Interview-based Internalizing Diagnoses: Same Genetic Factor Captured as by a Brief Self-report Scale?

Line C. Gjerde, Stud. Psychol.

Submitted as a Master Thesis at the Department of Psychology

UNIVERSITY OF OSLO

Autumn, 2009

Acknowledgements

Writing my master thesis has been a challenging and exciting experience. I would like to thank my two supervisors Espen Røysamb and Kristian Tambs for priceless help and guidance along the way. Espen - I greatly admire your methodological skills, entusiasm and pedagogic talent. Thank you for your time and patience, and for always managing to turn my feelings of despair into optimism and productivity. Kristian - thank you for answering my questions both thoroughly and at all times, and for sharing your expert knowlegde with me. Thanks also to Nikolai Czajkowski for valuable methodological contributions and proofreading.

The National Institute of Public Health has allowed me to use their data, and also provided me with an excellent workspace, and for that I am very grateful. My gratitude naturally extends to all the twins who have participated, and thereby made so many exciting studies possible.

I am also very grateful to my friends and family for their cheering and for their patience when I was absent, and to Magnus - my everyday hero, for his loving support.

Thanks also to Pål Ulleberg, Jon Martin Sundet, Dag Erik Eilertsen, Knut Inge Fostervold and Knut Hagtvet, whose inspiring and challenging lectures woke my interest for the field of quantitative research methodology.

Line C. Gjerde, Oslo, October 2009

Abstract

Author: Line C. Gjerde

Title: Interview-based Internalizing Diagnoses: Same Genetic Factor Captured as by a Brief Self-report Scale?

Supervisors: Drs Espen Røysamb, University of Oslo/National Institute of Public Health and Kristian Tambs, National Institute of Public Health

Internalizing disorders have a high prevalence in the population, and represent a major cost for the society and severe suffering for the individuals affected if not diagnosed and treated correctly. Traditionally, diagnostic interviews have been the main methodology for screening and diagnostic assessment. Different types of questionnaires intended to screen for symptoms of mental distress have also been developed to save time and resources, and it is necessary to have knowledge regarding these instruments' correspondence to psychiatric diagnoses. A host of studies have investigated the phenotypic correspondence between questionnaires and interview-based diagnoses. Typically, a moderate correspondence is found, which implies that the questionnaires do a mediocre job in detecting psychiatric cases. Twin modelling renders it possible to investigate the correspondence at a genotypic level, and hence is able to reveal more information about the correspondence than phenotypic studies. By decomposing the variance inherent in the instruments, twin modelling can delineate how much of the phenotypic correspondence that is due to genetic and environmental sources. In addition, it is possible to investigate the genetic and environmental correlation between the variables. To date, only one study has investigated the correspondence by the use of twin modelling (Foley, Neale & Kendler, 2001).

The present study makes use of both phenotypic analyses and twin modelling in order to investigate the correspondence between the Symptom Checklist-5 (SCL-5), and six internalizing disorders from the Computerized International Diagnostic Interview (CIDI). A sample of 7992 twins from the Norwegian Institute of Public Health Twin Panel (NIPHTP) replied to the questionnaire, and 2793 twins were interviewed. By including data from both males and females, and extending the analyses to including several internalizing disorders, the study represents an extensive replication of Foley, Neale and Kendler (2001).

The phenotypic correspondence found was moderate, with a tetrachoric correlation of .48, odds ratio of 4.75, and a relative risk of 2.95. The phenotypic results are in the area of

what have been found earlier. A confirmatory factor analysis justified collapsing of the six internalizing disorders into one factor. The twin modelling revealed that 81% of the phenotypic correspondence was due to genetic factors, and 19% was due to environmental factors. The main result was the finding of a strong genetic correlation of .82 (.61 - 1.0, 95% CI) between the genetic component in SCL-5 and internalizing disorder, which suggests that the genetic liability indexed by these instruments are practically the same. The environmental correlation was found to be .16 (.00 - .34, 95% CI), which implies mainly non-shared environmental factors between the SCL-5 and internalizing disorder.

The findings from the twin modelling cast light upon the findings from phenotypic studies and suggests that the results of Foley et al. may be generalized from the single disorder depression to internalizing disorder in general. In addition, the present study finds that the results hold for both males and females. The strong genetic correspondence in the instruments has implications for research and clinical utility of questionnaires measuring symptoms of anxiety and depression. In essence, the SCL-5 and CIDI tap onto the same genetic factor, and hence the finding implies that the SCL-5 may be used as a screener for genetic risk for internalizing disorder.

Table of Contents

Introduction

Depressive and anxiety disorders are among the most common psychiatric disorders, and affect most age groups (Feliciano & Arean, 2007). A psychiatric or mental disorder is described in DSM-IV as a clinically significant behavioural or psychological syndrome or pattern (APA, 2000). Apart from substance use disorders, anxiety disorders are the most prevalent disorders in the United States (Beidel & Stipelman, 2007). A national psychiatric comorbidity survey found a lifetime prevalence of any anxiety disorder to be 28.8% (Kessler et al., 2005a). Depressive disorders like major depressive disorder (MDD) are also highly prevalent, and are probably the most prevalent single disorder after substance use disorders. In Norway the lifetime prevalence of MDD was found to be 17.1% (9.9% for men and 24.0% for women) (Kringlen, Torgersen & Cramer, 2001), and a similar study in the United States found a lifetime prevalence of MDD to be 12% (21.3% for women and 12.3% for men) (Kessler, et al., 2005a). Anxiety and depression disorders cause significant distress for the individuals affected, and also for the individual's family and close ones, for instance in terms of emotional suffering and broken relationships (Pincus & Pettit, 2001). In addition, these disorders exert a great economical toll on the society in regard to lost manpower and primary health care treatments (Pincus & Pettit, 2001). Depressive disorders are found to be the second most disabling disorders after heart diseases, and cause significant reduction in life quality among the affected (Pincus & Pettit, 2001). As for depression disorders, individuals affected by anxiety disorders also experience significant impairment in many areas of their life (Beidel & Stipelman, 2007). Although eminently treatable (Pincus & Pettit, 2001), disorders like anxiety and depression are often unrecognized by primary care physiscans due to inadequate diagnostic procedures (Lecrubier, 2007; Williams, Noël, Cordes, Ramirez & Pignone, 2002). As many as 40% of the psychiatric cases may be unrecognized, and of those that are recognized as psychiatric cases, only a few receive a correct diagnosis (Lecrubier, 2007).

The Concept of Internalizing Disorders

Anxiety and depression are the prototypes of what has been called internalizing disorders, which are disorders characterized by negative-affect-laden mood and anxiety (Krueger & Markon, 2006). Internalization is a construct originally derived from the literature of child psychopathology (see e.g. Tandon, Cardeli & Lubi, 2009), but has later also been applied in

adult psychopathology (Acton, Kuhn, Wilson & Hall, 2005; Hettema, Neale, Myers, Prescott & Kendler, 2006; James & Taylor, 2008; Kendler & Prescott, 2006; Krueger, 1999; Krueger & Markon, 2006; Røysamb et al., 2009). Internalizing disorders are psychiatric disorders characterized by internal suffering, usually manifested as anxiety or depression (Kendler & Prescott, 2006). Examples of common internalizing disorders are major depressive disorder (MDD), dysthymia, generalized anxiety (GAD), social phobia, specific phobia, panic disorder, and agoraphobia. Internalizing disorders are often also referred to as emotional disorders. The disorders are diagnosed using symptom criteria listed in the diagnostic systems, the DSM-IV (APA, 2000) or the ICD-10 (WHO, 1992). Psychiatric interviews are the gold standard for diagnostic assessment, but various kinds of questionnaires are also used.

The Classificatory Systems for Psychiatric Disorders

The Diagnostic and Statistical Manual of Mental Disorders (DSM; APA, 2000) is the leading classificatory system, together with the International Classification of Diseases (ICD; WHO, 1992). The newest version, the DSM-IV published in 1994, and the text revision, the DSM-IV-TR released in 2000, is a nomenclature based on extensive empirical evidence (APA, 2000). The criteria for having a disorder are based on the phenomenology of the conditions, and not on underlying pathophysiological processes, as was the case with the first versions of the DSM. The nomenclature is further based on a hierarchical structure, in that one disorder may exclude the presence of another. Several diagnoses may be given as long as the hierarchic rules are followed (Torgersen, 2003). DSM is a categorical system, in that it divides mental disorders into types based on sets of criteria used to define the disorders (APA, 2000). The categories in the nomenclature system are prototypes of disorders, and a patient with a close approximation to a certain prototype, as measured by the number of criteria endorsed, is said to have that disorder (Maser & Patterson, 2002).

The development of the existing classificatory systems has a long history. Since the first nomenclature was introduced in the 1840s, a large number of different classificatory systems have developed. These systems were very different from each other in the emphasis put on etiology, phenomenology, and defining features of the included disorders (APA, 2000). The first version of DSM, the DSM-I, was published in 1952, as a result of the need for an instrument that could collect statistical information on psychiatric disorders (APA, 2000). The later DSM-II, published in 1968, was similar to the first version. Both were based on

dynamic, underlying processes, and put little emphasis on classification schemes, as it was reasoned that overt symptoms would not be informative in regard to the true essence of the disorders (Maes & Horwitz, 2005). The DSM-III (1980), however, represented a revolution in the classification of psychiatric disorders (Maes & Horwitz, 2005). Explicit criteria were included, along with a multiaxial system for classification. The explicit criteria were designed to be neutral regarding the etiology of the disorders, as opposed to the criteria in the DSM-I and -II. The criteria should be of such a nature that they could be directly observed or reported by the patient (Torgersen, 2003). The DSM-III was intended for use both in clinical practice and for research. The development of DSM-III was facilitated by the work on the development of criteria to be used in diagnostic interviews (APA, 2000), as a result of the 1970's growing interest for classification of psychiatric conditions. Until this point in time, there had been little agreement on which psychiatric disorders should be included in a classificatory system, and how they should be organized (APA, 2000). Feighner and co-workers (1972) published an influential article where they suggested a set of criteria for psychiatric disorders (Torgersen, 2003), and shortly after, Spitzer and co-workers developed a set of research criteria, named the research diagnostic criteria (RDC; Spitzer, Endicott & Robins, 1978). Inconsistencies in the DSM-III made a revision necessary, and in 1987, the DSM-III-R was published (APA, 2000). The five revisions that have been done have gradually adopted more diagnoses, while also changing some of them.

The diagnostic systems have among other things been criticized for having an artificial categorical approach, and for the lack of ability to explain the extensive co-occurrence of psychiatric disorders (Widiger & Mullins-Sweatt, 2009).

Psychiatric Interviews

Psychiatric interviews are required to set a psychiatric diagnosis (Torgersen, 2003). The development of the psychiatric interview stems from the development of criteria for psychopathological conditions (Feighner et al., 1972; Spitzer, Endicott & Robins, 1978). With the development of structured and semistructured interviews, the ability to reliably diagnose psychiatric disorders has been much improved. Based on the RDC, the first psychiatric interview, the Diagnostic Interview Schedule (DIS), was developed (Torgersen, 2003). Later, several other diagnostic interviews have made their way, like the Schedule for Affective Disorders and Schizophrenia (SADS; Endicott & Spitzer, 1978), the Structured Clinical Interview for DSM-IV axis I Disorders for clinical utilization (SCID-I clinician version;

Spitzer, Williams, Gibbon, & First, 1992), the Structured Clinical Interview for DSM-III for research utilization (Spitzer & Williams, 1984), the Present State Examination (PSE), today called the Schedules for Clinical Assessment in Neuropsychiatry (SCAN; Torgersen, 2003), and the Composite International Clinical Interview (CIDI; WHO, 1993; Wittchen & Pfister, 1997). The DSM-IV axis II personality disorders are classified and scored using other diagnostic interviews, like the SCID-II, and SIDP (Segal & Coolidge, 2007).

The interviews that are used for diagnostic purposes vary tremendously in the amount of structure imposed (Segal & Coolidge, 2007). Unstructured interviews depend on the questions the clinician finds suitable to ask (Segal & Coolidge, 2007), whereas structured interviews are conducted in strict accordance with a standardized sequence, wording, and scoring of the questions. The emergence of standardized diagnostic interviews has made it possible to conduct research on psychiatric disorders. Typically, such studies investigate the etiology, comorbidity, and prevalence rates of psychiatric disorders, as well as treatment effects (Segal & Coolidge, 2007). Structured interviews can be used to make reliable and valid current and lifetime diagnoses. However, the diagnoses derived from structured interviews are not any more valid than the system that the interview is based on, namely the DSM or ICD systems (Segal & Coolidge, 2007).

The Notion of Psychiatric Comorbidity

The DSM-IV and the ICD-10 have some important limitations despite widespread use and practical utility. The concept of having more than one condition, what is called diagnostic comorbidity, is one limitation that many clinicians and researchers have addressed (Kendall & Hammen, 1995; Widiger & Sankis, 2000; Kendler & Prescott, 2006; Krueger & Markon, 2006). Mental disorders are found to be significantly correlated (Krueger & Markon, 2006), and meeting the criteria for one disorder, tends to predict meeting criteria for another disorder. According to the diagnostic systems, having several conditions on the same axis is not supposed to occur other than by chance. A single diagnosis should capture the essentials of the patient's condition. If the patient has more than one diagnosis, the diagnoses received should be independent of each other (Kendall & Hammen, 1995). Unfortunately, the diagnoses are not so well behaved (Kendler, 2009, symposium at the Norwegian Institute of Public Health). Comorbidity does occur, and in fact seems to be more the rule than the exception (Kendler & Prescott, 2006). Some patterns of comorbidity are also more widespread than others, making the notion of independence and co-occurrence only by chance

rather dubious. The comorbidity between major depression and the anxiety disorders is the cardinal example of such comorbidity patterns (e.g. Boyd et al., 1984; Hettema, 2008; Kendler et al., 1992; Kendler, 1996; Kendler et al., 2006; Kessler et al., 2005b; Maser & Cloninger, 1990; Roy et al., 1995; Torgersen, 1990). In an extensive review on the topic, Hettema (2008) points out two broad hypotheses that are generally applied to explain the relationship. The first hypothesis proposed is that one of the disorders is caused by, or is an epiphenomenon of the other. The second hypothesis is that both disorders have common etiological factors and that these stem from either genetic or environmental sources. Epidemiological studies give insight into the magnitude of psychiatric comorbidity, but they are not able to reveal any causal mechanisms behind the patterns of comorbidity (Kendler & Prescott, 2006). A more thorough method is required.

Investigating Patterns of Comorbidity using CFA: Findings of an Internalizing Factor

What is interesting to date is not really whether or not there is comorbidity, as there are hardly any doubt of its occurrence. What is interesting is to find the origin of the comorbidity patterns, and thereby be able to explain the findings. Exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) are powerful methods to explore the patterns of comorbidity. As opposed to EFA, and as is the case for structural equation modelling (SEM) in general, CFA tests the fit of models positing an a priori specified structure to the structure of the measured data (Bollen, 1989). The structure in these models is based on continuous latent factors that account for covariation between disorders. It is assumed that underlying latent continuous variables can account for the structure of covariation among the observed dichotomous indicators, hence CFA is able to explain the pattern of comorbidity between disorders (Krueger, 1999). CFA is conducted using SEM software, like Lisrel (Jöreskog & Sørbom, 1996). For the purpose of testing hypotheses about the origin of comorbidity patterns, both bivariate (Neale & Kendler, 1995) and multivariate comorbidity models are employed (Krueger, 1999; Kendler et al., 2003; Røysamb et al., 2009). The multivariate models often converge on a specific hierarchical model for comorbidity, and often include two superordinate liabilities: internalizing, and externalizing (Krueger & Markon, 2006), as will be reviewed shortly. Externalizing is a general liability toward developing disinhibitory disorders such as substance use disorders and antisocial behaviour disorders (Krueger & Markon, 2006). A liability in this case is an unmeasured propensity for developing a certain

type of disorders, as a result of a concatenation of processes leading from genes and environment to behaviour (Kendler & Prescott, 2006). Often used in regard to the notion of liability is a liability-spectrum model. This model involves a dimensional conceptualization of vulnerability to a disorder, and posits disorders as manifestations of latent risk factors varying along a continuum (Røysamb et al., 2009). Studies investigating the causes of comorbidity tend to use this assumption, and often present models that index the relations between the included disorders. These presented models can often be treated as submodels to an associated liability model (Neale & Kendler, 1995), which posits that each of the disorders in the model are influenced by one or several latent liability factor and that these liability factors are correlated (Krueger & Markon, 2006). Variants of the associated liability model are also the general empirical finding in large-scale epidemiological studies. The concept of liability will be discussed more thoroughly in the methods section.

Several phenotypic studies of comorbidity using CFA have been conducted recently. Krueger (1999) conducted a confirmatory factor analysis (CFA) on 10 common DSM-III-R mental disorders. Two main factors were found; an internalizing factor and an externalizing factor, with the internalizing factor having two subfactors; anxious-misery and fear. Two main factors were also found in a study on common mental disorders in the DSM-IV and ICD-10 by Slade and Watson (2006). One of the factors was labelled internalizing and encompassed two subfactors, as in Krueger (1999), and the other was labelled externalizing. The subfactors of the internalizing factor were labelled distress and fear. The distress factor included the disorders MDD, dysthymia, GAD, and post traumatic stress disorder (PTSD), and the fear factor encompassed the disorders social phobia, agoraphobia, panic disorder and OCD. Support for the aforementioned findings is stated in an extensive meta-analysis on different comorbidity structure models conducted by Krueger and Markon (2006). In this analysis, two superordinate factors of externalizing and internalizing disorders were found, and with the internalizing factor encompassing two subordinated factors.

Røysamb et al. (2009) recently conducted the first study using EFA and CFA on a comprehensive set of DSM-IV axis I and axis II disorders, using interview data from the Norwegian Intitute of Public Health Twin Panel (NIPHTP). Evidence was found of five spectra of disorders, and a higher order structure of these five factors encompassed by a distinction between internalizing and externalizing disorders. The findings point to comorbidity as a result of associations between broad liability factors and not as a result of disorder-specific associations. The factors are associated with the big five model of personality factors (Costa & McCrae, 1992), and the association between personality trait of

neuroticism and internalizing disorders is particularly emphasized. The study is in accordance with other studies that have investigated the sources of comorbidity patterns, in that the axis I disorders seem to cluster under two factors of internalizing and externalizing disorders (and the PDs under the three remaining factors). The study also represents an extended replication of earlier findings because of the comprehensive set of disorders analyzed, and the higher order structure imposed.

The results from the above mentioned studies do not correspond to the expanding number of psychiatric disorders in the classificatory systems, the DSM and ICD. According to studies investigating comorbidity, there should be fewer diagnoses and not more. Røysamb et al. (2009) suggests that the DSM-V should take a meta-structure of internalizing and externalizing disorders into account.

It must be added that CFA is not the only option for studying comorbidity patterns. As will be explicated later, twin modelling is an even stronger methodology for this purpose.

