovascular-risk-reduction/tools/cohort

Norr, A. M., Albanese, B. J., Oglesby, M. E., Allan, N. P., & Schmidt, N. B. (2014). Anxiety sensitivity and intolerance of uncertainty as potential risk factors for cyberchondria. *Journal of Affective Disorders*, *174*, 64–69. http://doi.org/10.1016/j.jad.2014.11.023

Norr, A. M., Albanese, B. J., Oglesby, M. E., Allan, N. P., & Schmidt, N. B. (2015). Anxiety sensitivity and intolerance of uncertainty as potential risk factors for cyberchondria. *Journal of Affective Disorders*, *174*, 64–69. http://doi.org/10.1016/j.jad.2014.11.023

Norr, A. M., Allan, N. P., Boffa, J. W., Raines, A. M., & Schmidt, N. B. (2015). Validation of the cyberchondria severity scale (CSS): Replication and extension with bifactor modeling. *Journal of Anxiety Disorders*, *31*, 58–64. http://doi.org/10.1016/j.janxdis.2015.02.001

Norr, A. M., Capron, D. W., & Schmidt, N. B. (2014). Medical information

seeking: Impact on risk for anxiety psychopathology. *Journal of Behavior Therapy and Experimental Psychiatry*, *45*(3), 402–407. http://doi.org/10.1016/j.jbtep.2014.04.003

Norr, A. M., Oglesby, M. E., Raines, A. M., Macatee, R. J., Allan, N. P., & Schmidt, N. B. (2015). Relationships between cyberchondria and obsessive-compulsive symptom dimensions. *Psychiatry Research*, *230*(2), 441–446. http://doi.org/10.1016/j.psychres.2015.09.034

Noyes, R. (2001a). Epidemiology of hypochondriasis. In *Hypochondriasis: Modern perspectives on an ancient malady* (pp. 127–154). New York, NY: Oxford University Press.

Noyes, R. (2001b). Hypochondriasis: Boundaries and comorbidities. In *Health anxiety: Clinical perspectives on hypochondriasis and related conditions* (pp. 132–160). Chichester, UK: Wiley.

Noyes, R., Stuart, S., Longley, S. L., Langbehn, D. R., & Happel, R. L. (2002). Hypochondriasis and fear of death. *Journal of Nervous and Mental Disease*, *190*(8), 503–509. http://doi.org/10.1097/01.nmd.0000026619.27653.b4

Nunnally, J. C. (1978). *Psychometric Theory*. New York, NY: McGraw Hill.

Olatunji, B., Kauffman, B. Y., Meltzer, S., Davis, M. L., Smits, J. A. J., & Powers, M. B. (2014). Cognitive-behavioral therapy for hypochondriasis/health anxiety: A meta-analysis of treatment outcome and moderators. *Behaviour Research and Therapy*, *58*, 65–74. http://doi.org/10.1016/j.brat.2014.05.002

Olatunji, B. O. (2009). Incremental specificity of disgust propensity and sensitivity in the prediction of health anxiety dimensions. *Journal of Behavior Therapy and Experimental Psychiatry*, *40*(2), 230–239. http://doi.org/10.1016/j.jbtep.2008.10.003

Olatunji, B. O., Deacon, B. J., Abramowitz, J. S., & Valentiner, D. P. (2007). Body vigilance in nonclinical and anxiety disorder samples: Structure, correlates, and prediction of health concerns. *Behavior Therapy*, *38*(4), 392–401. http://doi.org/10.1016/j.beth.2006.09.002

Olatunji, B. O., Ebesutani, C., Haidt, J., & Sawchuk, C. N. (2014). Specificity of disgust domains in the prediction of contamination anxiety and avoidance: A multimodal examination. *Behavior Therapy*, *45*(4), 469–481.

Olatunji, B. O., Wolitzky-Taylor, K. B., Elwood, L., Connolly, K., Gonzales, B., & Armstrong, T. (2009). Anxiety sensitivity and health anxiety in a nonclinical sample: Specificity and prospective relations with clinical stress. *Cognitive Therapy and Research*, *33*(4), 416–424. http://doi.org/10.1007/s10608-008-9188-8

Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics*, *8*(2), 157–159. http://doi.org/10.2307/1164923

Owens, K. M. B., Asmundson, G. J. G., Hadjistavropoulos, T., & Owens, T. J. (2004). Attentional bias toward illness threat in individuals with elevated health anxiety. *Cognitive Therapy and Research*, *28*(1), 57–66. http://doi.org/10.1023/B:COTR.0000016930.85884.29

Page, M. J., Shamseer, L., Altman, D. G., Tetzlaff, J., Sampson, M., Tricco, A. C., … Moher, D. (2016). Epidemiology and reporting characteristics of systematic reviews of biomedical research: A cross-sectional study. *PLOS Medicine*, *13*(5), e1002028. http://doi.org/10.1371/journal.pmed.1002028

Pannekoek, J. N., & Stein, D. J. (2014). Diagnosis and classification of

hypochondriasis. In *Hypochondriasis and health anxiety: A guide for clinicians* (pp. 28–38). New York, NY: Oxford University Press.

Pauli, P., & Alpers, G. W. (2002). Memory bias in patients with hypochondriasis and somatoform pain disorder. *Journal of Psychosomatic Research, 52*(1), 45–53. http://doi.org/10.1016/S0022-3999(01)00295-1

Payne, K. A., Binik, Y. M., Pukall, C. F., Thaler, L., Amsel, R., & Khalife, S. (2007). Effects of sexual arousal on genital and non-genital sensation: A comparison of women with Vulvar Vestibulitis Syndrome and healthy controls. *Archives of Sexual Behavior, 36*(2), 289–300. http://doi.org/10.1007/s10508-006-9089-4

Peters, G. J. Y. (2014). The alpha and the omega of scale reliability and validity: Why and how to abandon Cronbach's alpha and the route towards more comprehensive assessment of scale quality. *European Health Psychologist, 16*(2), 56–69.

Peterson, R., & Reiss, S. (1987). *Anxiety sensitivity index manual*. Palos Heights, IL: International Diagnostic Systems.

Pilowsky, I. (1967). Dimensions of hypochondriasis. *British Journal of Psychiatry, 113*, 572–575.

Popper, K. (1972). *The logic of scientific discoveries*. Tiptree, UK: Anchor Press.

