Investigating differential symptom profiles in Major Depressive Episode with and without Generalized Anxiety Disorders: True comorbidity or symptom similarity?

Author/Contributor:
Sunderland, Matthew; Mewton, Louise; Slade, Tim; Baillie, Andrew J.

Publication details:
Psychological Medicine
v. 40
Chapter No. 7
pp. 1113-1123
0033-2917 (ISSN)

Publication Date:
2010

Publisher DOI:
http://dx.doi.org/10.1017/S0033291709991590

RUNNING HEAD: True co-morbidity or symptom similarity?

Word Count: 4.593 (excluding abstract and references), 2 Figure, 3 Tables.

Investigating differential symptom profiles in major depressive episode with and without generalized anxiety disorder: True co-morbidity or symptom similarity?

Matthew Sunderland [1]

Louise Mewton[2]

Tim Slade[2]

Andrew J. Baillie[3]

1. Clinical Research Unit for Anxiety and Depression (CRUfAD), University of New South Wales, Sydney, Australia.
2. National Drug and Alcohol Research Centre, University of New South Wales, Sydney, Australia.
3. Centre for Emotional Health, Department of Psychology, Macquarie University, Sydney Australia.

Address for correspondence:

Mr. Matthew Sunderland, Clinical Research Unit for Anxiety and Depression, 299 Forbes Street, Darlinghurst NSW 2010, Australia. Email: matthews@student.unsw.edu.au. Ph: 612 8382 1722. Fax: 612 8382 1721

ABSTRACT

Background: Large community based epidemiological surveys have consistently identified high co-morbidity between major depressive episode (MDE) and generalized anxiety disorder (GAD). Some have suggested that this co-morbidity may be artificial and the product of the current diagnostic system. Due to the added direct and indirect costs associated with co-morbidity it is important to investigate if methods of diagnostic classification are artificially increasing the level of observed co-morbidity.

Methods: The item response theory log-likelihood ratio procedure was used to test for differential item functioning of MDE symptoms between respondents with and without a diagnosis of GAD in the 2001-2002 National Epidemiological Survey on Alcohol and Related Conditions.

Results: The presence of GAD significantly increased the chances of reporting any symptom of MDE with odds ratios ranging from 2.54 to 3.60. However, there was no indication of significant differential item functioning of MDE symptoms in respondents with and without GAD.

Conclusions: The lack of any significant differential item functioning indicates that cases with GAD do not present with a distinct MDE symptom profile, one that is consistent with the endorsement of symptoms that are conceptually similar in nature between the two disorders, compared to cases without GAD. This does not support the hypothesis that co-morbidity between MDE and GAD is artificially inflated due to similar symptom criteria required by the current diagnostic system. Instead, MDE and GAD may be thought of as two distinct diagnostic entities that frequently co-occur due to a shared underlying trait.

Key words: Major depressive episode, generalized anxiety disorder, co-morbidity, differential item functioning, item response theory.

Epidemiological studies have consistently identified a high level of co-morbidity between DSM-IV major depressive episode (MDE) and generalized anxiety disorder (GAD) in large community-based samples (Kessler et al., 2005; Grant et al., 2005; Judd et al., 1998; Hunt et al., 2002; Hunt et al., 2004). Given that MDE and GAD are the most common form of anxiety-mood co-morbidity and they occur together beyond the level of chance (Sanderson et al., 1990), researchers have begun to question the source and nature of co-morbidity (Angold et al., 1999). For instance, does the observed co-morbidity between MDE and GAD represent a common etiology for the two disorders (Kendler et al., 2007; Hettema et al., 2006), or is co-morbidity merely a diagnostic artefact (Jorm, 2006; Maj, 2005; Frances et al., 1990)?

The issue of co-morbidity between mood and anxiety disorders has important implications for both diagnostic decision making and treatment planning that can lead to increased direct and indirect patient costs (Clarkin & Kendall, 1992). For instance, Andrews, Slade, and Issakidis (2002) identified increased service utilization and associated disability in survey respondents who reported more than one disorder in comparison to respondents who reported one disorder only. Furthermore, patients who receive more than one diagnosis report higher levels of distress, often have a poor prognosis (Brown et al., 1995), and increased heritability of co-morbid conditions (Carter et al. 2004). MDE is one of the most common mental disorders observed in communities worldwide (Kessler & Ustun, 2008), is one of the leading causes of global disease burden (Murrey & Lopez, 1996), and is associated with a decrement in health that is significantly greater than the decrement in health associated with angina, arthritis, asthma, and diabetes (Moussavi et al. 2007). Therefore, considering that co-

morbidity between MDE and GAD is quite often the rule rather than the exception, it seems imperative to investigate the source of this co-morbidity and to gather a greater understanding of how and why these two disorders co-occur so often in the population.

Researchers have argued whether the observed co-morbidity represents a common underlying etiology between the two disorders (Kendler et al., 2007; Hettema et al., 2006), or whether co-morbidity is an artefact of the current diagnostic system (Jorm, 2006; Maj, 2005; Frances et al., 1990). More specifically, there is the concern that co-morbidity between MDE and GAD is a product of the observed similarity or overlap between the DSM-IV diagnostic criteria required for both disorders (Brown & Barlow, 1992; Brown et al., 1995; Frances et al., 1992; Hunt, 2000; Maser, 1998; Slade & Watson, 2006). The DSM-IV symptom criteria required for MDE contains a strong anxiety component that is reflected by the inclusion of sleep disturbance, fatigue or loss of energy, psychomotor disturbance, and a diminished ability to think or concentrate (APA, 2000; see Table 1). Furthermore, the DSM-IV definition of MDE is polythetic in nature, which means that any five out of the nine symptoms, with at least one of the five being depressed mood or lack of interest/pleasure, can result in a positive diagnosis. Two individuals with different symptom profiles, therefore different clinical characteristics, could be classified under the same definition (Krueger & Bezdjian, 2009). In other words, the polythetic-categorical classification system implemented in the DSM-IV creates significant within-category heterogeneity. The symptom overlap exhibited between MDE and GAD combined with the polythetic nature of MDE begs the question of whether respondents with GAD differ in their MDE symptom profile compared to those without GAD. It seems

likely that patients with GAD may present with a distinct symptom profile, one that could be considered an anxious-depression, in comparison to those without GAD who present with a typical depression profile, one that is consistent with symptoms that are specific to MDE. Co-morbidity between GAD and MDE could therefore be considered, in part, artificial in the sense that these respondents could be better conceptualised as belonging to one distinct form of anxious-depression rather than receiving two diagnoses.

