Medical psychometrics

Lodder, Paul

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

PAUL LODDER

MEDICAL PSYCHOMETRICS

A psychometric evaluation of Type D personality and its predictive value in medical research

Medical Psychometrics

A psychometric evaluation of Type D personality and its predictive value in medical research

Paul Lodder

Author: Paul Lodder | www.paultwin.com

ISBN: 978-94-6458-789-0

Printing: Ridderprint | www.ridderprint.nl

Cover & design: JeRoen Murré | www.jeroenmurre.nl

The work presented in this dissertation was partly funded by Jelte Wichert's Consolidator Grant (IMPROVE) from the European Research Council (ERC; grant no. 726361).

Medical Psychometrics: A psychometric evaluation of Type D personality and its predictive value in medical research

Proefschrift

ter verkrijging van de graad van doctor aan Tilburg University
op gezag van de rector magnificus, prof. dr. W.B.H.J. van de Donk, in het
openbaar te verdedigen ten overstaan van een door het college voor
promoties aangewezen commissie in de Aula van de Universiteit op
vrijdag 27 januari 2023 om 13.30 uur

door

Paul Lodder

geboren te Middelharnis

Promotores

Prof. dr. J.M. Wicherts, Tilburg University

Prof. dr. J. Denollet[†], Tilburg University

Copromotores

Dr. W.H.M. Emons, Tilburg University

Dr. H.M. Kupper, Tilburg University

Promotiecommissie

Dr. S.N.C. Duijndam, Tilburg University

Prof. dr. B.M. Hughes, National University of Ireland, Galway

Prof. dr. T.W. Smith, The University of Utah

Dr. M.G.E. Verdam, Leiden University

Prof. dr. J.K. Vermunt, Tilburg University

[†] Prof. Dr. Johan Denollet passed away on October 26th 2019

TABLE OF CONTENTS

CHAPTER 1

General introduction

"On the whole, it is hard to believe that personality is not related to the development and progression of disease. A major difficulty, however, concerns the definition and measurement of personality. Until these problems are resolved, inconsistent outcomes are likely to persist in this area." – Johan Denollet (1993)

In 1978 the American Psychological Association formally acknowledged the field health psychology as a subdiscipline of psychology. This marked a turning point in the history of psychology, as its focus on health was no longer limited to mental health. Indeed, health psychologists started to study the psychological and behavioral processes related to physical health and healthcare. In the early days, research focused on health consequences of (negative) emotions and stress resulting from major life changes such as loss or bereavement (Dohrenwend & Dohrenwend, 1974; Levy, 1985) with less emphasis on more stable psychological characteristics such as personality. Howard Friedman's work paved to the road towards an increased emphasis on the role of personality in health. In a meta-analysis, Friedman & Booth-Kewley (1987) pointed to the existence of a "disease-prone" personality involving anxious, depressed, angry, and hostile traits. When meta-analytically relating these characteristics to several diseases based on prospective cohort studies, personality characteristics were found to be primarily associated with coronary artery disease (CAD), giving rise to the term "coronary-prone personality".

Friedman and Rosenman (1959) argued for the importance of Type A personality as a risk factor in the development and progression of CAD. Individuals with Type A personality are typically highly organized and ambitious, but also show anxious, angry, and hostile tendencies. They often set themselves high expectations, increasing job-related stress. The combination of these characteristics was found to be an important risk factor of both the incidence and progression of CAD (Barefoot et al., 1989; Haynes et al., 1980), though others found no support for this hypothesis (Shekelle et al., 1985). In light of these inconsistent findings, Denollet (1993) argued that Type A personality only partly explained the role of personality as a risk factor in CAD. By applying cluster analysis to a range of different personality traits, he identified several discrete personality subtypes and showed that both distressed (elevated stress levels) as well as inhibited (decreased self-expression) personalities were associated with various known behavioral correlates of CAD (e.g., hostility

and anger suppression). These distressed and inhibited personality constructs were subsequently reconceptualized as negative affectivity (NA) and social inhibition (SI), respectively. After transforming the scores on these two personality traits in four personality groups based on splitting the scores at their median value, Denollet, Sys & Brutsaert (1995) found that CAD patients who scored above the median on both traits had a six times higher risk on cardiac mortality than CAD patients in the three other personality groups combined. This study marks the beginning of research on Type D personality and at the time of this writing several hundreds of studies have been published on this distressed personality type and its association with various medical and psychosocial outcomes.

In this general introduction we will first discuss the procedure used to operationalize Type D personality from the NA and SI measurements. Then we will discuss the psychometric properties of the instrument used to measure NA and SI. Next, we will argue how to best conceptualize the construct Type D personality and its association with other constructs, followed by a review of the methods commonly used to statistically model such associations. We end this chapter by providing the general aims and outline of this dissertation.

Type D personality

In a landmark publication, Denollet and colleagues (1996) coined the term Type D personality to indicate a distressed personality type that is a combination of high NA and high SI. NA was defined as the stable tendency of experiencing negative thoughts and emotions, regardless of the time or situation. Individuals showing high NA tend to have a negative self-image and experience feelings of dysphoria. SI refers to the inhibited expression of emotions or behaviors during social interactions. Individuals with high SI are more reserved and tend to avoid social interaction. Denollet (2000) considered the combination of high SI and high NA a distressed personality type, because not only the negative emotions but also chronic distress resulting from failure to express those emotions was likely to harm an individual's health. Indeed, Denollet and colleagues (1996) found that Type D personality, operationalized as the combination of high scores on both NA and SI, predicts long-term mortality in a sample of 303 CAD patients. CAD is the most commonly occurring type of cardiovascular disease (CVD; American Heart Association, 2008). In high income countries, CVD and cancer are the leading causes of death for men and women alike

(Mahase, 2019), highlighting the importance of studying potential risk factors of this disease such as Type D personality.

The estimated prevalence of Type D personality depends on the definition used to classify people as having a Type D personality or not. To the best our knowledge, all recent studies that estimate Type D's prevalence are based on the same classification procedure. According to this procedure, an individual has a Type D personality if he/she scores above a predetermined cutoff on both NA and SI. Researchers generally use a cutoff score of 10 on both total scores when NA and SI are measured with the DS14 questionnaire (Denollet, 2005), which is currently the standard instrument used to measure NA and SI. This cutoff of 10 was originally chosen because the medians of NA and SI are approximately equal to 10 and at the time median splits were occasionally used to transform two continuous measures in a new grouping variable (York & John, 1992). In the context of Type D personality, the adequacy of this cutoff of 10 has been supported in earlier research, showing that the measurement precision of all items was highest across a wide range of NA and SI trait scores, including the cutoff score of 10 (Emons, Meijer & Denollet, 2007).

Mathematically, it is possible to calculate the expected prevalence of scoring above the median on two continuous variables while making assumptions about the bivariate distribution of these variables. When both variables are normally distributed and uncorrelated, then 25% of the individuals will score above the median on both variables. When the variables are normally distributed and show a correlation of .38 **(**see **Chapter 5)**, then the expected prevalence will increase to 31%. Adding to the distributions of these correlated variables a positive skewness of 1 (i.e., low scores occur relatively more often than high scores), will decrease the expected prevalence from 25% to 31%.

Empirically, in the general population the estimated prevalence ranges from 19% (Denollet, 2005), to 22% (Beutel et al., 2012) and 31% (Grande, Romppel, Glaesmer, Petrowski & Hermann-Lingen, 2010; Williams, Abbott & Kerr, 2015). The prevalence appears to decrease with higher age and is slightly higher in females than in males (Beutel et al., 2012). In the population of patients with cardiovascular disease, a meta-analysis indicated that the estimated prevalence of Type D personality ranges between 13.5% and 35% (Grande,

Romppel & Barth, 2012). The relatively high prevalence of Type D personality points to the importance of carefully studying this risk factor in the context of CVD.

Measuring Type D personality

The measurement of Type D personality has seen considerable development over the years. Strictly speaking, the construct Type D personality itself is not measured, but operationalized by applying the previously discussed cutoff procedure to the measurements of NA and SI. In the early days, there was no dedicated measurement instrument available, so NA and SI were measured by proxy using other instruments. Denollet, Sys & Brutsaert (1995) measured NA and SI using subscales of both the State-Trait Anxiety Inventory (STAI) and Heart Patients Psychological Questionnaire (HPPQ), respectively. Inspired by these measures, Denollet (1998) developed the Type-D scale-16 (DS16), measuring each of the NA and SI constructs with eight items. Several years later the DS16 was first revised in an extended version, the DS24 (De Fruyt & Denollet, 2002), before being adjusted into the slightly shorter DS14 instrument that provided a more balanced assessment of the various aspects of NA and SI (Denollet, 2005). The DS14 measures each of the NA and SI traits with seven items on a 0-4 Likert scale. To put less strain on the patients who are often filling out the DS14 in a healthcare setting, Emons, Mols, Pelle, Smolderen and Denollet (2012) converted the DS14 into a 0-2 Likert scale version called the DS$^{(3)}$. Nevertheless, from 2005 onwards the original DS14 became the standard instrument to measure Type D personality and has up to this date been translated into at least 28 languages. The measurement properties of NA and SI have been found invariant across various cultures, genders, and cardiac diagnoses (Kupper et al., 2013a). Measurement invariance has also been established when comparing the general population to clinical populations (Emons et al., 2007).

Psychometric characteristics of the DS14

The DS14 was designed to measure the two personality traits NA and SI. This two-factor structure has initially been confirmed by a principal component analysis and reliability analyses revealed Cronbach's alpha estimates of 0.88 for NA and 0.86 for SI in a combined sample of 1305 CVD patients and 2508 individuals from the general population (Denollet, 2005). Similarly high Cronbach's alpha estimates were reported for NA and SI in a large general population sample of 2495 individuals (Grande, Romppel, Glaesmer, Petrowski, &

Herrmann-Lingen, 2010). Both the factor structure and the adequate reliabilities have been corroborated in many follow-up studies using both exploratory as well as confirmatory factor analysis (e.g., Straat, van der Ark, & Sijtsma, 2012; Svansdottir et al., 2012).

The personality traits NA and SI are related to other personality traits. When comparing the Type D traits to the Big Five personality traits, De Fruyt and Denollet (2002) found strong correlations between NA and neuroticism (.68) and between social inhibition and extraversion (-.52) in a sample of 155 Belgian nurses and policemen. Denollet (2005) concluded convergent validity for NA and SI based on strong correlations with neuroticism (r = 0.68) and extraversion (r = -0.59) respectively. Similar correlations were found by Svansdottir and colleagues (2012), who additionally showed in a sample of 498 young healthy adults that NA correlates with both anxiety (r = 0.67) and depression (r = 0.55), and that SI correlates with emotional inhibition (r = 0.50). Both NA and SI showed absent to small correlations with agreeableness, openness to experience and conscientiousness, supporting the divergent validity of these Type D traits.

Important to the current dissertation is the moderate association between NA and SI themselves. Averaged across 18 published prospective cohort studies (***Chapter 5***), the correlation between NA and SI is estimated to be .38 in cardiovascular disease patients, though slightly higher correlations ranging between .4 and .5 were found in healthy populations (Ferguson et al., 2009; Williams, Bruce & Knapton, 2017; Horwood, Anglim, & Tooley, 2015). Earlier research has indicated an increased chance of a false positive effect when analyzing the combined effect of two dichotomized continuous predictors on a continuous outcome measure (Maxwell & Delaney, 1993), especially if the two predictors are correlated (MacCallum & Marr, 1995), or measured with error (Busemeyer & Jones, 1983). This body of research has implications to the field of Type D personality because researchers are often interested in the combined effect of the two correlated variables NA and SI that are dichotomized before using them in further analysis. In this dissertation, we show why some common methods used to estimate Type D personality effects can result in biased conclusions regarding the presence of a Type D effect, among others due to the positive correlation between NA and SI.

The correlations reported above were all estimated between the sum scores of the items measuring the personality traits. Such correlations are known to be attenuated due to the presence of measurement error in the questionnaire item scores (Spearman, 1904). Therefore, the correlations at the latent construct level are arguably even higher, suggesting tight relations between those personality constructs. Despite these high correlations, various studies have shown that Type D personality remains a predictor of various health outcomes, even after controlling for other personality traits such as neuroticism and extraversion (De Fruyt & Denollet, 2002; Howard & Hughes, 2012).

Personality traits are considered a relatively enduring set of thoughts, behaviors and feelings, typically thought to remain stable after reaching adulthood (McCrea & Costa, 1994). Denollet (2005) concluded temporal stability based on high correlations between baseline and 3-month follow-up measurements for both NA (r = 0.72) and SI (r = 0.82), in a combined sample of 1305 CVD patients and 2508 adults from the general population. Whereas the NA and SI scores on average did not change during follow-up, the negative affect *state* measurements decreased significantly due to a cardiac rehabilitation program, further supporting the temporal stability of the negative affect *trait*. Romppel, Hermann-Lingen, Vesper & Grande (2012) conducted a more thorough investigation by using various statistical approaches to assess the temporal stability of NA and SI in a sample of 679 cardiac patients across a six-year follow-up. They found that the factorial structure of the DS14 was stable across time and that the NA and SI sum scores correlated around 0.6 with their repeated measurements six year later. Other research has estimated the genetic stability of both Type D personality and its subcomponents NA and SI (Kupper et al., 2011), indicating that across nine years, the heritability of NA was stable and varied only slightly (between 40 and 45%). Similar genetic stability over time was found for SI, with heritability estimates varying between 42 and 49%. A limitation of these earlier studies is that they ignore the presence of measurement error in the DS14 item scores and do no test the essential assumption of longitudinal measurement invariance. ***Chapter 8*** of this dissertation aims to tackle these issues by investigating the temporal stability of NA and SI by applying several latent variable models to longitudinal data of 2625 cancer survivors.

Multilevel Exploratory Factor Analysis of the DS14

Although the two-factor structure underlying the 14 items of the DS14 has been confirmed in several studies (e.g., Denollet, 2005; Straat, van der Ark & Sijtsma, 2012), the factor structure and measurement characteristics of the DS14 might differ between studies. To integrate DS14 data from multiple studies in individual patient data meta-analysis (***Chapter 5***) and to meaningfully compare the DS14 scores across studies, we now investigate the invariance of its factor structure across a set of published prospective cohort studies. Specifically, we present the results of a multilevel exploratory factor analysis (EFA) based on data from 14 studies involving DS14 scores of 8058 cardiovascular disease patients that feature in Chapter 5. As those included studies differ in the type of cardiovascular disease sample, the aim of the multilevel EFA is to assess the cross-level measurement invariance of NA and SI across studies and to investigate whether we can replicate the two-factor structure at the patient-level and whether the use of NA and SI total scores in ***Chapter 5*** is warranted.