Porritt, J. M., Sufi, F., Barlow, A. P. S., & Baker, S. R. (2014). The role of illness beliefs and coping in the adjustment to dentine hypersensitivity. *Journal of Clinical Periodontology, 41*(1), 203–221. http://doi.org/10.1111/jcpe.12177

Priebe, S., Fakhoury, W. K. H., & Henningsen, P. (2008). Functional incapacity and physical and psychological symptoms: How they interconnect in chronic fatigue syndrome. *Psychopathology, 41*(6), 339–345. http://doi.org/10.1159/000152375

Roberts-Thomson, I., Youngchaiyud, U., Whittingham, S., & Mackay, I. (1974). Ageing, immune response, and mortality. *The Lancet, 304*(7877), 368–370. http://doi.org/10.1016/S0140-6736(74)91755-3

Roberts, K. E., Hart, T. A., & Eastwood, J. D. (2010). Attentional biases to social and health threat words in individuals with and without high social anxiety or depression. *Cognitive Therapy and Research, 34*(4), 388–399. http://doi.org/10.1007/s10608-009-9245-y

Robertson, N., Javed, N., Samani, N. J., & Khunti, K. (2008). Psychological morbidity and illness appraisals of patients with cardiac and non-cardiac chest pain attending a rapid access chest pain clinic: A longitudinal cohort study. *Heart, 94*(3), 5. http://doi.org/10.1136/hrt.2006.100537

Rode, S., Salkovskis, P., Dowd, H., & Hanna, M. (2006). Health anxiety levels in chronic pain clinic attenders. *Journal of Psychosomatic Research, 60*(2), 155–161. http://doi.org/10.1016/j.jpsychores.2005.07.005

Rodriguez, M. C., & Maeda, Y. (2006). Meta-analysis of coefficient alpha. *Psychological Methods, 11*, 306–322. http://doi.org/10.1037/1082-989X.11.3.306

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*(3), 638–641. http://doi.org/10.1037/0033-2909.86.3.638

Rust, J., & Golombok, S. (2009). *Modern psychometrics: The science of psychological assessment.* (3rd ed). Hove: Routledge.

Salkovskis, P. M., Rimes, K. A., Warwick, H. M., & Clark, D. M. (2002). The Health Anxiety Inventory: Development and validation of scales for the measurement of health anxiety and hypochondriasis. *Psychological Medicine*, *32*, 843–853.

Salkovskis, P. M., & Warwick, H. M. (2001). Making sense of hypochondriasis: A cognitive model of health anxiety. In G. J. G. Asmundson, S. Taylor, & B. J. Cox (Eds.), *Health anxiety: Clinical perspectives on hypochondriasis and related conditions* (pp. 46–64). Chichester, UK: Wiley.

Salkovskis, P. M., & Warwick, H. M. C. (1986). Morbid preoccupations, health anxiety and reassurance: A cognitive-behavioural approach to hypochondriasis. *Behaviour Research and Therapy*, *24*(5), 597–602. http://doi.org/10.1016/0005-7967(86)90041-0

Sánchez-Meca, J., López-López, J. A., & López-Pina, J. A. (2013). Some recommended statistical analytic practices when reliability generalization studies are conducted. *British Journal of Mathematical and Statistical Psychology*, *66*(3), 402–425. http://doi.org/10.1111/j.2044-8317.2012.02057.x

Savage, C. J., Vickers, A. J., Kats, J., & Molenaar, D. (2009). Empirical study of data sharing by authors publishing in PLoS journals. *PLoS ONE*, *4*(9), e7078. http://doi.org/10.1371/journal.pone.0007078

Scargle, J. D. (2000). Publication bias: The "File-Drawer" problem in scientific inference. *Journal of Scientific Exploration*, *14*(1), 91–106.

Schmidt, N. B., Lerew, D. R., & Trakowski, J. H. (1997). Body vigilance in panic disorder: Evaluating attention to bodily perturbations. *Journal of Consulting and Clinical Psychology*, *65*(2), 214–220. http://doi.org/10.1037/0022-006X.65.2.214

Schulz, K. F., Chalmers, I., Hayes, R. J., & Altman, D. G. (1995). Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA*, *273*(5), 408–12.

Schwind, J., Gropalis, M., Witthoft, M., & Weck, F. (2016). The effects of attention training on health anxiety: An experimental investigation. *Cognitive Therapy and Research*, *40*(2), 245–255. http://doi.org/10.1007/s10608-015-9745-x

Seivewright, H., Salkovskis, P., Green, J., Mullan, N., Behr, G., Carlin, E., … Tyrer, P. (2004). Prevalence and service implications of health anxiety in genitourinary medicine clinics. *International Journal of STD & AIDS*, *15*(8), 519–522. http://doi.org/10.1258/0956462041558122

Shahidi, S., Molaie, A., & Dehghani, M. (2012). Relationship between health anxiety and alexithymia in an Iranian sample. *4th World Conference on Educational Sciences*, *46*, 591–595. http://doi.org/10.1016/j.sbspro.2012.05.168

Shapiro, M. B. (1961). A method of measuring psychological changes specific to the individual psychiatric patient. *British Journal of Medical Psychology*, *34*(2), 151–155. http://doi.org/10.1111/j.2044-8341.1961.tb00940.x

Sharpe, D. (1997). Of apples and oranges, file drawers and garbage: Why validity issues in meta-analysis will not go away. *Clinical Psychology Review*, *17*(8), 881–901. http://doi.org/10.1016/S0272-7358(97)00056-1

Shields, C., & Murphy, K. (2011). Selective attention and health anxiety: Ill-health stimuli are distracting for everyone. *Cognitive Therapy and Research*, *35*(3), 241–252. http://doi.org/10.1007/s10608-011-9351-5

Singh, K., & Brown, R. J. (2014). Health-related Internet habits and health anxiety in university students. *Anxiety Stress and Coping*, *27*(5), 542–554. http://doi.org/10.1080/10615806.2014.888061

Sirri, L., & Fava, G. A. (2014). Clinical manifestations of hypochondriasis and related conditions. In *Hypochondriasis and health anxiety: A guide for clinicians* (pp. 8–27). New York, NY: Oxford University Press.