Item Response Theory (IRT) provides a method for examining whether the pattern of symptoms in MDE is consistent in cases with or without co-morbid GAD through the examination of Differential Item Functioning (DIF, Embretson & Reise, 2000, Holland & Wainer, 1993). Higher prevalence of MDE in the presence of GAD and the reverse, higher prevalence of GAD in the presence of MDE, might confound the comparison of symptom profiles as co-morbid cases may simply be more severe. However, DIF equates the two groups on their level of the underlying trait or theta (in this case the level of underlying MDE pathology) before comparing the performance of items, symptoms or criteria (here after symptoms). The performance of symptoms in IRT is modelled by one or more item response parameters that is best conceptualised in the item characteristic curves detailed in Figure 1. The first parameter (*b*), commonly called the difficulty or threshold parameter, represents the point on the underlying dimension of depression (x-axis) where the predicted probability (y-axis) of endorsing the symptom equals 0.5. The second parameter (*a*), commonly called the slope or discrimination parameter, represents the symptoms ability to discriminate between respondents with higher and lower levels of the underlying dimension of depression. The steeper the slope of the curve the better the

symptom is at discriminating between respondents (Hays et al. 2000). The third guessing parameter has no application in the current context and will be ignored. A given item is judged to display DIF when the threshold and/or the discrimination parameters differ significantly between the two examined groups. For example, if DIF is detected between respondents with GAD and without GAD it indicates that respondents from one group are more or less likely to endorse that particular symptom of MDE.

Small et al (2008) previously used DIF to examine different MDE symptom profiles in adolescents across cases without co-morbidity, with internalising co-morbidity, and with externalising co-morbidity. They found that psychomotor disturbance and cognitive impairments seemed to function differently between co-morbidity groups. They were able to conclude that adolescences with varying levels of co-morbidity have important differences in their symptom profile that may indicate distinct forms of depression and therefore the need to develop tailored treatments depending on the extent of co-morbidity.

In the current study, it is hypothesised that if the DSM-IV diagnostic system is artificially contributing to the level of co-morbidity observed between MDE and GAD then it is expected that cases with co-morbid GAD are more likely to have a distinct MDE symptom profile from cases without GAD, one that is consistent with the selection of overlapping anxiety symptoms outlined in Table 1 (sleep disturbance, fatigue or loss of energy, psychomotor disturbance, and diminished ability to think or concentrate). The finding of DIF implies that co-morbidity between the two disorders could be the result of a direct relationship between the overlapping MDE symptoms

and a GAD diagnosis, more specifically the MDE symptoms could be measuring both GAD and MDE diagnostic constructs. If on the other hand cases with co-morbid GAD experience a symptom profile that is no different to cases without GAD then we would expect to see no indication of DIF. This implies that an indirect relationship exists between MDE symptoms and a GAD diagnosis, therefore high co-morbidity between the two disorders may be better explained by an underlying relationship between the MDE and GAD diagnostic constructs instead of overlapping symptom criteria. To our knowledge no previous study has specifically examined DIF occurring in the symptom profile of MDE between cases with and without a diagnosis of GAD.

METHODS

Sample

Data were from the 2001-2002 National Epidemiological Survey of Alcohol and Related Conditions (NESARC). The NESARC uses a nationally representative household sample of 43 093 (overall response rate of 81%) non-institutionalised US civilians aged 18 years or older. The sample consists of 18 518 (43%) males and 24 575 (57%) females with a mean age of 46 (SD=18). A multistage stratified sampling design was used to stratify according to certain sociodemographic characteristics of the US general population with over-sampling of non-Hispanic Black households and young adults aged 18-24 (see Grant et al. 2003a for more information on the complex sampling design and methodology).

Diagnostic Assessment

The DSM-IV criteria for MDE and GAD were assessed using the Alcohol Use Disorder and Associated Disabilities Interview Schedule (AUDADIS). The AUDADIS is a fully-structured lay-administered diagnostic interview with good to

fair reliability for the common DSM-IV Axis I disorders (Grant et al., 2003b). Apart from the initial screening of depressed respondents, the AUDADIS assessed the remaining symptoms of MDE without the use of skip-outs. As a result of the initial screening of depressed respondents the DIF analysis was restricted to the remaining seven symptoms of depression. The following MDE symptoms assessed in this study include; 1) Weight disturbance, 2) Sleep disturbance, 3) Psychomotor disturbance, 4) Fatigue/Loss of energy, 5) Worthlessness or extreme guilt, 6) Trouble concentrating, 7) Thoughts of death or suicide. Although this is not ideal for the investigation of DIF it is argued that the benefits of using such a data-rich large scale community sample far outweigh the negatives if we do not analyse the data.

The AUDADIS assesses the presence of each DSM-IV MDE symptom occurring at the same period of time across the lifespan using aggregated self-report items. For example the symptom of Weight/Appetite disturbance is assessed via four binary self-report items; one for increased appetite, one for decreased appetite, one for weight gain, and one for weight loss. The responses to each of the four items are combined in computerized algorithms so that the respondent will be coded as either positive or negative for the symptom of Weight/Appetite disturbance. In the present example Weight/Appetite disturbance is coded as positive if at least one of the four items is endorsed by the respondent. A total of 19 items were aggregated in this manner to assess the presence of the seven DSM-IV symptoms under investigation.