Heritability of Internalizing Disorders

Studies conducted the past two decades have found evidence that there is a substantial genetic contribution to virtually all psychiatric disorders (Bouchard, 2004; Kendler & Prescott, 2006; Nes, Røysamb, Reichborn-Kjennerud, Harris, & Tambs, 2007; Plomin, DeFries, McClearn & McGuffin, 2001). The environmental stressors that interact with the genetic vulnerability eventually decide the phenotype, i.e. the type of disorder to be manifested. Evidence of genetic factors for different internalizing disorders has been investigated in a large number of studies. For instance have twin and family studies consistently shown that MDD tends to run in families (Kendler & Prescott, 2006), and genetic factors are found to account for 22 - 65% of the variance in MDD (42% in Edvardsen et al., 2009; 65% in Foley, Neale & Kendler, 2001, 37% in Bouchard, 2004; Sullivan, Nelae & Kendler, 2000; and 22% in Ørstavik, Kendler & Czajkowski, 2007). Studies conducted using the Virginia Adult Twin Study of Psychiatric and Substance Use Disorders (VATSPSUD) found similar results, in that 30-40% of the variation in liability to lifetime MDD is accounted for by genetic factors (Kendler & Prescott, 2006). In regard to various anxiety disorders, the VATSPSUD studies found different estimates for the different anxiety disorders. The genetic factors that account for variation in liability to panic disorder was found to be in the area of 20-35%. For GAD the genetic factors account for 20-32% of the variation, and for phobias 12-43% of the variance, depending on the type of phobia, with agoraphobia being the most heritable phobic disorder

and social phobia the least. Sundet, Skre, Okkenhaug and Tambs (2003) found heritability coeffisients ranging from .22 - .47 for different types of self-reported fears. The different types of fears had both shared and uniques genetic factors. In a review, Bouchard (2004) reports heritability coeffiecients ranging from .20 - .40 for different anxiety disorders. A recent study by Tambs et al. (2009) investigated the structure of the genetic and environmental risk factors in a comprehensive set of axis I anxiety disorders. They found that most of the genetic effects (54%) were shared between the anxiety disorders. The heritabilities of the single anxiety disorders were found to be variable and lower than the combined heritability estimate. Scherrer et al. (2000) found heritabilities of .32 for GAD and .43 for panic disorder, which is exactly the same estimates as were found in a meta-study of Hettema, Neale & Kendler (2001). In another study, symptoms defined as broad panic have been found to have a heritability of .33 (Reichborn-Kjennerud et al., 2004). The personality trait of neuroticism (N) is also found to share a substantial amount of the genetic variation with internalizing disorders (Hettema et al., 2006).

As seen, moderate heritability coefficients are typically found for the different internalizing disorders. In the extension of these findings it is natural to ask if the genetic effects between these disorders are shared to some extent. Kendler et al. (1992) and Kendler (1996) investigated the relationship between MDD and GAD using twin modelling (see later for explication), and found that the genetic correlation between these two disorders was not significantly different from unity (1.0). Middledorp, Cath, Van Dyck, and Boomsma (2005) also concluded in their extensive review of twin and family studies, that comorbidity between anxiety and depression disorders can be explained by shared genetic liability.

Questionnaire Measures for Symptoms of Anxiety and depression

Diagnostic interviews are required to give exact DSM-IV axis I and axis II and ICD-10 diagnoses. These interviews are extensive, thus requiring time and resources to be conducted. Having a measure of global mental health that is less resource demanding is therefore necessary for purposes such as routine screening in family practice and in large population studies. The idea is to only have to conduct extensive interviews with respondents scoring above an established threshold on a self-report scale, and hence are likely to be psychiatric cases. In addition to screening purposes, self-report scales may be used for monitoring effects of different forms of treatment (Derogatis, Lipman & Covi, 1973; Strand, Dalgaard, Tambs & Rognerud, 2003). A great number of such self-report scales for clinical symptom rating have

been developed over the past years. The scales are comprised of a number of items often stated as assertions that respondents are to rate to what extent they agree with. This methodology started after World War I, when shortage of psychiatrists created a need for an alternative to the traditional psychiatric interview (Derogatis, 1983).

One of the most widespread self-report scales for screening purpose is the Symptom Checklist (SCL), with its many variants of length. As symptoms of anxiety and depression are especially prevalent in primary care and general hospital settings, a large amount of questionnaires have been developed that screen specifically for these symptoms. Examples of such scales are SCL- 25 (Hesbacher et al., 1980), and its short form SCL-5 (Tambs & Moum, 1993), the Geriatric Depression Scale (GDS), Beck's Depression Inventory (BDI), the Center for Epidemiological Studies - Depression Scale (CES-D), the Dutch Anxiety and Depression Scale, (DADS), the Depression Scale (DEPS), the Primary Care evaluation of Mental Disorders (PRIME-MD), the Symptom Driven Diagnostic System - Primary Care (SDDS-PC), the Zung Self-assessment Depression Scale (SDS), the Single Question (SQ), the Patient Health Questionnaire (PHQ) (see e.g. Williams et al., 2002), and the Montgomery-Åsberg-Depression Rating Scale (MADRS) (see e.g. Fröjdh, Håkansson & Karlsson, 2004). These scales have not been found to index significant differences in ability to screen for clinical depression (Williams et al., 2002). For instance, the correlation between the MADRS and SCL-25 is found to be .77 (Fröjdh et al., 2004), and .78 with the Mental Health Index (MHI; Strand et al., 2003). Development of various short forms has been driven forward by the need for quicker and more convenient screening instruments (Winokur et al., 1984). Often, such self-report scales will be part of large questionnaires, or applied in a hectic primary care facility, and so it has been an imperative to make scales that are both short and easily administered, without having to sacrifice too much reliability and validity. The SCL-25 is one of the most applied scales for measuring anxiety and depression symptoms, and has been used as a symptom screener in a wide range of settings (see e.g. Hesbacher et al., 1980; Hansson et al., 1994; Winokur et al., 1984). It is comprised of the anxiety and depression subscales, which are two of the nine original subscales from the full scale SCL-90 (Derogatis, 1983; Derogatis, Lipman & Covi, 1973; Derogatis et al., 1974) and SCL-90-R (Derogatis, 1983). In addition to these items, the SCL-25 has two items measuring vegetative symptoms (appetite and sleep), also derived from the SCL-90. Some of its items could be interpreted as measuring somatoform symptoms (Sandanger et al., 1998). Because symptoms of anxiety and depression accompany many different mental disorders, some argue that the SCL-25 can measure global

mental health (Tambs & Moum, 1993). The SCL-25 has demonstrated acceptable reliability and validity as a measure of psychological distress (Derogatis et al., 1974; Glass et al., 1978). Its internal consistency is estimated to be in the area of .93 - .95, depending on the sample (Hesbacher et al., 1980; Joukamaa et al., 1994; Strand et al., 2003). The administration of the SCL-25 is estimated to take approximately five minutes.

For research purposes, the various SCL-versions have been used both as continuous and dichotomous variables. For ordinal or dichotomous usage of SCL it is necessary to establish one or several cut-off points. The cut-off should have the characteristic of being able to predict the true presence of a mental disorder (Strand et al., 2003). Scoring above the cut-off is defined as being a ”case”, which means that the person’s score indicate more symptoms than would be expected from daily hassles (Strand et al., 2003). For instance, the conventional cut-off point for SCL-25 for defining a case is a mean score of 1.75, somewhat depending on diagnosis and sex (Hesbacher et al., 1980; Winokur et al., 1984; Nettelbladt, Hansson, Stefansson, Borgquist, & Nordström, 1993; Hansson et al., 1994; Sandanger et al., 1998; Moum, 1998; Strand et al., 2003; Fröjdt et al., 2004).

With the needs of short, valid and reliable scales as a backdrop, the SCL-5 was developed. The SCL-5 is a short form of the SCL-25, thus also intended to measure symptoms of anxiety and depression (Strand et al., 2003). This scale was developed by Tambs and Moum (1993), by using stepwise multiple regression to find a combination of items that explain a maximum proportion of the variance in the full scale, the SCL-25. Two of the SCL-5 items are derived from the anxiety factor: (1) feeling fearful; (2) nervousness or shakiness inside, and three from the depression factor: (3) feeling hopeless about the future; (4) feeling blue; (5) worrying too much about things. In most of the SCL-versions, including the SCL-90 and the SCL-90-R, the instruction is to rate to what extent within the last seven days one has experienced the symptoms indicated by the items. It has been stated, however, that a time window of only seven days is rather stringent (Pedersen and Karterud (2004). Therefore, in the SCL-5 version the respondents were asked to indicate to what extent they had been experiencing the aforementioned problems within the last 14 days, the response categories being: 1 = "not at all"; 2 = "a little"; 3 = "quite a bit", or 4 = "extremely". The score on each item was summed and divided on the number of items to give a mean score for each respondent.

The internal consistency of the SCL-5, measured by Cronbach’s alpha, has been estimated to be .87 in one sample (Strand et al., 2003), and .85 in another sample (Tambs & Moum, 1993). The SCL-5 has been found to correlate .92 with the SCL-25, thus explaining

84% of the total variance in SCL-25 (Tambs & Moum, 1993). This correlation can be interpreted as the degree to which the instruments tap onto the same construct (Tambs & Moum, 1993; Strand et al., 2003). Sharing only 20% of the items with the SCL-25 and still obtaining a correlation with the SCL-25 of .92, should mean that these shared items make a good job tapping onto the construct. A suggested cut-off for the SCL-5 is 2.0, and was initially set to give approximately the same prevalence rate of cases as the SCL-25 with a cut-off at 1.75 (Strand et al., 2003).

Given the properties of brevity and high correlation with the full-scale version, it is advisable to use the shortest versions of the SCL (Williams et al., 2002; Strand et al., 2003). The SCL-5 is included as a measure in the present study.

Methodology, Findings and Implications from Studies that have Validated the SCL against Interview-based Diagnoses

The correspondence between various SCL versions and DSM-IV axis I disorders have been investigated by a host of studies. Many of these studies have focused on the full scale SCL-90 and -90R version (e.g. Aben et al., 2002; Kennedy et al., 2001; Olsen, Mortensen & Bech, 2004; Pedersen & Karterud, 2004; Peveler & Fairburn, 1990; Williams et al., 2002), and some on the SCL versions that are intended to measure symptoms of anxiety and depression (SCL-25; e.g. Feightner & Worrall, 1990; Fröjdh et al., 2004; Joukamaa et al., 1994; Nettelbladt et al., 1993; Sandanger et al., 1998; Winokur et al., 1984; the SCL-10; e.g. Foley, Neale & Kendler, 2001; and the SCL-5; e.g. Loke, Nicklason & Burvill, 1996). Most of the studies use receiver-operating characteristics (ROC) (see e.g. Green & Swets, 1966; Hudziak, Copeland, Stanger & Wadsworth, 2004; Zou, O'Malley & Mauri, 2007) for validation, as ROC is an efficient methodology for evaluating and visualizing the diagnostic accuracy of different psychiatric screening devices and diagnostic tools. ROC may also be applied in order to establish an optimal cut-off point in which as many true cases as possible are included, while also keeping the number of false alarms low. Sensitivity and specificity estimates index the cut-off point's ability to detect the true cases. The sensitivity rate indexes a scale's ability to identify the true positives, i.e. those respondents that score above cut-off on a disorder. The specificity of a scale indexes the scale's ability to correctly identify the true negatives, i.e. the respondents that do not have a disorder. Sensitivity and specificity estimates are also possible to calculate by other means than ROC, e.g. from a cross table of the frequency of scoring below and above cut-off on the two variables that are investigated. The sensitivity and

specificity are reciprocal validity coefficients, which implies that optimization of one of them will lower the other. A familiar example of this may be the balancing of risk of Type I and Type II errors in statistical significance testing. Therefore, having a scale that is intended to exclude all false positives will also yield a lower sensitivity than a less strict scale. The optimal cut-off will therefore be dependent on the intended use of the scale, for instance whether it is for clinical or research purpose.

Most of the studies that validate questionnaires against interview-based diagnoses found a moderate to strong correspondence, with sensitivity and specificity coefficients ranging from .50-.90 (Fröjdt et al., 1994; Loke, Nicklasson & Burvill, Nettelbladt et al., 1993; Olsen, Mortensen & Bech, 2004; Pedersen & Karterud, 2004; Sandanger et al., 1998; Veijola, 2003). The studies are not directly comparable, as the validity coefficients will be affected by factors like different sample size, length of the questionnaire, type of interview, and the different prevalence rates obtained in the various studies. However, the obtained validity coefficients give an indication of the phenotypic correspondence between questionnaires and diagnoses. Apart from sensitivity and specificity estimates, some of the studies show interesting results regarding phenotypic correspondence between the questionnaire versions and diagnoses that are worth pointing out.

Pedersen and Karterud (2004) investigated six of the subscales intended to cover the neurotic or internalizing aspects in the SCL-90, arguing that these were the complaints most commonly presented in their clinical sample. One of the main findings was that the subscale's associations with their respective diagnoses were not strong enough for screening purposes. For instance, the dysthymia diagnosis showed no relationship with the depression subscale, and those with somatoform disorder scored higher on the depression scale than on the somatoform scale. An exception was found for the phobic anxiety subscale, which showed a sensitivity of .69 and a specificity of .74 in respect to agoraphobia. Several studies found that the screening ability of the SCL versions was not specific to the various DSM-IV axis I disorders (Sandanger et al., 1998; Woody, Steketee & Chambless, 1995). When non-specificity is found, one should expect that the SCL-questionnaires index fewer factors than assumed. This is what is found in a large number of studies using item response models or factor analysis. The studies generally fail to replicate the nine factor structure originally suggested by Derogatis (1974; 1983; 1994) and instead found evidence of unidimensionality (Bonynge, 1993; Carpenter & Hittner, 1995; Cyr, McKennafoley, & Peacock, 1985; Evenson et al., 1980; Holcomb, Adams, & Ponder, 1983; Holi, Sammallahti, & Aalberg, 1998; Lipman, Covi, & Shapiro, 1979; Moum, 1998; Olsen, Mortensen & Bech, 2004; Rauter,

Leonard, & Swett, 1996; Ronan, Dreer, & Dollard, 2000; Shutty, Degood, & Schwartz, 1986). In addition to diagnostic non-specificity, findings of high intercorrelations between the subscales (as in Olsen, Mortensen & Bech, 2004) also supports the notion of unidimensionality. Kendler et al. (1986) found that twin's responses to a checklist reflected a single underlying dimension that influenced symptoms of both anxiety and depression.

A second important finding in Pedersen and Karterud (2004) was that those respondents scoring above the cut-off point on one subscale also tended to do so on several of the other subscales. In addition to supporting the notion of unidimensionality, this result may be an indication of comorbidity. If a scale is unidimensional it requires that the items of the scale share much of the variance with the superordinate factor, and hence a specific symptom does not have to indicate a specific disorder. Bech (2004) argues that the total symptom score on the SCL corresponds to the concept of comorbidity. An additive function is thus assumed, in that a low symptom score would index low comorbidity and a high symptom score would index high comorbidity. If a scale is unidimensional and the dimension is a subordinate of a multitude of DSM-IV axis I disorders, then a respondent endorsing symptoms from several subscales should have a high risk of being comorbid. The results of several studies underscore this assumption of comorbidity. Sandanger et al. (1998) found that the SCL-25 had an only moderate ability to select out the individuals that should be further investigated for diagnoses, but had far better sensitivity for the comorbid cases. Sandanger et al. (1998) also found that the cases detected by the SCL-25 and by CIDI were often discordant. However, the concordant cases were often the comorbid cases. Typical disorders for these cases were MDD, and panic disorders and GAD, all typical internalizing disorders. Haver (1997) found that SCL is a sensitive case-finder in a sample comprised of female alcoholics, which is a population found to have high psychiatric comorbidity (Di Sclafani, Finn, & Fein, 2008). Peveler and Fairburn (1990) found a close correspondence between the SCL-90R and the neurotic symptoms in the diagnostic interview PSE, indexed by a high correlation of .72-.82, depending on the sample used. This finding may not be surprising, as there is found a close, but not completely overlapping relationship between neuroticism and internalizing disorders (Hettema et al., 2006). Bech (2004) points out that no official list of neurotic symptoms exists in the DSM-IV system, and one therefore have to assume that the total symptom score corresponds to comorbidity. Sandanger et al. (1999) found that the cases detected by the SCL-25 had more illness classificators like loss of function and help seeking than the cases detected by CIDI. The results may at first seem surprising, as one would assume that the most serious cases were the ones that qualified to a diagnosis. The finding makes sense, however, if

taking into account that the cases detected with the SCL-25 were often comorbid cases (Sandanger et al., 1998). And further it is natural to assume that having two disorders would cause more suffering than having one. These findings can also be interpreted as poor differential diagnostic precision in the scales, instead of indications of comorbidity.

These studies indicate a relationship between the SCL versions and the diagnostic interviews as mainly moderate, with some exceptions indicating a strong relationship. There seems to be one single underlying dimension accounting for the variance in the scales, both for the short-versions measuring internalizing symptoms of anxiety and depression, and for the full-scale version that is intended to measure a broader spectrum of symptoms. Based on this knowledge it is tempting to suggest that internalizing symptoms are the core of what is measured by the SCL-questionnaires. Non-specificity is found, hence the SCL should not be recommended as screening instrument for specific disorders.

The findings of the these studies fit reasonably well with the CFA studies investigating the notion of comorbidity (Krueger, 1999; Krueger & Markon, 2006; Røysamb et al., 2009) as the symptoms patterns found in SCL scores seem to be subordinating to one factor. The CFA studies generally found two factors, however. The unidimensionality found in the SCL-studies thus seems to indicate that SCL taps onto the internalizing factor, and not the externalizing factor. This makes sense in regard to the SCL-25 and the short form SCL-5, as these are intended to measure anxiety and depression, but is a somewhat more sensational finding regarding the full scale SCL-90 and -90-R, as this is intended to measure a broad range of symptoms of mental distress.

Self-report symptom scales like the different versions of the SCL-90 were initially intended as state-measures as opposed to trait-measures (Reichborn-Kjennerud et al., 2002). However, several studies are indicating that this assumption might not hold true, and that the SCL-versions really are measuring symptoms that are temporally stable and therefore reflects more trait-like than state-like aspects of psychological functioning (Winokur et al., 1984; Kendler et al., 1995; Nes et al., 2007). This finding should have implications in regard to the self-rating scale's relationship with diagnoses, meaning that it might be that the instruments are in fact measuring more of the same than have been assumed.

In order to investigate the correspondence between SCL and interview-based diagnoses as thoroughly as possible, it is not sufficient to apply different phenotypic analyses like those extracting sensitivity and specificity estimates. A more thorough methodology is required that is able to analyse the correspondence at another level. Twin modelling is able to delineate the phenotypic correlation into shared and unique genetic and environmental effects,

and hence offers valuable information about the nature of the correspondence between SCL and diagnoses.

Heritability of SCL

In studies using the SCL-5, the liability towards experiencing symptoms of anxiety and depression is found to have a heritability ranging from .25 to .56 (Nes et al., 2007; Reichborn-Kjennerud et al., 2002; Tambs, Harris, & Magnus, 1997). Other versions of the SCL-questionnaire have also been used in studies investigating the heritability of symptoms of mental distress. In a study using the SCL-10 questionnaire intended to measure symptoms of depression, Foley, Neale and Kendler (2001) found heritability estimates of .61 for the depression symptoms. In two SCL-30 studies by Kendler et al. (1994; 1995), the questionnaire was split into four factors (depression, panic-phobia, somatization, and insomnia) after a factor analysis, and heritability estimates for each factor were calculated. The two factors relevant for the present study are the depression and panic-phobia factors. The panic-phobia factor of the questionnaire was found to have a heritability of .41 for men and .35 for females (Kendler et al., 1995), and the depression factor was found to have a heritability between .30 and .37 (Kendler et al., 1994). It is important to note that it is not given that the estimates from different studies can be compared, as there are differences in the items selected, and the size and nature of the sample used. A somewhat divergent result was found in Tambs and Moum (1993b), where the heritability estimates of SCL-25 were estimated to be .20 or lower, using data from first- and second-degree relatives. As a consequence of the limitations of the design in this study, the heritability could not be estimated precisely. The upper limit of the heritability is quite accurate though, as the sample size of first-degree relatives was large. A heritability of .20 is substantially lower than is common for studies investigating heritability of symptoms of mental distress. However, the authors point out that most other studies use twins only to estimate the heritability, and that the study's use of other types of relatives may explain some of the divergence in the results.

As seen in the studies investigating the heritability of internalizing disorders and symptoms of internalizing disorders, there is a substantial genetic component both in the liability towards experiencing symptoms of anxiety and depression, and in the liability towards having an internalizing disorder. To date little is known about the extent to which these two liabilities are overlapping. A study by Foley, Neale and Kendler (2001) has addressed this question by using twin modelling. Before presenting the results from the Foley

et al. (2001) study, however, the mode of thought in twin analysis and calculation of heritability are presented.