Multilevel EFA allows for separating the variance at the individual (i.e., patient) level from variance at the group (i.e., study) level (Dhaenens, van Damme & Onghena, 2010). After separating the individual level correlation matrix from the study level correlation matrix, an EFA is applied to each matrix. Not separating these two sources of variance risks a confounding of study-level differences when the factor structure is estimated at the individual level. Study-level factors that explain variation in the DS14 item scores in terms of differences between studies could be pointing to differential item functioning, or differences in the measurement properties of the DS14 across studies (i.e., measurement variance; Jak, Oort & Dolan, 2013). A multilevel EFA allows us to assess the invariance of the DS14 across studies. Cross-level measurement invariance requires an equal number of factors and similarly sized factor loadings at the patient and study levels (Schweig, 2013). We used oblique rotation (geomin) to allow for correlated factors. Mplus (Version 8; Muthén & Muthén, 1998-2010) with means and variance adjusted weighted least squares (WLSMV) estimation enabled us to estimate a multilevel EFA while modeling the DS14 Likert scale item scores at their ordered categorical measurement level. Model fit was evaluated in terms of RMSEA, CFI, and SRMR. The Mplus script is available in ***Appendix A***.

Figure 1 shows the scree plots of the eigenvalues of the factors that were extracted in the multilevel EFA. The individual-level scree plot on the left suggests a 2-level factor structure based on both Kaiser's criterium (eigenvalues should be larger than 1) and the clear bend in the curve at the third factor. This two-factor structure at the individual patient level corroborates the findings of earlier factor analyses, with factor loadings (Table 1) comparable to those reported in earlier studies (e.g., Denollet, 2005; Kupper et al., 2013a; Svansdottir et al., 2012).

The study-level factor structure was less clear because the bend in the curve at the second factor suggests a one-factor structure and the smaller bend at the fifth factor suggests a four-factor structure, while Kaiser's criterium points to a three-factor structure. To resolve these inconsistencies, Table 2 presents the model fit indices for EFA models with two individual-level (within) factors for varying numbers of study-level (between) factors. The chi-squared test is statistically significant for all models, but this test is very sensitive with high sample sizes. Fit measures RMSEA, CFI and TLI all suggest excellent model fit but no clear preference for any of the four models. However, the model with four study-level factors fitted the data best based on the SRMR fit index at the study level. Each of those Study-level factors represent variance in the DS14 item scores explained by phenomena occurring at the study-level.

Figure 1: Scree plots containing the eigenvalues of the factors extracted in the multilevel exploratory factor analysis on both the individual-level (within) and study-level (between).

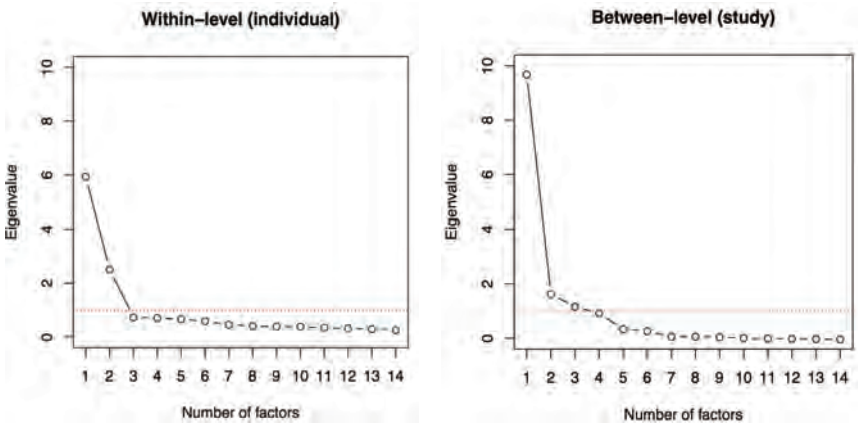


Table 1: For each DS14 item, the intraclass correlations (ICC) expressing the amount of variance explained by the study level (between). The last two columns indicate the estimated and geomin rotated factor loadings for the two-factor solution at the individual level (within).

				Patient-level factors		Study-level factors			
	Item	Label	ICC	Factor 1	Factor 2	Factor 1	Factor 2	Factor 3	Factor 4
	DS1	I make contact easily when I meet people	0.04	**0.972**	-0.482	0.019	**0.519**	**0.827**	0.019
	DS3	I often talk to strangers	0.04	**0.761**	-0.428	**0.425**	-0.011	-0.237	**0.618**
	DS6	I often feel inhibited in social interactions	0.06	**0.598**	0.267	**0.758**	0.245	0.035	0.120
SI	DS8	I find it hard to start a conversation	0.03	**0.77**	0.057	**1.115**	-0.086	-0.063	-0.282
	DS10	I am a closed kind of person	0.04	**0.718**	0.006	0.379	-0.002	**0.557**	0.260
	DS11	I would rather keep other people at a distance	0.05	**0.62**	0.009	**0.898**	0.059	-0.170	0.093
	DS14	When socializing, I don't find the right things to talk about	0.04	**0.809**	0.034	**0.954**	0.123	0.139	-0.227
	DS2	I often make a fuss about unimportant things	0.10	-0.068	**0.609**	-0.008	**0.691**	0.028	**0.630**
	DS4	I often feel unhappy	0.12	0.007	**0.723**	-0.060	**1.066**	-0.101	-0.135
	DS5	I am often irritated	0.10	0.004	**0.76**	**0.538**	**0.505**	0.056	0.033
NA	DS7	I take a gloomy view of things	0.10	0.212	**0.595**	0.060	**0.909**	0.078	-0.065
	DS9	I am often in a bad mood	0.12	0.118	**0.701**	0.006	**0.922**	0.056	-0.015
	DS12	I often find myself worrying about something	0.12	-0.033	**0.796**	0.026	**0.822**	-0.578	0.231
	DS13	I am often down in the dumps	0.10	0.136	**0.774**	0.041	**0.963**	-0.272	-0.009

* Factor loadings larger than 0.4 are printed in bold

Table 2: Fit indices for the multilevel exploratory factor analysis models with two individual-level (within) factors, varying across the number of study-level (between) factors. Bold faced cells indicate acceptable model fit.

Study-level factors	1	2	3	4
Parameters	111	124	136	159
χ^2	174.11*	162.28*	152.99*	147.38*
RMSEA	**0.005**	**0.006**	**0.006**	**0.007**
SRMR (patient level)	**0.069**	**0.069**	**0.069**	**0.069**
SRMR (study level)	0.154	0.120	0.097	**0.044**
95%CI	**[.002,.008]**	**[.003,.008]**	**[.003,.009]**	**[.004,.010]**
CFI	**0.996**	**0.996**	**0.996**	**0.995**
TLI	**0.995**	**0.994**	**0.993**	**0.991**

** $p < .05$*

Table 1 shows the estimated intraclass correlations that indicate the proportion of variance in the DS14 item scores attributable to differences between studies. The estimates are higher for NA item scores than for SI item scores, suggesting that NA items are better indicators of study-level phenomena than SI items. Approximately 5% of the variance in SI item scores and 10% of the variance in NA item scores is explained by differences between the studies included in these data, indicating the importance of separating these sources of variance by using a multilevel exploratory factor analysis.

Table 1 also shows the factor loadings for the solution involving two patient-level factors and four study-level factors. The two-factor structure at the patient level matches the structure reported in earlier DS14 factor analyses (e.g., Denollet, 2005; Straat, van der Ark & Sijtsma, 2012), with all NA and SI items clearly loading on their corresponding factor. Although the four-factor model showed the best fit at the study level, the first two of those factors correspond to the NA and SI factors at the patient level. The two additional study-level factors are each associated with two items. The items DS1 (I make contact easily when I meet new people) and DS10 (I am a closed kind of person) load on study-level factor 3, while the items DS2 (I often make a fuss about unimportant things) and DS3 (I often talk to

strangers) load high on study-level factor 4. These two additional study-level factors are caused by between-study differences such as differences across studies in how participants respond to these two item sets. Such differential item responses may for instance be caused by sociodemographic or linguistic and cultural differences between the study samples. The current findings suggest that the DS14 does not show full cross-level measurement invariance in this selection of 15 prospective cohort studies, supporting our choice to separate the study-level variance from the patient-level variance in our individual patient data meta-analysis on the relation between Type D and adverse events in cardiovascular disease (***Chapter 5*****).**

Figure 2: Best fitting two-factor model underlying the DS14 item scores. Bold arrows represent the standardized factor loadings, while the dashed arrow indicates the factor correlation. Items DS01 and DS03 are reversely coded.

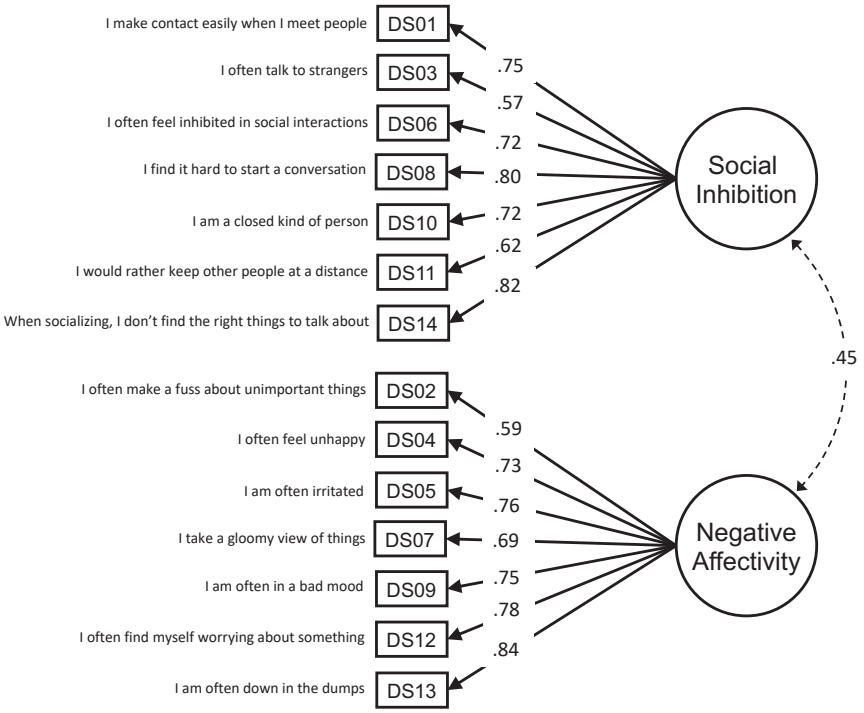


Figure 2 visualizes the two-factor model that best fitted the individual patient level in the multilevel EFA. The circles represent latent (unobserved) variables while the rectangles represent manifest (observed) variables. The arrows connecting the latent NA and SI traits causally point toward the observed item scores. We assume a reflective model for the relation between those latent constructs and the DS14 item scores, meaning that individual differences on these latent constructs are the main reason for individual differences in the corresponding item scores. This reflective model requires a realist ontology for NA and SI (Borsboom, 2003), in which these traits exist and affect the directly observable tendencies as measured with the DS14.

Conceptualizing Type D personality

An important question is how Type D personality relates to the latent traits of NA and SI in this model. A first option is to assume a reflective model for this particular relation, implying that individual differences in NA and SI are causally influenced by individual differences in Type D personality. This would mean that Type D precedes NA and SI both causally and developmentally. A second option is to assume a formative model for Type D. This would reverse the causal direction in such a way that individual differences in NA and SI causally influence (or create) individual differences on the latent Type D variable. According to a formative interpretation Type D personality would be akin to an index, because someone's position on the latent Type D variables depends on someone's position on the latent NA and SI variables, but not the other way around. This would make the construct Type D personality more similar to constructs such as socio-economic status (SES) or quality of life. Changes in SES do not cause individual differences in yearly income, but an increased income can result in a higher SES. Similarly, a formative model implies that changes in NA and SI can change the Type D status, but not vice versa. Such a formative model can still include Type D personality as a latent variable in the structural model as a predictor of important (health) outcomes over and beyond the predictive effects of NA and SI.

A third more pragmatic option is to argue that Type D personality does not exist ontologically and is merely a label used to describe the phenomenon that high scores on both NA and SI are predictive of a particular outcome. According to this interpretation it is not even necessary to add a latent Type D variable to Figure 2, because the latent variables

NA and SI are sufficient in capturing individual personality differences. Type D personality can then still be used to descriptively classify people in personality groups, but individual differences in the personality traits NA and SI are sufficient in explaining variation in a particular outcome. According to this pragmatic option the latent construct Type D personality does not causally influences other constructs and mainly serves an instrumental purpose, with its predictive validity resulting from the causal influence that both NA and SI exert on various outcomes.

Denollet (1993) argued that "since personality is conceived as a complex system that underlies regularities in human behavior, personality research should look beyond the traditional question of how single traits affect single behaviors to the way traits combine in the determination of behavior". Furthermore, when discussing this combination of NA and SI, Kupper and Denollet (2007) argued that "the combination of these two personality traits, called Type D personality, has shown to reliably predict adverse outcome in several groups of patients suffering from cardiovascular disease" (p. 118). These writings suggest that the Type D construct at least in part served a pragmatic purpose because it enabled estimation of a combined effect and showed predictive validity when used in statistical analyses. Later in this section we will return to the key question of how the effects of NA and SI combine when predicting an outcome.

The conceptualization of Type D personality also relates to the ongoing debate on whether psychological constructs can better be conceptualized as dimensions or typologies (Meehl, 1995; de Boeck, Wilson & Acton, 2005). A construct can be seen as fully dimensional if there exist gradual differences between individuals on a continuum ranging from low to high scores, with no major discontinuities in the score distribution. Personality traits have a long history of being considered dimensional constructs (Eysenck, 1967), for instance with the many individual differences in the scores on the personality trait neuroticism following a normal distribution.

Some researchers do not accept the black-and-white distinction between typologies and dimensions and argue that the difference between them is itself a matter of degree (De Boeck, Wilson & Acton, 2005). Indeed, recent advances in psychometrics have resulted in

mixture approaches that model individual differences on a dimension with a discontinuity in the distribution that is explained by the typology (Muthén & Muthén, 2006). An example of a mixture construct is coping styles. The different ways according to which individuals tend to cope with negative experiences suggests the existence of qualitatively distinct coping styles (e.g., avoidance, or problem solving). Nevertheless, individual differences exist in the extent to which individuals behave in accordance with a particular coping style (van Montfort, Kupper, Widdershoven & Denollet, 2018), and the emphasis on a particular coping style is also influenced by contextual factors such as the occurrence of continued environmental stress (Blount, Davis, Powers & Roberts, 1991).

Figure 3: Example of a dimensional, typological and mixture representation of a construct. Dimensions and mixtures involve within category heterogeneity, while typologies involve within-category homogeneity. Both typologies and mixtures involve qualitative differences between individuals, while both dimensions and mixtures involve quantitative differences.