Song, F., Parekh-Bhurke, S., Hooper, L., Loke, Y. K., Ryder, J. J., Sutton, A. J., … Harvey, I. (2009). Extent of publication bias in different categories of research cohorts: a meta-analysis of empirical studies. *BMC Medical Research Methodology*, *9*(79), 1–14. http://doi.org/10.1186/1471-2288-9-79

Sorensen, P., Birket-Smith, M., Wattar, U., Buemann, I., Salkovskis, P., Sørensen, P., … Salkovskis, P. (2011). A randomized clinical trial of cognitive behavioural therapy versus short-term psychodynamic psychotherapy versus no intervention for patients with hypochondriasis. *Psychological Medicine*, *41*(2), 431–441. http://doi.org/10.1017/s0033291710000292

Speckens, A. E. M., Spinhoven, P., Sloekers, P. P. A., Bolk, J. H., & Hemert, A. M. (1996). A validation study of the Whitely Index, the Illness Attidudes Scale and the Somatosensory Amplification Scale. *Journal of Psychosomatic Research*, *40*(1), 95–104.

Speckens, A. E. M., Spinhoven, P., Van Hemert, A. M., & Bolk, J. H. (2000). The Reassurance Questionnaire (RQ): Psychometric properties of a self-report questionnaire to assess reassurability. *Psychological Medicine*, *30*, 841–847. http://doi.org/doi:10.1017/S00332 91799002378

Speckens, A. E., Van Hemert, A. M., Spinhoven, P., & Bolk, J. H. (1996). The diagnostic and prognostic significance of the Whitely Index, the Illness Attitude Scales and the Somatosensory Amplification Scale. *Psychological Medicine*, *26*(5), 1085–90. http://doi.org/10.1017/S0033291700035418

Sperlinger, D. (2002). *Outcome assessment in routine clinical practice in psychosocial services measuring outcomes in routine clinical practice: Paper 1*. Leicester: British Psychological Society.

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Starcevic, V. (2014). Relationships with other psychopathology and differential diagnosis of hypochondriasis. In *Hypochondriasis and health anxiety: A guide for clinicians* (pp. 39–64). New York, NY: Oxford University Press.

Starcevic, V., & Lipsitt, D. R. (2001). *Hypochondriasis: Modern perspectives on an ancient malady*. Oxford: Oxford University Press.

Starcevic, V., & Noyes, R. (2014). Introduction and key issues. In *Hypochondriasis and health anxiety: A guide for clinicians* (pp. 1–7). New York, NY: Oxford University Press.

Sterne, J. A., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., … Higgins, J. P. (2016). ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*, *355*(4919), 1–7.

Stewart, S. H., Sherry, S., Watt, M., Grant, V., & Hadjistavropoulos, H. D. (2008). Psychometric evaluation of the multidimensional inventory of hypochondriacal traits: Factor structure and relationship to anxiety sensitivity. *Journal of Cognitive Psychotherapy*, *22*(2), 97–114.

http://doi.org/10.1891/0889-8391.22.2.97

Strauss, M. E., & Smith, G. T. (2009). Construct validity: Advances in theory and methodology. *Annual Review of Clinical Psychology*, *5*, 1–25. http://doi.org/10.1146/annurev.clinpsy.032408.153639

Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, *80*(1), 99–103. http://doi.org/10.1207/S15327752JPA8001_18

Stuart, S. (2104). Interpersonal psychotherapy for hypochondriasis and related disorders: An attachment based approach. In *Hypochondriasis and health anxiety: A guide for clinicians* (pp. 199–221). New York, NY: Oxford University Press.

Sulkowski, M. L., Mancil, T. L., Jordan, C., Reid, A., Chakoff, E., & Storch, E. A. (2011). Validation of a classification system of obsessive-compulsive spectrum disorder symptoms in a non-clinical sample. *Psychiatry Research*, *188*(1), 65–70. http://doi.org/10.1016/j.psychres.2011.01.015

Tabachnick, B. G., & Fidell, L. S. (2012). *Using Multivariate Statistics* (6th ed.). Harlow, UK: Pearson. http://doi.org/10.1037/022267

Tang, N. K. Y., Goodchild, C. E., Sanborn, A. N., Howard, J., & Salkovskis, P. M. (2012). Deciphering the temporal link between pain and sleep in a heterogeneous chronic pain patient sample: A multilevel daily process study. *Sleep*, *35*(5), 675–687. http://doi.org/10.5665/sleep.1830

Tang, N. K. Y., Salkovskis, P. M., Hodges, A., Soong, E., Hanna, M. H., & Hester, J. (2009). Chronic pain syndrome associated with health anxiety: A qualitative thematic comparison between pain patients with high and low health anxiety. *British Journal of Clinical Psychology*, *48*, 1–20. http://doi.org/10.1348/014466508x336167

Tang, N. K. Y., Wright, K. J., & Salkovskis, P. M. (2007). Prevalence and correlates of clinical insomnia co-occurring with chronic back pain. *Journal of Sleep Research*, *16*(1), 85–95. http://doi.org/10.1111/j.1365-2869.2007.00571.x

Taylor, S., & Asmundson, G. J. G. (2004). *Treating health anxiety: A cognitive-behavioral approach*. Guilford Press.

Taylor, S., & Cox, B. J. (1998a). An expanded Anxiety Sensitivity Index: Evidence for a hierarchic structure in a linical sample. *Journal of Anxiety Disorders*, *12*(5), 463–483.

Taylor, S., & Cox, B. J. (1998b). Anxiety sensitivity: Multiple dimensions and hierarchic structure. *Behaviour Research and Therapy*, *36*(36), 37–51.