Human Behaviour Genetics

Human behaviour genetics is broadly defined as the attempt to characterize and define the genetic or hereditary basis for human behaviour (Lorenz, 2003). The aim is to find out what makes individuals differ from the population mean on one or several traits. A trait is referred to as a phenotype, an observed characteristic about an individual resulting from the interplay of genes and environment (Plomin et al., 2001). By analyzing the phenotypic correlations between family members, inferences about the etiology of a trait can be made. The variance of the trait is then partitioned into what make family members similar to each other, and into what make them dissimilar. Monozygous (MZ) and dizygous (DZ) twins can provide powerful data for this purpose. In order to partition the variance into different components of genetic and environmental variance, the covariance between members of a twin pair on the trait is calculated. The variances and covariances of the trait scores can be represented in a variance-covariance matrix, where the diagonal represents the variable's variance with itself, and the off-diagonal elements represent the covariances. This kind of matrix is what is used for model fitting (as will be introduced later), and makes a mathematical model that is more powerful than the traditional calculation of heritability as presented below. The estimates provided are results of an iterative process, trying out a large number of parameter values before converging. The estimates are therefore more stable than the estimates obtained from the traditional method. In addition, the mathematical model makes estimation of standard errors possible.

The Single-gene Model: the Foundation of the Heritability Concept

When positing that a trait is heritable, it is implied that there is some genetic effect upon that trait (Plomin et al., 2001). A gene consists of at least two alleles, and is either homozygous (AA or aa) or heterozygous (Aa). Given random mating in a population, it is expected certain frequencies of the alleles A_1 and A_2 called p and q. These would be A_1A_1, A_2A_1, A_2A_2 at frequencies p^2, 2pq, q^2 (See e.g. Plomin et al., 2001, for evidence). Next, it is necessary to describe the effects of the alleles on a locus on the trait one is interested in. A locus is

associated with a trait if changing some of its alleles is associated with changes in the mean score of the trait. When the trait is quantitative, one needs to specify how much an allele contributes to the trait. If a locus has two alleles, A_1 and A_2, one can define the mean value for the homozygote A_1A_1 as *a*, and the other (A_2A_2) as *–a*. The value of the heterozygote A_1A_2 is *d* and is dependent on the dominance. If there is zero dominance, *d* is zero, and if A_1 is dominant over A_2, then *d* is bigger than zero. If the dominance is complete, the A_1A_2 heterozygote would be the same as A_1A_1 and *d* would be *a (Plomin et al., 2001)*.

Given what is known about the single-gene model, the genetic effects can be partitioned into additive and dominance effects. *Additive effects* (referred to as A*)* are the mean effects of the alleles. The additive genetic value is the genetic value that is expected from the number of alleles on the locus (either 0, 1 or 2). Every A_1 allele increases the individual's score with A units. If a parent has one copy of the A_1 allele, then there is a 50% chance that each offspring will inherit this allele. If the allele is inherited, then its effect on the phenotype will contribute the same amount as the parents' allele did to the phenotype, and therefore lead to increased parent-offspring similarity. *Dominance effects* (D) refer to the non-additive genetic effects. Inherent in non-additive effects is also epistasis. *Environmental effects* are not as easily partitioned as the genetic effects, but can be partitioned into what is shared and nonshared in a family. The shared family environment (C) contributes to family similarity, whilst the unshared environment (E) makes family members differ from each other.

The Polygenic Model

Complex traits such as psychiatric disorders are influenced by a large number of genes. In order to determine the effect of several genes on a phenotype, it is possible to extend the single-gene model to a polygenic model involving genes at many loci. This involves summing the additive and non-additive effects across the different loci (Plomin et al., 2001). The total genetic contribution to a phenotype is G, which encompasses the sum of all additive and non-additive effects, which is dominance, D, and epistatic interaction, I:

$$G = A + D + I \qquad (1)$$

Genetic and Environmental Contributions to a Phenotype

A phenotype, however, is not only the result of genes. Environmental contributions are also important, and the phenotype is therefore a function of both sources of variation. This can be written as an equation:

$$P = G + E \tag{2}$$

where P stands for phenotype, G stands for genetic effects, and E stands for environmental effects (Plomin et al., 2001). The equation represents an individual's deviation from the population mean.

Estimation of how much of the variance in P is attributable to the different G and E effects is possible by analyzing genetically related individuals (like the MZ and DZ twins in our example). The variance of the sum of two variables equals the sum of variances for both, plus two times the sum of their covariance. The variance for P is then:

$$Var(P) = Var(A + D + C + E) \tag{3}$$

where A is additive genetic effects, D is dominance effects, C is the shared family environment effects, and E is the unique environment effect which also contains measurement error (Plomin et al., 2001). And after writing out the sum of the variances and covariances, the equation is as follows:

$$\begin{aligned} Var(P) = {} & Var(A) + Var(D) + Var(C) + Var(E) + 2Cov(A, D) + 2Cov(A, E) \\ & + 2Cov(A, C) + 2Cov(D, E) + 2Cov(D, C) + 2Cov(C, E) \end{aligned} \tag{4}$$

(Plomin et al., 2001). This equation is not entirely correct, however, as A and D are independent of each other, like stated earlier. Also, it is assumed (somewhat incorrectly, as will be discussed shortly) that genes and environment are uncorrelated. In addition, the shared and unshared environment are uncorrelated by definition (Plomin et al., 2001). The equation should therefore be written as follows:

$$Var(P) = Var(A) + Var(D) + Var\ (C) + Var(E) \tag{5}$$

The Calculation and Definition of Heritability

MZ twins share all additive (A) and dominance (D) variance, all shared environment variance (C) and by definition none of the unshared environment (E) variance. DZ twins, on the other hand, share half of the variance in A, 25 % of the D-, all the C- and none of the E-effects (see e.g. Plomin et al., 2001 for evidence). The A and D together make up what one refers to as broad-sense heritability, and the A alone represents narrow-sense heritability, normally referred to as h^2 or a^2. Broad-sense heritability gives an indication of the extent to which genetic factors of any kind are responsible for trait variation in the population (Plomin et al., 2001). However, the narrow-sense heritability is the parameter most often reported in twin studies, as the structural equation models are unable to identify a model with all four of the variance components. The tradition is therefore to assume that only A adds to the genetic part of the variance. What contributes to the correlation or similarity between MZ twins on a trait is due to A and C, and is written as follows:

$$r_{MZ} = h^2 + c^2 \tag{6}$$

And for DZ twins:

$$r_{DZ} = h^2/2 + c^2 \tag{7}$$

Based on these equations, the equation for narrow-sense heritability (h^2) is:

$$h^2 = 2(r_{MZ} - r_{DZ}) \tag{8}$$

The rest of the variance that contributes to similarity is then allocated to c^2, as shown below:

$$c^2 = r_{MZ} - h^2 \tag{9}$$

Because a^2, c^2 and e^2 are proportions, they have to add up to 1.0. That is:

$$2(r_{MZ} - r_{DZ}) + r_{MZ} - 2(r_{MZ} - r_{DZ}) + e^2 = 1 \tag{10}$$

e^2 can then be found using the equation:

$$e^2 = 1 - r_{MZ} \quad (11)$$

because r_{MZ} is the h^2 and the c^2, and because every variance that is not shared between MZ twins must be due to variance from nonshared environment (e^2). These are some of the most important equations that twin modelling build upon, and that are applied in software for twin modelling.

Now that the calculation of heritability hopefully is clear, it is useful to address what the concept pertains to. Heritability is defined as the proportion of phenotypic variance that is attributable to genotypic variance (Plomin et al., 2001), and hence is a statistic used to describe the degree to which genetic differences between individuals cause differences in measured phenotypes. A common misconception about heritability is that it refers to individuals. However, the heritability statistic is based on the measured variance in a phenotype, and this variance is calculated based on scores from a sample. It is not the individuals per se that are interesting. Instead, what is interesting is the sample as a means to infer the true parameter of heritability, in the population that the sample is assumed to represent.

In addition to be applicable only on the group level, the heritability of a certain phenotype is not a constant and generalizable statistic. It varies over time and across groups and contexts (Plomin et al., 2001). It applies only to one population at one point in time, and for that particular composition of environmental factors present at the time. A typical example is height and its changing heritability in different countries. Individual differences in nutrition and other environmental factors may account for a larger part of individual differences in height in developing countries than they do in industrialized countries. This pattern is a result of the nature of the components of variance inherent in a phenotype. The components are proportions, and hence together add up to 1. If one of the components is low, then the remaining variance is attributed to the other components.

The preciseness of heritability is also a misconception. Heritability is distorted by unreliability of measurement instruments and other measurement errors (Plomin et al., 2001). This limitation implies that the confidence intervals of the heritability estimate are important for its interpretation.

In spite of the limitations of the concept, however, heritability is a frequently applied statistic in the human behaviour genetics area. It gives an indication of the importance of

genes in a phenotype, and from this information it is possible to derive answers to several questions. Typically, one may be interested in the relative contribution of genetic and environmental sources to the variance of a certain phenotype, or, in a multivariate design, to estimate to what extent these sources are shared between phenotypes.

Assumptions

The genetic and environmental effects upon a phenotype do not work independently of each other. Effects like assortative mating, gene-environment correlation and gene-environment interaction complicate the analyses. In order to be statistically able to estimate the different amounts of genetic and environmental contributions to the variance of a trait, several assumptions therefore have to be made. These assumptions are visible in the model fitting.

The genetic effects are not always straightforward, as non-additive effects like dominance and epistatis also may contribute to phenotypic variation. One assumption that is often made is that there are no dominance effects, only additive effects. The assumption is empirically based, as the phenotypic correlation matrices are first inspected in order to find out if there is reason to drop the D in the equations. If it is not found that the MZ correlation is less than twice the DZ correlation, which is a result likely due to dominance effects, then fitting an ACE model is justified (Plomin et al., 2001). The assumption is visible in the model fitting in that the D is normally dropped (as it is seldom found a pattern of correlation that indicates non-additive effects), and an ACE-model is stated. The reason for dropping either the D or C parameter is that it is not possible to identify a model requiring that many parameters to be estimated (see later for explication of model identification). If the assumption of no dominance effects does not hold true, and an ACE model is fitted to the data, then the heritability estimates would be biased (Plomin et al., 2001).

Moving over to the environmental assumptions, it is often assumed that there is an equal environment for MZ and DZ twins, which means that the MZ twins are not assumed to be treated more equally by the environment than the DZ twins. This implies, as in equation 6 and 7, that the c^2 is of equal size for both MZ and DZ twins. Equal behaviour within MZ twin pairs can be attributed to their shared genes. If the assumption of equal treatment from the environmental for both MZ and DZ twins is violated, and MZ twins are treated more equal than DZ twins, this would lead to overestimations of heritability (Kendler & Prescott, 2006; Sundet, Eriksen & Tambs, 2008). However, this assumption has been the object of empirical

investigation, and seems to hold true for most traits (e.g. Bouchard & Propping, 2001; Derks, Dolan & Boomsma, 2006; Kendler et al., 1995).

The assumption of random mating posits that the genotypes of mates are uncorrelated (Kendler & Prescott, 2006; Neale & Maes, 2000). The opposite of random mating is assortative mating, that is mating someone genetically similar to one self. Assortative mating would tend to bias the estimates toward weaker genetic impact and stronger impact from the shared environment (Kendler & Prescott, 2006). This assumption can be accounted for by including appropriate parental information (Plomin et al., 2001). This has been done for instance in a study by Tambs et al. (1993b), who could not find evidence of assortative mating.

Genes and environment are assumed to be uncorrelated, as indicated by equation 5. However, this is a simplifying assumption, as humans often change their environments in part for genetic reasons (Plomin et al., 2001). Gene-environment correlation reflects a distribution of environment among different genotypes that are non-random (Neale & Maes, 2000). Several types of gene-environment correlations are possible, one being passive correlation, when the environment in which individuals develop is provided by their biological relatives (Neale & Maes, 2000). An example may be a home of scholars, which filled with books may encourage reading for the offspring with genetic potential to intellectual activities (Plomin et al., 2001). The other types of gene-environment correlation are evocative and active correlation. An example of evocative gene-environment correlation is the MZ twins that evoke a more similar treatment from the environment as a cause of their shared genes. Active gene-environment correlation is when an individual with a particular genotype seeks out particular environments (Larsen & Buss, 2008). Gene-environment correlations have by Kendler and Eaves (1986) been described as genetic control of the environments. Adoption data provide a test for the presence of passive gene-environment correlation (Neale & Maes, 2000).

Gene-environment interaction is present when genes control the sensitivity to differences in the environment (Neale & Maes, 2000). This is asserted for instance in individual differences susceptibility to nicotine. Adoption data are required to detect gene-environment interaction.

New Applications of Twin Analyses - the Multitrait Approach

Twin research is increasingly moving away from the question of to what extent a single trait is heritable. Single trait analyses have been conducted in order to understand the sources of variability among people in risk of developing for instance psychiatric disorders. This is done by investigating the extent to which genetic and environmental differences can account for the variability. Twin research today is more concerned about investigating the sources of covariation among people (see e.g. Kendler & Prescott, 2006). One important example of this focus is the research done on the sources of patterns of comorbidity. Even though CFA methodology gives valuable information on this issue, twin modelling is an even more powerful approach. As for the CFA methodology, the question of sources of comorbidity necessitates investigation of several traits or disorders at a time (called bivariate or multivariate analyses). In addition to find common factors that account for the comorbidity between variables, like CFA does, twin modelling renders possible delineation of the etiological sources into genetic and environmental contributions, and estimates the degree of correlation between these factors. If a strong correlation is found between parameters, it is assumed that these parameters are shared to some extent between the traits (see Figure 1). Inherent in this mode of thinking is the idea of a shared liability between disorders that can account for common patterns of comorbidity (see e.g. Krueger & Markon, 2006). The findings in the comorbidity research are thus also adding support to the emerging liability spectrum conceptualization of psychopathology, which to date is not inherent in the existing categorical diagnostic systems, the DSM-IV and the ICD-10.

In the case of the SCL-5 and internalizing disorders, we have some information about the heritability of these variables, as referred earlier. What has not been investigated, however, is the extent of shared genetic and environmental effects. A study by Foley et al. (2001) addresses this question.

Genetic and Environmental Overlap in a Questionnaire Measure of Depression and an Interview-based Depression Diagnosis

The Foley, Neale and Kendler (2001) study is, to our knowledge, the only study thus far that have investigated the relationship between SCL and diagnoses using twin modelling.

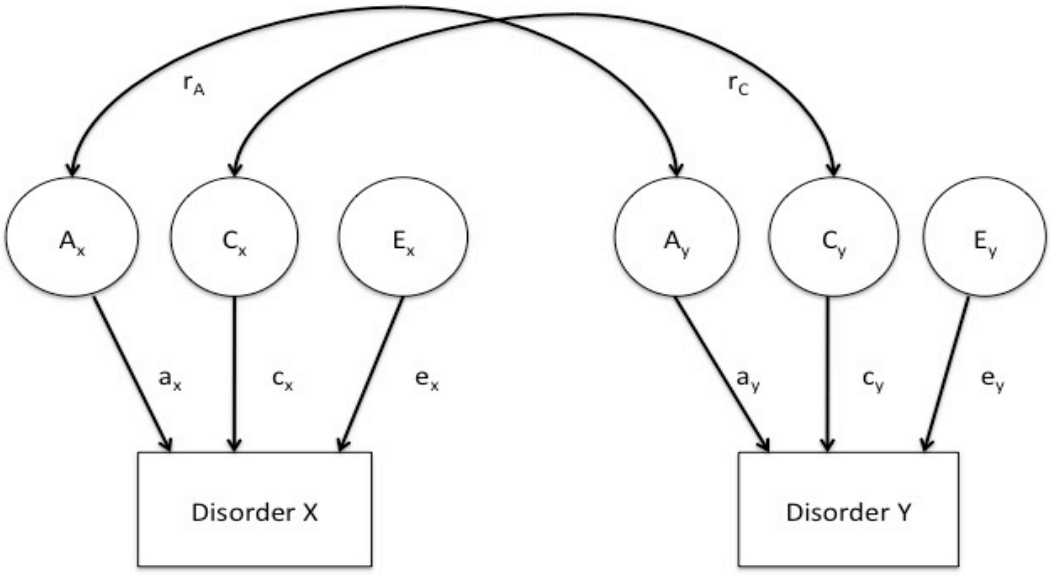

Figure 1: a bivariate twin model, indexing paths for correlation between the parameters of the disorders (r_A and r_C) (from Kendler & Prescott, 2006). For the sake of simplicity, the model indexes only one of the twins in a pair.

By estimating parameters of shared and non-shared genetic and environmental variance, they wanted to find out whether a respondent's self-rated depression symptoms index the genetic or environmental liability to major depressive disorder. Ten depression items from the SCL-90 were used to tap depression by self-report, and the SCID was chosen as structured interview. Only the relation to the DSM-IV diagnosis major depression (MDD) was investigated, using a sample of female twins. The genetic correlation between the SCL-10 and MDD was found to be .83 and the non-familial environmental correlation was .17, which means that the genetic overlap is almost complete. They concluded that the SCL-10 is a useful screening device for many of the genetic risk factors that influence the liability to MDD. These results give a new impression of the relationship between the SCL and diagnoses. Instead of finding the relation to be moderate at best, as the phenotypic studies tend to, it is now indexed as strong. However, because this study is the only study to have investigated the relationship between the SCL and diagnoses using a twin modelling approach, it is necessary to determine whether these results can be replicated in other samples, and for other diagnoses. For instance, is the pattern the same in a male sample, or in a Scandinavian sample? Would a larger sample be able to detect more shared or non-shared variance due to more statistical power? And will the results extend to internalizing disorders in general?

Summing Up and Aim of the Present Study: What do we Know and What do we Not Know

There are three premises for the correspondence between symptoms of anxiety and depression as measured by various forms of the SCL-questionnaire and diagnostic interviews. The first premise is the correspondence found in studies validating the SCL against psychiatric diagnoses. The correspondence that is found is not impressive. However, the nature of the correspondence cannot be fully revealed by phenotypic studies alone. The second premise is the results from the investigations of patterns of comorbidity in DSM-IV and ICD-10 disorders, and unidimensionality of the SCL scale. These patterns suggest that these instruments may both be explained by a superordinate, internalizing factor. To what extent it is the same factor however, is not known a priori. The third premise is the heritability estimates found in both SCL and internalizing disorders in that they both show substantial heritability. This premise casts some light on the first and second premise, in that there is a phenotypic correspondence, and also a substantial amount of genetic factors in both instruments. Is it so that the factors encompassing the patterns of comorbidity and unidimensionality in the two instruments share the same underlying genetic liability? Till now, we have not been in position to answer the question of genetic overlap, although we know that Foley et al. (2001) found a strong genetic correlation between depression and the SCL-10.

The aim of the present study is to investigate the correspondence between the SCL-5 and internalizing disorders using both phenotypic analyses and twin modelling. Twin modelling can delineate the contributions from genetic and environmental sources to the phenotypic correspondence. In light of the comorbidity findings regarding internalizing disorders, it will be tested in the present study whether the strong genetic correspondence found in Foley, Neale and Kendler (2001) can be generalized from depression to internalizing disorders in general. We will also investigate the relationship using both sexes in the analyses and hence test for sex differences in the parameter estimates. Our hypothesis is that the scores on the two instruments are a result of the same underlying genetic liability.

Methods

Sample

The data for the analyses in the present study come from the Norwegian Institute of Public Health Twin Panel (NIPHTP). The twins are identified through information contained in the

national Birth Registry, established January 1, 1967, which receives mandatory notification of all live- and stillbirths of at least 16 weeks of gestation (Medical Birth Registry, 1987). The current panel consists of information from all 15370 like- and unlike-sexed twins born between 1967 and 1979. Two questionnaires have been collected, the first one in 1992 (twins born 1967 and 1975), and the second in 1998 (twins born from 1967 and 1979).

The first questionnaire was sent to 3996 twin pairs, and the responses were received from 5864 twins after one reminder (2570 pairs and 740 twins whose co-twin did not respond). This gives a response rate of 74%. The responses from this first questionnaire are not analyzed in the present study.