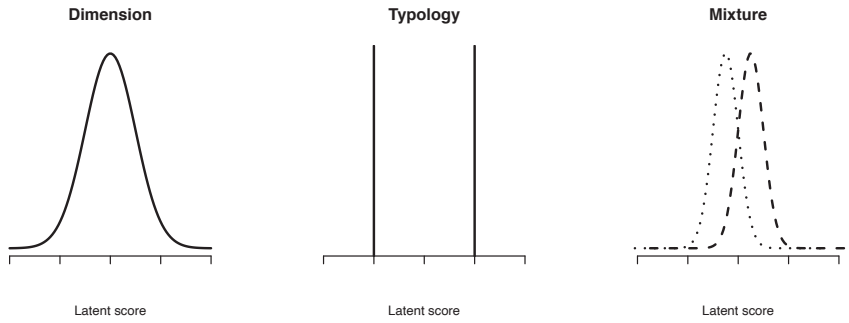


Figure 3 visualizes the distinction between dimensional, typological and mixture constructs. The figure highlights the implied within-group homogeneity of the typological approach and the relaxation of this restriction for the mixture approach. When relating this typology versus dimensions debate to Type D personality, the word “type” suggest that this construct involves classifying individuals in a personality type (i.e., those scoring high on NA and SI). The question becomes whether this personality type shows within-category homogeneity or whether individuals are allowed to differ in NA and SI *within* the Type D category. Denollet, Schiffer and Spek (2010) argued that such within category individual differences are

possible: “Dimensional and categorical approaches to personality are not mutually exclusive but represent two ways of capturing psychological tendencies of individuals. Type D refers to individuals who are more similar to their subgroup’s personality profile than other personality profiles, but, of course, individuals belong only probabilistically to these subgroups” (p. 553).

Classifying individuals in subgroups also does not necessarily require that this classification exists in an ontological sense. Classification based on shared characteristics can be one of various ways to capture psychological tendencies of individuals (Denollet, Schiffer & Spek, 2010). Classification in personality groups is a *person-centered* approach (Asendorpf & Denissen, 2006). Such a person-centered classification of people in a Type D or non-Type D group was considered to have several advantages. First, the results of statistical tests could conveniently be interpreted as the effect on a particular outcome of having a Type D personality versus not having a Type D personality. As such, the Type D effect could conveniently be assessed by interpreting the results of a single statistical test. Furthermore, a typological approach allows for classifying people into categories, where each person for instance receives the label "Type D" or "no Type D". Such an approach assigns each person to exactly one personality subgroup and enables further characterization of these subgroups in terms of sociodemographic or medical characteristics. Third, if Type D personality is a risk factor for adverse health outcomes, then convenient screening individuals for risk factors requires information about whether an individual scores “high enough” on NA and SI to be at risk. From a clinical perspective, person-centered classification makes the screening and medical decision making more convenient. Lastly, the person-centered approach resonates with the idea that only high scores on *both* NA and SI are predictive of various aspects of people's life (Denollet, 2005).

The person-centered approach can be contrasted with a *variable-centered* approaches that focuses on the associations between (often dimensional) variables rather than on classifying individuals in groups. The variable-centered approach is relevant to research on Type D personality because NA and SI are generally considered to be dimensional. Similar to other personality traits such as neuroticism and introversion, gradual individual differences exist on the latent NA and SI dimensions, and they are approximately normally distributed

without an obvious discontinuity (***Chapter 7***, figure 2). If individual differences in negative affectivity and social inhibition are possible, then one could imagine individual differences in the extent to which people have a Type D personality. This would imply that Type D is either a dimensional construct or a mixture of a typology and a dimension (Hillen, 2017).

Indeed, researchers have argued that a typology is not the appropriate way to conceptualize Type D personality (Ferguson et al., 2009; Smith, 2011). These authors claim that the personality types are constructed to be categorical, by reducing the scores on two dimensional personality traits to a limited number of personality types. Reifying these artificially constructed personality types would be committing the fallacy of misplaced concreteness: mistakenly assigning concrete existence to abstract concepts (Whitehead, 1997). This argument questions the realist ontology assigned to personality types and implies that such a typology can better be seen as instrumental according to a formative model. The burden of proof therefore lies with those who want to use a reflective model and assign ontological existence to personality types above and beyond personality dimensions such as NA and SI. Indeed, MacCallum and colleagues (2002) argue that claiming the existence of types, and consequently dichotomizing continuous variables in groups, requires compelling support from taxometric analyses. In the context of Type D personality, a taxometric analysis showed that Type D can better be seen as a dimensional construct (Ferguson et al., 2009).

In support of a mixture conceptualization, Denollet (1993; 2000) argued that individuals within each cluster are more similar to each other than to individuals in other clusters. This suggests that individuals with Type D personality should be clearly distinguishable from those without Type D personality. As a test of this assumption, Hillen (2017) applied a latent variable mixture model to a general population sample of 1587 adults to investigate whether individual differences in NA and SI can be reduced to two latent classes representing those with and those without Type D personality. Although those in the Type D class turned out to have higher NA and SI scores than those in the non-Type D class, these differences were too small to validly distinguish these two classes based on Meehl's (1995) criterion for sufficient class separation.

The empirical inadequacy of both the typological and mixture approaches suggests that the dimensional approach may be the most adequate conceptualization of Type D personality. Individuals can differ in their position on the latent NA and SI dimensions. But should this dimensional conceptualization be limited to the traits NA and SI, or is there also a role for a construct called Type D personality? We consider the construct Type D personality a useful label referring to individuals with high scores on both NA and SI. This label in part originated to facilitate interpretation of the empirical phenomenon that high scores on *both* NA and SI are predictive of various outcomes. However, the finding that high scores on both traits predict various outcomes does not necessarily imply that a latent construct exists that represents high scores on these two constructs. Many other examples exist of two constructs for which the combined effect is predictive of an outcome, such as *mixed states* in bipolar disorder (high scores on both manic and depressive symptoms; Goldberg et al., 1998) or *comorbid anxiety and depression* (May-Ling, Loxton & McLaughlin, 2015). For each of those examples it is possible to classify individuals in a group representing high scores on both constructs and to empirically show that this classification is predictive of various outcomes. However, such an empirical finding would not necessarily be a reason to assume that *comorbid anxiety and depression* is a new latent construct that *causally* influences various outcomes. The causal influence more likely resides in the underlying constructs anxiety and depression. In other words: the label *comorbid anxiety and depression* is empirically convenient, but causally irrelevant. Furthermore, proposing a new latent construct for all constructs that show a combined effect on an outcome would result in a proliferation of many new latent constructs and is therefore not parsimonious.

Relating this discussion to Type D personality, we argue that it is not the construct Type D personality, but the personality traits NA and SI that are causally related to various outcomes. Although classifying individuals in personality groups can be a convenient way to describe various psychological tendencies, we argue that NA and SI should be the main focus in statistical analyses in research on Type D personality. How to estimate this combined influence of NA and SI on an outcome is a statistical issue. In the Type D personality literature various methods have been used to model its prediction on outcomes.

Modeling the Type D effect

In research on Type D personality researchers predominantly investigate the association between Type D personality and an outcome using statistical models that are part of the generalized linear model family (e.g., logistic regression; linear regression; ANOVA), though other models are also often used (e.g., cox regression; repeated measures ANOVA). Considerable debate exists on how to best model Type D effects in statistical analyses (Ferguson et al., 2009; Smith, 2011; Coyne & de Voogd, 2012). The conceptualization of Type D personality determines what statistical models are appropriate for analyzing the predictive value of Type D. The fact that different statistical models can produce inconsistent findings stresses the importance of a clear conceptual definition of Type D personality. In the previous section we have argued for conceptualizing Type D personality as a label representing the empirical phenomenon that high scores on both NA and SI causally predict various outcomes. This implies that the statistical focus should be on detecting how the dimensional NA and SI traits causally influence an outcome. Kupper and Denollet (2007) state that it is the combined presence of both NA and SI that has been found as a risk factor for various outcomes. From this point onward, when we speak of a *Type D effect,* we refer to this assumed combined causal influence of NA and SI on an outcome. How to model this combined effect of NA and SI is a statistical question. This dissertation aims to shed more light on this issue by studying various methods used to model Type D effects. We will now discuss the methods most commonly used in the literature.

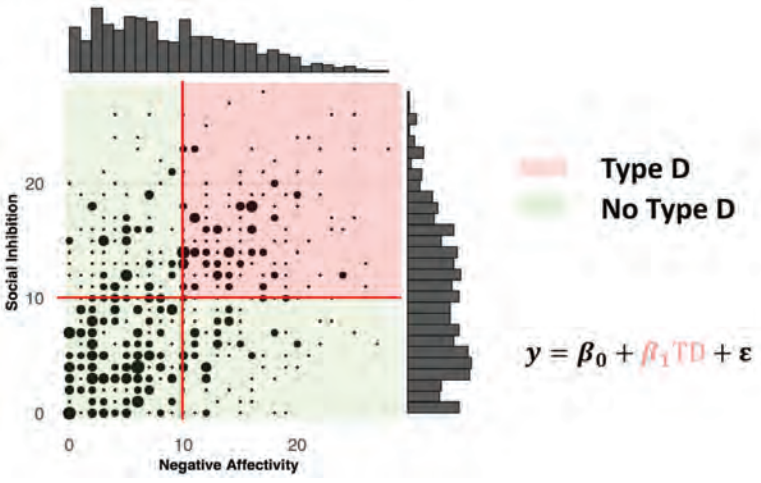
2-group method

Most approaches to model Type D effects first compute the total NA and SI scores by summing the scores on the items measuring each of those constructs. Subsequently, the methods start to diverge. The oldest and most used approach classifies individuals in two personality groups based on whether they score above a predetermined cutoff on both the NA and SI scales. Figure 4 shows a scatterplot and histograms of the NA and SI sum scores (Figures 4-6 are based on data from Denollet et al., 2013b). The size of a dot in the scatterplot represents the number of participants with that score combination. The figure shows how the 2-group method transforms the continuous NA and SI scores into a dichotomous Type D variable. Each participant who scores equal to or above the cutoff score of 10 on both total scores (two red lines) receives the status “Type D”, while all other

participants receive the status “No Type D”. The resulting dichotomous variable is then often included in a regression equation to predict some outcome variable.

This 2-group method inherits all the advantages of a person-centered approach discussed above. The 2-group method has likely been motivated by both practical and clinical considerations. The cutoff score of 10 was initially based on a median split of the NA and SI scores. Although the current consensus is to avoid the use of median splits (Royston, Altman & Sauerbrei, 2006), in the early days of research on Type D personality this practice was still commonly used in medical and psychological research (e.g., York & John, 1992). Dichotomies such as the presence of a diagnosis such as anxiety disorder are commonly used in the clinical literature. Furthermore, approaches have been used to study the influence of other personality types such as Type A personality (Barefoot et al., 1989) or defensive hostility (Helmers et al., 1995). A more practical motivation is that the 2-group results in a single statistic that intuitively represents the effect of having a Type D personality versus not having a Type D personality.

Figure 4: Scatterplot and histogram of the NA and SI sum scores (based on data from ***Chapter 7***), indicating how the 2-group method transforms these scores into a dichotomous Type D variable using cutoff scores of 10 (two red lines) and how this resulting variable is typically included in a regression equation.

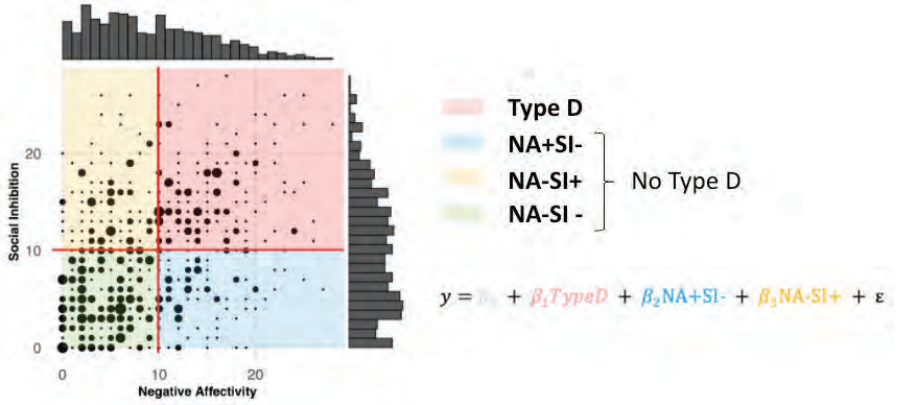


The 2-group method has been criticized in earlier publications (Ferguson et al., 2009; Smith, 2011). Many studies have pointed to the drawbacks of dichotomizing continuous variables. First, dichotomization results in lower statistical power by reducing individual differences in personality traits to two homogeneous groups (Cohen, 1983; Aiken & West, 1991). Second, bivariate dichotomization of two correlated continuous variables in a limited number of groups can also result in spurious main- and interaction effects of these two variables on an outcome measure (Maxwell & Delaney, 1993; MacCallum, Zhang, Preacher & Rucker, 2002; Royston, Altman & Sauerbrei, 2006) when only one of the variables is causally related to the outcome. In the context of Type D personality, Smith (2011) warned that significant Type D effects could also be found when either NA or SI alone were causally driving the outcome measure. If only NA is causally related to the outcome, then the 2-group method is expected to still suggest a Type D effect, because Figure 4 clearly shows that the NA scores are expected to be higher on average in the "Type D" group than in the "No Type D" group. Despite this criticism, the 2-group method is still used in empirical studies on Type D personality, yet often complemented with the continuous interaction method (see below). In ***Chapter 2*** we aim to show the consequences of using the 2-group method to the bias and false positives in the estimated Type D effects. In light of those findings, ***Chapter 4*** compares the findings reported in the Type D literature based on both the 2-group and continuous interaction method by reviewing studies that have reported the results according to both methods.

4-group method

The 4-group method constructs four rather than two personality groups based on the bivariate dichotomization of NA and SI. Figure 5 shows how this 4-group approach classifies individuals in four personality groups and how the resulting variable is dummy coded before including it in a regression equation. By modeling each of the four cells in the scatterplot, it was assumed that the 4-group method, as opposed to the 2-group method, could distinguish the causal mechanism that only NA *or* SI is related to an outcome, from the mechanism that high scores on both traits are causally efficacious. However, in ***Chapters 2 and 3*** we show that this is only true when two dichotomized continuous variables are uncorrelated. Due to the positive correlation between NA and SI spurious effects for both personality traits arise when only one trait has causal influence.

Figure 5: Scatterplot and histogram of the NA and SI sum scores (based on data from ***Chapter 7***), indicating how the 4-group method transforms these scores into a nominal personality group variable using cutoff scores of 10 (two red lines). This nominal variable is typically dummy coded before including it in a regression equation.



Like the 2-group approach, the 4-group approach suffers from the previously discussed limitations associated with dichotomizing continuous variables. Both the 2-group and 4-group approach implicitly assume that individual differences in NA and SI above or below the cutoff score of 10 are not relevant in explaining variation in an outcome measure. To prevent ignoring such individual differences, it could be worthwhile to investigate statistical models that take a dimensional approach.