A separate module containing a series of binary self-report questions was used to assess the lifetime prevalence of GAD. These questions were designed to assess periods of feeling intense nerves, anxiety, or worry and the associated symptom

profile, excessiveness, onset, recency, and duration. The anxiety or worry was then checked against the relevant DSM-IV exclusion criteria such as; the anxiety or worry must not be the direct effect of a substance or physical illness or occur exclusively during a mood disorder, psychotic disorder, or pervasive developmental disorder; the focus of the worry is not confined to features of another axis I disorder; and the anxiety or worry must cause clinically significant distress. The items were combined using computerized algorithms that were designed to specifically match the DSM-IV diagnostic decision rules to produce the final variable for a lifetime diagnosis of GAD. The final diagnostic variable was used to stratify the population into those with GAD and those without GAD.

Statistical Analyses

The analysis proceeded in three phases – first the symptom prevalence rates and standard errors were estimated, next the dimensionality of the MDE symptoms was examined, and finally differential item functioning was examined. The prevalence rates were weighted for the US general population and the standard errors and confidence intervals were estimated using the Taylor series linearization method to adjust for the complex sampling design of the NESARC. The odds of endorsing a particular MDE symptom given the respondent has GAD in comparison to the respondent not having GAD were calculated using separate univariate logistic regression models. The prevalence and logistic regression analyses were conducted using the SUDAAN statistical software package (Shah et al., 1997).

The IRT model implemented to detect DIF in the current study relies on the underlying assumption that all examined items load significantly on one underlying

factor, commonly referred to as unidimensionality. It is generally acknowledged that unidimensionality, in the strictest sense, may not be required to satisfy the IRT assumption but instead for most practical applications the notion of "essential" unidimensionality will suffice (Drasgow & Parsons, 1983). Essential unidimensionality holds when the data exhibit one dominate factor that is strong enough so that it is not influenced by the presence of smaller factors (Hays et al., 2000). This assumption was examined in the current study by conducting a confirmatory factor analysis (CFA) of tetrachoric correlations and fitting a one-factor solution in the whole sample and in the two subsamples under examination. The one factor model was fitted to the symptom data using the weighted least squares mean and variance adjusted (WLSMV) method of estimation.

The choice of model fit statistics was informed by previous studies that indicate the chi-square goodness of fit statistic is overly sensitive to minor differences in very large samples (Browne et al., 2002). Yu (2002) recommends, as an alternative, the use of the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI), and the Root Mean Square Error of Approximation (RMSEA). Good model fit is evidence by a CFI score ≥ 0.96, TFI ≥ 0.95, and RMSEA ≤ 0.05. Previously, these fit statistics have been reliably applied to estimate the factor structure of alcohol and cannabis use disorders in a large community sample (Proudfoot et al., 2006; Teesson et al., 2002). The CFA and model fit statistics were generated using Mplus version 5 (Muthen & Muthen, 2007).

A parametric approach, known as the Item Response Theory Log-likelihood Ratio procedure (IRTLR; Thissen et al., 1993), was implemented to detect DIF (see Teresi

et al., 2007 for a detailed description of IRTLR procedure). IRTLR uses a 2-parameter logistic item response model. To clarify some specific terms, *a*-DIF involves the identification of DIF in the slope parameters while *b*-DIF involves the identification of DIF in the difficulty parameters between the two examined groups.

The DIF procedure implemented in the current study requires that members from each group are matched on an estimate of theta (the underlying trait). To achieve a valid estimate of theta it is best to match members of the two groups using items that are presumed to be group-invariant. This set of items is referred to as the anchor set and, if there is no prior knowledge that certain items do not exhibit DIF, several methods have been developed to "purify" items to assist with the selection of appropriate anchors (Woods, 2009). A simulation study found that purification of the anchor items prior to DIF detection significantly improved the accuracy rate and reduced false positive results, particularly in IRT-based methods of DIF detection (Navas-Ara & Gomez-Benito, 2002).

As a means of generating the purified anchor set, it was decided to implement an iterative purification process demonstrated in a recent study by Edelen et al. (2006). All symptoms were first entered into the IRTLR-DIF program to test for the presence of DIF. At this stage no specific anchor set had been identified therefore all symptoms were tested using all the other symptoms as the temporary anchor set. For each item the likelihood-ratio test statistic (G^2) was produced, which compares a model with the parameter estimates constrained to be equal between the two examined groups with a model that frees the parameters to be estimated separately between the two groups. If the omnibus G^2 statistic exceeded the threshold point of 3.84 (the critical value

associated with α=0.05 in a chi-square distribution with 1 degree of freedom), then the item is judged to display DIF in one or both items parameters. The choice of $G^2 \geq 3.84$ as the threshold point, regardless of the degree of freedom associated with the omnibus test and multiple comparisons, was made to protect against the possibility of type II error and therefore ensuring a conservative selection of anchor items (Thissen, 2001). The symptoms exhibiting possible DIF were removed from the analysis and the remaining symptoms were re-analysed. This purification process was repeated until no symptoms exhibited any indication of DIF. The final DIF-free set of symptoms formed the purified anchor set that estimates theta between the two groups while the remaining symptoms formed the test symptoms for the subsequent DIF analysis.

The next step of the DIF analysis involved conducting the final IRTLR procedure for all target symptoms to detect DIF while using the purified anchor set to estimate theta. The log-likelihood ratio test statistic (G^2) was again used as an omnibus test for significant DIF in the difficulty and slope parameters. Assuming the theoretical probability that all DIF may be located in one parameter the statistical significance of the omnibus test statistic was calculated using 1 degree of freedom as suggested by Thissen (2001). If the omnibus test was significant at the 0.05 level (i.e. a G^2 statistic greater than 3.84), follow-up tests (df=1) were then conducted on the individual parameters to identify if the DIF was present in the difficulty or slope parameter. It was at this stage that the Benjamini-Hochberg method (Benjamini, & Hochberg, 1995; Thissen et al. 2002) was used to adjust the critical p-value for multiple comparisons to reduce the possibility of type I error. The log-likelihood ratio tests were conducted using the IRTLR-DIF version 2.0 software package (Thissen, 2001).