Terry, M. L., Leary, M. R., Mehta, S., & Henderson, K. (2013). Self-compassionate reactions to health threats. *Personality and Social Psychology Bulletin*, *39*(7), 911–926. http://doi.org/10.1177/0146167213488213

The EuroQol Research Group. (1990). EuroQol - a new facility for the measurement of health-related quality of life. *Health Policy*, *16*(3), 199–208. http://doi.org/10.1016/0168-8510(90)90421-9

The Mendeley Support Team, Birchwood, M., Iqbal, Z., Chadwick, P., Trower, P., Birchwood, M., … Hermann, C. (2013). Illness Behavior. *Cognitive Therapy and Research*, *27*(3), 68–78. http://doi.org/10.1016/j.janxdis.2015.02.001

Tyrer, P., Cooper, S., Salkovskis, P., Tyrer, H., Crawford, M., Byford, S., … Barrett, B. (2014). Clinical and cost-effectiveness of cognitive behaviour

therapy for health anxiety in medical patients: A multicentre randomised controlled trial. *Lancet*, *383*(9913), 219–225. http://doi.org/10.1016/s0140-6736(13)61905-4

Ucar, M., Sarp, U., Karaaslan, O., Gul, A. I., Tanik, N., & Arik, H. O. (2015). Health anxiety and depression in patients with fibromyalgia syndrome. *Journal of International Medical Research*, *43*(5), 679–685. http://doi.org/10.1177/0300060515587578

Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement*, *58*, 6–20. http://doi.org/10.1177/0013164498058001002

Vacha-Haase, T., & Thompson, B. (2011). Score reliability: A retrospective look back at 12 years of reliability generalization studies. *Measurement and Evaluation in Counseling and Development*, *44*, 159–168. http://doi.org/10.1177/0748175611409845

Veale, D., Ellison, N., Boschen, M. J., Costa, A., Whelan, C., Muccio, F., & Henry, K. (2013). Development of an inventory to measure specific phobia of vomiting (Emetophobia). *Cognitive Therapy and Research*, *37*(3), 595–604. http://doi.org/10.1007/s10608-012-9495-y

Veale, D., Hennig, C., & Gledhill, L. (2015). Is a specific phobia of vomiting part of the obsessive compulsive and related disorders? *Journal of Obsessive-Compulsive and Related Disorders*, *7*, 1–6. http://doi.org/10.1016/j.jocrd.2015.08.002

Veddegjaerde, K.-E., F, F. E., Sivertsen, B., Wilhelmsen, I., & Skogen, J. C. (2014). Confirmatory factor analysis and item response theory analysis of the whiteley index. Results from a large population based study in Norway. The Hordaland Health Study (HUSK). *Journal of Psychosomatic Research*, *77*(3), 213–218. http://doi.org/10.1016/j.jpsychores.2014.06.011

Vermeersch, D. a, Lambert, M. J., & Burlingame, G. M. (2000). Outcome questionnaire: Item sensitivity to change. *Journal of Personality Assessment*, *74*(2), 242–61. http://doi.org/10.1207/S15327752JPA7402_6

Vranceanu, A. M., Safren, S. A., Cowan, J., & Ring, D. C. (2010). Health concerns and somatic symptoms explain perceived disability and idiopathic hand and arm pain in an orthopedics surgical practice: A path-analysis model. *Psychosomatics*, *51*(4), 330–337.

Walmyr Publishing Company. (1992). *The Walmyr Assessment Scales Scoring Manual*. Tempe, AZ.

Warwick, H. M. C., & Salkovskis, P. M. (1990). Hypochondriasis. *Behaviour Research and Therapy*, *28*(2), 105–117. http://doi.org/10.1016/0005-7967(90)90023-C

Wattar, U., Sorensen, P., Buemann, I., Birket-Smith, M., Salkovskis, P. M., Albertsen, M., & Strange, S. (2005). Outcome of cognitive-behavioural treatment for health anxiety (hypochondriasis) in a routine clinical setting. *Behavioural and Cognitive Psychotherapy*, *33*(2), 165–175. http://doi.org/10.1017/s1352465804002000

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale—Fourth Edition: Technical and interpretive manual*. San Antonio, TX: Pearson.

Weck, F., Neng, J. M. B., Richtberg, S., & Stangier, U. (2012a). Dysfunctional beliefs about symptoms and illness in patients with hypochondriasis. *Psychosomatics*, *53*(2), 148–154.

http://doi.org/10.1016/j.psym.2011.11.007

Weck, F., Neng, J. M. B., Richtberg, S., & Stangier, U. (2012b). The restrictive concept of good health in patients with hypochondriasis. *Journal of Anxiety Disorders*, *26*(8), 792–798. http://doi.org/10.1016/j.janxdis.2012.07.001

Weck, F., Neng, J. M. B., & Stangier, U. (2013). The effects of attention training on the perception of bodily sensations in patients with hypochondriasis: A randomized controlled pilot trial. *Cognitive Therapy and Research*, *37*(3), 514–520. http://doi.org/10.1007/s10608-012-9482-3

Welch, P. G., Carleton, R. N., & Asmundson, G. J. G. (2009). Measuring health anxiety: Moving past the dichotomous response option of the original Whiteley Index. *Journal of Anxiety Disorders*, *23*(7), 1002–1007. http://doi.org/10.1016/j.janxdis.2009.05.006

Wells, A., & Hackmann, A. (1993). Imagery and core beliefs in health anxiety: Content and origins. *Behavioural and Cognitive Psychotherapy*, *21*(3), 265–273. http://doi.org/10.1017/S1352465800010511

Wheaton, M. G., Abramowitz, J. S., Berman, N. C., Fabricant, L. E., & Olatunji, B. O. (2012). Psychological predictors of anxiety in response to the H1N1 (swine flu) pandemic. *Cognitive Therapy and Research*, *36*(3), 210–218. http://doi.org/10.1007/s10608-011-9353-3

Wheaton, M. G., Berman, N. C., Franklin, J. C., & Abramowitz, J. S. (2010). Health anxiety: Latent structure and associations with anxiety-related psychological processes in a student sample. *Journal of Psychopathology and Behavioral Assessment*, *32*(4), 565–574. http://doi.org/10.1007/s10862-010-9179-4

White, P. D., Pinching, A. J., Rakib, A., Castle, M., Hedge, B., & Priebe, S. (2002). A comparison of patients with chronic fatigue syndrome attending separate fatigue clinics based in immunology and psychiatry. *Journal of the Royal Society of Medicine*, *95*(9), 440–444. http://doi.org/10.1258/jrsm.95.9.440

Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *The American Psychologist*, *61*(7), 726–728. http://doi.org/10.1037/0003-066X.61.7.726

Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*(8), 594–604. http://doi.org/10.1037/0003-066X.54.8.594

Wilson, D. B. (2010). Meta-analysis macros for SAS, SPSS, and Stata. Retrieved July 5, 2017, from http://mason.gmu.ed/dwilson/ma.html

Witthöft, M., Mier, D., Ofer, J., Müller, T., Rist, F., Kirsch, P., … Diener, C. (2013). Neuronal and behavioral correlates of health anxiety: Results of an illness-related emotional Stroop task. *Neuropsychobiology*, *67*(2), 93–102. http://doi.org/10.1159/000345545

Wood, L., Egger, M., Gluud, L. L., Schulz, K. F., Jüni, P., Altman, D. G., … Sterne, J. A. C. (2008). Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ*, *336*(601).