Continuous interaction method

Considering Type D personality to be an interplay between the two dimensional constructs NA and SI requires statistical analyses that model how individual differences in Type D personality relate to individual differences in the dependent measure. The difficulty in modeling a Type D effect according to a dimensional operationalization resides in the fact that the effect concerns an interplay between the two constructs NA and SI (Kupper & Denollet, 2007; Denollet, 2010). But how should one model such a combined influence in statistical analysis? The answer to this question depends on whether one considers the Type D effect to be additive or synergistic. An *additive Type D effect* implies that NA and SI are each independent predictors of an outcome, but that these traits do not interact in

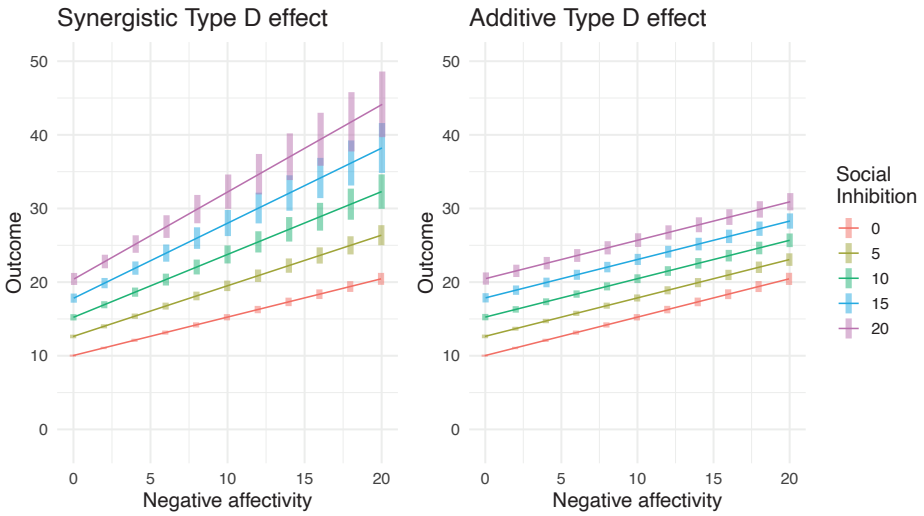
increasing each other's effect on the outcome. The panel at the right in Figure 6 indicates that in such a scenario, the predicted values on the outcome of interest are higher for those scoring high on both NA and SI than for those scoring high on only one of these traits.

However, many researchers have argued that the Type D effect is more than the additive NA and SI effects (Denollet et al., 2006; Denollet, Pedersen, Vrints, & Conraads, 2013; Denollet, Sys, & Brutsaert, 1995; Kupper & Denollet, 2007; Pedersen & Denollet, 2003). For instance, Kupper and Denollet (2007) explicitly stated that Type D personality is a *synergy* between NA and SI. Statistically, synergy can best be modeled in terms of a statistical interaction between the two constructs having a synergistic effect (Smith, 2011). When the interaction between NA and SI is significant in such a way that across the entire observed NA and SI score range, the conditional effect of each trait increases with higher scores on the other trait, then there is a *synergistic Type D effect*. Note that the direction of this synergistic effect can be either positive or negative, depending on whether the hypothesized association between Type D personality and an outcome measure is positive or negative. As the left panel in Figure 6 illustrates, a synergistic effect implies that the predicted values on the outcome of interest resulting from an interaction model are higher than the predicted values resulting from a model only including the NA and SI main effects. This requirement excludes interaction effects that for instance imply that the effect of NA on some outcome becomes smaller at higher SI scores than at lower SI scores. Note that the first-order effects of NA and SI were equal in Figure 6, so plotting SI scores on the x-axis given separate values of NA would have produced the same figure. However, in empirical data these first-order effects are not necessarily equal, making it important to separately visualize both conditional NA and SI effects.

Both additive as well as synergistic Type D effects can be estimated in regression models including the NA and SI total scores, as well as their interaction effect. Compared to the 2-group or 4-group approaches, this continuous interaction approach is expected to both have more statistical power because individual differences in NA and SI are included in the model. Furthermore, the continuous interaction approach is also able to distinguish synergistic Type D effects from additive Type D effects and effects of NA or SI only (Ferguson et al., 2009; Smith, 2011). Researchers have been using this continuous interaction method since 2009.

Although this method does not suffer from the main problems that trouble the 2-group and 4-group methods, it has problems of its own when not modeled adequately. Earlier research has shown that the presence of a quadratic effect for one of two correlated continuous variables can masquerade as interaction effects between those two predictors (Busemeyer & Jones, 1983; Belzak & Bauer, 2019). Inspired by these studies, the aim of ***Chapter 3*** is to investigate the consequences of not investigating whether a significant interaction between NA and SI can better be seen as a quadratic effect of either NA or SI.

Figure 6: Simulated data examples of the predicted scores on an outcome given various NA and SI scores. The panel at the left shows an example of a synergistic Type D effect while the panel at the right shows an additive Type D effect.



Latent variable models

Most studies in the Type D personality literature have taken an observed score approach by modeling the Type D effect based on the NA and SI sum scores (continuous interaction method) or on dichotomizations of these sum scores (2-group & 4-group methods). By using the NA and SI sum scores, these observed score methods assume that the NA and SI scores are perfectly reliable measurements of the latent NA and SI constructs. However, both classic (Traub, 1997) and modern (Hambleton, Swaminathan & Rogers, 1991) test theory

assume that item score variation is caused both by variation in the true scores on a latent construct, as well as by other random influences called measurement error. The variation in the item scores of reliable measurement instruments is mostly caused by the latent construct rather than by measurement error. Not considering the presence of measurement error assumes perfectly reliable measurement of latent constructs, which is seldom the case in psychology. Because sum scores do not separate the true score variance from the error score variance, the noisy measurement error variance attenuates the true association between the latent constructs, resulting in underestimated effects, a phenomenon called attenuation bias (Spearman, 1904). This problem is especially relevant in the context of modeling interaction effects because the measurement error in the two interacting variables is compounded in their product term.

Latent variable methods such as item response models (Rasch, 1960; Birnbaum, 1968) or structural equation models (Jöreskog & Sörbom, 1993) use a *measurement model* to estimate the associations between the latent constructs and the scores on the items designed to measure these constructs. Each latent variable can have its own measurement model and the association(s) between two or more latent variables are expressed in what is called the *structural model*. By directly modeling the observed item scores, these methods can separate the measurement error variance from the true score variance, preventing attenuation bias. They also do not have to assume that all items measure the latent construct equally well, while this is implicitly assumed when including unweighted sum scores as variables in a model. Therefore, we expect that latent variable methods provide less biased estimates of the Type D personality effects than methods based on the sums of observed item scores.

Compared to traditional regression modeling, latent variable modeling is still a relatively young field. For many complex models (e.g., latent variable interaction models) it remains unclear whether they produce unbiased parameter estimates and to what extent they result in false positive or false negative conclusions. The simulated datasets used to develop these novel methods often have good statistical properties, such as a reasonably large sample size and normally distributed latent variables and item scores. In contrast, the empirical data are often less ideal, with small sample sizes and ordinal item scores that are often not normally

distributed (e.g., the DS14 item scores). This stresses the importance of studying the performance of these models when applied to data with suboptimal characteristics.

In ***Chapters 6 and 7*** we argue for modeling the Type D personality effect as an interaction between latent variables using structural equation modeling. Several methods to model latent interaction effects have been proposed in the literature (Kenny & Judd, 1984; Ping, 1995; Klein & Moosbrugger, 2000; Marsh, Wen & Hau, 2004), but it remains unclear which of those performs best when item scores are ordinal and positively skewed like those of the DS14. To determine this, each of those chapters also includes a simulation study that compares the performance (in terms of bias, power and false positives) of various latent interaction models in a wide range of datasets with varying characteristics. Latent variable models are especially beneficial when testing interactions or quadratic effects. As these effects are often based on multiplications of individual item scores, the measurement error in these scores also gets multiplied and is therefore larger than in the original item scores. This highlights the importance of separating the measurement error variance using a latent variable model.

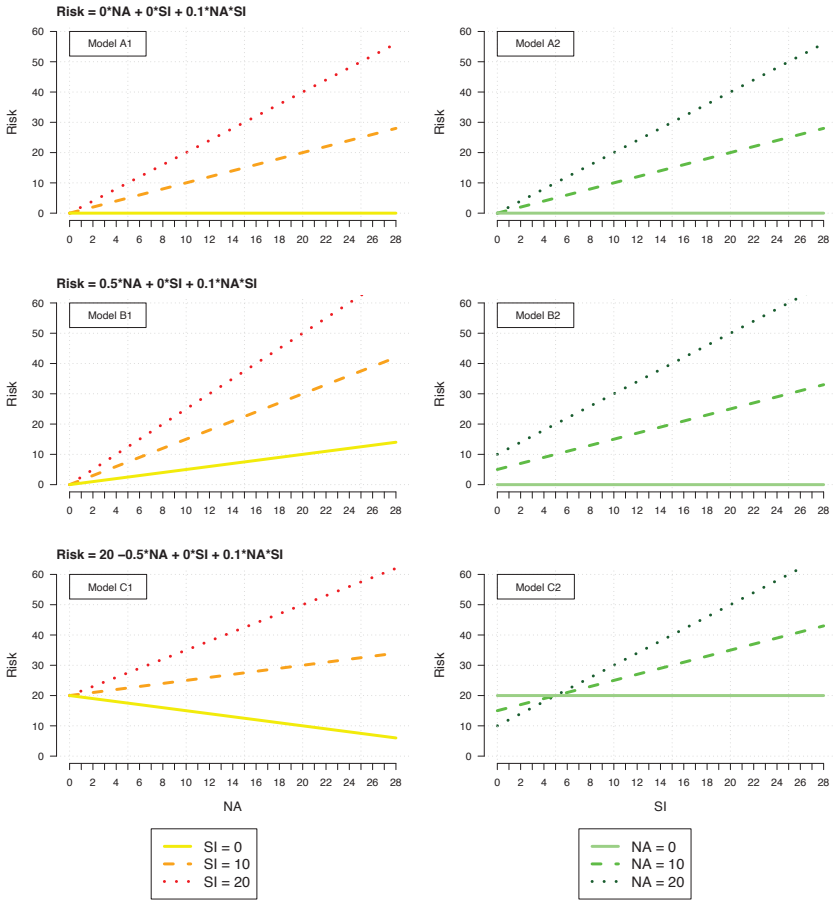
The conclusions of ***Chapters 2, 3, 6, and 7*** in this dissertation are partly based on computer simulations. Such simulation studies first generate many datasets that vary across a wide range of design factors, such as the effect size, sample size, the reliability of the measurement instrument, the measurement level of the item scores, or the skewness in item scores. Next, they analyze each of those datasets using various methods. Accordingly, simulation studies can show for what types of datasets a method performs adequately or, more importantly, in what situations it produces biased effect estimates or inflated false positive or false negative rates.

Interpreting interaction effects

Whenever a significant interaction effect between NA and SI is found, visual inspection of this interaction to facilitate interpretation is often recommended (see Hayes (2017), or Loftus (1978) for a more in-depth discussion of interaction effect interpretation). In linear interaction models, conditional regression lines always have a cross-over point when visualizing them across an infinite score range. For most researchers it may not be

straightforward to infer the location of this cross-over based on the estimated regression coefficients alone. Therefore, we recommend researchers to visualize significant interactions between NA and SI to decide whether there is a synergistic Type D effect. Visual inspection is especially helpful in the context of Type D personality research to assess whether the NA and SI effects are synergistic across the entire observed score range. When the conditional effect size of trait does not increase at higher values of the other trait, then this significant interaction effect does not represent a synergistic Type D effect.

Figure 7: Visualization of three different interaction effects between NA and SI on the risk on some outcome. The separate lines in the left panel show the effect of NA conditional on SI, while the lines in the right panel show the effect of SI conditional on NA.



For each row in Figure 7, we simulated data according to a different continuous interaction model. The left panels in Figure 1 show the association between NA and the outcome given various SI scores (separate lines), while the right panels show the association between SI and the outcome given various NA scores (separate lines). In each model, the interaction coefficient is the same (0.1), while the intercept and first-order regression coefficients differ across models.

Model A involves a synergistic Type D effect, because each personality trait's effect on the outcome increases with higher scores across the entire score range of the other trait. Because both first-order effects are equal, the conditional regression lines in panels A1 and A2 are the same. Model B also involves a synergistic Type D effect, but the different first-order coefficients result in an asymmetric visualization of the interaction effect. Although the interaction effect is the same, the effect of NA on the outcome conditional on SI (left panel) is not equal to the effect of SI on the outcome conditional on NA (right panel). If the first-order coefficients are not equal, then we recommend researchers to visualize the conditional NA and SI effects separately.

So far, all effects are in line with the idea that the combination of high scores on these traits carries a higher risk of an adverse outcome than a high score on only one of the traits. Across the entire NA and SI score range, there is only an increase in the effect of one trait conditional on the other trait. However, interaction model C shows an example of an interaction effect that does not represent a fully synergistic Type D effect. First, in the left panel, the effect of NA on the outcome is positive for high SI scores, yet negative for low SI scores. The visualization in the right panel differs from the left panel because the NA and SI first-order coefficients are not equal. The right panel involves a cross-over interaction within the observed NA and SI score range. The crossover point in model C2 implies that individuals with the highest risk on the outcome are *both* those with the highest NA and SI scores. This is still consistent with what Type D theory predicts. However, at lower SI scores the pattern starts to reverse, because there the risk on an outcome increases with *lower* NA scores. Such an interaction would not be in line with the prediction of Type D theory that *only* the combination of high scores on both traits produce an increased risk on an adverse outcome.

Aims and outline of dissertation

This dissertation aims to study how the construct Type D personality can best be operationalized and modeled in statistical analyses. In ***Part I*** we aim to evaluate the methods that are predominantly used in the literature to model a synergistic Type D effect. We set out to investigate whether several methods commonly used to estimate Type D effects can detect various causal mechanisms relating NA and SI to an outcome measure. ***Chapters 2 and 3*** use simulation studies to investigate the performance of the commonly used 2-group, 4-group, and continuous interaction methods in estimating a Type D effect. The findings of ***Chapters 2 and 3*** indicate that the 2-group and 4-group method cannot adequately distinguish between various causal mechanism relating NA and/or SI to outcome measures and therefore do not specifically test a synergistic Type D effect. Therefore, conclusions in the published Type D literature based on these methods may have to be reconsidered.

Part II of this dissertation provides a first start at reconsidering the Type D literature. The general aim is to investigate whether synergistic Type D effects have been concluded prematurely based on methods that inadequately detect such effects. In ***Chapter 4***, we first estimate the discrepancy in the conclusions of the 2-group and continuous interaction method, by means of a systematic review of the published Type D literature including all studies that have used both these methods to estimate a Type D effect. Our finding that half of the published Type D effects are likely effects of NA or SI only stressed the importance of reanalyzing the published Type D literature. Therefore, in ***Chapter 5*** we present the findings of an individual patient-data meta-analysis, reanalyzing 18 published prospective cohort studies investigating Type D personality as a risk factor for adverse events in cardiovascular disease patients.

In ***Part III*** we investigate the potential benefit of using latent variable models in research on Type D personality. Given the measurement errors and skewness in the ordinal item scores measuring NA and SI, we investigate whether the synergistic Type D effect can better be estimated using latent variable models than observed score methods. As it is not straightforward to model interaction effects between latent variables, ***Chapters 6 and 7*** use a combination of simulations and empirical applications to investigate the performance of

several latent interaction models in the context of structural equation modeling. Both chapters focus on continuous latent predictor variables, but ***Chapter 6*** studies continuous latent outcome variables, while ***Chapter 7*** studies manifest dichotomous outcome variables. Lastly, ***Chapter 8*** illustrates how various latent variable models can be used to assess the temporal stability of psychological constructs, and more specifically that of NA and SI across a four-year follow-up. In conclusion, ***Chapter 9*** summarizes the dissertation's main findings and ***Chapter 10*** provides an in-depth discussion of the key findings, alongside the main implications, potential limitations, and future considerations.