RESULTS

Of the total sample, 13,753 respondents (32%) reported feeling sad, blue, depressed or lacked interest/pleasure for a period of 2 weeks at some point in their life. Respondents who did not experience two weeks of depressed mood or lack of interest or pleasure were excluded from the current analysis, as they were not assessed for the remaining MDE symptoms. Of the studied sample there were 1,645 (11.9%) respondents with a lifetime diagnosis of GAD. The prevalence estimates and odds ratios of MDE symptoms stratified by GAD are presented in Table 2. In respondents with GAD, the odds of reporting any MDE symptom were higher than in respondents without GAD.

Table 3 shows the results of the confirmatory factor analysis in the total sample and in the two subsamples. The CFI, TLI, and RMSEA statistics provide evidence of a reasonably well fitting 1-factor model in the total sample and the two sub-samples. Loadings on the one factor were sufficiently high for all symptoms, ranging from 0.59 to 0.82 in the total sample, and 0.55 to 0.81 in the two sub-samples. Taking these results into account we are able to conclude that the symptom data showed an adequate level of essential unidimensionality required by the IRT model.

In the purification process prior to the DIF detection, two symptoms, psychomotor disturbance and worthlessness/guilt, were identified as the DIF-free anchor items and the remaining five symptoms identified as the target symptoms for the DIF analysis. Figure 2 presents the results of the DIF analysis for each symptom in the form of ICCs for the two studied groups. Assuming a theoretical probability that all DIF is

located in one parameter, using 1 degree of freedom indicates that DIF may be present in the *a* or *b* parameters in two symptoms, fatigue/loss of energy ($G^2 = 4.8$, $p = 0.03$), and thoughts of death/suicide ($G^2 = 4.1$, $p = 0.04$). In follow-up comparisons before adjustment, only one test symptom, fatigue/loss of energy, indicated significant *b*-DIF (the *difficulty* parameters were different between the two groups but the *slope* parameters remained the same; $G^2 = 4.8$, $p = 0.03$). However, after adjusting the critical p-value for multiple comparisons using the Benjamini-Hochberg method no symptom of depression indicated significant *b*-DIF or *a*-DIF between respondents with and without GAD.

DISCUSSION

The aim of the current study was to further understand the nature and source of co-morbidity between MDE and GAD by comparing MDE symptom functioning in those with co-morbid GAD and those without. The current study analysed data from a large community based epidemiological survey to examine if conceptually similar symptoms required for MDE and GAD may artificially contribute to co-morbidity. The data did not support the hypotheses of the current study that symptoms thought to be similar in wording and definition would exhibit significant DIF. After controlling for multiple comparisons no symptoms of MDE were found to exhibit significant DIF between respondents with and without GAD. This result indicates that there is no evidence to suggest that cases with GAD differ in their MDE symptom profile to cases without GAD.

Many researchers have previously suggested that co-morbidity between MDE and GAD may be, in part, a function of the high degree of similarity in diagnostic criteria

operationalised by the DSM-IV classification system and the associated endorsement of MDE symptoms is biased as a result (Brown & Barlow, 1992; Brown et al., 2001; Brown et al., 1995; Frances et al., 1992; Hunt, 2000; Maser, 1998; Slade & Watson, 2006). However, the findings from the current study do not offer further support for this hypothesis. If it were true we would have expected to see a different symptom profile, one associated with the endorsement of overlapping anxiety symptoms, emerge in cases with GAD compared to cases without GAD. However, some caution is required when drawing this conclusion since DIF detection using the IRTLR procedure is only one method to investigate differential symptom profiles between experimental groups and there is ongoing debate regarding which methodology is the most accurate (Millsap & Everson, 1993). Furthermore, the procedures of IRT and particularly DIF are relatively new to the field of psychiatry and clinical psychology and there is still a need for further refinement to the procedures used in this study. It is argued that these findings need to be replicated using alternative methods of DIF and in a variety samples in order to provide additional support for the current findings.

Despite the above concerns the current findings can tentatively provide some conclusions regarding the source and nature of co-morbidity between MDE and GAD. As there was no finding of DIF this seems to indicate that there is no interaction between the endorsement of anxiety symptoms in MDE and the presence of GAD. The presentation and experience of MDE could therefore be considered the same in cases with and without GAD. Why then do these two disorders co-occur, and why are the observed odds of endorsing MDE symptoms increased with the presence of GAD? One possible explanation stems from the literature addressing the latent structure of mental disorders. Clark and Watson (1991) first argued that mood and anxiety

disorders could be understood in terms three underlying factors (negative affect, positive affect, and physiological hyperarousal). This model has since been extended to investigate the higher order factor structure of the common DSM mental disorders, which has indicated that anxiety and unipolar mood disorders exist together on one underlying construct or dimension, similar to negative affect but now labelled the internalizing factor (Krueger, 1999; Clark & Watson, 2006; Slade & Watson, 2006).

Krueger and Finger (2001) fitted a 2 parameter IRT model to examine the relationship between the seven anxiety and unipolar mood disorders and the internalizing latent dimension. They found that MDE tended to occupy the lower end of the internalizing spectrum whilst GAD occupied the upper end. This model implies that MDE is an easy disorder to meet criteria for and respondents with GAD (indicating a high level of internalizing) should more often meet criteria for the disorders that are indicators of a lower level of internalizing. Integrating this conception of co-morbidity with the results of the current study it could be argued that the symptoms of MDE function as similar indicators of the underlying trait regardless of the presence of GAD. Respondents with GAD however, are more likely to have endorsed the full range of MDE symptoms, and therefore a reduced chance of having pure GAD, as they are considered higher up on the internalizing factor.

There are several limitations of the current study that should be considered in light of the results. First, the data were based on retrospective self-reports of MDE and GAD. There is some concern that the retrospective nature can influence the results due to recall bias (Kruijshaar et al., 2005; Simon et al., 1995; Simon & Von Korff, 1992). Certain MDE symptoms may or may not have been reported depending on the limits

of each respondent's memory. This is a common problem face in the field of psychiatry and clinical psychology as there are relatively few and in some cases no strict laboratory tests that can be applied to corroborate the patients self-report. As a result, the current study must assume that the majority of patient self-report responses are an accurate reflection of past life events.