The second questionnaire was sent to 12700 twin pairs, and responses were received from 8045 twins after one reminder (3334 pairs and 1377 twins whose co-twin did not respond). This gives a response rate of 63%. Of the 8045 twins that responded, 7992 had valid responses that could be used in the analyses of the present study. Invalid questionnaire responses include non-complete responding, in addition to scanning - and registration errors (Knudsen, 2009, personal communication). 9478 twins responded to at least one of the questionnaires. Data used in the present study come from the second questionnaire, and from a diagnostic interview of axis I and axis II Psychiatric Disorders, conducted between June 1999 and May 2004.

The interview participants were recruited from a sample of 3153 complete twin pairs from the second questionnaire study who had given consent to be contacted again later, and 68 twin pairs drawn directly from the NIPHTP. Twins were not approached for interview until preliminary consent had been contained from both members of a pair. 2801 twins were interviewed with the Composite International Diagnostic Interview (CIDI) and the Structured Interview for DSM-IV Personality (SIDP-IV). Of these 2801 respondents, only 2793 responses were valid, and hence 2793 are the number of twins from the interview sample that are analyzed in the present study. The response rate was 44%. Non-participants consisted of 0.8% pairs not willing or able to participate, 16.2% pairs in which only one twin agreed to participate, and 38.2% pairs in which none responded after reminders. The high rate of attrition from the questionnaire studies to the interview study has recently been investigated by Tambs et al. (2009). The attrition was found not to affect twin analyses of mental health related variables, in spite of a moderate selection effect towards good mental health. 2562 of the interviews were conducted face-to-face, and for practical reasons 231 were interviewed over the phone. The interviews were mainly conducted by psychology students late in their training and psychiatric nurses. The interviewers received training by one psychiatrist and two

psychologists with extensive experience with the instrument. Members of a pair were assessed by different interviewers that were blind to the information obtained from the co-twin. The number of twins that completed both the second questionnaire and the CIDI interview was 2758.

Zygosity was determined using questionnaire items previously shown to classify correctly 97.5% of the twin pairs (Magnus, Berg, & Nance, 1983), and molecular markers for a subgroup of the sample, based on genotyping 24 microsatellite markers. Discrepancy between classification based on the questionnaire and DNA markers implied an expected misclassification rate of 0.67% for the whole sample. The NIPHTP is described thoroughly in Harris, Magnus, and Tambs (2002).

Approval was received from the Regional Ethical Committee and the Norwegian Data Inspectorate, and written informed consent was obtained from the participants after complete description of the study.

Measures

The short form SCL-5 consists of 5 items on a four-point scale ranging from 1 = "not at all" to 4 = "extremely". In this investigation the threshold established by Strand et al. (1998) at 2.0 was employed. The choice of dichotomizing the SCL-5 variable will be explicated later in the methods section. The SCL-5 correlates 0.92 with the SCL-25 and has satisfying reliability properties with a Cronbach's alfa of .85 in one sample (Tambs & Moum, 1993) and .87 in another (Strand et al., 2003).

The participants were assessed using the Norwegian version of the computerized Munich Composite of International Diagnostic Interview (CIDI) (Wittchen, & Pfister, 1997). This is a comprehensive structured diagnostic interview developed by the World Health Organization for the assessment of DSM-IV axis I diagnoses and ICD-10 lifetime diagnoses. The interview has previously shown good test-retest and interrater reliability (Wittchen, Lachner, Wunderlich & Pfister, 1998; Wittchen, 1993). A built-in algorithm in the computerized scoring program dichotomized the responses to either 0 or 1. The following disorders were selected for the present study: Major depressive disorder (MDD), dysthymia, generalized anxiety disorder (GAD), social phobia, panic disorder, and agoraphobia. The diagnoses were assigned without diagnostic hierarchical rules in order to examine co-occurrence without exclusions, in accordance with previous studies (Kessler et al., 2005b, Røysamb et al., 2009).

Statistical Analyses

Before describing the statistical procedures conducted in the present study, the logic behind the main statistical procedure, structural equation modelling and structural equation modelling using twin data, are presented.

The Essence of Structural Equation Modelling (SEM)

SEM is a multivariate statistical procedure that encompasses factor analysis and regression analysis, and subordinates confirmatory factor analysis (CFA). Like CFA, SEM in general deals with both latent variables, which are abstract constructs that cannot be observed directly, and observable, measured variables. The values of the latent variables can only be inferred from the measured variables. If there are multiple indicators, i.e. tests or measures of the latent variable, and these indicators covary, this common source of variance is attributed to the latent variable. The latent variable is often said to cause the values in the measured variables. This is an assumption, because there might be other causes as to why indicators covary (e.g. one indicator causing another). Structural equation models are divided into two parts: a measurement model, which deals with the relationship between the measured variables and the latent variables, and a structural model, which deals with the relationship between the latent variables only (Bollen, 1989). The inclusion of a structural model, in addition to the measurement model also found in CFA, renders it possible to elaborate the causal relationships between the included variables to a greater degree than is possible in CFA. The parameters in SEM are variances, regression coefficients and covariances between variables, and the data used to estimate these parameters are either sample variances and covarianses, correlations, or raw data taken from a population. SEM is based on path analysis developed by Sevall Wright in the 1930s (Neale, Boker & Maes, 2003), which among other things contributes to make an easy to grasp graphic representation of the models. Path analysis also includes a number of tracing rules for obtaining information about the relationship between the variables in a model (see e.g. Plomin et al., 2001).

Mx

Several software programs using SEM have been developed, for instance LISREL (Jöreskog & Sørbom, 1996), and Mx (Neale, Boker & Maes, 2003). Mx is a software developed specifically for twin data analysis. As for SEM in general, the data that can be used to estimate parameters are either sample variances and covariances, correlations, or raw data. In Mx, the raw data option has several advantages, such as including the variance of single responders (Neale et al., 2003). Using raw data implies that Mx tests the model against every case in the sample, instead of the observed variance-covariance matrices, resulting in a large number of degrees of freedom. However, using raw data is extremely computationally demanding, requiring a running time of hours and even days for multivariate analyses.

Mx is also efficient in conducting multi-group analyses, that is analysis of several data sets simultaneously. This is because in the context of twin data analyses, one operates with several groups of subjects as some are MZ and some are DZ, where some are like-sexed and some are unlike-sexed. The formal model describing the variance-covariance structure in terms of genetic and environmental variance is therefore allowed to differ between the groups (Posthuma & Boomsma, 2005). The data on opposite sex twins renders it possible to investigate whether genetic and environmental effects exercise different effects on the phenotypes for males and females (Czajkowski et al., 2008). There are at least two types of sex differences. When it is assumed that the same genes influence a trait in both genders, but to a different degree, this is referred to as quantitative sex differences (also referred to as non-scalar or general sex differences). When different genes are assumed to affect the same trait, this is referred to as qualitative sex differences (or scalar or common sex differences). It is possible to test for qualitative sex differences by freeing the parameter that specifies the correlation between the DZ-twin groups to be lower than .5. The genes shared between DZ twins are per definition 50%, and if they are allowed to correlate less than .5, it would imply the effects of different genes. Testing for quantitative sex differences is done by allowing the A, C and E parameters to vary between the male and female twins.

The Mode of Thought in Model Fitting on Twin Data

Model fitting in general involves construction of one or several models that attempt to describe the observed data as closely as possible, while also striving for keeping the model as parsimonious as possible, i.e. finding the best fitting model with the fewest parameters. The

model is often formulated as variance-covariance matrices, consisting of several parameters, most often the a^2, c^2 and e^2 components of variance. Various combinations of different values on these components will produce different expected variance-covariance matrices (Plomin et al, 2001). Below is an example of the variance-covariance matrices for a hypothetical measured phenotype for MZ and DZ twins, with the variances in the diagonal, and the covariances on the sides:

$$\begin{bmatrix} Var_1^{MZ} & \\ Cov_{12}^{MZ} & Var_2^{MZ} \end{bmatrix} \quad \begin{bmatrix} Var_1^{DZ} & \\ Cov_{12}^{DZ} & Var_2^{DZ} \end{bmatrix}$$

It is first assumed that this is a trait that is caused by different genetic and environmental components of variance, more precisely by A, C and E. It is also often assumed that there are no dominance effects (D), only additive genetic effects. Expressed as variances, this is written:

$$\sigma^2_P = \sigma^2_A + \sigma^2_C + \sigma^2_E \qquad (10)$$

To construct a twin model, every element of the variance-covariance matrix in form of the model parameters has to be explicitly written out. Covariance between twins is a function of the components of variance, and to what extent these are shared between the twins. As stated earlier in equations 4 and 5, MZ twins share all the a^2 and c^2 variance, and DZ twins share half the a^2 variance and all the c^2 variance. This can be stated in two variance-covariance matrices, one for the MZ and one for the DZ twins, respectively:

$$\begin{bmatrix} \sigma^2_A + \sigma^2_C + \sigma^2_E & \\ \sigma^2_A + \sigma^2_C & \sigma^2_A + \sigma^2_C + \sigma^2_E \end{bmatrix} \quad \begin{bmatrix} \sigma^2_A + \sigma^2_C + \sigma^2_E & \\ \frac{\sigma^2_A}{2} + \sigma^2_C & \sigma^2_A + \sigma^2_C + \sigma^2_E \end{bmatrix}$$

These two matrices represent a univariate model. The univariate model is also often displayed in path analysis formalism, and may be referred to as the classic twin model (see Fig. 2).

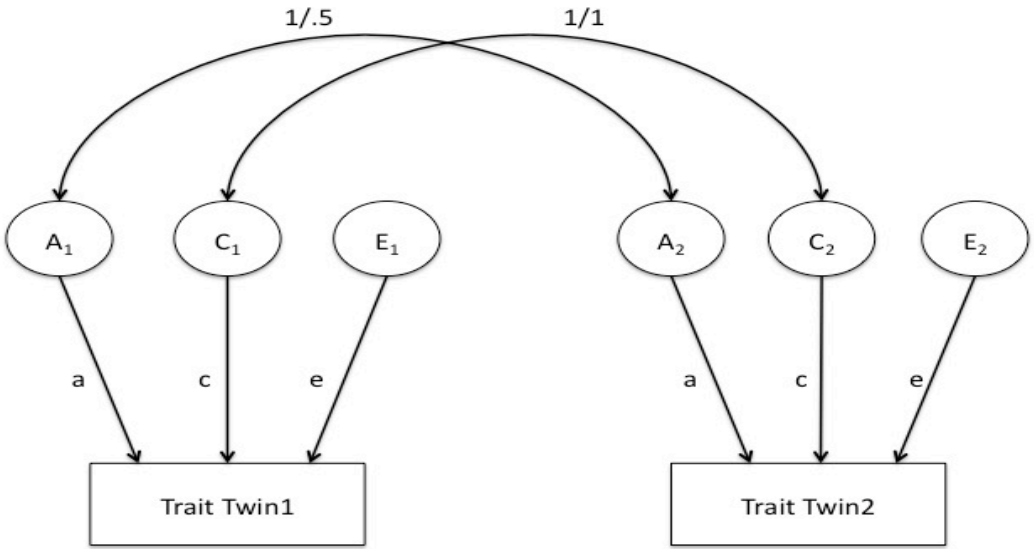

Figure 2: The classic twin model displayed in path analysis formalism. The variance components for both twins are displayed as the a, c and e paths. The correlation between the additive genetic components (A_1 and A_2) is by definition 1.0 for MZ twin pairs, and .5 for DZ twin pairs.

After a model is constructed, the next step is to specify the model into a script. The script states which of the parameters in the model that are free to be estimated and which of the parameters that are fixed to zero. Stating the model tells the SEM software which relationships in the model that generate the observed variance-covariance matrices, and is therefore an important step. Running the model then results in estimates of what set of parameter values would best reproduce the observed variance-covariance matrices, and therefore give the best fit. The process of finding the best fitting parameter values is called optimization and is highly computationally demanding. For multivariate analyses, this parameter fitting process often requires days and even weeks of computer running to converge. The estimation process is iterative in nature, trying out a great number of parameter values that might fit. Specifying good starting values in the syntax makes the analysis quicker and less computationally demanding, and reduces the chance of getting trapped in local minima during parameter estimation. The starting values could be determined from parameter values found in similar studies, or generated from theory, or by observing the correlation matrix. A good fit indicates that the observed covariance matrix (S) very much looks like the model's estimated covariance matrix (Σ) and hence the residual matrix ($S-\Sigma$) has small values. Or, said in a different manner: the null hypothesis $\Sigma = \Sigma\theta$, generated in the model specification step, where Σ is the estimated covariance matrix, and θ is the fixed and free

parameters, is compared to the observed covariance matrix S, and found not to differ significantly (Bollen, 1989). The process of finding the best fitting parameter can use different functions, like the maximum likelihood (ML), the generalized least squares (GLS) or the unweighted least squares (ULS). ML is often preferred, as it is found to be robust, even with small samples (Hoyle, 1995). ML expresses the likelihood of a model as a function of observed data and model parameters. This is measured as log-likelihood, and abbreviated LL. The difference in LL approaches a Chi-square distribution, and this way one can check for significant deterioration in χ^2 in two nested submodels, calculated as $\Delta\chi^2$-Δdf. A Chi-square distribution is a distribution of standardized scores, applied in the case of skewed samples (see e.g. Witt & McGrain, 1986). If the difference in χ^2 is non-significant, then the model has an acceptable fit. However, to be able to find the best model among several good-fitting ones, the Akaike Information Criterion (AIC) calculated as χ^2-2df may be applied (Neale et al., 2003). This way of calculation penalizes the models that are not parsimonious (Bollen, 1989). The best model is then reflected by the lowest AIC-value (Akaike, 1987). The root mean square error approximation (RMSEA) fit index is an often used fit index in structural equation modelling, but is not available using the raw data approach, as the models are tested against each case and not against a basis model.

To have confidence in the goodness of fit test requires a sample of at least 100 (Hoyle, 1995). If the sample gets very large, however, this might give the analysis so much power that almost any model is able to reject the null-hypothesis (see e.g. Neale & Maes, 2000 for details).

Model Trimming: the Saturated Model and Nested Submodels

To be able to estimate the specified parameters, the model has to be identified. An identified model is a model that has as many or more parameters as there are observations, and is also referred to as a saturated model (for identification rules, see Bollen, 1989). A saturated model will have zero degrees of freedom and always gives a perfect fit (χ^2 goodness of fit = 0). This is because there are no other sets of parameter values that can produce the same expected variance-covariance matrices. Having a perfect fit, though, does not mean that these parameters are the true values in the population. They will only represent the true values if the model is a good model (Plomin et al., 2001). The model fits out of pure statistical necessity, and this perfect fit is therefore not really valuable. However, if a model fits the data after dropping one or more of the latent variables, one could have more confidence in this being a

good estimation of the true parameter values. This is what is referred to as model trimming. By dropping parameters from the analysis, nested submodels are obtained. These are the result of fixing one or more of the pathways to zero, in order to test for significant contribution of genetic and environmental effects. For instance, fixing the shared environmental contribution C to zero yields an AE model, and fixing the genetic contribution A to zero yields a CE model. Unfortunately, with an unsaturated model, there is no unique set of parameter values that fits the observed data perfectly, but several that will fit the data reasonably well. Theory should therefore accompany the numbers in choosing the best model.

At the end of the analysis, parameter estimates are standardized for final model presentation. When parameter estimates are standardized, they can be interpreted with reference to other parameters in the model and relative strength of pathways within the model can be compared. Hence for univariate analyses, one can answer questions like; what is most important for this specific trait – genes or the environment, and how much does each contribute to the phenotypic variance? For multivariate analyses, one can also check for correlations between the latent variables, and thereby find out if the covariance between two or more traits is due to shared genes or shared environment. This ability renders it possible to answer questions about the etiology of different traits.

Twin Pair Similarity and the Liability Threshold Model

One of the core features in analyses on twin data is correlation statistics. Different types of variables require different types of correlation estimation procedures. Pearson product-moment correlation is used for continuous variables, and tetrachoric correlation and polychoric correlation are used for dichotomous and ordinal variables, respectively. Examples of dichotomous traits are mental disorders as conceptualized by the diagnostic system DSM-IV and ICD-10, where established cut-offs related to a certain number of diagnostic criteria endorsed determine whether one has a disorder or not. This is an operational definition in the existing diagnostic classificatory systems. However, many would argue that diagnoses, in particular axis II disorders, are best described using a dimensional approach (Widiger, Costa & McRae, 2002). In the liability threshold model, mental disorders are assumed to arise as a cause of several risk or liability factors adding up, and that those with liability above the threshold expresses the trait (Plomin et al., 2001). The underlying liability distribution of a disorder is further assumed to be normally distributed (Kendler & Prescott, 2006). Estimation

of tetrachoric correlation relies on this assumption. Measuring the variance of a dichotomous trait is more difficult than with a continuous trait. As one cannot observe the liability, one assumes that it has a variance of 1 (unity) (Plomin et al., 2001). With changes in the variance, the number of individuals under and above the threshold will also change. Higher thresholds result in lower prevalence rates than low thresholds. Tetrachoric correlation is vulnerable to low sample sizes. With a small sample, the 95% confidence intervals are wide. Therefore, in order to have enough power to estimate parameters with reasonable confidence, the sample sizes should be large. Tetrachoric correlation is applied in the present study.

The correlations between twins and the measured traits give valuable information. The within-twin cross-trait correlations are the same as phenotypic correlations for any individual regardless of zygosity. This is because when the twin is compared to him- or herself, the degree of genetic resemblance is not included, as it is in the cross-twin analyses, where the two twins in a pair and across pairs are compared to each other. It is possible for two traits to have a phenotypic correlation of zero, and still be related because of shared common genetic or environmental sources of variance. The cross-twin within-trait correlation compares twin1 and twin2's scores on the same variable, and hence it is possible to inspect the correlations for increased similarities between the zygosity groups. If MZ-twins generally obtain higher cross-twin within-trait correlations, it is implied that the trait is heritable, to some extent. If the MZ twins' cross-twin within-trait correlation is less than twice the DZ correlation, then this is a pattern that suggests effects from the shared environment. If the MZ correlation is more than twice the DZ correlation, then this may imply that there are non-additive genetic effects on the trait. The cross-twin cross-trait correlations encompass the correlation between twin1 and trait2 and twin2 and trait1. The information from the cross-twin cross-trait correlations is used to understand the sources of comorbidity (Kendler & Prescott, 2006).

The Present Study

Assuming that the liability to experience symptoms of anxiety and depression is normally distributed, the SCL-5-variable was dichotomized to 0 for those under the threshold and to 1 for those equal to and above the threshold 2.0. We further assume that those scoring below and above the threshold reflect different degrees of severity of an underlying continuum of liability to experience symptoms of anxiety and depression. The dichotomization of the SCL-5 is discussed under limitations in the discussion section. However, in a preliminary analysis we found that the phenotypic correlation with the internalizing factor did not change

significantly whether we applied a dichotomous SCL-variable or an ordinal SCL-variable with two thresholds.

As initial measures of the importance of genetic and environmental influences, within-twin correlations and cross-twin correlations were calculated for all the variables. As seen above, when applying dichotomized variables, tetrachoric and polychoric correlations should be applied, as Pearson product-moment correlations would tend to underestimate the true measure of association.

Confirmatory Factor Analysis

A confirmatory factor analysis (CFA) with the six DSM-IV axis I disorders was performed in Lisrel 8.80 (Jöreskog & Sörbom, 1996), in order to test the validity of the assumption of a one-factor structure for the internalizing disorders. PRELIS was used to create tetrachoric correlation matrices and asymptotic covariance matrices from the six internalizing disorders MDD, GAD, dysthymia, panic disorder, social phobia and agoraphobia. A one-factor structure was tested, in which all disorders were presumed to be indicators of a single, unitary propensity to experience internalizing disorders. The method of Maximum Likelihood (ML) was used to estimate parameters. For estimation of the model's fit to the data, we used multiple fit-indexes: the goodness of fit statistic (GFI), the comparative fit index (CFI) and the root-mean-square-error of-approximation (RMSEA) (Hoyle, 1995). GFI above .90 and CFI scores above .95 indicate good fit. RMSEA scores below .08 are acceptable, and scores below .05 are good (Hoyle, 1995).