PART I

Type D personality effects

CHAPTER 2

Modeling synergy: How to assess a Type D personality effect?

P. LODDER

Journal of Psychosomatic Research

ABSTRACT

Introduction: In research on Type D personality, its subcomponents negative affectivity (NA) and social inhibition (SI) are hypothesized to have a synergistic effect on various medical and psychosocial outcomes. As some methods to analyze Type D personality have been criticized, this study investigated whether these methods adequately detect a Type D effect.

Method: We used a simulation and two empirical illustrations to investigate each method's performance (bias, power and false positives) in detecting the Type D effect.

Results: Our simulation showed that the two most commonly used methods to assess the Type D effect (subgroup methods) did not specifically test a synergistic Type D effect. These methods could not distinguish between situations where only NA, only SI, or both NA and SI were causally related to an outcome. The method that best detected synergistic Type D effects modeled the continuous NA/SI main effects and their statistical interaction in a regression analysis. Reanalysis of two empirical Type D personality datasets confirmed the patterns found in our simulation.

Conclusion: This study showed that Type D effects should be modeled with a continuous interaction approach. Other approaches either showed lower power or could not distinguish Type D effects from effects of NA or SI only. We recommend against using subgroup approaches to assess Type D effects, regardless of whether the Type D effect is synergistic or additive in nature.

INTRODUCTION

Scientific models are often multidimensional, where variation in a particular quantitative or qualitative outcome is explained by more than one predictor. There are several ways to conceptualize the relation between two predictors and one outcome. For instance, one predictor can confound, mediate, or moderate the association between the other predictor and the outcome (see Bauman, Sallis & Dzewaltowski (2002) for a review). The focus of the present article is a specific type of moderating effect, called synergy.

Two predictors synergistically influence an outcome when the conditional effect of each predictor increases across the observed score range of the other predictor. In the absence of synergy, two predictors may influence an outcome, but the conditional effect of one predictor does not change across the score range of the other predictor. In such an additive model, the main effects of the two predictors generally capture their influence on an outcome. To test synergy between these predictors the model can be extended by including an interaction effect between the two predictors. For concluding a synergistic effect, a significant interaction is necessary but not sufficient. An additional requirement is that the conditional effects of each predictor on the outcome should all be positive and increasing across the score range of the other predictor. This requirement for instance excludes interaction effects that for instance imply that the effect of one predictor on some outcome becomes smaller at higher scores on the other predictor.

Type D personality

Research on Type D personality arguably serves as a perfect case study for modeling synergy (Smith, 2011). People with a Type D (Distressed) personality type score high on the two personality traits negative affectivity (NA) and social inhibition (SI). People with negative affectivity have a tendency to experience negative thoughts and behaviors and socially inhibited persons have difficulty expressing their thoughts and emotions, especially in social situations (Denollet et al., 1996). It is the *combined* presence of high scores on both personality traits NA and SI that has been found as a risk factor for various outcomes (Smith, 2011). Earlier writings on Type D personality suggested that the Type D effect is *synergistic*.

For instance, Kupper & Denollet (2007) explicitly stated that Type D personality is a synergy between NA and SI (see also Pedersen & Denollet (2003, p. 245)). Furthermore, Denollet, Sys and Brutsaert (1995) claimed that the *interaction* [emphasis added] of emotional distress and inhibition of one's feelings can be viewed as a form of stress that may create or exacerbate serious health problems'' (p. 583). Similarly, Denollet and colleagues (2006a; 2013b) more than once suggest that social inhibition *modulates* [emphasis added] the effect of negative emotions on cardiac prognosis. These findings point to a synergistic Type D effect.

Type D personality has been associated with various adverse medical and psychosocial outcomes. For instance, a systematic review showed people with Type D personality to have a 3-fold increased risk on cardiac events compared to people with no Type D personality (Denollet, Schiffer & Spek, 2010). Furthermore, in a population of patients with cardiovascular disease, a meta-analysis concluded that Type D's show a higher all-cause mortality than non Type D's (Grande, Romppel & Barth, 2012). However, the size of these effects appeared to decrease over time because more recent studies (Coyne et al., 2011; Grande et al., 2011; Condén et al., 2017) failed to corroborate earlier findings. Some have argued that these inconsistencies can in part be explained by the different approaches used to operationalize the Type D effect (Smith, 2011; Ferguson et al., 2009; Suls, 2014). Next, we discuss each of those methods.

How to assess the Type D effect?

Although later in this dissertation (***Chapter 6 and 7***) we argue that latent variable methods are very useful alternatives in modeling a Type D effect, for the purpose of the present discussion we focus on methods that do not specify a measurement model (also called observed score methods), as these are most commonly used in the Type D literature. Though observed score methods are relatively easy to model, they fail to consider measurement error in the item scores or other aspects of the measurement model.

To the best of our knowledge, four operationalizations of Type D personality have been reported and used in the literature. Table 1 shows an example dataset required to operationalize Type D personality using each of the four methods. NA and SI, the two traits

underlying Type D personality, are each measured with seven items on a 0 to 4 Likert scale in the DS14 questionnaire (Denollet, 2005). The four methods have in common that the seven item scores measuring each construct are first summed, resulting in NA and SI sum scores ranging from 0 to 28. However, from this point onwards the observed score methods start to diverge.

Table 1: Example of data required to analyze Type D personality according to four methods

ID	NA	SI	NA+	SI+	NA+SI+	NA+SI-	NA-SI+	NA-SI-	NAc	SIc	Nac * SIc
1	20	17	1	1	1	0	0	0	8	7	56
2	7	9	0	0	0	0	0	1	-5	1	5
3	18	3	1	0	0	1	0	0	6	-7	-42
4	5	12	0	1	0	0	1	0	-7	2	-14
5	10	9	1	0	0	1	0	0	-2	-1	2

*NA = negative affectivity sum score; SI = social inhibition sum score; NA+ = NA score equal to or above cutoff; SI+ = SI score equal to or above cutoff; NA-SI- = NA and SI score below cutoff; NA+SI- = NA score equal to or above cutoff, and SI score below cutoff; NA-SI+ = NA score below cutoff, and SI score equal to or above cutoff; NA+SI+ = NA and SI score equal to or above cutoff (Type D group); NAc = mean centered NA score; SIc = mean centered SI score; NAc * SIc = multiplication of mean centered NA and SI scores.*

The two most widely used methods first dichotomize the NA and SI sum scores using a fixed cutoff score of 10, to indicate whether people score high or low on the NA and SI traits (NA+ and SI+ in Table 1). Initially this cutoff score was derived from a median split of the NA and SI total scores. Though IRT analyses suggested that the NA and SI traits are reliably measured around the cutoff score (Emons, Meijer & Denollet, 2007), others have questioned the validity of the assumptions underlying that analysis (Ferguson et al., 2009). In any case, all people who score equal to or above the predetermined cutoff on both traits are assumed to have a Type D personality (NA+SI+ in Table 1). Similarly, three other classifications can be made using the two dichotomized NA and SI scores: people scoring equal to or above the cutoff on NA but lower than the cutoff on SI (NA+SI-); people who score lower than the cutoff on NA but equal to or above the cutoff on SI (NA-SI+); lower than the cutoff on both NA and SI (NA-SI-).

2-group approach

The first method used to assess a Type D effect is called the *2-group* approach and includes the dichotomous NA+SI+ variable as a predictor in a regression model. This method estimates the effect of people who score high on both constructs versus people who do not score high on both constructs. Despite being the most commonly used operationalization, Smith (2011) has argued that the 2-group approach does not appropriately assess a synergistic effect, as such a 2-group effect could also result from patterns other than synergy. For instance, if in reality *only* NA is causally related to some outcome, then comparing a Type D group (NA+SI+) with a non Type D group (NA+SI-, NA-SI+, NA-SI-) will falsely suggest that both NA and SI are causally related. Both the Type D and non-Type D group contain a high-risk group of people with high NA scores (NA+SI+ vs. NA+SI-). However, in the non Type D group the effect of the high-risk group (NA+SI-) is averaged with that of the two low risk groups (NA-SI+ & NA-SI-). We expect the *2-group* approach to result in significant Type D effects when the underlying causal mechanism is a synergistic Type D effect, additive Type D effect, or when only NA or SI is causally related to the outcome.

4-group approach

The second method is called the *4-group* approach and includes the NA+SI+, NA+SI- and NA-SI+ variables as predictors in a dummy-coded regression model. In this way the effect of each of these three dummy variables is estimated relative to that of the reference group NA-SI-. This 4-group approach is expected to show larger Type D effects than the 2-group approach when focusing on the contrast of the Type D group with the lowest risk group only (see for instance: Denollet et al., 2018), while for the 2-group approach Type D is contrasted with the three non Type D groups combined. Note that in practice, researchers sometimes apply a stricter criterion before concluding a significant Type D effect based on the 4-group approach, where the Type D group needs to show a significant effect relative to *each* of the three other groups separately.

Continuous interaction approach

The third method to assess a Type D effect has been advocated by various critics as the method of choice (Smith, 2011; Ferguson et al., 2009; Suls, 2014; Coyne & de Voogd, 2012). This *continuous interaction* approach includes both the NA and SI sum scores, as well as their

interaction term in a regression model. When constructing the interaction term, the sum scores are typically mean-centered before multiplying them (NAc, SIc, NAc * SIc in Table 1). The continuous interaction approach differs from the three other operationalizations in that it uses the NA and SI sum scores, rather than their dichotomized values. This is also the main reason the critics prefer this approach above the others. Various authors have argued against the practice of dichotomizing continuous variables, not only because it reduces the power in statistical tests (Cohen, 1983; Royston, Altman & Sauerbrei, 2006), but also because under some circumstances it increases the risk on spurious findings (Maxwell & Delaney, 1993; MacCallum, Zhang, Preacher & Rucker, 2002; Thoresen, 2019). For instance, if predictors A and B are correlated and only predictor A has a causal effect on an outcome, then dichotomizing both predictors before including them in a regression model results in an increased false positive rate for predictor B and for the interaction between A and B (Lubinski & Humphreys, 1990). In the context of research on Type D personality, because the continuous NA and SI scores typically show a moderate correlation (r = 0.5), critics have argued against using approaches based on dichotomized NA and SI variables (Smith, 2011; Suls, 2014).

Adjusted 2-group approach

As opposed to the first three approaches, the fourth operationalization of Type D personality is almost never used in practice (see Condén and colleagues (2017) for an exception). This approach is similar to the continuous interaction approach, but models the dichotomized instead of continuous NA and SI scores, resulting in less statistical power to detect the Type D effect. When multiplying the dichotomized NA and SI scores, the resulting interaction term corresponds to the dichotomous Type D variable in the 2-group approach (NA+SI+). However, the *adjusted 2-group* approach differs from the regular 2-group approach because it also includes the first-order dichotomized NA and SI effects in the regression model. Consequently, as opposed to the regular 2-group approach, the *adjusted 2-group* approach can distinguish Type D effects from effect of NA or SI.

Conflicting Type D results

Several studies have reported the effect of Type D personality on some outcome measure, using multiple operationalizations of Type D personality. Some of these studies showed

significant effects for the 2-group approach, while the continuous interaction approach failed to reach significance. For instance, Dulfer and colleagues (2015) used the 2-group approach and reported that people with Type D personality had a larger odds on all-cause mortality than people without Type D personality (OR=1.58, 95%CI=1.22, 2.03), while the effect according to the continuous interaction approach failed to reach significance (OR=0.95, 95%CI=0.78, 1.17). An imaging study by Wang and colleagues (2016) showed that Type D's, compared to non-Type D's, were at increased odds of having lipid artery plaque according to the 2-group approach (OR=4.87, 95%CI=1.41, 11.14), while the effect based on the continuous interaction approach did not reach significance (OR=0.66, 95%CI=0.17, 2.51). Further research by Wang and colleagues (2016) reported that Type D's, compared to non-Type D's, were at increased odds of having In-stent restenosis (OR=2.82, 95%CI= 1.26, 6.3) according to the 2-group approach, while the continuous interaction approach did not show such a significant effect (OR=1.13, 95%CI=0.45, 3.10). Lastly, Williams, O'Connor, Grubb and Carroll (2012) reported based on the 2-group approach that people with Type D personality had a lower quality of life compared to non-Type D's (d=-1.52, 95%CI=-1.86, -1.19), while the effect according to the continuous interaction approach failed to reach significance. A limitation of these reports, however, is that the effect sizes (e.g., odds ratios) were not calculated using predictors on a standardized scale, making it difficult to compare them in size. Nevertheless, p-values are not affected by the standardization process and these indicate that the dichotomous approaches were statistically significant while the continuous interaction approaches were not.

Taken together, these findings stress the importance of assessing the consequences of using each of those four statistical methods to assess a Type D effect. Although these methods in essence test different hypotheses, they are typically used to answer the same question, namely whether Type D personality is related to a particular outcome. Given that in earlier studies on Type D reported above, the conclusions were conditional on the chosen operationalization of Type D, it is paramount to uncover which of these methods accurately detect a true Type D effect (i.e., a synergy between NA and SI). In the present study we aim to investigate this by using both a Monte Carlo simulation, as well as a reanalysis of earlier published data investigating the link between Type D personality and various medical outcomes.

METHOD

Procedure

In our simulation study we generated 75000 datasets to test the association between Type D personality and cardiac events under varying circumstances. We varied these simulated datasets across two parameters: (1) the size of the NA & SI main effects on cardiac events (odds ratio = 0.8, 0.85, 0.9, 0.95, 1.00, 1.05, 1.10, 1.15, 1.20, 1.25, 1.30, 1.35, 1.40, 1.45, 1.50) and (2) the size of the NA & SI interaction effect on cardiac events (odds ratio = 0.50, 0.75, 1.00, 1.25, 1.50), resulting in 15*5 = 75 different simulation conditions. In each of these 75 conditions we generated 1000 datasets, where each dataset contained simulated DS14 item scores for 500 (fictious) participants and a simulated dichotomous outcome.

In the second step of our simulation study, we analyzed each of those 75000 datasets according to the four Type D personality operationalizations. Within each condition we aggregated the results of the 1000 replications by averaging the estimated Type D effects and by computing the percentage of statistically significant effects. To assess the performance of each method, we reported the absolute and relative bias in the estimated Type D effects, as well as the percentage of significant Type D effects. The latter allowed us to determine both the statistical power and the percentage of false positives in detecting a Type D effect. Note that for all approaches, slight negative bias in the estimated Type D effects was expected because all approaches do not consider the measurement error in the NA and SI item scores.