World Health Organisation. (1992). *International Classification of Diseases* (10th ed.). Geneva: Author.

Zarifeh, J. A., Mulder, R. T., Kerr, A. J., Chan, C. W., & Bridgman, P. G. (2012). Psychology of earthquake-induced stress cardiomyopathy, myocardial infarction and non-cardiac chest pain. *Internal Medicine*

Journal, *42*(4), 369–373. http://doi.org/10.1111/j.1445-5994.2012.02743.x

Zhang, Y. Q., Liu, R., Li, G. H., Mao, S. Q., & Yuan, Y. G. (2015). The reliability and validity of a Chinese-version Short Health Anxiety Inventory: an investigation of university students. *Neuropsychiatric Disease and Treatment*, *11*, 1739–1747. http://doi.org/10.2147/ndt.s83501

Zhang, Y. Q., Zhao, Y. Q., Mao, S. Q., Li, G. H., & Yuan, Y. G. (2014). Investigation of health anxiety and its related factors in nursing students. *Neuropsychiatric Disease and Treatment*, *10,* 1223–1234. http://doi.org/10.2147/ndt.s61568

Appendix A HAQ and SHAI

The Health Anxiety Inventory

HAI

name: _____   date: _____

Each question is this section consists of a group of four statements. Please read each group of statements carefully and then select the one which best describes your feelings, over the past six months (or other agreed time period). Identify the statement by ringing the letter next to it, i.e. if you think that statement *a.)* is correct, ring statement *a.).* It may be that more than one statement applies, in which case, please ring any that are applicable.

1.
    a.) I do not worry about my health.

    b.) I occasionally worry about my health.

    c.) I spend much of my time worrying about my health.

    d.) I spend most of my time worrying about my health.

2.
    a.) I notice aches/pains less than most other people (of my age).

    b.) I notice aches/pains as much as most other people (of my age).

    c.) I notice aches/pains more than most other people (of my age).

    d.) I am aware of aches/pains in my body all the time.

3.
    a.) as a rule, I am not aware of bodily sensations or changes.

    b.) sometimes I am aware of bodily sensations or changes.

    c.) I am often aware of bodily sensations or changes.

    d.) I am constantly aware of bodily sensations or changes.

4.   *a.)* resisting thoughts of illness is never a problem.

       b.) most of the time I can resist thoughts of illness.

       c.) I try to resist thoughts of illness but am often unable to do so.

       d.) thoughts of illness are so strong that I no longer even try to resist them.

5.   *a.)* as a rule, I am not afraid that I have a serious illness.

       b.) I am sometimes afraid that I have a serious illness.

       c.) I am often afraid that I have a serious illness.

       d.) I am always afraid that I have a serious illness.

6.   *a.)* I do not have images (mental pictures) of myself being ill.

       b.) I occasionally have images of myself being ill.

       c.) I frequently have images of myself being ill.

       d.) I constantly have images of myself being ill.

7.   *a.)* I do not have any difficulty taking my mind off thoughts about my health.

       b.) I sometimes have difficulty taking my mind off thoughts about my health.

       c.) I often have difficulty in taking my mind off thoughts about my health.

       d.) Nothing can take my mind off thoughts about my health.

8.   *a.)* I am lastingly relieved if my doctor tells me there is nothing wrong.

       b.) I am initially relieved but the worries sometimes return later.

c.) I am initially relieved but the worries always return later.

d.) I am not relieved if my doctor tells me there is nothing wrong.

9.   *a.)* if I hear about an illness I never think I have it myself.

b.) if I hear about an illness I sometimes think I have it myself.

c.) if I hear about an illness I often think I have it myself.

d.) if I hear about an illness I always think I have it myself.

10.   *a.)* if I have a bodily sensation or change I rarely wonder what it means.

b.) if I have a bodily sensation or change I often wonder what it means.

c.) if I have a bodily sensation or change I always wonder what it means.

d.) if I have a bodily sensation or change I must know what it means.

[cont.]

11.   a.) I usually feel at very low risk for developing a serious illness.

b.) I usually feel at fairly low risk for developing a serious illness.

c.) I usually feel at moderate risk for developing a serious illness.

d.) I usually feel at high risk for developing a serious illness.

12.   *a.)* I never think I have a serious illness.

b.) I sometimes think I have a serious illness.

c.) I often think I have a serious illness.

d.) I usually think that I am seriously ill.

13. *a.)*   if I notice an unexplained bodily sensation I don't find it difficult to think about other things.

     b.)   if I notice an unexplained bodily sensation I sometimes find it difficult to think about other   things.

     c.)   if I notice an unexplained bodily sensation I often find it difficult to think about other things.

     d.)   if I notice an unexplained bodily sensation I always find it difficult to think about other things.

14.   *a.)*   my family/friends would say I do not worry enough about my health.

     b.)   my family/friends would say I have a normal attitude to my health.

     c.)   my family/friends would say I worry too much about my health.

     d.)   my family/friends would say I am a hypochondriac.

For the following questions, please think about what it might be like if you had a serious illness of a type which particularly concerns you (e.g. heart disease, cancer, multiple sclerosis & so on). Obviously, you cannot know for definite what it would be like; please give your best estimate of what you *think* might happen, basing your estimate on what you know about yourself and serious illness in general.

15.   *a.)*   if I had a serious illness I would still be able to enjoy things in my life quite a lot.

     b.)   if I had a serious illness I would still be able to enjoy things in my life a little.

     c.)   if I had a serious illness I would be almost completely unable to enjoy things in my life.

     d.)   if I had a serious illness I would be completely unable to enjoy life at all.

16.   *a.)*   if I developed a serious illness there is a good chance that modern medicine would be able to cure me.

   b.)   if I developed a serious illness there is a moderate chance that modern medicine would be able to cure me.

   c.)   if I developed a serious illness there is a very small chance that modern medicine would be able to cure me.

   d.)   if I developed a serious illness there is no chance that modern medicine would be able to cure me.

17.   *a.)*   a serious illness would ruin some aspects of my life.

   b.)   a serious illness would ruin many aspects of my life.

   c.)   a serious illness would ruin almost every aspect of my life.

   d.)   a serious illness would ruin every aspect of my life.