Mx Analyses

Liability-threshold models were fitted using maximum likelihood (ML) estimation procedure on raw data in Mx (Neale et al., 2003). The SCL-5 scores and the internalizing disorders, which were collapsed into one factor in the present study, were both treated as dichotomous factors in these analyses, and assumed to have a latent continuous distribution. Univariate analyses were performed in order to estimate the magnitude of the genetic and environmental contributions to the two variables. The univariate analyses were also performed to determine whether there were sex differences in the genetic and environmental effects on the SCL-5 and the internalizing factor. The univariate scripts specified a model that allowed for qualitative

sex differences, a model that allowed for quantitative sex differences, and a model that allowed for no sex differences by constraining all the estimated parameters to be equal for males and females. The established approach in univariate twin modelling of selecting a saturated ACE-model as the full model, and comparing this with nested submodels (i.e. an AE-model, CE-model, and E-model) was followed within all the sex difference scripts.

A bivariate analysis was performed after the univariate analysis in order to investigate for shared variance between the SCL-5 and internalizing disorder parameters, and to test whether the sex differences from the univariate analyses would hold. Often, a Cholesky decomposition is applied in bivariate twin modelling. However, according to Neale, Røysamb & Jacobson (2006) it is not straightforward to model a Cholesky decomposition that allows for sex differences in bivariate and multivariate analyses. Direct extension of the sex limitation modelling from the univariate decomposition to a multivariate composition may result in biased fit estimates. A correlational approach may be applied instead to circumvent the problem (Neale et al., 2006). Therefore, a correlated factor model was fit to the bivariate data, allowing for sex differences (Figure 5). In this approach, each variable is separately decomposed into its genetic and environmental components, and the correlation between the components are estimated (Loehlin, 1996). A correlated factor model will reveal the same fit estimates as a Cholesky model, but makes use of a different parameterization of the relation between the variables than the Cholesky approach. The Cholesky approach decomposes the components of the variables into what is shared and what is unique to each variable, whereas the correlated factor approach only indexes what is unique to a variable. The extent to which the parameters are shared between the variables is indexed by the correlation between the parameters.

Results

Life-time Prevalences

The lifetime prevalences for having one or several internalizing disorders, and the prevalence of scoring above cut-off on the SCL-5 are presented in Table 1. Out of the 7992 individuals in the questionnaire sample, 968 (12.1%) scored above cut-off on the SCL-5. Among the 2793 individuals in the interview sample, 560 (20.1%) had one or several internalizing disorders. 394 individuals (14.1%) had MDD, 48 (1.7%) had dysthymia, 112 (4%) had social phobia, 87 (3.1%) had panic disorder, 134 (4.8%) had agoraphobia, and 55 (2%) had GAD.

Table 1. ***Prevalences for scoring above cut-off on SCL-5 and on one or several internalizing disorders (Int) for the sample as a total, and for males and females***

Model	N	N Above Cut-off	% Above Cut-off
SCL_{Total}	7992	968	12.1
SCL_{Male}	3415	300	8.8
SCL_{Female}	4577	668	14.6
Int_{Total}	2793	560	20.1
Int_{Male}	1019	144	14.1
Int_{Female}	1774	416	23.4

Phenotypic Correspondence between the SCL-5 and Internalizing Disorder

The phenotypic correspondence between two variables can be operationalized using several types of analyses. In the present study, we have chosen to represent the correspondence using tetrachoric correlations calculated in Mx, odds ratio calculated from binary logistic regression and relative risk calculated from a cross tabulation in SPSS 17.0. The tetrachoric correlation between SCL-5 and the internalizing factor was .48 (95% CI = .41 -.55). The odds ratio of having an internalizing disorder if scoring above cut-off on the SCL-5 was 4.75 (95% CI = 3.71 - 6.07). When splitting on sex, the results show that females have an odds ratio of 4.82 (95% CI = 3.65 - 6.38) to develop an internalizing disorder if scoring above cut-off if scoring above the cut-off on the SCL-5, whilst the males have an odds ratio of 3.02 (95% CI = 1.70 - 5.36). The relative risk for developing an internalizing disorder if scoring above cut-off on the SCL-5 is 2.95 (2.74 for men and 3.60 for females). When validating SCL-5 against CIDI for internalizing disorders, we found sensitivity and specificity estimates of 28% and 93%, respectively, which suggest that the SCL-5 does a good job in excluding non-cases but have a mediocre ability to detect cases. However, it was not the focus in the present study to optimize the values of the validity coefficients. A lower cut-off would result in a higher sensitivity. The comparison between a variable that was measured as a lifetime prevalence and a variable that indexed a prevalence from the last 14 days of a respondent's life also contribute to make the coefficients low. The cross tabulation that the odds ratio, relative risk and polychoric correlations are based on is presented in Table 2.

Table 2. ***Cross tabulation of the variables SCL-5 and internalizing disorder.***

		Internalizing	
		0	**1**
SCL	**0**	2045	399
	1	163	151

0 = scoring below the cut-off, 1 = scoring above the cut-off

Confirmatory Factor Analysis

In order to investigate the unidimensionality of the six internalizing disorders, MDD, dysthymia, GAD, social phobia, agoraphobia and panic disorder, a confirmatory factor analysis stating a one-factor structure was tested on the data. The one-factor model yielded an excellent fit: χ^2 = 4097,8 (p = 0.0), df=9, GFI = 0.922, CFI= 0.998, RMSEA= 0.026, and hence justifies collapsing the disorders into one sum score (internalizing disorder) in the remaining analyses. Factor loadings ranged from .55 for MDD, to .83 for panic disorder, with a mean loading of .70. Figure 3 shows the CFA model with the factor loadings for each disorder to the latent variable.

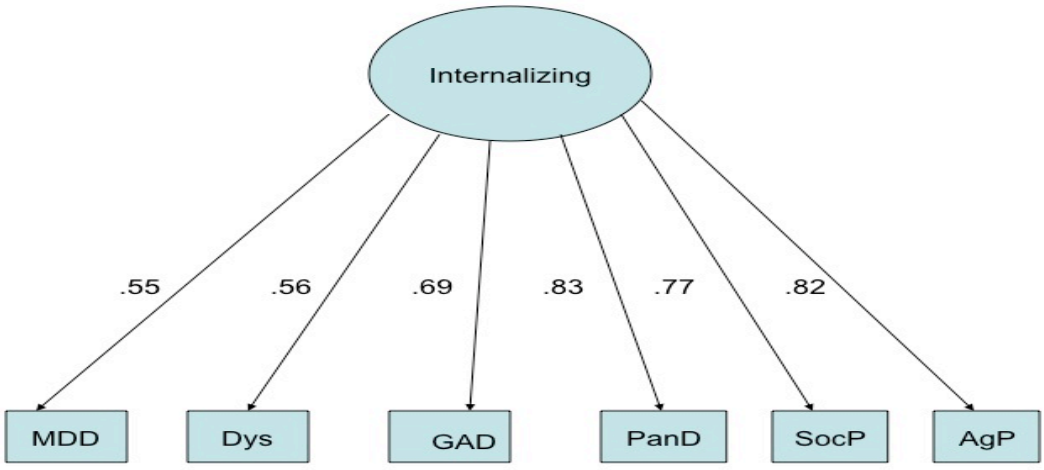

Figure 3: The CFA model, showing the six internalizing disorder's factor loadings onto the latent factor internalizing. Dys = Dysthymia, PanD = panic disorder, SocP = social phobia, AgP = agoraphobia.

Twin correlations

Tetrachoric twin correlations with 95% confidence intervals were estimated for SCL-5 and internalizing disorder. The phenotypic correlation (the within-twin cross-trait correlation) between SCL-5 and internalizing disorder was found to be .48 (.41 - .55). Table 3 displays the cross-twin within-trait correlations for SCL-5 and internalizing disorder, and the cross-twin cross-trait correlations by zygosity groups.

The total MZ correlations (combining males and females) were generally high, and more than twice the corresponding DZ correlations. Thus there were clear indications of substantial genetic effects on both phenotypes, and on their covariation. No indication of common environmental effects was found. This pattern was generally also found for the sexes separately for both SCL-5 and internalizing. There are some exceptions that might suggest the presence of shared environmetal effects and non-additive genetic effects on the traits, but the wide confidence intervals do not provide any grounds to conclude whether these effects are present or not. No indication of sex differences in the SCL-5 variable is found, as the same sex and unlike sex DZ correlations are approximately the same. The negative correlations in the opposite sex DZ group as opposed to the same sex DZ groups for internalizing disorder may imply qualitative sex differences.

The wide and overlapping confidence intervals imply that the analyses might lack power to identify minor effects in the pattern of correlations.

Table 3. ***Twin correlations with 95% confidence intervals for the SCL-5 and internalizing disorder by gender and zygosity***

	Cross-twin SCL-5	Cross-twin Internalizing	Cross-twin cross-trait
MZ both sexes	**.49** (.40 - .56)	**.46** (.37 - .54)	**.40** (.28 - .52)
DZ both sexes	**.21** (.14 - .29)	**.15** (.04 - .25)	**.16** (-.04 - .32)
MZ males	**.42** (.15 - .64)	**.46** (.16 - .70)	**.24** (.02 - .44)
DZ males	**.28** (-.07 - .57)	.**23** (-.23 - .61)	**.07** (-.24 - .37)
MZ females	**.50** (.37 - .62)	**.47** (.31 - .60)	**.40** (.31 - .48)
DZ females	.**21** (.05 - .37)	**.37** (.16 - .55)	**.16** (.04 - .28)
DZ unlike sex	**.21** (.05 - .36)	**-.11** (-.33 - .11)	**.06** (-.7 - .19)

Univariate Model Fitting

In order to determine the heritability and significance of additive genetic, shared environment and unique environmental effects on the SCL-5 and the internalizing disorder variables, univariate analyses were performed in Mx on the data. In these univariate analyses, it was also tested for quantitative and qualitative sex differences on the genetic and environmental estimates of the variables. Table 4 and 5 show the χ^2 goodness of fit index measured as a maximum likelihood function (ML), the degrees of freedom (df), the p-value and the AIC for the SCL-5 and internalizing disorder models, respectively.

SCL-5. First tested against the SCL-5 data was the fully saturated ACE model allowing for both quantitative and qualitative sex differences. This was the model to which the subsequent nested submodels were tested against. The fit statistics are listed in Table 4. Model 3 (CE) could be firmly discarded, due to the significant deterioration compared to the full model ($\Delta\chi^2 = 11$, $\Delta df = 2$, $p < .01$). This indicates the importance of genetic contributions to the scores on the questionnaire. Model 4 (E) could also be firmly discarded, due to significant deterioration in fit compared to the full model ($\Delta\chi^2 = 70.3$, $df = 4$, $p < .01$), also establishing that the additive genetic effects are significant and therefore can not be left out of the analyses. Fixing the genetic correlation between the genders to .5 did not result in significant deterioration in fit in model 5 ($\Delta\chi^2 = 0.2$, $\Delta df = 1$, $p = .70$), and hence evidence of qualitative sex differences was not found. Neither did we find evidence of quantitative sex differences as the ACE parameters could be constrained to be equal without significant deterioration in fit in model 9 ($\Delta\chi^2 = .3$, $\Delta df = 3$, p = n.s.). The best fitting model to the data was model 10, which is an AE-model specifying no sex differences ($\Delta\chi^2 = 0.3$, $\Delta df = 4$, p = n.s., AIC = - 7.7). Model 10 did not yield significant deterioration in fit compared to model 9 ($\Delta\chi^2 = 0$, $\Delta df = 1$, p = n.s) and 6 ($\Delta\chi^2 = .1$, $\Delta df = 1$, p = n.s). The χ^2 goodness of fit was nearly identical same for model 6, 9 and 10, but model 10 was able to account for the data in a more parsimonious way, and was therefore preferred. The result of an AE-model being the best model indicates that familial resemblance for SCL-5 scores could be explained solely by additive genetic effects. The parameter estimates for the best fitting model (Fig. 4) were estimated to .48 for the additive genetic effects (95% CI = .38 - .58), and to .52 for the unique environment (95% CI = .42 - .62).

Table 4. ***Univariate model fits for the SCL-5, with general-, common-, and no sex limitations***

Model	-2LL	df	p	AIC
Qualitative sex differences				
1 ACE	5764.6	7978	-	-
2 AE	5764.6	7980	n.s.	- 4.0
3 CE	5775.6	7980	*	10.0
4 E	5834.9	7982	*	62.3
Quantitative sex differences				
5 ACE	5764.8	7979	n.s	-1.8
6 AE	5764.8	7981	n.s.	-5.8
7 CE	5775.6	7981	*	5
8 E	5834.9	7983	*	60.3
No sex differences				
9 ACE	5764.9	7981	n.s.	-5.7
10 **AE**	**5764.9**	**7982**	**n.s.**	**-7.7**
11 CE	5777.3	7982	*	4.7
12 E	5834.9	7983	*	60.3

Best fitting model in bold type * = sign. at p < .01 n.s. = non-significant

Internalizing Disorder. As for the SCL-5, the fully saturated ACE model was first tested against the data on internalizing disorder, allowing for both quantitative and qualitative sex differences. This was the model to which the subsequent nested submodels were tested against. The fit statistics are listed in Table 5. When comparing the models allowing for qualitative, quantitative and no sex differences, the model allowing for qualitative sex differences (model 2) indicated the best fit to the data ($\Delta\chi^2 = 1.6$, $\Delta df = 2$, p = n.s., AIC = -2.4). This finding implies that there may be different genes operating for males and females, and hence also most probably differences in the size of the genetic and environmental parameters between males and females for internalizing disorder. Finding that the best model was an AE model by testing each of the models to the saturated model, it also made sense to

Table 5. ***Univariate model fits for internalizing disorder, with qualitative-, quantitative-, and no sex differences***

Model	-2LL	df	p	AIC
Qualitative sex differences				
1 ACE	2708.2	2786	-	-
2 AE	**2709.8**	**2788**	**n.s.**	**- 2.4**
3 CE	2718.7	2788	*	6.5
4 E	2762.6	2790	*	46.4
Quantitative sex differences				
5 ACE	2715.5	2787	*	5.3
6 AE	2717.1	2789	*	4.9
7 CE	2718.7	2789	*	6.5
8 E	2762.6	2791	*	44.4
No sex differences				
9 ACE	2720.0	2789	*	5.8
10 AE	2720.0	2790	*	3.8
11 CE	2729.6	2790	*	13.4
12 E	2762.6	2791	*	44.4

Best fitting model in bold type

* sign. at $p < .01$ n.s. = Not significant

test whether this result would hold when testing across the different AE models (model 2, 6 and 10). Model 6, which allows for quantitative sex differences, yielded a significantly worse fit compared to model 2 ($\Delta\chi^2 = 7.3$, $\Delta df = 2$, $p < .01$). This result suggests that the potentially different genes that are operating for males and females on the phenotypic variance in internalizing disorder may have approximately the same effect on the different sexes. Comparing model 10 to model 2 also yielded a significantly worse fit ($\Delta\chi^2 = 10.2$, $\Delta df = 1$, $p < .01$). The additive genetic effects for the best fitting model (Fig. 4) were estimated to .47 (95% CI = .18 - .70) for males, and .49 (95% CI = .35 - .61) for females. The unique environmental effects were estimated to .53 (95% CI = .30 - .82) for males and .51 (95% CI = .39 - .65) for females.

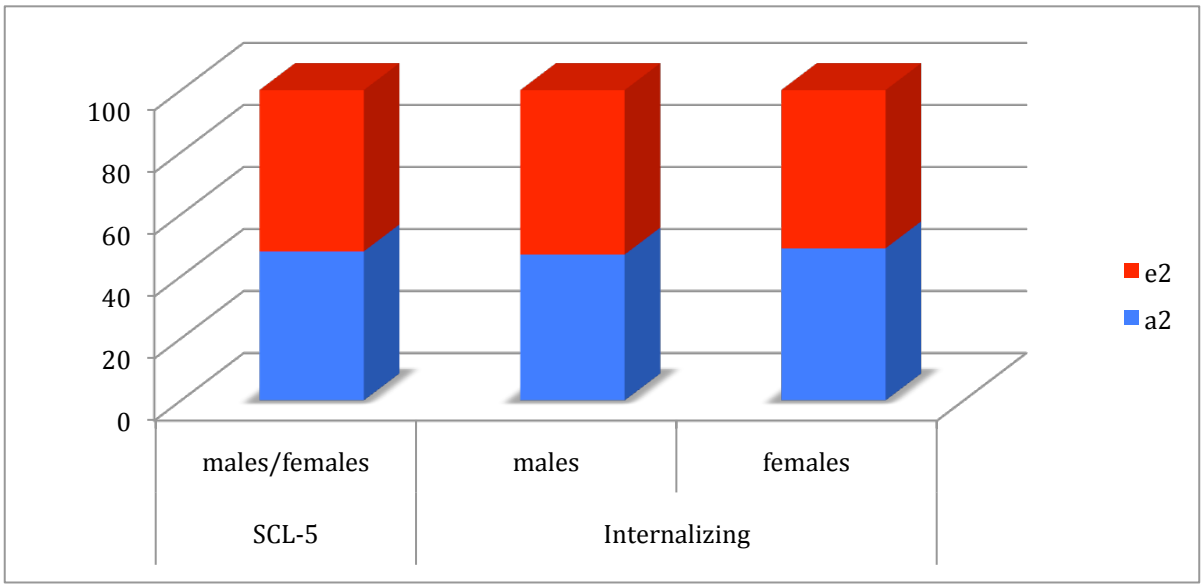

Figure 4: the standardized a^2 and e^2 parameters from the best fitting models for the SCL-5 and internalizing factor data. The parameters for the SCL-5 variable are the same for males and females, as the univariate analyses did not show evidence of sex differences. The additive genetic effects (a^2) may be regarded as narrow-sense heritability, and index moderate heritability for both phenotypes.

Bivariate Model Fitting

As the univariate analyses did not support shared environmental effects for the variables, these effects were left out of the bivariate analyses, and only AE-models were fitted to the data. The fit statistics for the bivariate modelling is shown in Table 6. Even though there was found evidence of sex differences in the univariate model fitting, the best-fitting bivariate model was model 3, which is an AE-model allowing for no sex differences ($\Delta\chi^2 = 12.622$, $\Delta df = 7$, $p = < .08$, AIC = -1.378). The additive genetic contribution to the variance in SCL-5 and internalizing disorders are specified by the path coefficients A_1 and A_2, respectively. For SCL-5, the path coefficient was estimated to .70 (.62 - .76, 95% CI), which gives a heritability of .48. For the internalizing disorder, the path coefficient was estimated to .66 (.56 - .74, 95% CI), which gives a heritability of .44. The unique environmental contribution to the variance in SCL-5 and internalizing is specified by the path coefficients E_1 and E_2, respectively. For SCL-5, the path coefficient was estimated to be .72 (.66 - .79, 95% CI), which gives an environmental effect of .52. For internalizing disorder, the path coefficient was estimated to .74 (.67 - .83, 95% CI), which gives an environmental effect of .56. The estimates of heritability and environmental effects are more precise than those obtained in the univariate

Table 6. ***Model Fits for the bivariate model fitting***

Model	**-2LL**	**df**	**p**	**AIC**
Qualitative sex differences				
1 AE	8336.423	10762	-	-
Quantitative sex differences				
2 AE	8345.204	10767	n.s (.1185)	-1.228
No sex differences				
3 **AE**	**8349.054**	**10769**	**n.s (.0818)**	**-1.378**

Best fitting model in bold type

n.s = non-significant

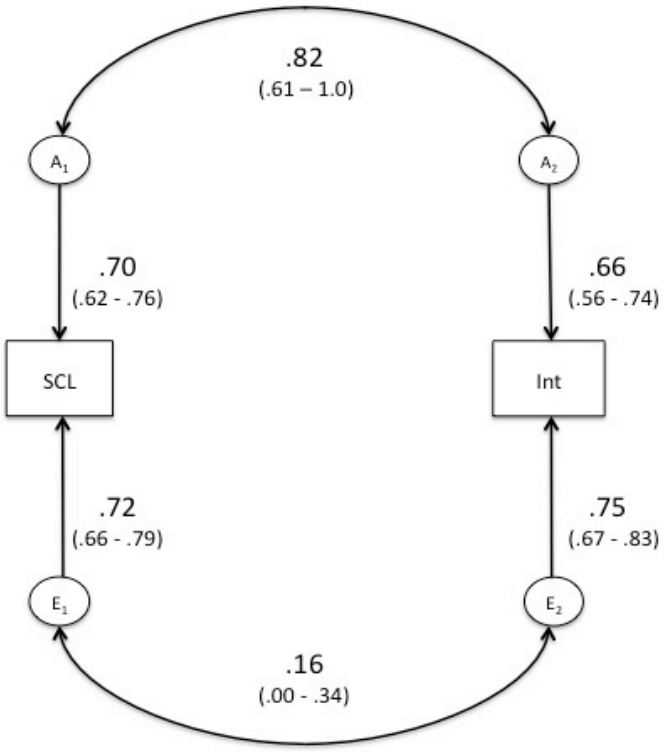

Figure 5: Best fitting correlated factor model, drawn for one half of the twin pair as the parameters are constrained to be equal for each twin in a pair.

modelling. The bivariate model fitting also reveals that the phenotypic correlation can be decomposed into genetic and environmental components. The phenotypic correlation predicted by the model (see Figure 5) was .46 (.70 x .82 x .66 + .72 x .16 x .75). Of this model predicted phenotypic correlation, 81% was explained by genetic factors (.70 x .82 x .66 = .379), and 19% was explained by environmental factors (.72 x .16 x .75 = .086). The genetic correlation in the best fitting model (between A_1 and A_2) was estimated to be .82 (.61 – 1.0, 95% CI). According to the CI-estimates, the correlation is not significantly different from unity, a finding indicating that the genetic factors in the two variables are identical. The environmental correlation (between E_1 and E_2) was estimated to be .16 (.00 - .34).