Data generation

As formula 1 shows, the data generating mechanism in this simulation study was a latent logistic interaction model, where a dichotomous outcome (cardiac events) is regressed on the latent variables NA (ξ_1) and SI (ξ_2), and their interaction ($\xi_1\xi_2$):

$$\ln\left(\frac{p(\xi)}{1-p(\xi)}\right) = \beta_1\xi_1 + \beta_2\xi_2 + \beta_3\xi_1\xi_2 \qquad (1)$$

This regression model contained three parameters: one regression coefficient for the main effect of the latent NA construct (β_1), one for the main effect of the latent SI construct (β_2), and one for the interaction between the latent NA and SI constructs (β_3). The magnitudes of these regression coefficients were varied across the 45 simulation conditions by taking the natural logarithm of the odds ratio effect sizes reported above.

The latent NA and SI constructs both followed a bivariate standard normal distribution with a mean vector of zero and correlation matrix: $(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) \sim \mathcal{N}(0, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$. These latent NA and SI scores were used to simulate continuously distributed DS14 item scores based on a standard two factor model, with the seven factor loadings of NA and SI ranging between 0.7 and 0.8, corresponding to an estimated Cronbach's alpha of 0.89 for NA and 0.86 for SI. For all items the intercepts were fixed at -0.9 to keep the prevalence of Type D personality at approximately 25%. The resulting 14 continuous item scores were transformed to 14 ordinal item scores on a 0-4 Likert scale, using four threshold parameters (Muthén & Kaplan's (1985) Case 1 thresholds: [-1.645, -0.643, 0.643, 1.645]). As a sensitivity analysis, we transformed the continuous item scores to *positively skewed* ordinal item scores using alternative threshold parameters (Muthén & Kaplan's (1985) Case 3 thresholds: [-0.05, 0.772, 1.341, 1.881]).

In the last step of the data generation, for each participant a cardiac event score was drawn from a binomial distribution with a probability resulting from filling in formula one the participant's latent NA and SI score, as well as the three regression coefficients that varied across the simulation conditions. As another sensitivity analysis, scores on an observed *continuous* outcome were generated by filling in formula one (without the logit link function) and by adding a random error term with mean zero and variance one.

Data analysis

For every simulation condition, each of the 1000 datasets was analyzed using the four Type D operationalizations. A binary logistic regression model was used for the dichotomous outcomes, while a linear regression model was used for continuous outcomes. The regression terms were specified based on the description in the synergy assessment section.

In the main simulation, the Type D effect according to the 4-group approach was investigated using the contrast between the Type D group and the reference group with scores below the cutoff on both NA and SI. An additional simulation was conducted investigating for the 4-group approach the contrasts of the Type D group with all three other groups separately. In this simulation, the correlation between NA and SI was varied (0 or 0.5) because this will illustrate why the 4-group approach did not perform adequately in detecting Type D effects.

In the analyzed logistic regression models, the estimated regression coefficients were standardized by multiplying the unstandardized coefficients with the standard deviation of the predictor variable (Agresti, 2018), and subsequently exponentiated to compute the odds ratio. For continuous outcomes, the regression coefficients were standardized by multiplying them with the standard deviation of the predictor divided by the standard deviation of the continuous outcome.

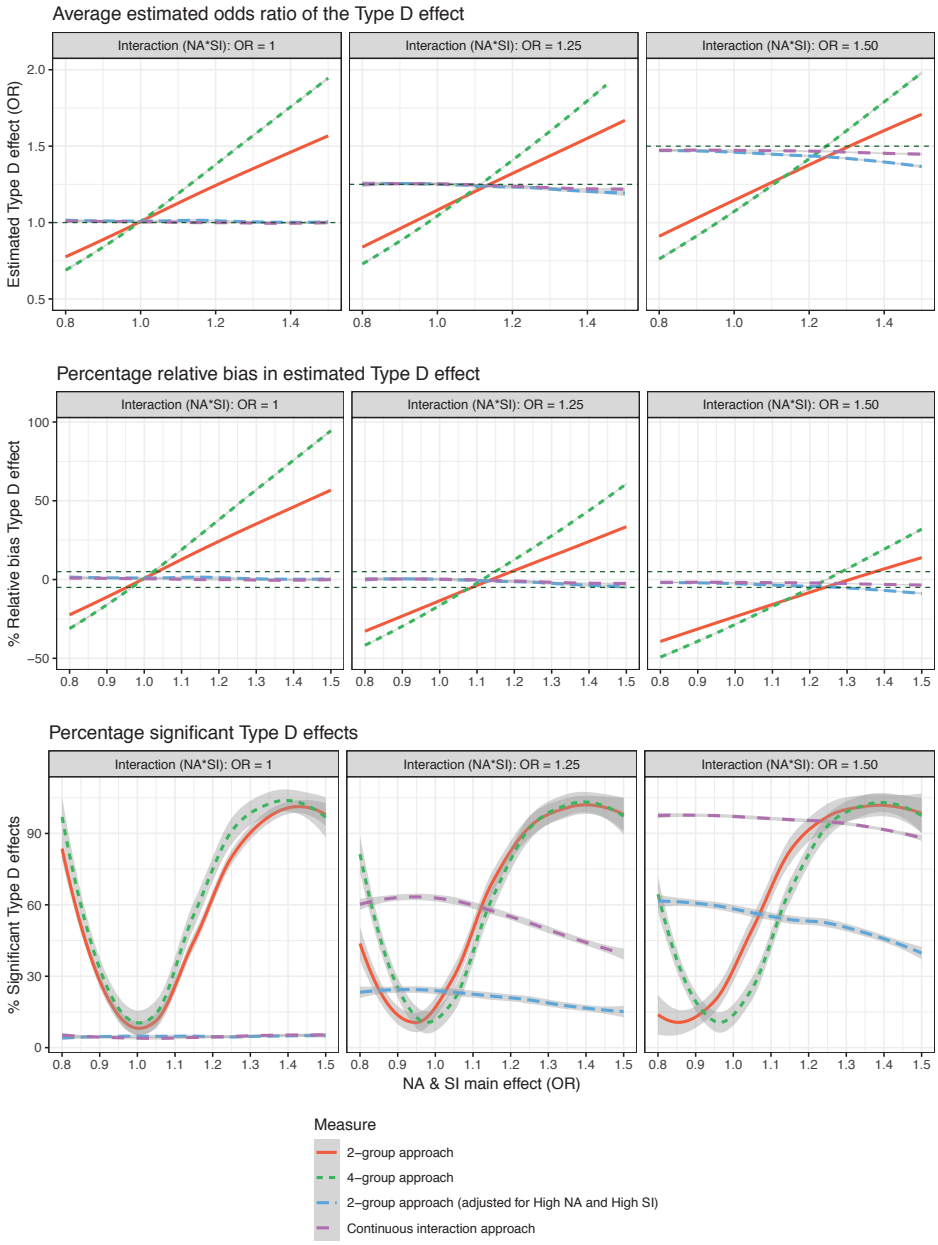
In each condition, the estimated effects were averaged across the 1000 replications. The absolute bias was determined by subtracting from this average the true condition specific Type D effect (e.g., the size of the interaction between NA and SI). The relative bias was determined by dividing the absolute bias by the true condition specific Type D effect. The percentage of significant effects (based on a wo-sided test with a significance level of 0.05) was calculated by dividing the total number of significant effects in a condition by 1000 (replications) and multiplying by 100%. The R-script of this simulation can be found on this project's open science framework page: https://osf.io/qdgkr/.

RESULTS

Figure 1 shows the results of our simulation study according to three outcome measures: mean estimated Type D effect (top row), relative bias (center row) and percentage of significant effects (bottom row). The three columns represent varying sizes of the Type D effect, where the interaction effect is absent in the first column and positively increasing in the second and third column. In each plot the outcome measure (y-axis) is plotted against the size of the NA/SI first-order effects (x-axis) to investigate whether the size of the first-order effects may influence the detection of Type D effects. Note that the scale of these main effects is standardized. Therefore, in the absence of an interaction, a standardized NA effect of 1.2 implies that someone with an average NA score, has a 1.2 higher odds of the outcome event, than someone scoring one standard deviation below average on NA. Lastly, in all plots each line has its own color and type, representing the different approaches to operationalize the synergistic Type D effect.

Ideally, for testing a synergistic Type D effect the lines in each plot of these figures should be perfectly horizontal, as this would indicate that the estimated Type D effects and corresponding significance tests are independent of the size of the NA/SI main effects. Inspection of the top row of Figure 1 shows that this was indeed the case for the continuous interaction and the adjusted 2-group approaches, both estimating the Type D effects close to the true underlying effects (i.e. the black dotted line). As the NA/SI main effects became larger, these two approaches slightly underestimated only the largest Type D effects. The 2-group and 4-group approaches were both unable to estimate the underlying Type D effect correctly. When a true synergistic effect was absent, both methods started to increasingly overestimate the Type D effect more as the size of the NA/SI effects increased, suggesting these approaches also result in significant Type D effects when the causal mechanism relating NA and SI to an outcome is not synergistic.

Figure 1: For each Type D operationalization and for varying levels of the Type D effect and the NA and SI main effects, the mean estimated odds ratio (upper), percentage relative bias (middle) and percentage significant results (lower) of the Type D effect.



In fact, the correlation between these NA/SI main effects and the estimated Type D effect was *r*=0.97 (95%CI=0.95, 0.99; *p*<.0001) for the 2-group approach and *r*=0.99 (95%CI= 0.99, 1.00; *p*<.0001) for the 4-group approach, compared with *r*=-0.10 (95%CI=-0.38, 0.20; *p*=.537) for the adjusted 2-group approach and *r*=-0.05 (95%CI=-0.34, 0.25; *p*=.764) for the continuous interaction approach, indicating that in the absence of a synergistic Type D effect, the bias in the 2-group and 4-group methods was almost perfectly correlated with the size of the NA and SI effects. When adjusting for these main effects, by either using the adjusted 2-group or continuous interaction approach, this correlation reduced to zero. These findings suggest that the 2-group and 4-group approaches cannot distinguish a synergistic Type D effect from effects of NA or SI only.

The adjusted 2-group and continuous interaction approaches show much less bias. Whether this bias was any reason for concern can be seen in the relative bias plots in the center row of Figure 1. The two black dotted lines mark the interval between +5% and -%5 relative bias. The results of adequately performing methods should fall within this interval. It turned out that only the continuous interaction approach was unbiased based on this criterion. The adjusted 2-group approach crossed the -5% border when both the NA/SI main effects as well as their interaction effect was large.

Inspection of the percentage significant Type D effects in the bottom row of Figure 1 indicated that both the continuous interaction and adjusted 2-group approach had an adequately controlled the false positive rate of 5% when a true effect was absent. The 2-group and 4-group approaches, however, showed a similar false positive rate when the NA/SI main effects were absent (OR=1). However, the percentage of significant Type D effects increased as the first-order NA and SI effects became larger, up to 90% significant effects. The middle and right columns of the bottom row show the power to detect a Type D effect when there was a true underlying interaction effect. The 2-group and 4-group approaches showed curves similar to when the interaction effect was absent, indicating that these approaches could not distinguish synergistic Type D effects from effects of NA or SI only. With respect to the other two approaches, the continuous interaction approach consistently showed higher power to detect Type D effects than the adjusted 2-group

approach, suggesting that the continuous interaction approach performed best, both in terms of minimizing bias, minimizing false positives as well as in maximizing power.

Figure 2 shows the simulation results when one of the NA/SI traits was fixed to an odds ratio of one. For all approaches, the results were similar to our main results, though less extreme. Interestingly, the fact that the 2-group and 4-group approaches still showed bias indicates that they may falsely conclude a Type D effect even if only one of the NA/SI main effects is significant. This suggests that even if the Type D effect is not synergistic, but additive in nature, the 2-group approaches still remains biased in detecting the Type D effect, because it cannot distinguish between the following three scenarios: (1) only SI shows an effect; (2) only NA shows an effect; (3) both SI and NA show an effect.

Figure 3 illustrates the simulation results for the 4-group approach where the Type D group was contrasted with each of the three other groups. The difference between Figure 3 and the previous figures is that it includes an additional simulation condition where NA and SI did not correlate, whereas in the main simulation this correlation was fixed to the value typically observed in the literature (i.e. $r = 0.5$). This correlation appeared to be important in explaining the bias in the Type D effects resulting from the 4-group approach. The leftmost column of Figure 3 illustrates the scenario with no synergistic Type D effect, no SI effect and no correlation between NA and SI. The results indicated that as the size of the NA effect increased, the contrasts between the Type D group and both the reference group (NA-SI-) and the High SI group (NA-SI+) also increased, while the contrast with the High NA group (NA+SI-) remained zero on average, with an expected false positive rate similar to the chosen significance level of 5%. In this condition the 4-group method performed as expected, because when only NA is related to the outcome, the two groups containing elevated NA scores (the Type D & High NA groups) should differ significantly from the two groups with low NA scores (the reference group and High SI group), but not from each other.

Figure 2: For each Type D operationalization and for varying levels of the Type D effect and the NA main effect on a dichotomous outcome (SI main effect is fixed at OR=1), the mean estimated standardized regression coefficient (upper), percentage relative bias (middle) and percentage significant results (lower) of the Type D effect.