18.   *a.)*   if I had a serious illness I would not feel that I had lost my dignity.

   b.)   if I had a serious illness I would feel that I had lost a little of my dignity.

   c.)   if I had a serious illness I would feel that I had lost quite a lot of my dignity.

   d.)   if I had a serious illness I would feel that I had totally lost my dignity.

---

all groups are scored 0, 1, 2 or 3 depending on the statement selected;

if more than statement is selected, use the highest-scoring statement of those chosen.

main section score (questions 1 to 14) =

negative consequences score (questions 15 to 18) =

total score =

scoring the 18 item HAI

In the 2002 paper describing the development of both the full Health Anxiety
Inventory and this current shortened 18 item version, the following scores were
reported for the shortened form in a series of different populations. The table
below gives means (and standard deviations):

	health anxiety	*anxiety sufferers*	*controls*	*students*	*Gp*	*gastro patients*
main section	30.1 (5.5)	14.9 (6.2)	9.4 (5.1)	9.6 (4.5)	11.2 (4.6)	11.4 (6.3)
negative consequences	7.8 (2.8)	3.6 (2.2)	2.2 (2.1)	3.0 (1.8)	3.2 (2.0)	2.4 (1.9)
total score	37.9 (6.8)	18.5 (7.3)	12.2 (6.2)	12.6 (5.0)	14.5 (5.9	13.9 (7.4)

At an initial assessment, it is probably appropriate to ask these questions about
the last six months. When monitoring treatment, applying the scale questions to
the last week is more usual.

Salkovskis P.M., Rimes K.A., Warwick H.M.C. & Clark D.M. *The health
anxiety inventory: development and validation of scales for the measurement of
health anxiety and hypochondriasis* Psychological Medicine 2002;32:843-853.

Appendix 2. The Health Anxiety Questionnaire

HAQ

name: _____   date: _____

Below is a list of questions about health anxiety. Please carefully read each item on the list. Indicate how often you have been bothered in this way ***during the past week, including today*** (or other agreed time period), by placing an ***x*** in the appropriate space in the columns to the right of each question.

		not at all or rarely	*some-times*	*often*	*most of the time*
1	do you ever worry about your health?				
2	are you ever worried that you may get a serious illness in the future?				
3	does the thought of a serious illness ever scare you?				
4	when you notice an unpleasant feeling in your body, do you tend to find it difficult to think of anything else?				
5	do you ever examine your body to find whether there is something wrong?				
6	if you have an ache or pain do you worry that it may be caused by a serious illness?				
7	do you ever find it difficult to keep worries about your health out of your mind?				
8	when you notice an unpleasant feeling				

	in your body, do you ever worry about it?				
9	when you wake up in the morning do you find you very soon begin to worry about your health?				
10	when you hear of a serious illness or death of someone you know, does it ever make you more concerned about your own health?				
11	when you read or hear about an illness on tv or radio does it ever make you think you may be suffering from that illness?				

		not at all or rarely	some-times	often	most of the time
12	when you experience unpleasant feelings in your body do you tend to ask friends or family about them?				
13	do you tend to read up about illness and diseases to see if				

	you may be suffering from one?				
14	do you ever feel afraid of news that reminds you of death (such as funerals, obituary notices)?				
15	do you ever feel afraid that you may die soon?				
16	do you ever feel afraid that you may have cancer?				
17	do you ever feel afraid that you might have heart disease?				
18	do you ever feel afraid that you may have any other serious illness? which illness? …………………………… …………				
19	have your bodily symptoms stopped you from working during the past six months or so?				
20	do your bodily symptoms stop you from concentrating on what you are doing?				
21	do your bodily symptoms stop you from enjoying yourself?				

for each question score *0* for *"not at all or rarely"*, *1* for *"sometimes"*, *2* for *"often"*, and

3 for *"most of the time"*

add scores for 1, 4, 6 – 9, 11 & 18 to give *health worry & preoccupation* score =

add scores for 2, 3, 10 & 14 – 17 to give *fear of illness and death* score =

add scores for 5, 12 & 13 to give *reassurance-seeking behaviour* score =

add scores for 19, 20 & 21 to give *interference with life* score =

total score =

Lucock MP & Morley S *The health anxiety questionnaire*

Br J Health Psychol 1996;1:137-50

Appendix B: Data capture form

Author name: **Study ID:**
                                              HAQ 14 / 18
                                              SHAI:
                                              Scoring: 0-3 / 1-4
Title: **Introduction:** Y / N

Precis

N:
Sample description (& nationality):

Sample/quality criteria

	Yes	**No**	**Other (CD, NA, NR)***
Sample composition:			
-Homogenous and			
-Sample from the same population?			
-Represent population of interest?			
-Well described?			
-Medically unwell included?			
Recruitment bias:			
-Non-biased sampling strategy?			
-Participation rate at least 50%?			
-Uniform inclusion/exclusion criteria?			
Testing conditions			
Standardised?			
Test administered correctly?			
-Measurement context likely to elevate anxiety?			

-Measurement point likely to elevate anxiety?			
-Researcher 'blinded' to participant grouping?			
Data analysis			
-Missing data reported?			
-Adequately dealt with?			
-How:			

*CD = could not determine, NR = not reported, NA = not applicable

Any other info:

Data

Sample name	**n**	**Age** Mean (s.d)	**% Fem**	**SHAI** Mean (s.d.)	**Alpha**	**SHAI** Range

Correlations:

Measure name	**Coefficient**

<u>Other info</u>

Appendix C: SHAI and HAQ mean, standard deviations and reliability coefficients

HAQ mean score, SD and reliability coefficients by population for all studies.