Discussion

The aim of the present study was to investigate the correspondence between the SCL-5 and internalizing disorders as measured by CIDI. In order to analyze the correspondence, both phenotypic analyses and twin modelling were performed. The correspondence between short self-report scales and interview-based diagnoses have previously been investigated using phenotypic analyses, with the exception of one study investigating the genetic and environmental effects and the correlation between these effects. Little is therefore known about the correspondence at a genotypic level.

The lifetime prevalence rate for internalizing disorder was 20.1%, with MDD as the most prevalent single disorder (14.1%). These numbers support the notion of internalizing disorders as being prevalent, and hence should be considered a substantial health problem in the population. The prevalence of internalizing disorder is similar to prevalence estimates obtained in population studies (Kessler et al., 2005a; Kringlen, Torgersen & Cramer, 2001). The similarities in prevalence rates in the present study and population studies may be considered a validation of the twin sample as a means to detect general tendencies in the population. The prevalence of scoring above cut-off on the SCL-5 was 12.1%, and is very close to the prevalence found in a population study by Strand et al. (2003), which was 12.5% for the SCL-5, and 11.1% for SCL-25. The similarities in the present and Strand et al.'s study may be an indication that we have a good picture of the prevalence in the Norwegian population. The difference that is found in prevalence rates between SCL-5 and internalizing disorder may not be surprising when considering the response interval between the measures. The respondents to the SCL-5 questionnaire are asked to indicate to what extent the

symptoms had been present during the last 14 days, whereas the diagnostic interview involves a lifetime perspective. It makes sense that more symptoms are reported during a lifetime than during two weeks.

Phenotypic Correspondence

The phenotypic analyses conducted in the present study included phenotypic correlations, odds ratio, relative risk, and estimates of sensitivity and specificity. When investigating the phenotypic correlation (the within-twin cross-trait correlation) between the SCL-5 and internalizing disorder the relation was found to be moderate (.48). Several differences between the measures contribute to lower the correlation. For instance, the SCL-5 only includes five items, whilst the CIDI is an extensive interview. There is also a substantial time lag between the two measures, as the questionnaire was conducted in 1998, and the CIDI was conducted between 1999 and 2004.

The finding of a moderate phenotypic correlation between the measures fits with other studies that have investigated phenotypic correlation between brief self-report questionnaires and interview-based diagnoses. Sandanger et al. (1998) found a phenotypic correlation between MDD and the SCL-25 of .45, a very similar estimate to the correlation in the present study. Sandanger et al. (1998) included a different sample of internalizing disorders than in the present study, namely MDD, GAD/panic disorder, somatoform disorder, and phobias. Some of these disorders (somatoform disorder and phobias) are found to be more peripheral to the internalizing factor in the form of weaker factor loadings, as found in for instance Røysamb et al. (2009). The correlation between these internalizing disorders combined and the SCL-25 was .37, and therefore lower than was found using the present study's sample of internalizing disorders. Haver (1997) investigated the correspondence between the SCL-90 and the diagnostic interview SCID. A phenotypic correlation of .47 was found between MDD and the depression subscale. The present study obtained a correlation between internalizing disorder and the SCL-5 of .48. As depression is one of the core disorders in internalizing disorders, and the depression subscale in the SCL-90 share some of the items with the SCL-5, it is therefore meaningful to compare these two correlations, although with caution.

Other studies typically report only one statistics to index the phenotypic correspondence. The present study has a broader operationalization of the phenotypic correspondence, and has applied odds ratio, relative risk and sensitivity and specificity estimates in addition to the correlation estimate. It was found that respondents scoring above

cut-off on the SCL-5 have a 4.75 times higher odds of scoring above cut-off on an internalizing disorder than respondents that do not score above cut-off on the SCL-5. This is in accordance with the study conducted by Sandanger et al. (1998), where an odds ratio between SCL-25 and the study's selection of internalizing disorders was found to be 3.89. The more peripheral selection of disorders may explain the lower correspondence found in Sandanger et al. The relative risk for developing an internalizing disorder was found to be 2.95 in the present study. This implies that the probability for developing an internalizing disorder if one scores above cut-off on the SCL-5 is about three times higher than if one scores below the cut-off. A relative risk of 3.30 was found in a study by Williams et al. (2002), which investigated the correspondence between several self-rating scales for anxiety and depression to the interview-based depression diagnosis. Williams et al. also found that differences between the different scales' ability to detect psychiatric cases was non-significant. The similarities between the present study's results and other studies may be interpreted as a validation of the present study's methods and generalizability of the twin sample.

The sensitivity and specificity coefficients were found to be .28 and .93, respectively. With the chosen cut-off value of 2.0 for SCL-5 the sensitivity becomes lower than is usually found in similar studies. However, the object of the present study was not to obtain the best possible sensitivity and specificity, as is often the case in studies investigating the correspondence between diagnoses and self-report scales. The sensitivity and specificity were investigated as a means to operationalize the correspondence using other measures than correlation, and the cut-off of 2.0 was chosen to give approximately the same prevalence rate as is found using the SCL-25 (Strand et al., 2003). Other studies often seek to find the cut-off point on a self-rating scale that optimizes the balance between the sensitivity and specificity. The different aims of the present and other studies that apply this instrumental operationalization of the correspondence may explain the discrepancies.

Internalizing Tendencies are Expressed Mainly as Symptoms of Anxiety and Depression

Individuals often express internalizing tendencies as symptoms of anxiety and depression. It was argued in the present study that anxiety and depression disorders share a common liability, which may be labelled internalizing. The CFA tested a one-factor structure of the included anxiety and depression disorders, and obtained an excellent fit to the data. The finding fits with previous studies investigating the structure of common psychiatric disorders,

which also found evidence of anxiety and depression disorders sharing a common liability (Krueger, 1999; Krueger & Markon, 2006; Røysamb et al., 2009; Slade & Watson, 2006). The common finding of two superordinate factors has been referred to as the IE-model (Internalizing and Externalizing; Røysamb et al., 2009). The mean loading on the factor was .70, which is quite high, but still implies that some of the variance in the disorders is unique. A similar pattern and strength of factor loadings were obtained in Røysamb et al. (2009), which is not a surprising finding, as mainly the same data as Røysamb et al.'s were analyzed in the present study. Another finding from the CFA conducted in the present study was that the different disorders included in the internalizing factor varied in their degree of reflecting core characteristics, with anxiety disorders loading stronger onto the internalizing factor than depression. Only a few studies exist that have tested the factor structure of common mental disorders, and hence the present study is a contribution to this research.

The finding of a one-factor structure in the studies investigating the structure of anxiety and depression disorders also fits with the findings in studies investigating the factor structure of the SCL-25 and the short form version SCL-5. For instance, Moum (1998) found that much of the variance in SCL-25 could be explained by a one-factor structure. This finding is opposed to the notion of the brief self-report scales like the SCL-25 and SCL-5 as measuring symptoms of anxiety and depression, and may be interpreted in several ways. To begin with, the one-factor structure may imply that the SCL is unable to distinguish between anxiety and depression. This would imply weak validity as a differential diagnostic instrument. Another possibility may be that the anxiety and depression symptoms are not specific to different anxiety and depression disorders, but instead good indicators of general mental distress. The fact that symptoms of anxiety and depression tend to accompany many different mental disorders supports this notion. Therefore, high scores on the anxiety and depression subscales do not have to indicate a specific disorder but rather that it is likely that the respondent has some kind of DSM-IV disorder. The unidimensionality may also be understood in the light of anxiety and depression as different phenotypic expressions of the same genetic liability or propensity to develop a mental disorder. If the anxiety and depression disorders are in fact parts of the same genetic liability, it may seem odd that they are expressed as quite phenomenologically different traits. MDD and GAD are classic examples of traits that are phenomenologically quite distinct, and still are found to share a substantial amount of genetic factors. Different phenomenology does not have to imply that traits cannot share a substantial amount of genetic factors (Kendler, 2009, personal communication). Even low phenotypic correlations between traits do not exclude the

possibility of substantial genetic or environmental correlation. One explanation to this could be that the type of environmental exposures decides the phenotypic expression of the genetic liability. When modelling the relation between anxiety and depression, studies have found that the genetic sources are shared, and that the environmental effects are mostly specific and unique (Kendler et al., 1992; Kendler, 1996). Findings like these accentuate the importance of the environment to decide the phenotypic expression of a common genetic liability.

The study by Røysamb et al. (2009 investigated the structure of both axis I and axis II disorders and found evidence of five factors. What is unique to this study compared to other studies investigating the structure of common mental disorders is not only that the included disorders are rather comprehensive. The study also investigated the higher order structure of all the included disorders. Evidence of two factors was found that accounted for the axis I internalizing and externalizing disorders and the axis II disorders, and two higher order factors encompassing both axis I and axis II disorders. The authors argue that the higher order factors that was found can integrate the structure of axis I and axis II disorders, in addition to normal personality traits.

There has long been a debate concerning the proximity of normal personality traits, particularly the Big Five model (Costa & McCrae, 1992), and axis I and axis II personality disorders (e.g. Krueger & South, in press; Widiger & Mullins-Sweatt, 2009). The personality trait of neuroticism is particularly central in this context. Both internalizing disorders and the personality trait of neuroticism are characterized by negative affect-laden mood and emotional instability (Røysamb et al., 2009). Neuroticism is found to be a strong predictor and correlate with a spectre of different internalizing axis I disorders like depression, anxiety, anorexia, panic disorders and phobia (Clark, Watson & Mineka, 1994; Hettema, Neale, Myers, Prescott & Kendler, 2006; Kendler et al., 1993; Kendler, Kuhn, & Prescott, 2004; Lahey, 2009; Røysamb et al., 2009). Hettema et al. (2006) found that there is a significant amount of shared genetic and environmental factors between neuroticism and internalizing disorders, and that this helps explain the comorbidity between anxiety and depression. Future research should investigate the correspondence between personality factors and mental disorders further. It may be that the normal personality trait of neuroticism may explain a large amount of the shared variance both within and between several axis I and axis II mental disorders, as was found in Røysamb et al. (2009). The assumption should be investigated further to test whether it is a replicable pattern. If it is indeed replicable, the finding is in support of a dimentional approach in the classification systems, and should perhaps have

implications for the new DSM-V (see below: reflections on the nature of the existing classification systems).

Genetic and Environmental Effects

The phenotypic correspondence between brief self-report scales and interview-based diagnoses has been found to be moderate. In the present study, the more thorough methodology of twin modelling was applied. It was hypothesized that this methodology would reveal a higher correspondence than phenotypic analyses are able to. One of the aims of the present study was to investigate whether the results from Foley, Neale & Kendler (2001) of substantial genetic correlation between the interview-based depression diagnosis and the brief self-report questionnaire SCL-10 could be replicated using several internalizing disorders and both sexes in the analyses.

The first step in the analyses was to investigate the tetrachoric correlations on the different zygosity groups. By inspecting the correlational patterns between the zygosity groups, assumptions regarding the importance of different sources of variance can be made. The patterns obtained in the present study had wide confidence intervals, but clearly suggested genetic effects on both the SCL-5 and internalizing disorder variables.

The twin modelling was performed both univariate and bivariate. The univariate modelling was conducted in order to determine the importance of the different components of variance in the variables, namely the genetic, shared environmental and unique environmental effects. The univariate analyses also investigated whether there were sex differences in the genetic architecture of the two phenotypes, which might give reason to suspect sex differences in the correspondence between the variables. The bivariate analyses investigated the correlation between the variables' latent factors, and also investigated potential sex differences. The twin modelling confirmed the pattern from the tetrachoric correlations, as the final heritability coefficients were found to be .48 and .44 for the SCL-5 scores and the scores on internalizing disorder, respectively. The heritability estimates obtained in the present study are in accordance with previous findings. Typically, heritability estimates of internalizing disorders are found to be in the area of .20 - .45 (Bouchard, 2004; Edvardsen et al., 2009; Hettema, Neale & Kendler, 2001; Kendler & Prescott, 2006; Reichborn-Kjennerud et al., 2004; Scherrer et al., 2000; Sullivan, Neale & Kendler, 2000; Sundet et al., 2003; Tambs et al., 2009; Ørstavik, Kendler & Czajkowski, 2007), with one exception finding a heritability of .65 (Foley, Neale & Kendler, 2001). The heritability of symptoms of anxiety and depression,

as measured by SCL, is found to be in the same area (Kendler, 1994; 1995; Nes et al., 2007; Reichborn-Kjennerud et al., 2002; Tambs, Harris, & Magnus, 1997; Tambs & Moum, 1993), with the exception of Foley et al. (2001) who found a heritability of .61. There was found no evidence of common environmental effects for internalizing disorders or SCL-scores in the present study. This finding is in accordance with most twin studies on internalizing disorders, which are typically unable to find evidence of shared environmental effects (e.g. Plomin & Daniels, 1987; Kendler & Prescott, 2006). The rest of the variance in internalizing disorders and SCL-5 was explained by unique environmental factors. The finding in the present study of a combined heritability of .44 for internalizing disorders is in the upper area of what has been found earlier. Psychometrically, this is not surprizing, as measures of single disorders are more prone to measurement error and hence unreliability than is measures of several disorders combined (Tambs et al., 2009). Unreliability of measurement may in some circumstances contribute to lower the estimates of heritability. The same tendency was present in Tambs et al. (2009), who found a substantially higher estimate of heritability for the combined anxiety disorders compared to single anxiety disorders. Edvardsen et al. (2009) also found higher heritability estimates for depression when several depression disorders were included in the analyses. The finding of a higher heritability for combined disorders than for single disorders may be interpreted as support to the notion that we have captured a general underlying liability. This interpretation is also supported by Edvardsen et al. (2009), who found evidence for mainly shared genetic effects for various depressive disorders, and hence evidence of a shared genetic liability to depressive disorders.

The bivariate twin modelling results show that practically all of the genetic effects are shared between the SCL-5 and the internalizing factor. The genetic correlation between liability to internalizing disorders and SCL-5 symptoms of anxiety and depression is estimated to be .82 (95% CI = .61 - 1.0), with a confidence interval reaching unity. It thus seems that the genes inherent in interview-based internalizing diagnoses are the same genes that are inherent in the brief self-report questionnaires. The environmental correlation between the instruments was found to be low in the present study (.16, 95% CI = .00 - .34), and thus implying that the E-effects were generally unique to the scores from the two instruments. This finding is in accordance with Foley et al.'s results.

The present study may thus be regarded as an extended replication of the Foley et al. (2001) study as it converge on very much the same results by using several disorders and both sexes in the analyses.

Can the SCL-5 be a Useful Screener for Internalizing Disorders?

At first sight, the SCL-5 is not very impressive as a screening instrument for the primary care clinician, as the sensitivity is low (.28). Diagnostic interviews like the CIDI reveal substantially more information than a short self-rating scale like the SCL is able to regarding various symptoms of psychiatric disorders. Using only a five-item version like the SCL-5 to manage approximately the same job might therefore seem rather optimistic at first glance. However, despite seemingly large differences, it is indicated that the SCL versions are able to some degree to tap onto the same construct as the diagnostic interviews in studies investigating phenotypic correspondence. Better still, the SCL-5 is found to capture the same genetic variance as internalizing disorder. This finding implies that the SCL-5 may function as an excellent index of the genetic risk factors for developing an internalizing disorder. Williams et al. (2002) found no significant differences in ability to reliably screen for clinical depression in a selection of various scales. This result may indicate that they are all measuring the internalizing construct. Based on this finding, it is plausible that the genetic risk factors are essentially the same in internalizing disorders and other brief self-report scales intended to measure symptoms of anxiety and depression. The information of shared genetic risk factors is valuable both for research and clinical purposes.

Reflections on the Nature of the Existing Classification Systems

Many psychiatric disorders are found to be comorbid. As is indicated in the present study and in studies investigating the origins of patterns of comorbidity, psychiatric axis I disorders cluster into a few factors and are found to share important etiologic sources, as is found in studies investigating the genetic and environmental effects on the disorders (Kendler et al., 1992; Kendler, 1996). These findings do not support the notion in the existing diagnostic system of many independent, categorical disorders. Instead, the evidence is in favour of a diagnostic system with fewer and broader categories that do not drift into each other by sharing a large amount of symptoms. Broader categories may contribute to the independency of the disorders being better attended to than in the present system. The idea of broader diagnoses is also relevant for a dimensional approach for the future diagnostic system, as categories and dimensionality in a classification system is not mutually exclusive (e.g. Krueger & Bezdjian, 2009). Support in favour of a dimensional approach is found in studies that point to the correspondence between personality traits, like the Big Five (Costa &

McCrae, 1992) and for psychiatric disorders like depression and alcohol dependence (e.g. Kendler, Gardner, & Prescott, 2006; Krueger & South, in press; Widiger & Mullins-Sweatt, 2009). The correspondence is found to be particularly strong for the personality trait neuroticism and psychiatric disorders (Hettema et al., 2006; Røysamb et al., 2009).

Limitations

The time interval between the measures is most probably a factor that contributes to lower the correspondence between the SCL-5 and internalizing disorder, as the SCL-5 data were gathered around 1998, and the interviews around 2000. It is possible to investigate the impact of the time interval in future studies, by applying a longitudinal design. The NIPHTP already has this design, as there have been gathered data on the SCL-5 in two waves of questionnaires. By investigating the differences in phenotypic correlation between the first questionnaire and the interview, and the second questionnaire and the interview, one may have an indication of the measurement intervals' impact on the correspondence. Preliminary analyses in the present study revealed a slightly lower correlation between internalizing disorder and the questionnaire conducted in 1992 than between internalizing disorder and the questionnaire conducted in 1998. It is therefore reasonable to assume that the correspondence would be higher if the measurements were closer in time.

There was a significant dropout from the questionnaire measure and to the interview measure. Less than 44% of the twins that responded to the questionnaire participated in the interview. However, empirical investigation of the attrition did not find any grounds to suspect that selection bias could affect the estimates of correspondence (Tambs et al., 2009). In addition to the dropout rate, the present sample consisted of a young age cohort. It may therefore not extrapolate to other age cohorts of the population (Nes et al., 2007).

The present study chose to dichotomize the responses on the SCL-5. There are several limitations associated with dichotomizing continuous data (MacCallum et al., 2002). Among these are loss of information and lowered statistical power. In addition, the estimates obtained in the present study would probably be more stable and have smaller CIs if the SCL-5 variable had been treated as continuous. This was difficult to accomplish, as Mx does not allow one of the variables to be continuous and the other to be ordinal or dichotomous. Also, it makes sense to dichotomize the SCL-scores, as the internalizing disorders are dichotomized

according to the existing diagnostic system. However, preliminary analyses revealed almost identical results whether the SCL-5 was dichotomized or ordinal.