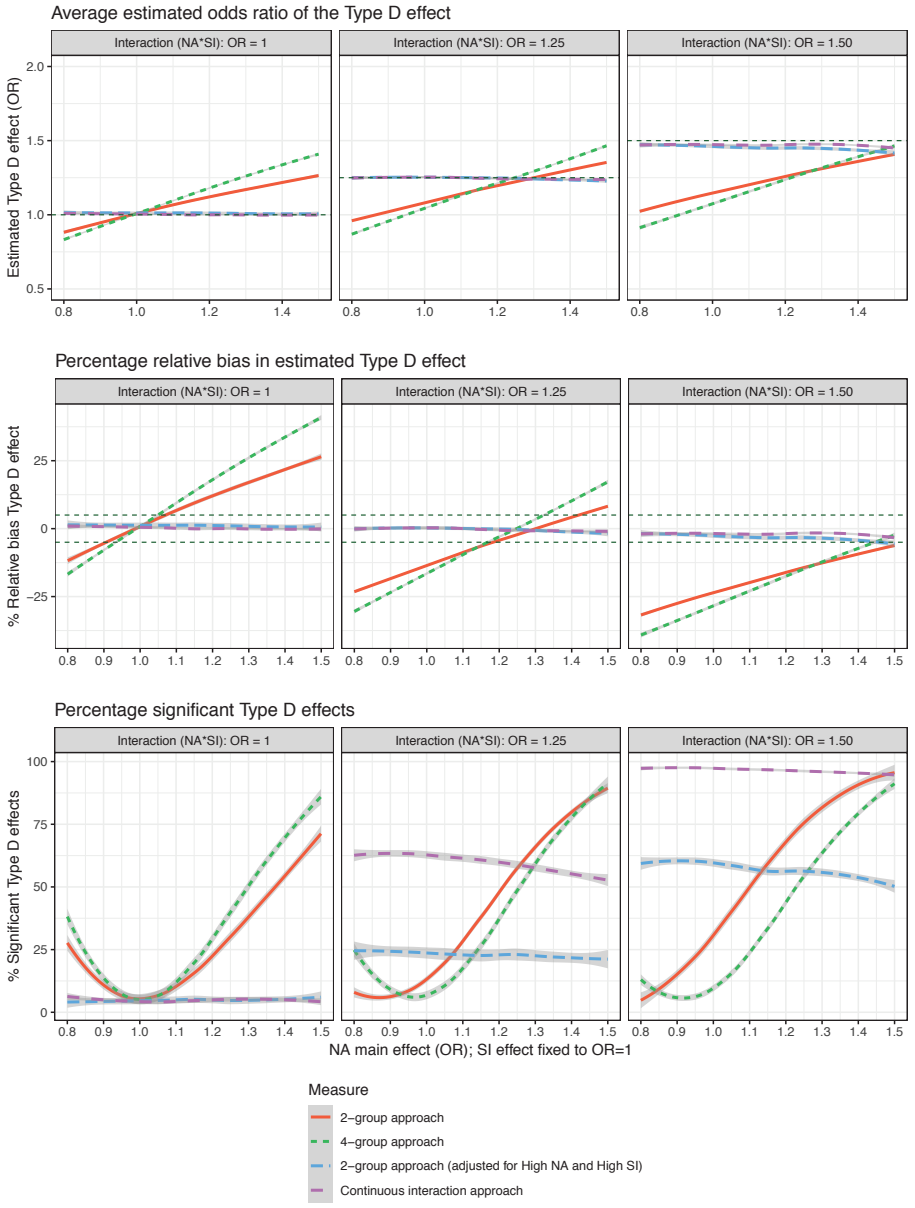
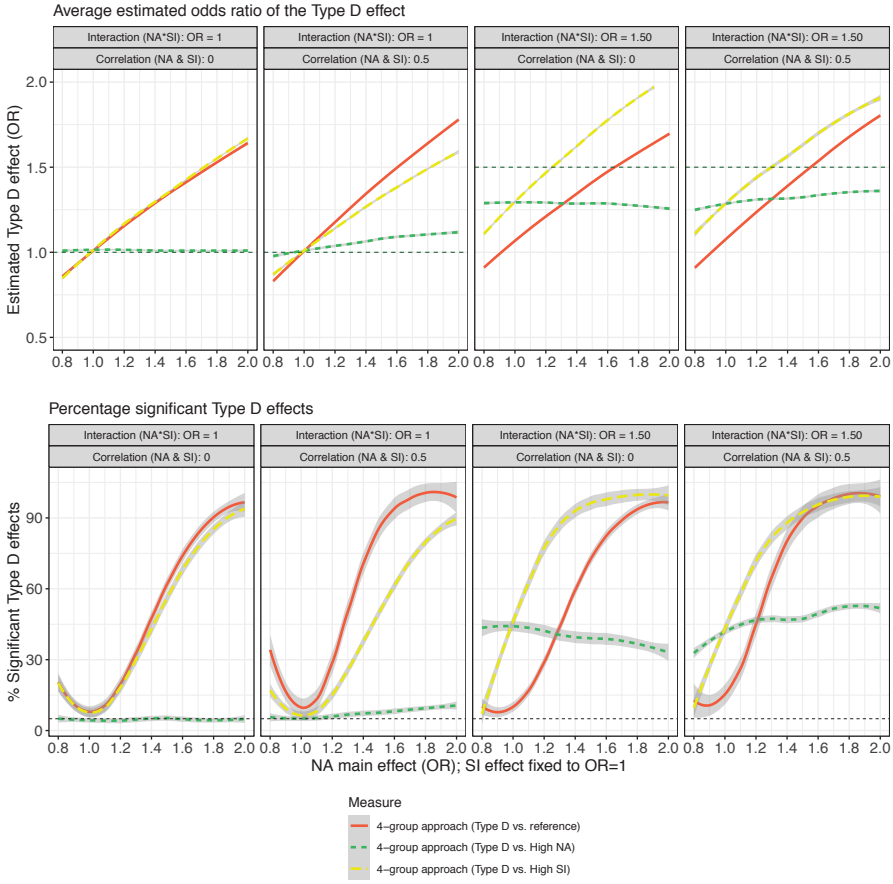


Figure 3: For varying levels of the synergistic Type D effect and NA main effect on a dichotomous outcome (SI main effect is fixed at OR=1), the mean estimated standardized regression coefficient (upper), percentage relative bias (middle) and percentage significant results (lower) of the Type D effects estimated by the 4-group approach.



The second column of Figure 3 shows the results when increasing the correlation between NA and SI to 0.5. This change resulted in overestimated Type D effects relative to both the reference and high NA groups, and this bias got stronger as the size of the NA main effect increased. This is not desirable, because when only NA is related to the outcome, people in the Type D group should not differ from people in the High NA group. It also turned out that

the effect of the Type D group relative to the reference group became larger than the effect of the Type D group relative to the High SI group. This is also not desirable, because the reference and High SI groups should show equal effects when only NA is related to the outcome. The explanation for these biased results is that the positive correlation between NA and SI causes spurious effects for the groups with elevated SI scores (the Type D group and High SI group) when only NA is causally related to an outcome. In continuous analyses this does not happen because there the effect of SI is controlled for the effect of NA. However, this adjustment is no longer adequately performed when using the 4-group approach.

Sensitivity analyses

In ***Appendix B***, three additional figures show the simulation results for several sensitivity analyses, to show that are results are not contingent on specific design choices in our simulation. Figure B1 shows the simulation results for less than additive interaction effects. Although Type D theory predicts the interaction between NA and SI to be more than additive, the inclusion of this sensitivity analysis may improve the generalizability of our results to fields where less than additive interactions are of interest. The results in Figure A1 are similar to those in Figure 1: The Type D effects estimated by the 2-group and 4-group varied with the size of the NA and SI main effects, while the Type D effect estimated with the continuous interaction method and adjusted 2-group approach did not. The 2-group and 4-group approaches showed high false-positive rates when a synergistic Type D effect was absent and the NA or SI main effects were present. Although the false-positive rates of both the continuous interaction method and adjusted 2-group approach were adequate, the continuous method outperformed the adjusted 2-group method in terms of power to detect the Type D effect.

Figure B2 shows the simulation results for continuous outcomes. These results are largely similar to those found for dichotomous outcomes. Both the 2-group and 4-group approaches failed to detect the Type D effect and merely picked up the presence of NA/SI main effects. The adjusted 2-group and continuous interaction approaches both showed much less biased, yet they slightly underestimated the true Type D effect as the size of the NA/SI main effect

increased. The 2-group and 4-group approaches showed very high false positive rates. The power to detect true Type D effects was best for the continuous interaction approach.

Figure B3 shows the simulation results of the sensitivity analysis where the ordinal NA/SI item scores were positively skewed rather than normally distributed. The results were largely similar to normally distributed item scores. However, all methods showed attenuated estimates of the Type D effects as the NA/SI main effects became larger. For the 2-group and 4-group approaches this attenuated the earlier found positive bias, while for the adjusted 2-group and continuous interaction approaches it resulted in somewhat underestimated Type D effects. As a result, the power of these approaches to detect a true Type D effect decreased as the NA/SI main effects became larger. Thus, large main effects for skewed variables obfuscated the presence of true interaction effects.

Empirical reanalysis

To illustrate the implications of our simulation study in empirical data we have reanalyzed earlier published data of two empirical studies investigating the association between Type D personality and various dichotomous outcomes. If the results of our simulation are accurate, one would expect to find similar patterns when analyzing empirical data using the four operationalizations of Type D personality. For each of the two datasets and for each of the four operationalizations we have reported the odds ratios of the Type D effect in Table 2. To allow for comparison across operationalization methods, we have calculated the odds ratios based on the standardized logistic regression coefficients.

Using the first dataset (Pelle et al., 2010), we reanalyzed the association between Type D and elevated depressive symptoms in a sample of 650 outpatients with chronic heart failure. In the second dataset (Denollet et al., 2018) we reanalyzed the association between Type D personality and endothelial dysfunction in a sample of 180 patients with coronary artery disease. Reanalysis of these two datasets showed results strikingly similar to the patterns found in our simulation study. First, for each dataset the 2-group and 4-group approach showed statistically significant effects, while the adjusted 2-group and the continuous interaction approach did not. Second, in both datasets the 4-group approach resulted in larger odds ratios than the 2-group approach, similar to our simulation results. Interestingly,

in both datasets only one of the NA/SI main effects was significant. Similar to the results of our simulation study, the reanalysis suggests that the 2-group and 4-group methods suggest a synergistic Type D effect, whereas, as suggested by the continuous method, in reality only NA or SI are linearly related to an outcome.

Table 2: Study characteristics of two different datasets investigating Type D personality, and the estimated odds ratio (95% CI) of the Type D effect, according to four operationalizations of Type D personality.

Dataset	Pelle et al. (2010)	Denollet et al. (2018)
Study characteristics		
Sample size	641	180
Outcome	Elevated depression	Endothelial dysfunction
NA main effect	**2.18 (1.92, 2.48)**	1.10 (0.80, 1.49)
SI main effect	1.01 (0.92, 1.11)	**1.46 (1.06, 2.02)**
Type D operationalization		
2 groups	**1.49 (1.38, 1.62)**	**1.41 (1.04, 1.90)**
4 groups	**1.75 (1.58, 1.94)**	**1.60 (1.11, 2.33)**
2 groups (adjusted for NA+ & SI+)	0.99 (0.86, 1.15)	1.07 (0.79, 1.44)
Continuous interaction (NA * SI)	1.00 (0.99, 1.01)	1.05 (0.79, 1.41)

CAD = coronary artery disease; NA = negative affectivity sum score; NA+ = NA score above cutoff (yes/no); SI = social inhibition sum score; SI+ = SI score above cutoff (yes/no).

Note: all Type D effects are odds ratios (95%CI) based on the standardized regression coefficients of the Type D effect. Bold faced results indicate a statistically significant odds ratio with a p-value smaller than 0.05.

DISCUSSION

In this study we showed that the most commonly used methods to model a Type D personality effect failed to detect the presence of a synergistic Type D effect. These results apply to models with either a dichotomous or a continuous observed outcome. When a true synergistic effect was absent, the chance that the 2-group and 4-group approaches found significant effects increased as the size of the NA/SI main effects became larger, rendering the methods unsuited for studying synergistic effects. Our simulations showed that this problem occurs even when only one of the NA/SI main effects was present and the other absent. Regardless of whether the Type D effect is synergistic or additive in nature, the most commonly used 2-group approach did not assess how NA and SI interact or combine; it was merely sensitive to the presence of *any* main effect. Interestingly, the 2-group and 4-group approaches showed an almost perfect correlation between their estimated Type D effects and the true size of the NA/SI main effects. These findings support Smith's (2011) hypothesis that the 2-group approach may falsely conclude the presence of a Type D effect when only NA/SI main effects are present.

The continuous interaction and adjusted 2-group approaches were both relatively unbiased and showed acceptable false positive rates. However, the statistical power to find a Type D effect based on the adjusted 2-group approach was only 50 to 70% the size of the power to detect such effects using the continuous interaction approach. As noted by Smith (2011), this might have been caused by the fact that the 2-group approach uses dichotomized predictors. Indeed, earlier research has indicated that dichotomization of continuous predictors may decrease precision to 65% of when using continuous predictors (Lagakos, 1988).

Regarding the 4-group approach, our study showed that not only the contrast between the Type D group and reference group did not adequately test for synergy, but neither do the contrast between the Type D groups and the two other groups. These results imply that when only NA is causally related to an outcome, the 4-group approach not only results in significant Type D effects relative to the reference group and high SI group, but also relative

to the High NA group. Such results may lead researchers to falsely infer the presence of a synergistic Type D effect, when in reality only NA was causally related to the outcome. When only one of two correlated traits is causally related to an outcome, then a spurious association may arise between the other trait and the outcome, suggesting elevated scores for people scoring high on this other trait. This problem does not occur when the two traits are not correlated. This suggests that two correlated continuous variables should not be categorized in groups based on the different combinations of scoring high or low on these variables. This recommendation does not only apply to research on Type D personality, but to any field where two correlated continuous variables are categorized in subgroups.

The patterns found in our simulation study were corroborated in our reanalysis of empirical studies on Type D personality. For all analyses, the continuous interaction approach failed to reach significance, with odds ratios close to one, while the 2-group and 4-group approaches showed a significant Type D effect with odds ratios varying between 1.2 and 1.6. In light of the results of our simulation one would expect this pattern in a scenario where NA and/or SI have positive main effects, yet no synergy. Each reanalysis suggested that not Type D personality, but merely one of its subcomponents was related to the outcome. Although these reanalyses seem to suggest there is no synergistic Type D effect underlying these empirical studies, we cannot exclude the possibility of a true but small synergistic effect that could not be detected due to low statistical power.

This highlights the importance of reanalyzing other published research on Type D personality using the 2-group and 4-group approaches. Echoing earlier recommendations (Smith, 2011, Coyne & de Voogd, 2012), we therefore encourage the authors of those publications to reanalyze their data using the continuous interaction approach. Future meta-analyses on Type D personality should use individual patient data rather than aggregated study level data, to allow for assessing the overall Type D effect according to the continuous interaction approach. Future clinical studies on Type D personality should be sufficiently power to detect the Type D effect according to the continuous interaction approach. More importantly, we recommend against using either the 2-group or 4-group approach in future studies on Type D personality. Under the assumption that the Type D effect is synergistic, we advise researchers to always use the continuous interaction approach, as it outperformed all

other approaches in terms of minimizing bias, minimizing false positives and maximizing power. Although the adjusted 2-group approach performed comparably in terms of bias, it showed lower power to detect additive or synergistic Type D effects. Furthermore, a recent simulation study showed that including two dichotomized predictors in a regression analysis may result in spurious interaction effects (Thoresen, 2019), suggesting that continuous variables should always be assessed on their original scale. In line with suggestions by Smith (2011), a re-analysis of published studies using the correctly specified continuous interaction approach may clarify whether the significant Type D effects reported in the literature are validly interpreted as a synergistic interaction between NA and SI, or an additive combination of NA and SI, or a main effect of NA or SI only.

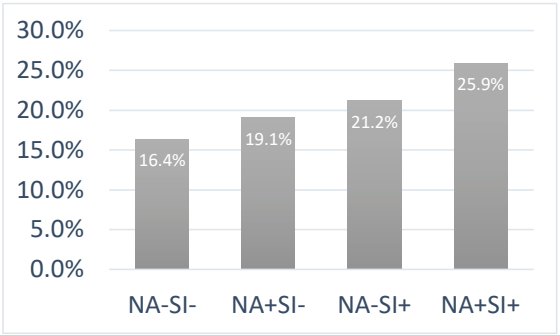
Common misconceptions

We will now rebut several arguments that could be raised against our conclusions. First, the results of the 4-group method are typically represented using bar plots. Figure 4 shows such a plot, illustrating that the Type D group does in fact have higher mean scores or percentages than the other three NA&SI subgroups. Many researchers seem to conclude from this figure that there is a synergistic Type D effect, but we argue that bar plots such as Figure 4 may raise a false impression of synergy between two measures. Let's assume that both NA and SI have a causal effect on having no romantic partner. Let's further assume there is no synergistic effect between NA and SI. Such an additive model can sufficiently explain the differences between the personality groups visualized in Figure 4. If an outcome is causally related to both NA and SI but both effects are independent of each other (i.e. no interaction), then people who score high on both NA and SI already have a higher chance on living without a partner, than people who score high on only one of those traits. The effect is the sum of the individual effects of NA and SI. That is, when predicting an individual outcome using the regression equation, the individual contributions of the NA and SI effects are already combined due to their additive nature. A synergistic effect is said to exist if the percentage is even higher than predicted from first-order effects. For this reason, though Figure 4 may lead some to interpret the presence of synergy, such bar plots are in fact not able to visualize the difference between synergistic models and causal models containing main effects only. Instead, we recommend researchers to use line diagrams to visualize the

association between an outcome (y axis) and one personality trait (x-axis), given various scores on the other personality trait (separate lines).