Second, the interview contained embedded gate/skip questions that automatically partition the respondents into smaller sub-samples depending on their responses. For the current study, the sample was partitioned into respondents who reported at least one of the first two symptoms of depression. An accurate analysis of these two symptoms was not possible nor was it possible to examine respondents who might endorse symptoms of depression without these two symptoms. A study by Zimmerman et al. (2006) using a skip-free diagnostic assessment found that almost all patients (98.5%) who endorsed five or more symptoms of MDE, therefore meeting the DSM-IV symptom criteria, indicated at least low mood or lack of interest/pleasure as one of their symptoms. Therefore, while an accurate analysis of the first two symptoms was not able to be attained, we are confident that the sample examined in this study is representative of a sub-population at greater risk of developing MDE. Furthermore, the benefits of examining data-rich community based epidemiological studies despite the restricted sample far outweigh the negatives if we do not. One major benefit includes the ability to investigate the symptom profiles in a large sample of respondents from the wider community with a diverse range of socio-demographic features.

Third, despite the significant strengths of the IRTLR, including the ability to detect differences in groups that are matched according to a common underlying trait, there are some unresolved limitations that are often neglected in previous studies that implement the same procedure. Foremost is the IRT assumption that the underlying trait is normally distributed in the population under investigation. In the field of psychopathology, where the focus is on abnormal or maladaptive behaviour, the underlying latent trait is often unlikely to have a normal distribution. Woods (2008a), in a simulation study, found that violations of this assumption can lead to greater type I error rates, particularly when no anchor set is specified. The current study attempted to address this issue by first, setting an explicit DIF-free anchor set and second, by using the Benjamini-Hochberg method to control type I error at 5% across the study. Woods (2008b) has recently developed a method that can reliably detect DIF using IRTLR methods in non-normal distributions. However, at this point in time the procedure has been developed assuming a non-normal distribution in the focal group only and is not expected to work well when the reference group is non-normal yet presumed normal. We have no reason to believe that in the current study the focal group has a non-normal distribution but the reference group is normal therefore it was decided not to follow-up with this additional testing.

The current study was the first to specifically investigate the hypothesis that conceptually similar symptoms contained in the DSM-IV diagnostic system may artificially contribute to the co-morbidity between MDE and GAD. Using an IRT based procedure our results show that there is no evidence of differential reporting of MDE symptoms between survey respondents with and without GAD. Nevertheless, high co-morbidity between MDE and GAD remains. Fortunately, the current findings

possibly indicate that co-morbidity is less likely to be an artefact of the diagnostic system and more likely to be the result of a true underlying association between diagnostic constructs. Additional research is required to support and confirm the findings of the current study.

ACKNOWLEDGEMENTS

This study was funded in part by an Australian Postgraduate Award doctoral scholarship and NHMRC project grant no. 510137. The authors declare no conflict of interest.

REFERENCES

American Psychiatric Association (2000). *The Diagnostic and Statistical Manual of Mental Disorders - Fourth Edition Text Revision*. American Psychiatric Association: Washington, DC.

Andrews G, Slade T, Issakidis C (2002). Deconstructing current comorbidity: data from the Australian National Survey of Mental Health and Well-Being. *British Journal of Psychiatry* **181**, 306-314.

Andrews G, Stewart G, Morris-Yates A, Holt P, Henderson S (1990). Evidence for a general neurotic syndrome. *British Journal of Psychiatry* **157**, 6-12.

Angold A, Costello EJ, Erkanli A. (1999). Comorbidity. *Journal of Child Psychology and Psychiatry* **40**, 57-87.

Benjamini Y, Hochberg Y (1995). Controlling for false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B* **57**, 289-300.

Brown TA, Antony MM, Barlow DH (1995). Diagnostic comorbidity in panic disorder: effect on treatment outcome and course of comorbid diagnoses following treatment. *Journal of Consulting and Clinical Psychology* **63**, 408-418.

Brown TA, Barlow DH (1992). Comorbidity among anxiety disorders: Implications for treatment and DSM-IV. *Journal of Consulting and Clinical Psychology* **60,** 835-844.

Brown TA, Marten PA, Barlow DH (1995). Discriminant validity of the symptoms constituting the DSM-III-R and DSM-IV associated symptom criterion of generalized anxiety disorder. *Journal of Anxiety Disorders* **9**, 317-328.

Browne MW, MacCallum RC, Kim CT, Anderson BL, Glaser R (2002). When fit indices and the residuals are incompatible. *Psychological Methods* **7**, 403-421.

Carter AS, Pollock RA, Suvak MK, Pauls DL (2004). Anxiety and Major Depression comorbidity in a family study of Obsessive-compulsive disorder. *Depression and Anxiety* **20**, 165-174.

Clark LA, Watson D. (1991). Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology* **100**, 316-336.

Clark LA, Watson D. (2006). Distress and fear disorders: an alternative empirically based taxonomy of the 'mood' and 'anxiety' disorders. *British Journal of Psychiatry* **189**, 481-483.

Clarkin JF, Kendall PC (1992). Comorbidity and treatment planning: summary and future directions. *Journal of Consulting and Clinical Psychology* **60**, 904-908.

Drasgow F, Parsons CK (1983). Application of unidimensional item response theory models to multidimensional data. *Applied Psychological Measurement* **7**, 189-199.

Edelen MO, Thissen D, Teresi JA, Kleinman M, Ocepek-Welikson K (2006). Identification of differential item functioning using item response theory and the likelihood-based model comparison approach. *Medical Care* **44**, S134-S142.

Embretson SE, Reise SP (2000). *Item Response Theory for Psychologists*. Lawrence Erlbaum Associates: Mahwah.

Frances AJ, Manning D, Marin D, Kocsis J, McKinney K, Hall W, Kline M (1992). Relationship of anxiety and depression. *Psychopharmacology* **106**, S82-S86.