Some reliability and validity issues should be mentioned. The SCL-5 has been found to have satisfactory reliability. A study by Strand et al. (2003) found a Cronbach alpha of .87 using a population-based Norwegian sample. Regarding the internal consistency of CIDI disorders, this has not been investigated because of the hierarchical structure of this instrument. That is most respondents only answer a small number of questions because the responses to "entry items" often indicate that they are free from symptoms and thus further inquiry is not necessary. Other measures of reliability have been conducted, such as temporal stability and interrater reliability. The temporal stability and interrater reliability of CIDI disorders are found to be satisfactory in many studies (Wittchen, 1993). One potential problem to the estimates of correspondence in the present study is that the CIDI is conducted as an interview and the SCL-5 is based on self-report. It may be that these differences are a source of systematic error variance (Moum, 1998). High internal reliability and test-retest reliability in self-report scales cannot rule out the possibility that response biases are operating (Moum, 1998). One consequence of increased error variance is inflated Es in the twin modelling. Therefore, an important question in the present study concerns the extent to which reported symptoms change as a function of the administration mode. Social desirability is a response bias that may affect the results of the present study. Generally, it is found that respondents to self-report measures tend to report more symptoms and higher severity than in interviews (Moum, 1998), whilst face-to-face interviews tend to be susceptible to socially desirable responding and hence under-report (Moum, 1998). Kendler et al. (1993) found that the assessment of lifetime history of major depression was almost twice as reliable using a self-rating scale as when assessed using interviewers, an indication that a bias may be operating on the scores, whereas Amodei et al. (2003) found that the differences in responding to a questionnaire that was administered by an interviewer and a questionnaire that was responded by self-report were only negligible. Moum (1998) also found that the young and well-educated respondents are particularly prone to social desirability in the form of under-report of symptoms when interviewed. Our sample is relatively young (twins born between 1967 and 1979), and so there is reason to suspect that social desirability is a bias worth attention. Whether social desirability will contribute to lower or enhance the correspondence between SCL and CIDI scores depends on whether the bias is operating on one or both of the instruments. Increased uncorrelated measurement error, due to social desirability only affecting one of the measures or due to other factors, reduces the phenotypic

correlation between the measures. Both the genetic and environmental contribution to the phenotypic correlation will be lowered. The correlation between A_1 and A_2 in Figure 4 will remain the same, but the heritabilities (the path coefficients from A_1 and A_2 to the phenotypes) will decrease. The estimated environmental effects (from E_1 and E_2) will increase, but the correlation between E_1 and E_2 will decrease even stronger, resulting in a net reduction of the contribution from environmental pathway between SCL-5 and internalizing disorder. The relative values of the genetic and environmental contributions to the phenotypic correlation will remain unchanged. If the bias is operating the same way on both measures, however, the correlation between them and the Es may be enhanced, and the contribution from the phenotypic correlation from genetic sources will decrease (Reichborn-Kjennerud et al., 2002). As discussed in regard to the mode of questioning and social desirability bias, it has been found that the bias is only operating on one of the modes, namely the interview mode (Moum, 1998). Hence there is reason to assume that the relative contributions from A and E to the phenotypic correlation remain unbiased by measurement errors.

Several assumptions were mentioned in the section regarding the logic of twin analyses. The assumption of equal environment (EEA) and the assumption of assortative mating are especially relevant for the present sample. Even though there exist evidence that claims that the assumptions do not represent a threat to the validity and reliability of the heritability estimates of many studies, there is no guarantee that these assumptions hold true in this particular study. However, if violated, the consequences are probably small in our sample (Tambs, Harris, & Magnus, 1995) especially for the estimates of the genetic and environmental correlations.

If genes inherent in different traits correlate strongly, is it safe to say that they are in fact the same genes? The phenomenon of the same gene affecting different phenotypes is called pleiotropy (Neale & Maes, 2000). Even though the tradition within the research field, which the present study is part of, is to assume that genetic correlation implies shared genes, other explanations may also be plausible. One possibility is linkage, which is a phenomenon that occurs when there is a very close proximity of loci on a chromosome (Plomin et al., 2001). This phenomenon leads to a particular pair of genes to be inherited together and not independent of each other. We cannot rule out the possibility that linkage may account for the pattern found in the bivariate twin modelling in the present study. It is, however, possible to model the relation of allele sharing between siblings to test for the presence of linkage on a trait (Plomin et al., 2001). This is beyond the reach of the present study, but may be investigated further in future studies.

Finally, the results of the univariate analyses suggested qualitative sex differences for internalizing disorder, meaning that different genes may be influencing the liability to internalizing disorders for males and females. The differences were found not to be significant in the bivariate analysis. However, the tendency may hold if analyzed with a larger sample and hence with larger statistical power, as the differences in AIC values between the models allowing for sex differences and the ones that did not allow for sex differences were small. Very large sample sizes are required to be able to reliably detect sex differences when analyzing on dichotomous data (Nes et al., 2007; Neale, Eaves & Kendler, 1994). Middledorp, Wray, Andrews, Martin and Boomsma (2006) found no sex differences in symptoms of depression, where the genes for depression seemed to be expressed in the same way for both sexes. Mackinaw-Koos and Vasey (2000) in their review of sex differences in anxiety disorders, report that there are substantial sex differences in anxiety disorders in symptoms reported and prevalence rates across the life span. However, the authors point out that the sources of these sex differences are poorly understood, and that they may reflect methodological problems. Studies reporting sex differences mainly use self-report measures, and it may be that the findings are an artefact of the methods used. One reasonable explanation for the sex differences in the present study could be that we have modelled a factor encompassing several internalizing disorders, many of which have low prevalences. The SCL-5 has only a moderate sensitivity, and it may therefore be left to the coincidence which of the participants end up above or below the cut-off. This bias could then be transferred to the twin modelling, and hence affect the parameter estimates. At the same time, claiming that the sex differences is a result of coincidence is contrary to the results of the significance testing, which show that the chance of the estimates being the result of coincidence is less than one percent. The most likely explanation for sex specific genetic effects may be that the risk for developing internalizing disorders is a result of the different gender roles in the society. Different genes underlie the different desirable characteristics in males and females, and hence it is possible to envisage that internalizing may be the result of failing on different areas requiring different characteristics (Røysamb, Harris, Magnus, Vittersø, & Tambs, 2002). Anyway, the findings are interesting and should be investigated further with larger samples.

Conclusion

The correspondence between brief self-report questionnaires and interview-based diagnoses seem to be closer than has previously been assumed. The more thorough methodology applied in the present study reveals a genetic correlation between the SCL-5 and internalizing disorder that is not significantly different from unity. There is reason to assume that this correspondence is generalizable to similar self-report scales for symptoms of anxiety and depression. The present study also extends the Foley et al.'s (2001) results by finding that it is the same genetic factor that is captured by a brief self-report scale and internalizing disorder in general, as for the single disorder depression. Whereas Foley et al. only investigates females, the present study finds that the results hold for both males and females. The finding of a shared genetic factor is important for both clinical and research purposes and implies that the SCL-5 may be an excellent screener for genetic risk for internalizing disorders.

References

Aben, I., Verhey, F., Lousberg, R., Lodder, J., & Honig, A. (2002). Validity of the Bech Depression Invetory, Hospital Anxiety and Depression Scale, SCL-90, and Hamilton depression rating scale as screening instruments for depression in stroke patients. *Psychosomatics, 43,* 386-393.

Acton, G. S., Kunz, J. D., Wilson, M., & Hall, S. H. (2005). The construct of internalization; conceptualization, measurement and prediction of smoking treatment outcome. *Psychological Medicine, 35,* 395-408.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrica, 52,* 317-332.

Amodei, N., Katerndahl, D. A., Larme, A. C., & Palmer, R. (2003). Interview versus self-answer methods of assessing health and emotional functioning in primary care patients. *Psychological Reports, 92,* 937-948.

American Psychiatric Association (2000). *DSM-IV-TR: Diagnostic and Statistical Manual of Mental Disorders. Fourth Edition.* American Psychiatric Publishing Inc.

Bech, P. (2004). The impact of the SCL-90 on the validity of the DSM-IV neurotic or stress-related disorders. Editorial. *Acta Psychiatrica Scandinavica, 110,* 161-162.

Beidel, D. C., & Stipelman, B. (2007). Anxiety disorders. In: Hersen, M., Turner, S. M., & Beidel, D. C. (Eds.) (2007). *Adult Psychopathology and Diagnosis.* New York: John Wiley & Sons, Inc.

Bollen, K. A. (1989). S*tructural Equations with Latent Variables.* New York: John Wiley.

Bonynge, E. R. (1993). Unidimensionality of SCL-90-R scales in adult and adolescent crisis samples. *Journal of Clinical Psychology, 49,* 212-215.

Bouchard, T. J. (2004). Genetic influence on human psychological traits. A survey. *Current Directions in Psychological Science, 13,* 148-151.

Bouchard, T. J., and Propping, P. (eds) (1993). Twins as a tool of behaviour genetics. In Plomin R., DeFries J. D., McClearn, G. E., and McGuffin, P. (2001). *Behavioural Genetics.*Worth Publishers, 4th Edition.

Boyd, J. H., Burke, J. D., Gruenberg, E., Holzer, C. E., III, Rae, D. S., George, L. K., Karno, M., Soltzman, R., McEvoy, L., & Nestadt, G. (1984). Exclusion criteria of DSM-III: A study of co-occurence of hierarchy-free syndromes. *Archives of General Psychiatry, 41,* 983-989.

Carpenter, K. M., & Hittner, J. B. (1995). Dimensional characteristics of the SCL-90-R: evaluation of gender differences in dually diagnosed inpatients. *Journal of Clinical Psychology, 51,* 383-390.

Costa, P. T., & McCrae, R. R. (1992). The 5-factor model of personality and its relevance to personality-disorders. *Journal of Personality Disorders, 6,*343-359.

Cyr, J. J., McKennafoley, J. M., & Peacock, E. (1985). Factor structure of the SCL-90-R - is there one. *Journal of Personality Assessment, 49,* 571-578.

Clark, L. A., Watson, D. & Mineka, S. (1994). Temperament, personality, and the mood and anxiety disorders. *Journal of Abnormal Psychology, 103,* 103-116.

Czajkowski, N., Kendler, K. S., Jacobson, K., Tambs, K., Røysamb, E., & Reichborn-Kjennerud, T. (2008). Passive-aggressive (negativistic) personality disorder: a population-based twin study. *Journal of Personality Disorders, 22,* 109-122.

Derks, E. M., Dolan, C. V., & Boomsma, D. I. (2006). A test of the equal environment assumption (EEA) in multivariate twin studies. *Twin Research and Human Genetics, 9* 403-411.

Derogatis, L. R. (1983). *SCL-90-R: Administration, scoring and procedures manual II for the revised version and other instruments of the psychopatology rating scales series:* Baltimore, MD: Clinical Psychometric Research.

Derogatis, L. R. (1994). *SCL-90-R: Symptom Checklist-90-R. Administration, scoring and procedures manual.* Minneapolis, MN: National Computer Systems.

Derogatis, L. R., Lipman, D. S. And Covi, R. S. (1973). SCL-90: an outpatient psychiatric rating scale – preliminary report. *Psychopharmacology Bulletin, 9,* 13-28.

Derogatis, L. R., Lipman, R. S., Rickels, K., Uhlenluth, E. H., & Covi, L. (1974). The Hopkins Symptom Checklist (HSCL): a selfreport symptom inventory. *Behavior Science, 19,* 1-15.

Di Sclafani, V., Finn, P., & Fein, G. (2008). Treatment-naive active alcoholics have greater psychiatric comorbidity than normal controls but less than treated abstinent alcoholics. *Drug and Alcohol Dependence, 98,* 115-122.

Edvardsen, J., Torgersen, S., Røysamb, E., Lygren, S., Skre, I., Onstad, S., & Oien, P. A. (2009). Unipolar depressive disorders have a common genotype. *Journal of Affective Disorders, 117,* 30-41.

Endicott, J., & Spitzer, R. L. (1978). Diagnostic Interview - schedule for affective disorders and schizophrenia. *Archives of General Psychiatry, 35,* 837-844.

Evenson, R. C., Holland, R. A., Mehta, S., & Yasin, F. (1980). Factor-analysis of the Symptom Checklist-90. *Psychological Reports, 46,* 695-699.

Feliciano, L., & Arean, P. A. (2007). Mood disorders: Depressive disorders. In: Hersen, M., Turner, S. M., & Beidel, D. C. (Eds.) (2007). *Adult Psychopathology and Diagnosis.* New York: John Wiley & Sons, Inc.

Foley, D., Neale, M. C, & Kendler, K. S. (2001). Genetic and environmental risk factors for depression assessed by subject-rated Symptom Check List versus structured clinical interview. *Psychological Medicine, 31,* 1413-1423.

Fröjdh, K., Håkansson, A., & Karlsson, I. (2004). The Hopkins Symptom Checklist-25 is a sensitive case-finder of clinically important depressive states in elderly people in primary care. *International Journal of Geriatric Psychiatry, 19,* 386-390.

Glass, R. M., Allan, A. T., Uhlenhuth, E. H., Kimball, C. P., & Borinstein, D. I. (1978). Psychiatric screening in a medical clinic. An evaluation of a self-report inventory. *Archives of General Psychiatry, 35,* 1189-1195.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* Oxford, England: Viley.

Hansson, L., Nettelbladt, P., Borgquist, L., & Nordström, G. (1994). Screening for psychiatric illness in primary care. A cross-sectional study in a Swedish health district. *Social Psychiatry and Psychiatric Epidemiology, 29,* 83-87.

Harris, J., Magnus, P. & Tambs, K. (2002). The Norwegian Institute of Public Health Twin Panel: a description of the sample and program of research. *Twin Research, 5,* 415-423.

Haver, B. (1997). Screening for psychiatric comorbidity among female alcoholics: The use of a questionnaire (SCL-90) among women early in their treatment programme. *Alcohol and Alcoholism, 32,* 725-730.

Hesbacher, P. T., Rickels, R., Morris, R. J., Newman, H, & Rosenfeld, M. D. (1980). Psychiatric illness in family practice. *Journal of Clinical Psychiatry, 41,* 6-10.

Hettema, J. M. (2008). What is the genetic relationship between anxiety and depression? *American Journal of Medical of Medical Genetics Part C (Seminars in Medical Genetics), 148C,* 140-146.

Hettema, J. M., Neale, M. C., Kendler, K. S. (2001). A review and meta-analysis of the genetic epidemiology of anxiety disorders. *American Journal of Psychiatry, 158,* 1568-1578.

Hettema, J. M., Neale, M. C., Myers, J. M., Prescott, C. A., & Kendler, K. S. (2006). A population-based twin study of the relationship between neuroticism and internalizing disorders. *The American Journal of Psychiatry, 163,* 857-864.

Holcomb, W. R., Adams, N. A., & Ponder, H. M. (1983). Factor structure of the Symptom Checklist-90 with acute psychiatric-inpatients. *Journal of Consulting and Clinical Psychology, 51,* 531-538.

Holi, M. M., Sammalllahti, P. R., & Aalberg, V. A. (1998). A Finnish validation study of the SCL-90. *Acta Psychiatrica Scandinavica, 97,* 42-46.

Hoyle, R. H. (ed.) (1995). *Structural Equation Modelling*. SAGE Publications, Inc. Thousand Oaks, CA.

Hudziak, J. J., Copeland, W., Stanger, C. & Wadsworth, M. (2004). Screening for DSM-IV externalizing disorders with the Child Behavior Checklist: a receiver-operating characteristic analysis. *Journal of Child Psychology and Psychiatry, 45,* 1299-1307.

James, L. M., & Taylor, J. (2008). Revisiting the structure of mental disorders: Borderline personality disorder and the internalizing/externalizing spectra. *British Journal of Clinical Psychology, 47,* 361-380.

Joukamaa, N., Lehtinen, V., Karlsson, H., & Rouhe, E. (1994). SCL-25 and recognition of mental disorders reported by primary health care physisians. *Acta Psychiatrica Scandinavica, 89,* 320-323.

Jöreskog, K., & Sørbom, D. (1996). *Lisrel 8: User's Reference Guide.* Chicago: SSI, Inc.

Kendall, P. C., & Hammen, C. (1995). *Abnormal Psychology.* Boston, MA: Houghton Mifflin Company.

Kendler, K. S. (1996). Major depression and generalized anxiety disorder: same genes, (partly) different environments – revisited. *British Journal of Psychiatry, 168,* 68-75.

Kendler, K. S., & Eaves, L. J. (1986). Models for the joint effects of genotype and environment on liability to psychiatric illness. *American Journal of Psychiatry, 143,* 279-289.

Kendler, K. S., Gardner, C. O., Gatz, M., & Pedersen, N. L. (2006). The sources of co-morbidity between major depression and generalized anxiety disorder in a Swedish national twin sample. *Psychological Medicine, 37,* 453-462.

Kendler, K. S., Gardner, C. O., & Prescott, C. A. (2006). Toward a comprehensive developmental model for major depression in men. *American Journal of Psychiatry, 165,* 115-124.

Kendler, K. S., Heath, A. C., Martin, N. G, & Eaves, L. J. (1986). Symptoms of anxiety and depression in a volunteer twin population: the etiologic role of genetic and environmental factors. *Archives of General Psychiatry, 43,* 213-221.

Kendler, K. S., Kuhn, J., & Prescott, C. A. (2004). The interrelationship between neuroticism, sex and stressful life events in the prediction of episodes of major depression. *American Journal of Psychiatry, 161,* 631-636.

Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C., & Eaves, L. J. (1992). Major depression and generalized anxiety disorder: same genes, (partly) different environments? *Archives of General Psychiatry, 49,* 716-722.

Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C., & Eaves, L. J. (1993). A longitudinal twin study of personality and major depression in women. *Archives of General Psychiatry, 50,* 853-862.

Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C., & Eaves, L. J. (1993). The lifetime history of major depression in women: reliability of diagnosis and heritability. *American Journal of Psychiatry, 50,* 863-870.

Kendler, K. S., & Prescott, C. A. (2006). *Genes, environment, and psychopathology: Understanding the causes of psychiatric and substance use disorders.* New York: Guilford Press.

Kendler, K. S., Walters, E. E., Truett, K. R., Heath, A. C., Neale, M. C., Martin, N. G., & Eaves, L. J. (1994). Sources of individual differences in depressive symptoms: analyses of two samples of twins and their families. *American Journal of Psychiatry, 151,* 1605-1614.

Kendler, K. S., Walters, E. E., Truett, K. R., Heath, A. C., Neale, M. C., Martin, N. G., & Eaves, L. J. (1995). A twin-family study of self-report symptoms of panic-phobia and somatization. *Behavior Genetics, 25,* 499-515.

Kennedy, B. L., Morris, R. L., Pedley, L. L., & Schwab, J. J. (2001). The ability of the Symptom Checklist SCL-90 to differentiate various anxiety and depressive disorders. *Psychiatric Quarterly, 72,* 277-288.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005a). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry, 62,* 593-602.

Kessler, R. C., Chiu, W. T., Demler, O., Merikangas, K. R., & Walters, E. E. (2005b). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry, 62,* 617-627.

Kringlen, E., Torgersen, S., & Cramer, V. (2001). A Norwegian psychiatric epidemiology study. *The American Journal of Psychiatry, 158,* 1091-1098.

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

Krueger, R. F., & Bezdjian, S. (2009). Enhancing research and treatment of mental disorders with dimensional concepts: towards DSM-V and ICD-11. *World Psychiatry, 8,* 3-6.

Krueger R. F. & Markon, K. E. (2006). Reinterpreting comorbidity: A model-based approach to understanding and classifying psychopatology. *Annual Review of Clinical Psychology, 2*, 111-133.

Krueger, R. F., & South, S. C. (in press). Externalizing disorders: Cluster 5 of the proposed meta-structure for DSM-V and ICD-11. *Psychological Medicine.*

Lahey, B. B. (2009). Public health significance of neuroticism. *American Psychologist, 64,* 241-256.