Author name	Diagnosis	Mean	Adjusted mean*	SD	Alpha coefficient
Asbury, Creed and Collins (2004)	Healthy control	8.00	-	9	-
Meana and Lykins (2009)	Healthy control	16.83	-	10.20	-
Melli, Carraresi, Poli and Bailey, (2016)	Healthy control	32.47	11.47	8.70	.92
Asbury, Chandrruangphen and Collins (2006)	Post menopausal women	10.15	10.15	5.35	-
Lucock and Morley (1996)	Healthy control	8.26	-	7.96	-
Lucock and Morley (1996)	Students	12.10	-	6.71	-
Asbury, Webb and Collins (2011)	Cardiac syndrome X	13.25	-	8.10	-
Asbury, Creed and Collins (2004)	Cardiac syndrome X	13.50	-	10	-
Asbury et al. (2008)	Cardiac syndrome X	11.25	-	5.80	-

Asbury, Creed and Collins (2004)	Coronary heart disease	14.00	-	11	-
Asbury et al. (2012)	Angina	19.15	-	12.59	-
Zarifeh, Mulder, Kerr, Chan and Bridgman (2012)	Myocardial infarction	14.00	-		-
Zarifeh et al. (2012)	Stress cardiomyopathy	7.00	-		-
Zarifeh et al. (2012)	Attenders cardiac clinic	13.00	-		-
Lucock, Morley, White and Peake (1997)	Gastroscopy attender	12.73	-	4.22	-
Meechan, Collins, Moss-Morris and Petri (2005)	Breast symptom diagnostic test	13.15	-	9.56	.79
Fiddler, Jackson, Kapur, Wells and Creed (2004)	Frequent attenders at Cardiology and neurology	16.37	-	1.12	-
Hewison et al. (2006)	Amniotic test	31.73	10.73	7.48	-
Aggarwal et al. (2014)	General medicine	7.00	-		-

Noyes, Stuart, Longley, Langbehn and Happel (2002)	General medicine	14.60	-	9.30	-
Jackson, Kincey, Fiddler, Creed and Tomenson (2004)	Attenders cardiology	19.00	-	12.80	-
Jackson, Kincey, Fiddler, Creed and Tomenson (2004)	Attenders neurology	15.00	-	9.10	-
					-
Graff et al. (2006)	Active Crohn's disease	17.98	-	10.60	-
Graff et al. (2006)	Active Ulcerative colitis	15.07	-	9.16	-
Esfandyari and Harewood (2007)	Gastro-internal symptoms	20.60	-	1.40	-
Graff et al. (2006)	Inactive Crohn's disease	11.85	-	7.70	-
Graff et al. (2006)	Inactive ulcerative colitis	9.29	-	5.59	-
Meana and Lykins (2009)	Intercourse pain	21.46	21.46	9.60	-
Dolin et al. (2003)	Osteoarthritis	11.20	11.20	8.71	-
Eastin and Guinsler (2006)	Mixed medical	49.14	28.14		0.95

Lucock and Morley (1996)	Mixed medical and mental health	-	-	-	0.91
White et al. (2002)	Chronic fatigue	36.50	15.50	9.50	-
Priebe, Fakhoury and Henningsen (2008)	Chronic fatigue	37.80	16.80	9.50	-
Aggarwal et al. (2014)	Medically unexplained symptoms	12.00	-	-	-
Aggarwal et al. (2014)	Medically unexplained sympoms	14.00	-	-	-
Noyes, Stuart, Longley, Langbehn and Happel (2002)	Hypochondriasis	31.90	31.90	10.60	-

*Adjusted mean refers to scores adjusted because incorrect scoring used in article

SHAI-14 mean score, SD and reliability coefficients by population for all studies.

Author name	Population	Mean	Adjusted mean	SD	Alpha coefficient
Alberts, Sharpe, Kehler and Hadjistavropoulos, (2011)	Healthy control	9.19	-	4.86	-
Bailer et al., (2014)	Healthy control	5.80	-	2.80	-
Boston and Merrick (2010)	Healthy control	7.04	-	4.30	-
Fergus (2013b)	Healthy control	10.97	-	6.62	.91
Fergus (2014a)	Healthy control	11.42	-	6.71	.90
Fergus (2013a)	Healthy control	10.95	-	6.83	.92
Goodwin, Fairclough and Poole (2013)	Healthy control	10.01	-	5.82	-
Kehler and Hadjistavropoulos (2009)	Healthy control	9.37	-	5.12	-
Kowalyk, Hadjistavropoulos, and Jones (2009)	Healthy control	9.02	-	4.66	-
Payne et al., (2007)	Healthy control	11.70	-	6.95	-
Tang, Wright and Salkovskis (2007)	Healthy control	6.90	-	4.50	-
Terry, Leary, Mehta, and Henderson (2013)	Healthy control	25.23	11.23	5.81	.88

Study	Group				
Veale et al., (2013)	Healthy control	7.70	-	5.10	-
Kowalyk and Hadjis-Tavropoulos (2007)	Pregnant women	8.22	-	4.33	.83
Fergus (2014b)	Healthy control	-	-	-	.91
Boschen, Veale, Ellison and Reddell (2013)	Healthy control	-	-	-	-
Dzhambov and Dimitrova (2014)	Healthy control	15.02	-	4.51	-
Alberts and Hadjistavropoulos (2014)	Student	13.03	-	5.60	.86
Goodwin et al., (2013)	Students				.87
Jeffers et al., (2015)	Students	32.91	18.91	7.25	.86
Karademas, Christopoulou, Dimostheni, and Pavlu (2008). Study 1.	Students	24.86	-	5.06	.81
Karademas, Christopoulou, Dimostheni, and Pavlu (2008). Study 2.	Students	25.20	-	5.33	.82
Macatee and Cougle (2013)	Students	10.44	-	4.65	.82
Norr, Capron, and Schmidt (2014)	Students	10.12	-	6.60	.91
Singh and Brown (2014)	Students	14.13	-	5.80	-
Bailer et al., (2014)	Hypochondriasis	30.10	-	4.60	-
Barsky and Ahern (2004)	Hypochondriasis	37.71	23.71	-	-

Lovas and Barsky, (2010)	Hypochondriasis	28.90	-	-	-
Sorensen et al., (2011)	Hypochondriasis	27.60	-	5.20	-
Tyrer et al., (2014)	Hypochondriasis	25.00	-	4.35	-
Wattar et al., (2005)	Hypochondriasis	30.50	-	-	-
Bailer et al., (2014)	Depression	9.60	-	3.60	.97
Bailer et al., (2014)	Mixed anxiety disorder	9.60	-	3.60	-
					-
Tang et al., (2007)	Chronic pain	12.50	-	6.90	-
Tang, Goodchild, Sanborn, Howard and Salkovskis (2012)	Chronic pain & insomnia	14.60	-	7.40	-
Tang et al., (2007)	Chronic pain & insomnia	16.40	-	7.20	-
Boschen et al., (2013)	Vomiting phobia		-	-	0.91
Veale, Hennig and Gledhill, (2015)	Vomiting phobia	28.60		8.90	

Veale et al., (2013)	Vomiting phobia	19.80		8.60	.95
Jones, Hadjistavropoulos and Gullickson, (2014)	Breast cancer	15.51		6.93	.90
Porritt, Sufi, Barlow and Baker, (2014)	Dentine sensitivity	26.30		8.60	
Claude, Hadjistavropoulos,= and Friesen, (2014)	Diabetes	14.78		7.05	.88
DeMarinis, Barsky, Antin, and Chang (2009)	Stem cell transplant	40.84	26.84	9.02	
Payne et al., (2007)	Vulvar vestibular syndrome	19.45		8.24	

*Adjusted mean refers to scores adjusted because incorrect scoring used in article

SHAI-18 mean score, SD and reliability coefficients by population for all studies.