Frances AJ, Widiger TA, Fyer MR (1990). The influence of classification methods on comorbidity. In *Comorbidity of Mood and Anxiety Disorders* (eds. Maser J, Cloninger CR). American Psychiatric Press: Washington, DC.

Grant BF, Dawson DA, Stinson FS, Chou PS, Kay W, Pickering R (2003b). The Alcohol Use Disorder and Associated Disabilities Interview Schedule-IV (AUDADIS-IV): Reliability of alcohol consumption, tobacco use, family history of depression and psychiatric diagnostic modules in a general population sample. *Drug and Alcohol Dependence* **71,** 7-16.

Grant BF, Hasin DS, Stinson FS, Dawson DA, Ruan WJ, Goldstein RB, Smith SM, Saha TD, Huang B (2005). Prevalence, correlates, co-morbidity, and comparative disability of DSM-IV generalized anxiety disorder in the USA: results from the National Epidemiologic Survey of Alcohol and Related Conditions. *Psychological Medicine* **35**, 1747-1759.

Grant BF, Kaplan K, Shepard J, Moore T (2003a). *Source and Accuracy Statement for Wave 1 of the 2001-2002 National Epidemiologic Survey on Alcohol and Related Conditions.* National Institute on Alcohol Abuse and Alcoholism: Bethesda, MD.

Hays RD, Morales LS, Reise SP (2000). Item response theory and health outcomes measurement in the 21st century. *Medical Care* **39**, II28-II42.

Hettema JM, Neale MC, Myers JM, Prescott CA, Kendler KS (2006). A population-based twin study of the relationship between neuroticism and internalizing disorders. *American Journal of Psychiatry* **163**, 857-864.

Holland PW, Wainer H (1993). *Differential Item Functioning*. Lawrence Erlbaum Associates: Mahwah.

Hunt C (2000). The diagnosis and nature of generalized anxiety disorder. *Current Opinion in Psychiatry* **13**, 157-161.

Hunt C, Issakidis C, Andrews G (2002). DSM-IV generalized anxiety disorder in the Australian National Survey of Mental Health and Well-being. *Psychological Medicine* **32**, 649-659.

Hunt C, Slade T, Andrews G (2004). Generalized anxiety disorder and major depressive disorder comorbidity in the National Survey of Mental Health and Well-Being. *Depression and Anxiety* **20**, 23-31.

Jorm AF (2006). National surveys of mental disorders: are they researching scientific facts or constructing useful myths? *Australian and New Zealand Journal of Psychiatry* **40**, 830-834.

Judd LL, Kessler RC, Paulus MP, Zeller PV, Wittchen HU, Kunovac JL (1998). Comorbidity as a fundamental feature of generalized anxiety disorders: Results from the National Comorbidity Study (NCS). *Acta Psychiatrica Scandinavica* **98**, 6-11.

Kendler KS, Gardner CO, Gatz M, Pedersen NL (2007). The sources of co-morbidity between major depression and generalized anxiety disorder in a Swedish national twin sample. *Psychological Medicine* **37**, 453-462.

Kessler RC, Chiu WT, Demler O, Walters EE (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry* **62**, 617-627.

Kessler RC, Ustun B eds. (2008). *The WHO World Mental Health Survey: Global perspectives on the epidemiology of mental disorders*. Cambridge University Press: New York, NY.

Krueger RF. (1999). The structure of common mental disorders. *Archives of General Psychiatry* **56**, 921-926.

Krueger RF, Bezdjian S. (2009). Enhancing research and treatment of mental disorders with dimensional concepts: toward DSM-V and ICD-11. *World Psychiatry* **8**.

Krueger RF, Finger MS. (2001). Using item response theory to understand comorbidity among anxiety and unipolar mood disorder. *Psychological Assessment* **13**, 140-151.

Kruijshaar ME, Barendregt J, Vos T, de Graaf R, Spijker J, Andrews G. (2005). Lifetime prevalence estimates of major depression: An indirect estimation method and a quantification of recall bias. *European Journal of Epidemiology* **20**, 103-111.

Maj M (2005). 'Psychiatric comorbidity': an artefact of current diagnostic systems? *British Journal of Psychiatry* **186**, 182-184.

Maser JD (1998). Generalized anxiety disorder and its comorbidities: dispute at the boundaries. *Acta Psychiatrica Scandinavica* **98**, 12-22.

Millsap RE, Everson HT (1993) Methodology Review: Statistical approaches for assessing measurement bias. *Applied Psychological Measurement* **17**, 297-334.

Moussavi S, Chatterji S, Verdes E, Tandon A, Patel V, Ustun B (2007). Depression, chronic diseases, and decrements in health: results from the World Health Surveys. *The Lancet* **370**, 851-858.

Murry CJL, Lopez AD eds. (1996). *The Global Burden of Disease*. Harvard University Press: Harvard, MA.

Muthen LK, Muthen BO (2007). *Mplus Users' Guide Fifth Edition*. Muthen & Muthen: Los Angeles, CA.

Navas-Ara MJ, Gomez-Benito J (2002). Effects of ability scale purification on the identification of DIF. *European Journal of Psychological Assessment* **18**, 9-15.

Proudfood H, Baillie AJ, Teesson M (2006). The structure of alcohol dependence in the community. *Drug and Alcohol Dependence* **81**, 21-26.

Sanderson WC, Beck AT, Beck J. (1990). Syndrome comorbidity in patients with major depression or dysthymia: Prevalence and temporal relationships. *American Journal of Psychiatry* **147**, 1025-1028.

Shah BV, Barnwell BG, Biegler GS. (1997). *SUDAAN User's manual*. Research Triangle Institute: Research Triangle Park, NC.

Simon GE, Von Korff M (1992). Reevaluation of secular trends in depression rates. *American Journal of Epidemiology* **135**, 1411-1422.

Simon GE, Von Korff M, Ustun TB, Gater R, Gureje O, Sartorius N (1995). Is the lifetime risk of depression actually increasing? *Journal of Clinical Epidemiology* **48**, 1109-1118.

Slade T, Watson D (2006). The structure of common DSM-IV and ICD-10 mental disorder in the Australian general population. *Psychological Medicine* **36**, 1593-1600.