Larsen, R. J., & Buss, D. M. (2008). Personality psychology: domains of knowledge about human nature. (3rd edition). New York: McGraw-Hill

Lecrubier, Y. (2007). Widespread underrecognition and undertreatment of anxiety and mood disorders: Results from 3 European studies. Journal of Clinical Psychiatry, 68, 36-41.

Lipman, R. S., Rickels, K., Coci, L., Derogatis, L. R., & Uhlenhuth, E., H. (1969). Factors of symptom distress. Doctors ratings of anxious neurotic outpatients. *Archives of General Psychiatry, 21*, 328-338.

Lipman, R. S., Covi, L., & Shapiro, A. K. (1979). The Hopkins Symptom Checklist (HSCL). Factors derived from the HSCL-90. *Journal of Affective Disorders, 1,* 9-24.

Loehlin, J. C. (1996). The Cholesky approach: a cautionary note. *Behavior Genetics, 26,* 65-69.

Loke, B., Nicklasson, F., & Burvill, P. (1996). Screening for depression: clinical validation of geriatrician's diagnosis, the brief assessment schedule depression cards and the 5-item version of the Symptom Check List among non-demented geriatric inpatients. *International Journal of Geriatric Psychiatry, 11,* 461-465.

Lorenz, D. R. (2003). Behavioral genetics: scientific and social acceptance. *Journal of Biolaw and Business, 6,* 30-39.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7,* 19-40.

Mackinaw-Koos, B., & Vasey, M. W. (2000). Considering sex differences in anxiety and its disorders across the life span: a construct-validation approach. *Applied and Preventive Psychology, 9,* 191-209.

Maes, R. & Horwitz, A. V. (2005). DSM-II and the revolution in the classification of mental disorders. *Journal of the History of the Behavioral Sciences, 41,* 249-267.

Magnus, P., Berg, K., & Nance, W. E. (1983). Predicting zygosity in Norwegian twin pairs born 1915-1960. *Clinical Genetics, 24,* 103-112.

Maser, J. D., & Cloninger, C. R. (eds.) (1990). *Comorbidity of mood and anxiety disorders.* Washington DC: American Psychiatric Press.

Maser, J. D., & Patterson, T. (2002). Spectrum and nosology: implications for DSM-V. *Psychiatric Clinic of North America, 25,* 855-885.

Medisinsk Fødselsregister. (1987). *Fødsler og perinatale helseproblemer 1967-1984.* Bergen, Norway: Medical Birth Registry of Norway, University of Bergen.

Middledorp, C. M., Cath, D. C., Van Dych, R., & Boomsma, D. I. (2005). The co-morbidity of anxiety an depression in the perspective of genetic epidemiology: a review of twin and family studies. *Psychological Medicine, 35,* 611-624.

Middledorp, C. M., Wray, N. R., Andrews, G., Martin, N. G., & Boomsma, D. I. (2006). Sex differences in symptoms of depression in unrelated individuals and opposite-sex twin and sibling pairs. *Twin Research and Human Genetics, 9,* 632-636.

Moum, T. (1998). Mode of administration and interviewer effects in self-reported symptoms of anxiety and depression. *Social Indicators Research, 45,* 279-318.

Neale, M. C., Boker, M. S., & Maes, H. H. (2003). *Mx: statistical modelling.* VCU box 900126 Richmond, VA 23298: Department of Psychiatry. 6th edition.

Neale, M. C., Eaves, L. J., Kendler, K. S. (1994). The power of the classical twin study to resolve variation in threshold traits. *Behavior Genetics, 24,* 239-258.

Neale, M. C. and Maes, H. H. M. (2000). *Methodology for genetic studies of twins and*

families. Kluwer Academic Publishers B.V.

Neale, M. C, Røysamb, E., & Jacobson, K. (2006). Multivariate genetic analysis of sex limitation and G x E interaction. *Twin Research and Human Genetics, 9,* 481-489.

Nes, R. B., Røysamb, E., Reichborn-Kjennerud, T., Harris, J., & Tambs, K. (2007). Symptoms of anxiety and depression in young adults: genetic and environmental influences on stability and change. *Twin Research and Human Genetics, 10,* 450-461.

Nettelbladt, P., Hansson, L., Stefansson, C. G., Borgquist, L., & Nordstöm, G. (1993). Test characteristics of the Hopkins Symptom Checklist-25 (HSCL-25) in Sweden, using the Present State Examination (PSE-9) as a caseness criterion. *Social Psychiatry and Psychiatric Epidemiology, 28,* 130-133.

Olsen, L. R., Mortensen, E. L. & Bech, P. (2004). The Scl-90 and the SCL-90R versions validated by item response models in a Danish community sample. *Acta Psychiatrica Scandinavica, 110,* 225-229.

Pedersen, G. & Karterud, S. (2004). Is SCL-90R helpful for the clinician in assessing DSM-IV symptom disorders? *Acta Psychiatrica Scandinavica, 110,* 215-224.

Peveler, R., & Fairburn, C. (1990). Measurement of neurotic symptoms by self-report questionnaire - validity of the SCL-90R. *Psychological Medicine, 20,* 873-879.

Pincus, H. A., & Pettit, A. R. (2001). The societal costs of chronic major depression. *Journal of Clinical Psychiatry, 62,* 5-9.

Plomin, R., & Daniels, D. (1987). Why are children in the same family so different from each other? *Behavioural Brain Science, 10,* 44-54.

Plomin, R., DeFries J. D., McClearn, G. E., and McGuffin, P. (2001). *Behavioural Genetics*. Worth Publishers, 4th Edition.

Posthuma, D., & Boomsma, D. I. (2005). Mx scripts library: structural equation modelling scripts for twin and family data. *Behavior Genetics, 35,* 499-505.

Rauter, U. K., Leonard, C. E., & Swett, C. P. (1996). SCL-90-R factor structure in an acute, involuntary, adult psychiatric inpatient sample. *Journal of Clinical Psychology, 52,* 625-629.

Reichborn-Kjennerud, T., Røysamb, E., Tambs, K., Torgersen, S., Kringlen, E., Magnus, P., & Harris, J. R. (2004). Genetic and environmental influences on the association between smoking and panic attacks in females: a population-based twin study. *Psychological Medicine, 34,* 1271-1277.

Reichborn-Kjennerud, T., Stoltenberg, C., Tambs, K., Roysamb, E., Kringlen, E., Torgersen, S., & Harris, J. (2002). Back-neck pain and symptoms of anxiety and depression: a population-based twin study. *Psychological Medicine, 32,* 1009-1020.

Ronan, G. F., Dreer, L. E., & Dollard, K. M. (2000). Measuring patient symptom change on rural psychiatry units: utility of the Symptom Checklist-90 revised. *Journal of Clinical Psychiatry, 61,* 493-497.

Roy, M. A., Neale, M. C., Pedersen, N. L., Mathe, A. A., & Kendler, K. S. (1995). A twin-study of generalized anxiety disorder and major depression. *Psychological Medicine, 25,* 1037-1049.

Røysamb, E., Kendler, K. S., Tambs, K., Ørstavik, R. E., Neale, M. C., Aggen, S. H., Torgersen, S., & Reichborn-Kjennerud, T. (2009). The joint structure of axis I and axis II disorders. *Journal of Abnormal Psychology,* submitted.

Røysamb, E., Harris, J., Magnus, P., Vittersø, R., & Tambs, K. (2002). Subjective well-being. Sex-specific effects of genetic and environmental factors. *Personality and Individual Differences, 32,* 211-223.

Sandanger, I., Moum, T., Ingebrigtsen, G., Dalgard, O. S., & Bruusgaard, D. (1998). Concordance between symptom screening and diagnostic procedure: the Hopkins Symptom Checklist-25 and the Composite International Diagnostic Interview I. *Social Psychiatry and Psychiatric Epidemiology, 33,* 345-354.

Sandanger, I., Moum, T., Ingebrigtsen, G., Sørensen, T., Dalgard, O. S., & Bruusgaard, D. (1999). The meaning and significanse of caseness: the Hopkins Symptom Checklist-25 and the Composite International Diagnostic Interview II. *Social Psychiatry and Psychiatric Epidemiology, 34,* 53-59.

Scherrer, J. F., True, W. R., Xian, H., Lyons, M. J., Eisen, S., A., Goldberg, J., Lin, N., &Tsuang, M. T. (2000). Evidence for genetic influences common and specific to symptoms of generalized anxiety and panic. *Journal of Affective Disorders, 57,* 35.45.

Shutty, M. S., Degood, D. E., & Schwartz, D. P. (1986). Psychological dimensions of distress in cronic pain patients - a factor-analytic study of Symptom Checklist-90 responses. *Journal of Consulting and Clinical Psychology, 54,* 836-842.

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

Spitzer, R. L., Endicott, J., & Robins, E. (1978). Research diagnostic criteria. *Archives of General Psychiatry, 35,* 773-782.

Spitzer, R., L., Williams, J. B. W., Gibbon, M., & First, M. B. (1992). The structured clinical interview for DSM-III-R (SCID). 1. History, rationale and description. *Archives of General Psychiatry, 49,* 624-629.

Strand, B. H., Dalgard, O. S., Tambs, K. and Rognerud, M. (2003). Measuring the mental health status of the Norwegian population: A comparison of the instruments SCL-25, SCL-10, SCL-5 and MHI-5 (SF-36). *Nordic Journal of Psychiatry, 57,* 113-118.

Sullivan, P. F., Neale, M. C., & Kendler, K. S. (2000). Genetic epidemiology of major depression: review and meta-analysis. *American Journal of Psychiatry, 157,* 1552-1562.

Sundet, J. M., Eriksen, W., & Tambs, K. (2008). Intelligence correlations between brothers decrease with increasing age difference. *Psychological Science, 19,* 843-847.

Sundet, J. M., Skre, I., Okkenhaug, J. J., & Tambs, K. (2003). Genetic and environmental causes of the interrelationships between self-reported fears. A study of a non-clinical sample of Norwegian identical twins and their families. *Scandinavian Journal of Psychology, 44,* 97-106.

Tambs, K., Czajkowski, N., Røysamb, E., Neale. M. C., Reichborn-Kjennereud, T., Aggen, S., Harris, J. R., Ørstavik, R., & Kendler, K. (2009). Structure of genetic and environmental risk factors for dimensional representations of DSM-IV anxiety disorders. *The British Journal of Psychiatry, 195,* 301-307.

Tambs, K. and Moum, T. (1993). How well can a few questionnaire items indicate anxiety and depression? *Acta Psychiatrica Scandinavica, 87,* 364-367.

Tambs, K., Harris, J., & Magnus. P. (1995). Sex-specific causal factors and effects of common environment for symptoms of anxiety and depression in twins. *Behaviour Genetics, 25,* 33-44.

Tambs, K., Harris, J. R., & Magnus, P. (1997). Genetic and environmental contributions to the correlation between alcohol consumption and symptoms of anxiety and depression. Results from a bivariate analysis of Norwegian twin data. *Behavior Genetics, 27,* 241-250.

Tambs, K., Rønning, T., Prescott, C. A., Kendler, K. S., Reichborn-Kjennerud, T., Torgersen, S., & Harris, J. (2009). The Norwegian Institute of Public Health Twin Study of mental health: examining recruitment and attrition bias. *Twin Research and Human Genetics, 12,* 158-168.

Tandon, M., Cardeli, E., & Lubi, J. (2009). Internalizing disorders in early childhood: a review of depressive and anxiety disorders. *Child and Adolescent Psychiatric Clinics of North America, 18,* 593-610.

Torgersen, S. (1990). The comorbidity of major depression and anxiety disorders in twin pairs. *American Journal of Psychiatry, 147,* 1099-1202.

Torgersen, S. (2003). Det strukturerte psykistriske intervjuet. In Rønnestad, M. H. & von der Lippe, A. (eds.) (2003). *Det Kliniske Intervjuet*. Oslo: Gyldendal Akademisk.

Williams, J. W., Noël, P. H., Cordes, J. A., Ramirez, G., & Pignone, M. (2002). Is this patient clinically depressed? *Journal of the American Medical Association, 287,* 1160-1170.

Widiger, T. A., Costa, P. T., & McCrae, R. R. (2002). A proposal for axis II: diagnosing personality disorders using the Five-Factor model of personality. In: Costa, P. T., & Widiger, T. A. (Eds.). *Personality disorders and the Five-Factor Model of Personality* (2nd ed.). Washington DC: American Psychological Association.

Widiger, T. A., & Mullins-Sweatt, S. N. (2009). Five-factor model of personality disorder: a proposal for DSM-V. *Annual Review of Clinical Psychology, 5,* 197-220.

Widiger, T. A., & Sankis, L. M. (2000). Adult psychopatology: Issues and controversies. *Annual Review of Psychology, 51,* 377-404.

Winokur, A., Winokur, D. F., Rickels, K. R., & Cox, D. S. (1984). Symptoms of emotional distress in a family planning service: stability over a four week period. British Journal of Psychiatry, 144, 395-399.

Witt, P. L., & McGrain, P. (1986). Non-parametric testing using the chi-square distribution. *Physical Therapy, 66,* 264-268.

Wittchen, H. U. (1993). Reliability and validity studies of the WHO-Composite International Diagnostic Interview (CIDI): a critical review. *Journal of Psychiatric Research, 28*, 57-84.

Wittchen, H. U., Lachner, G., Wunderlich, U. & Pfister, H. (1998). Test-retest reliability of the computerized DSM-IV version of the Munich-Composite International Diagnostic Interview (M-CIDI). *Social Psychiatry and Psychiatric Epidemiology, 33*, 568-578.

Wittchen, H. U., & Pfister, H. (1997). DIA-A Interviews (M-CIDI). Frankfurt, Germany: Swets & Zeitlinger.

Woody, S. R., Steketee, G., & Chambles, D. L. (1995). The usefulness of the obsessive compulsive scale of the symptom checklist-90-revised. *Behavior Research and Therapy, 33,* 607-611.

World Health Organization (1992). *The ICD-10 classification of Mental and Behavioural Disorders. Clinical Descriptions and diagnostic guidelines.* Geneva: World Health Organization.

World Health Organization (1993). *The ICD-10 Classification of Mental and Behavioural Disorders. Diagnostic criteria for research.* Geneva: World Health Organization.

Zou, K. H., O'Malley, A. J., & Mauri, L. (2007). Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation, 115,* 654-657.

Ørstavik, R. E., Kendler, K. S., & Czajkowski, N. (2007). The relationship between depressive personality disorder and major depressive disorder: a population-based twin study. *American Journal of Psychiatry, 164,* 1866-1872.

Appendix

A. Example of a univariate Mx-script

This is a univariate Mx-script used to analyze the internalizing disorder variable.

```
!*********************************
! Univariat GSL
! *********************************
#define nvar 1
#define mt 1

G1:Model parameters
CALCULATION NGroups = 9
Begin Matrices;
  X lower nvar nvar free ! genetic path coefficients
  Y lower nvar nvar !free      ! FREE !  shared
environment path coefficients
  Z lower nvar nvar free ! nonshared environment
path coefficients
 T FUll mt nvar Free ! THRESH
 D Lower mt mt  ! to make t2>t1>t0 etc.
 End Matrices;

Begin Algebra;
  A=X*X';              ! genetic variance components
  C=Y*Y';              ! shared environment variance
components
  E=Z*Z';              ! nonshared environment
variance components
 End Algebra;
Value 1 D 1 1 to D mt mt
Start .65 X 1 1
Start .7 Z 1 1
start 1.0  T 1 1
interval A 1 1
interval E 1 1
BOund  .8  1.2  T 1 1    ! set boundaries  threshold

End
! *********************************
G2:Model parameters
CALCULATION

 Begin Matrices;
  X lower nvar nvar free ! genetic path coefficients
  Y lower nvar nvar ! FREE !  shared environment
path coefficients
  Z lower nvar nvar free ! nonshared environment
path coefficients
 S FUll mt nvar Free ! THRESH
 D Lower mt mt  ! to make t2>t1>t0 etc.
 End Matrices;

Begin Algebra;
  A=X*X';              ! genetic variance components
  C=Y*Y';              ! shared environment variance
components
  E=Z*Z';              ! nonshared environment
variance components
 End Algebra;

Value 1 D 1 1 to D mt mt
start .7 S 1 1
BOund .5  1.0 S 1 1    ! set boundaries  threshold
Start .6 X 1 1
Start .7 Z 1 1
interval A 1 1
interval E 1 1
Option jiggle

End
! *********************************
G3: MALE MZ TWINS
Data NI=13 NO=0
Missing=-9
ORdinal_data file=LinexSE1(ny).DAT
Labels Group scl2 scl2dic scl2tri int1 int1dic int1tri
tscl2 tscl2dic tscl2tri tint1
tint1dic tint1tri
Select if group = 1;
Select  int1dic tint1dic ;
Begin Matrices = Group 1
THresholds D*(T|T) ;
Covariance A+C+E | A+C _
          A+C   | A+C+E ;
Option Rsiduals
Option jiggle

End
! *********************************
G4: MALE DZ TWINS
Data NI=13 NO=0
Missing=-9
ORdinal_data file=LinexSE1(ny).DAT
Labels Group scl2 scl2dic scl2tri int1 int1dic int1tri
tscl2 tscl2dic tscl2tri tint1
tint1dic tint1tri
Select if group = 2;
Select  int1dic tint1dic ;

Begin Matrices = Group 1
H FULL 1 1
End Matrices;
Matrix H .5
THresholds D*(T|T) ;
```

```
 Covariance A+C+E | H@A+C _
        H@A+C | A+C+E ;
Option jiggle
Option Rsiduals
End
! ******************************
G5: FEMALE MZ TWINS
Data NI=13 NO=0
Missing=-9
ORdinal_data file=LinexSE1(ny).DAT
Labels Group scl2 scl2dic scl2tri int1 int1dic int1tri
tscl2 tscl2dic tscl2tri tint1
tint1dic tint1tri
Select if group = 3;
Select  int1dic tint1dic ;
Begin Matrices = Group 2
THresholds D*(S|S) ;
Covariance A+C+E | A+C _
        A+C   | A+C+E ;
Option Rsiduals
Option jiggle

End
! ******************************
G6: FEMALE DZ TWINS
Data NI=13 NO=0
Missing=-9
ORdinal_data file=LinexSE1(ny).DAT
Labels Group scl2 scl2dic scl2tri int1 int1dic int1tri
tscl2 tscl2dic tscl2tri tint1
tint1dic tint1tri
Select if group = 4;
Select  int1dic tint1dic ;
Begin Matrices = Group 2
H Full 1 1
End Matrices;
Matrix H .5
 THresholds D*(S|S) ;
 Covariance A+C+E | H@A+C _
        H@A+C | A+C+E ;
Option Rsiduals
Option jiggle

End
! ******************************
G7: MALE-FEMALE DZ TWIN PAIRS
Data NI=13 NO=0
Missing=-9
ORdinal_data file=LinexSE1(ny).DAT
Labels Group scl2 scl2dic scl2tri int1 int1dic int1tri
tscl2 tscl2dic tscl2tri tint1
tint1dic tint1tri
Select if group = 5;
Select  int1dic tint1dic ;
BEGIN MATRICES= GROUP 1
S Full mt 1 =S2
H Full 1 1
U lower  1 1 = X1              !female a
F lower 1 1 = Y1                   !female c
W lower 1 1 = Z1             !female e
K CO 1 1  = A1                   !female A
G co 1 1 = C1                  !female C
O CO 1 1 = E1                   !female E
H lower 1 1 free
End Matrices;
interval H 1 1
bound .0 .5 H 1 1
THresholds D*(T|S) ;
Covariances A+C+E     |  H@(X*U') + (Y*F') _
            H@(U*X') + (F*Y')  | K+G+O ;
Option jiggle
OPTION RSIDUAL

END
! ******************************
Const8
Constraint
Matrices=Group 1;
M Unit 1 1
End matrices;
Constrain M=\d2v(A+C+E);
End

Const9: to standardize vars
Constraint
Matrices=Group 2;
M Unit 1 1
End matrices;
Constrain M=\d2v(A+C+E);
End
```