Study author	Population	Mean	SD	Alpha coefficient
Annagur, Avci, Demir and Uygur (2015)	Myocardial infarction (early age)	18.62	9.34	-
Annagur, Avci, Demir and Uygur (2015)	Myocardial infarction (late age)	14.25	8.90	-
Robertson, Javed, Samani and Khunti (2008)	Attenders cardiac clinic	13.01	7.49	-
Seivewright et al., (2004)	Attenders GUM clinic	11.02	6.09	-
Brown et al., (2012)	Endoscopy MUS	14.50	10.50	-
Brown et al., (2012)	Endoscopy organic illness	17.60	14.00	-
Hedman et al., (2013)	Health anxiety	33.70		-
Hedman et al., (2013)	Health anxiety	34.00		-
Abramowitz, Deacon and Valentiner (2007)	Hypochondriasis	31.40	9.26	-
Deacon and Abramowitz (2006)		31.22	9.20	-

	Hypochondriasis			
McManus, Surawy, Muse, Vazquez-Montes and Williams (2012)	Hypochondriasis	33.23	7.22	.89
Abramowitz, Deacon and Valentiner (2007)	Anxiety disorder mixed including hypochondriasis	18.60	13.22	.96
Olatunji, Deacon, Abramowitz and Valentiner (2007)	Anxiety disorder mixed including hypochondriasis	37.45	11.27	-
Aydemir, Kirpinar, Sati, Uykur, and Cengisiz (2013)	Depression	14.30	7.90	-
Abramowitz, Deacon and Valentiner (2007)	GAD	13.09	11.91	-
Abramowitz, Deacon and Valentiner (2007)	OCD	17.44	16.07	-
Deacon and Abramowitz (2006)	OCD	15.43	12.87	-
Abramowitz, Deacon and Valentiner (2007)	Panic	21.67	12.28	-
Aydemir, Kirpinar, Sati, Uykur, and Cengisiz (2013)	Panic	23.90		-

Deacon and Abramowitz (2006)	Panic	19.08	10.32	-
Abramowitz, Deacon and Valentiner (2007)	Social phobia	11.62	9.13	-
Abramowitz, Deacon and Valentiner (2007)	Somatoform disorder	24.80	1.30	-
Abramowitz, Deacon and Valentiner (2007)	Specific phobia	14.93	11.68	-
Morales, Espada, Carballo, Piqueras and Orgiles (2015)	Adolescents	10.48	5.70	.93
Koteles, Freyler, Kokonyei and Bardos (2015)	Adolescents	13.70	5.78	.78
Aydemir, Kirpinar, Sati, Uykur, and Cengisiz (2013)	Healthy controls	16.60	6.70	-
Gerolimatos and Edelstein (2012b)	Healthy controls	11.23	6.97	.90
Gerolimatos and Edelstein (2012a)	Healthy controls	14.68	6.97	.86
Kobori, Okita, Shiraishi, Hasegawa and Iyo (2014)	Healthy controls	13.21	7.02	-

Koteles, Freyler, Kokonyei and Bardos (2015)	Healthy controls	14.40	5.43	.76
Rode, Salkovskis, Dowd and Hanna (2006)	Healthy controls	8.50	4.40	-
Ucar et al., (2015)	Healthy controls	9.38	4.24	-
Abramowitz and Moore (2007)	Students	10.79	6.38	.86
Brady, Cisler and Lohr (2014)	Students	11.54	6.19	-
Fan and Olatunji (2013)	Students	14.70	6.13	.86
Fergus and Valentiner (2011)	Students	13.83	8.37	.91
Fetzner et al., 2014)	Students	12.13	6.96	.88
Goetz, Lee and Cougle (2013)	Students	9.69	5.77	.86
Abramowitz, Olatunji and Deacon (2007)	Students	14.70	6.02	.84
Olatunji et al., (2007)	Students	10.79	6.37	-
Olatunji et al., (2009)	Students	12.85	6.01	.86
Olatunji, Ebesutani, Haidt and Sawchuk (2014)	Students	13.57	7.08	-
Olatunji et al., (2009)	Students	13.35	6.40	-

Shahidi, Molaie and Dehghani (2012)	Students			0.78
Sulkowski et al., (2011)	Students	12.20	6.46	.87
Wheaton, Berman, Franklin and Abramowitz (2010)	Students	12.48	6.79	.88
Wheaton, Abramowitz, Berman, Fabricant and Olatunji (2012)	Students	12.03	6.75	.89
Zhang, Liu, Li, Mao and Yuan (2015)	Students	11.98	5.78	.74
Zhang, Zhao, Mao, Li and Yuan (2014)	Students	11.14	-	-
Bot, Anderson, Neuhaus and Ring (2013)	Arm pain	11.00	6.80	.88
Vranceanu, Safren, Cowan and Ring (2010)	Arm pain non-traumatic	9.38	5.45	-
Rode, Salkovskis, Dowd and Hanna (2006)	Chronic pain	16.30	7.80	-
Ucar et al., (2015)	Fibromyalgia	23.50	10.78	-
Rode, Salkovskis, Dowd and Hanna (2006)	Pain	10.50	4.50	-

Kobori, Okita, Shiraishi, Hasegawa and Iyo (2014)	Chronic health problems	18.93	10.55	-
Kobori, Okita, Shiraishi, Hasegawa and Iyo (2014)	Mixed	16.07	-	.93
Muse, McManus, Leung, Meghreblian and Williams (2012)	Mixed students and clinic	14.35	7.91	.90