Small DM, Simons AD, Yovanoff P, Silva SG, Lewis CC, Murakami JL, March J. (2008). Depressed adolescents and comorbid psychiatric disorders: are there differences in the presentation of depression? *Journal of Abnormal Child Psychology*, 1-14.

Teesson M, Lynskey M, Manor B, Baillie A (2002). The structure of cannabis dependence in the community. *Drug and Alcohol Dependence* **68**, 255-262.

Teresi JA, Ocepek-Welikson K, Kleinman M, Cook KF, Crane PK, Gibbons LE, Morales LS, Orlando-Edelen M, Cella D (2007). Evaluating measurement equivalence using the item response theory log-likelihood ratio (IRTLR) method to assess difference item functioning (DIF): applications to measures of physical functioning ability and general distress. *Quality of Life Research* **16**, 43-68.

Thissen D (2001). *IRTLR DIF v2.0*: Software for the computation of the statistics involved in item response theory likelihood-ratio tests for differential item functioning. Available from http://www.unc.edu/~dthissen/dl.html.

Thissen D, Steinberg L, Kuang D (2002). Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. Journal of Educational and Behavioral Statistics 27, 77-83.

Thissen D, Steinberg L, Wainer H (1993). Detection of differential item functioning using the parameters of item response models. In *Differential Item Functioning* (ed. PW Holland and H Wainer). Lawrence Erlbaum Associates: Hillsdale, NJ.

Yu CY (2002). Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes (Doctoral thesis). University of California: Los Angeles.

Woods C (2009). Empirical selection of anchor for tests of differential item functioning. *Applied Psychological Measurement* **33**, 42-57.

Woods C (2008a). Likelihood-ratio DIF testing: effects of nonnormality. *Applied*

Psychological Measurement **32**, 511-526.

Woods C (2008b). IRT-LR-DIF with estimation of the focal-group density as an empirical histogram. *Educational and Psychological Measurement* **88**, 571-586.

Zimmerman M, McGlinchey JB, Young D, Chelminski I (2006). Diagnosing major depressive disorder IV: Relationship between number of symptoms and the diagnosis of disorder. *Journal of Nervous and Mental Disease* **194**, 450-453.

FIGURES

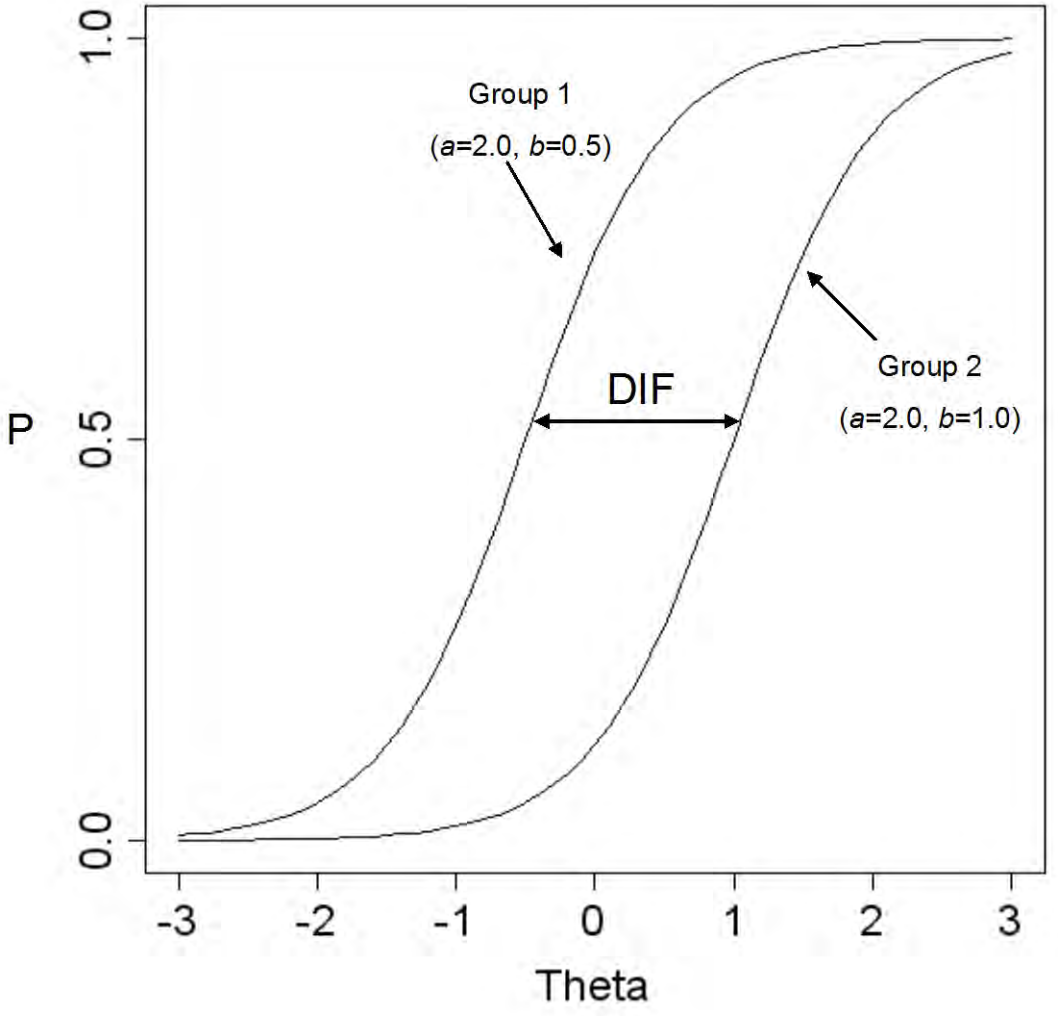

Figure 1: Graphical representation of uniform DIF between two groups.

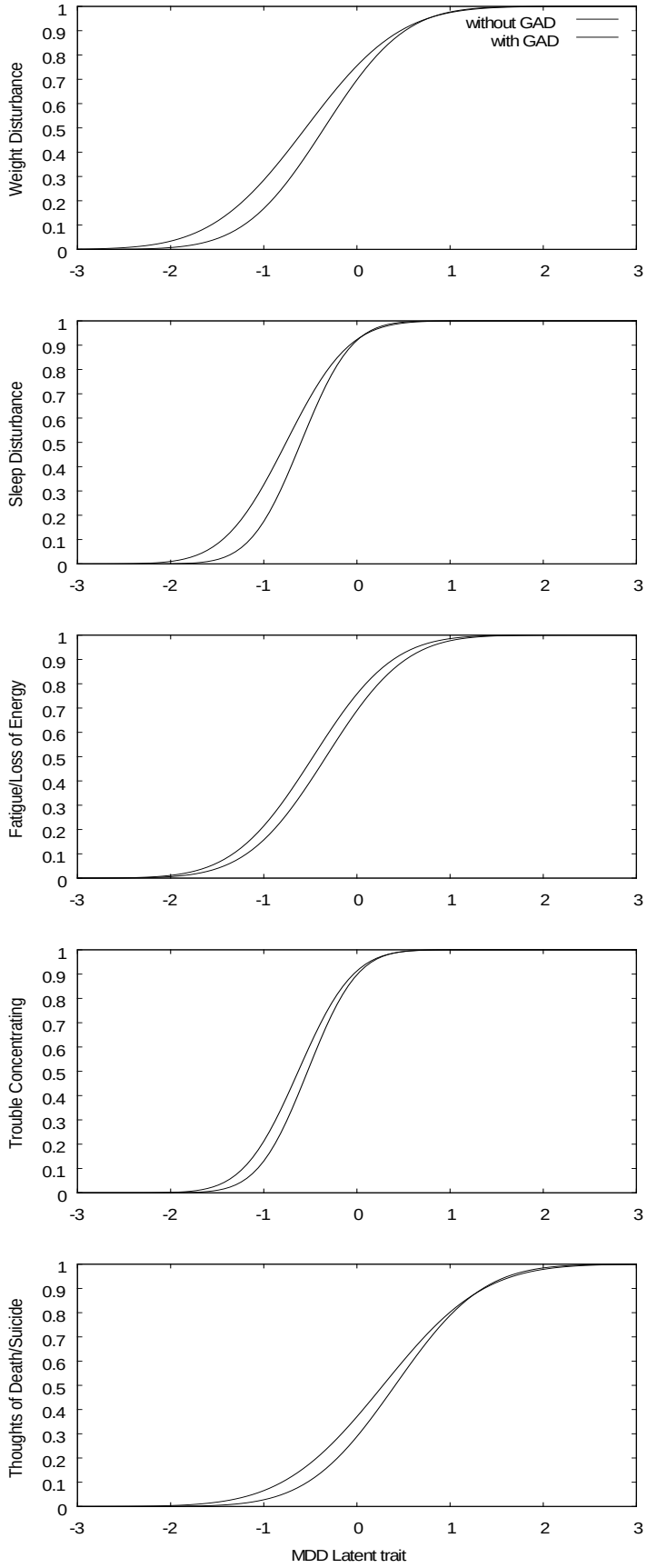

Figure 2: Item characteristic curves (ICCs) for the five examined MDE symptoms. Note: psychomotor disturbance and worthlessness/guilt were anchor items and assumed to be DIF free.

TABLES

Table 1: DSM-IV MDE and GAD Symptom Comparison

MDE Symptoms	GAD Symptoms
Psychomotor agitation or retardation nearly every day.	**Restlessness or feeling keyed up or on edge.**
Fatigue or loss of energy nearly every day.	**Being easily fatigued.**
Diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or as observed by others).	**Difficulty concentrating or mind going blank.**
Insomnia or hypersomnia nearly every day.	**Sleep disturbance (difficulty falling or staying asleep, or restless unsatisfying sleep).**
Depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g., feels sad or empty) or observation made by others (e.g., appears tearful).	Irritability.
Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by either subjective account or observation made by others).	Muscle tension.
Significant weight loss when not dieting or weight gain (e.g., a change of more than 5% of body weight in a month), or decrease or increase in appetite nearly every day.	
Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick).	
Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide.	

Note: Symptoms similar in respect to wording/definition are highlighted in bold.

Table 2: Prevalence and Odds Ratios of MDE Symptoms Stratified by Co-morbid GAD

Symptom	No GAD		GAD		Odds Ratio	
	%	SE	%	SE	Est.	95% CI
Weight Disturbance	61	0.7	80	1.3	2.54	2.15 - 2.99
Sleep Disturbance	71	0.6	90	1.1	3.57	2.81 - 4.53
Psychomotor Disturbance	46	0.6	74	1.3	3.36	2.92 - 3.87
Fatigue/loss of energy	60	0.7	84	1.1	3.46	2.92 - 4.11
Worthlessness/Guilt	52	0.7	80	1.2	3.60	3.10 - 4.19
Trouble concentrating	67	0.6	92	0.8	5.36	4.36 - 6.59
Thoughts of death/suicide	40	0.6	64	1.3	2.69	2.39 - 3.02

Note: Figures were estimated using the weight and sampling design of the NESARC survey.

Table 3: Confirmatory Factor Analysis of Symptom Criteria for DSM-IV MDE

Symptom	***Total Sample*** **Factor Loadings**	***With GAD*** **Factor Loadings**	***Without GAD*** **Factor Loadings**
Weight Disturbance	0.625	0.594	0.613
Sleep Disturbance	0.816	0.810	0.808
Psychomotor Disturbance	0.733	0.697	0.718
Fatigue/Loss of energy	0.740	0.649	0.730
Worthlessness/Guilt	0.765	0.703	0.752
Trouble Concentrating	0.792	0.675	0.784
Thoughts of death/Suicide	0.590	0.550	0.567
	Fit Statistics	**Fit Statistics**	**Fit Statistics**
CFI	*0.978*	*0.982*	*0.974*
TFI	*0.979*	*0.982*	*0.976*
RMSEA	*0.059*	*0.034*	*0.062*