A second objection against our findings could be that Type D's causal effect on outcome measures is not best captured by a synergistic effect between NA and SI, but rather by the presence of both an NA and SI effect. One could argue that our conclusions only hold when the underlying causal mechanism is synergistic. Indeed, our findings suggest that observed Type D effects from the 2-group and 4-group approaches reported in Type D studies were to a large extent capturing the additive influence of the NA and SI main effects, rather than their interaction. However, we argue that these methods are also not adequate in detecting additive NA and SI effects for three reasons: [1] our stimulation showed that the 2-group and 4-group approaches cannot distinguish whether the effect is caused by NA only, or by SI only, or by both NA and SI; [2] using dichotomized variables in regression analyses results in lower power to detect significant effects (Cohen, 1983); [3] using dichotomized variables in regression analyses risks spurious main effects and interactions, especially when the two variables are correlated (Maxwell & Delaney, 1983; MacCallum et al., 2002) or when measurement error in the item scores is not taken into account (Jaccard & Wan, 1995). So even if the causal influence of NA and SI on an outcome is additive rather than synergistic, the continuous method is more suitable to estimate this association than the 2-group or 4-group methods.

Figure 4: Example of a bar plot resulting from the 4-group approach, showing for each of the four personality subgroups the percentage of participants having a romantic partner. Underlying this simulated data are main effects for both NA and SI, but no synergistic effect.



Moreover, this second objection proposes that the Type D effect is additive rather than synergistic, but is this in line with the literature on Type D personality? As noted in our introduction, earlier writings on Type D personality do suggest that the Type D effect is *more* than additive (Kupper & Denollet, 2007; Pedersen & Denollet, 2003; Denollet, Sys & Brutsaert, 1995; Denollet et al., 2006). Moreover, the most commonly used Type D vs. non Type D operationalization suggest an interaction between those traits because individuals only score 1 on this variable when they score above a cutoff on *both* NA and SI. Taken together this would suggest that the Type D effect is synergistic rather than additive. However, if researchers would on second thought conclude that the Type D effect is better seen as additive than synergistic, then this would require all further analyses to restrict their focus on the additive NA/SI effects. These main effects should then be entered as *continuous* variables in regression analyses, because using their dichotomized versions (i.e. the 2-group and 4-group approach) will not only results in lower power, but also risks both spurious main- (Maxwell & Delaney, 1983) and interaction effects (Thoresen, 2019). Note that such an additive model does not necessarily have to be a linear model. Non-linear models such as quadratic, spline, or threshold regression models are perhaps more suitable in testing the additive continuous NA and SI effects, while taking into account the idea that these traits are only influential above a cut-off score of 10 (Denollet, 2005).

Another counterargument against our findings could be that the data generating mechanism in our simulation study was a continuous interaction model and that other data generating mechanisms for a Type D effect would have resulted in different conclusions. A continuous interaction model is a variable centered approach and can be contrasted with a person-centered approach (Asendorpf & Denissen, 2006; Muthén & Muthén, 2000). Variable centered approaches investigate the association between two or more dimensions (variables) within a population. In contrast, person centered approaches aim at classifying individuals in distinct subgroups or classes of people with similar characteristics. One could argue that the validity of our simulation results would be threatened if the assumed data generating mechanism differs from the true data generating mechanism. It could be that the true mechanism underlying the Type D effect is a set of distinct latent personality classes giving rise to the different score patterns on the DS14 questionnaire, with differences between the classes in their scores on an outcome measure.

Our first response to this counterargument is that we agree that if our simulations do not represent the actual data generating mechanism for synergistic Type D effects, the results of the simulations may not provide a valid benchmark. However, we believe we have good reasons to believe that the simulated way of doing things is realistic. First, the individual difference literature suggest that personality traits are dimensional (Ferguson et al, 2009; Suls, 2014). Second, the personality type variables are *constructed* to be categorical by reducing the scores on two or more personality traits to a limited number of personality types. Subsequently concluding without empirical evidence that these typologies have concrete existence (i.e., reification) would be committing the fallacy of misplaced concreteness (Whitehead, 1997). Moreover, while the existence of personality types would imply a bimodal NA and SI distribution, the empirical results indicate that these constructs show unimodal distributions (***Chapter 6***). Lastly, several empirical studies offer reasons to doubt that Type D is a *categorical* latent construct (Ferguson et al., 2009; Hillen, 2017).

Hence, the burden of proof is on those claiming that Type D personality is a categorical latent construct. But even when its categorical nature would be empirically supported, we would argue for analyzing these personality types as continuous variables. Categorizing continuous traits may result in heterogeneous groups of people as most would not fit the prototypical personality type. Modeling the personality types as continuous scores in statistical analyses, by expressing them as the person's similarity to the prototypical personality type, would more accurately capture such heterogeneity (Chapman, Roberts & Duberstein, 2011).

A limitation of our simulation study is that we merely focused on an interaction between the NA and SI main effects, thereby excluding other nonlinear (e.g., quadratic, cubic, spline) effects and their Interactions. Future research could build upon our work by studying how the various Type D operationalizations perform in detecting such nonlinear NA and SI effects and the interactions between them. A limitation of our empirical study is that we only reanalyzed two datasets. This is only a small part of the many studies that have investigated the association between Type D personality and a wide range of medical and psychosocial outcomes. Future research could therefore focus on replicating our findings by comparing the results of the four Type D operationalizations in as many datasets as possible.

Conclusion

This study showed that the continuous interaction approach is the appropriate method for studying synergistic Type D effects. Other approaches showed either more bias, more false positive findings or lower power. We showed that the most commonly used operationalizations of Type D personality do not adequately distinguish between situations where only NA *or* SI, or *both* NA and SI are causally related to an outcome. Reanalysis of two earlier published Type D studies showed that the significant effects according to these commonly used methods were no longer significant when using the correct continuous interaction approach. However, there remain plenty of studies showing significant Type D effects using the continuous interaction approach (Denollet et al., 2013; Williams, O'Connor, Grubb & O'Carroll, 2011; Nefs et al., 2015; Kupper & Denollet, 2016). Regardless of whether the Type D effect is synergistic or additive, our results imply that findings of earlier research on Type D personality should be reconsidered if they were based on dichotomized subgroup approaches. The present study served as a first step in separating the wheat from the chaff.

CHAPTER 3

A re-evaluation of the Type D personality effect

P. LODDER

Personality and Individual Differences

ABSTRACT

Introduction: Type D personality has been associated with various medical and psychosocial outcomes. Type D's underlying personality traits negative affectivity (NA) and social inhibition (SI) are hypothesized to either additively or synergistically affect an outcome. As some of the methods used to assess a Type D effect have been criticized in the past, this study aimed to investigate for all commonly used methods their tendency of falsely suggesting a Type D effect.

Method: 486000 datasets were generated using a Monte Carlo Simulation. Each dataset was analyzed using various methods to assess a Type D effect. Each method's performance was assessed in terms of absolute bias in the regression estimates and the percentage of significant findings. An online application was developed where readers can easily experiment with this simulation.

Results: Our simulation showed that all commonly used methods under certain circumstances produce findings that could be falsely interpreted as Type D effects. Some of these methods were only biased when NA and SI were correlated.

Conclusion: All commonly used methods to assess a Type D personality effect produce findings that may falsely be interpreted as Type D effects. All earlier research based only on these methods should be reconsidered.

INTRODUCTION

The construct Type D ("distressed") personality (Denollet, Rombouts, Gillebert, Brutsaert, & Sys, 1996; Denollet, Sys, & Brutsaert, 1995) is characterized by the combination of its two subcomponents negative affectivity (NA) and social inhibition (SI). Negative affectivity represents the tendency to experience negative thoughts, emotions and behaviors, while social inhibition refers to the difficulty in expressing thoughts and emotions, particularly in a social context. The combined presence of these traits is called Type D personality and has been linked to various medical and psychosocial outcomes, such as an increased risk of cardiac events (for a meta-analysis, see Grande, Romppel, & Barth, 2012) or poor medication adherence (Williams, O'Connor, Grubb & O'Carroll, 2011). Type D theory states that the *combined* influence of NA and SI is essential, because the combination of experiencing emotional distress and not being able to express these feelings is especially stressful to individuals and may result in serious health problems (Denollet, Sys & Brutsaert (1995).

Considerable debate exists on how to statistically model this combined influence of two personality traits. Commonly used methods classify people in personality subgroups based on whether they score above or below a particular cut-off on the continuous NA and SI traits. Such subgroup approaches should result in Type D effects when both NA and SI are important in explaining an outcome, but various authors have argued that these approaches may cause researchers to falsely conclude Type D effects when only one of the Type D personality traits is related to the outcome (Ferguson et al., 2009; Smith, 2011). In ***Chapter 2*** we provided the first empirical support of this criticism by showing that if *only one* of the personality traits (NA or SI) is related to the outcome, these methods commonly show results that are interpreted as Type D personality effects. The last decade, researchers have started to estimate Type D personality effects based on the main and interaction effects of the continuous NA and SI scores. However, in this article we argue that even such continuous analyses may falsely suggest a Type D effect when the regression model is not correctly specified. This conclusion is substantiated by a computer simulation, where we have generated a wide variety of empirically plausible datasets and have compare the performance of various commonly used methods in estimating a Type D effect.

Commonly used methods

Although latent variable methods are arguably preferred when modeling constructs measured with error such as psychological questionnaire scores (**Chapters 6, 7 & 8**), the present chapter focuses on methods that directly model the observed scores, as these are usually used in the Type D literature. What the commonly used Type D operationalizations have in common is that they use the DS-14 questionnaire (Denollet, 2005) to measure Type D's subcomponents NA and SI, both measured with seven items on a 0-4 Likert scale. These two sets of seven items are then summed to get the NA and SI sum scores. From that point onward the methods start to diverge.

2-group method

The *2-group method* is most commonly used and classifies persons as "Type D" when they score equal to or higher than the predetermined cutoff of 10 on both the NA and SI sum scores. All other people are classified as "Not Type D". The resulting binary/dichotomous variable is subsequently used as an independent variable in statistical analyses to investigate its association with some outcome. For an example of this method see Denollet (2005). This approach resonates with the idea that people need to score high enough on both underlying NA and SI traits (i.e., above the cut-off) before their personality can start to negatively influence various aspects of people's life (Denollet, 2005).

4-group method

The *4-group method* is similar to the 2-group method, but further classifies the "Not Type D" people in three categories based on whether they score higher or lower than the cutoff of 10 on the NA and SI sum scores: "NA+SI-", "NA-SI+", "NA-SI-". As this method classifies people in 4 groups, using this variable in regression analyses requires transforming this 4-group variable into three dummy variables that indicate whether people belong to the Type D group (NA+SI+), the High NA group (NA+SI-) or the High SI group (NA-SI+). Using these three dummy variables as predictors in regression analyses by default causes the remaining fourth group (NA-SI-) to become the reference group. The effects of the three other groups are thus expressed relative to this reference group. Testing whether the Type D group differs from the NA+SI- or NA-SI+ groups requires a slightly different dummy coding with the Type D group as reference group instead of the NA-SI- group. Using the 4-group method to identify a Type D

effect requires three separate tests to be statistically significant. Strictly speaking, the Type D group should not only differ significantly from the reference group, but also from the NA+SI- and NA-SI+ groups, as this would show the added value of scoring high on both personality traits. For an example of this method, see Nefs and colleagues (2014).

Limitations of subgroup methods

The 2-group and 4-group methods have been criticized by various scholars for several reasons (***Chapter 2***; Ferguson et al., 2009; Smith, 2011). First, by using a cutoff to classify people into high or low scores on NA and SI, these methods destroy valuable information on individual differences on these personality traits (Jaccard, Wan, & Turrisi, 1990; MacCallum, Zhang, Preacher & Rucker, 2002; Maxwell & Delaney, 1993). Consequently, this categorization forces researchers to assume that the effect on an outcome is similar for every member of a subgroup. A second criticism of these methods is that the Type D effects are not only sensitive to true Type D effects (i.e., both NA and SI affect the outcome), but also result in significant effects when just one of the two traits is related to the outcome (Coyne & de Voogd, 2012; Smith, 2011). Consequently, these methods are not very specific in their conclusions. At most, they tell us that some aspect of Type D personality is related to the dependent measure, but they do not inform about the nature of the association (e.g., additive, quadratic, synergistic) or whether the Type D effect is caused by NA, SI, or both. A computer simulation study found support for this criticism, showing that subgroup methods resulted in significant Type D effects even when only one of the Type D personality traits was related to the outcome (***Chapter 2***).

One could argue that the spurious Type D effects that occur when using the 2-group method can be prevented by modeling the effect of the Type D group versus the three other groups using the 4-group method, because if only NA is causally related to the outcome, then surely the comparison between the Type D group and NA+SI- group would be non-significant. Similarly, if only SI is causally related to the outcome, then the Type D group should not differ from the NA-SI+ group. However, we argue that this is not necessarily true when the two constructs involved in the classification are correlated, because dichotomizing two continuous correlated variables may cause spurious effects for one variable (e.g., SI) when only the other variable (e.g., NA) is related to an outcome (Maxwell & Delaney, 1993).

Continuous method

To tackle the problems of these subgroup methods, several researchers have argued to use assess Type D effects using a *continuous method* (***Chapter 2***; Ferguson et al., 2009; Smith, 2011), using the continuous NA and SI sum scores as predictors in a regression analysis together with their interaction (i.e., the product of the mean-centered sum scores). When the NA and SI sum scores are both independently related to an outcome in the same direction, there is an *additive Type D effect*. In such a scenario, the predicted values on the outcome of interest are higher for those scoring high on both NA and SI than for those scoring high on only one of these traits. A *synergistic Type D effect*, on the other hand, is present when the interaction between NA and SI is significant in such a way that across the entire observed NA and SI score range, the conditional effect of each trait is positive and increases with higher scores on the other trait. In that case the predicted values on the outcome of interest are even higher than the predicted values resulting from a model including only the NA and SI main effects but not their interaction effect. Many researchers have argued that the Type D effect is synergistic (Denollet et al., 2006; Denollet, Pedersen, Vrints, & Conraads, 2013; Denollet, Sys, & Brutsaert, 1995; Kupper & Denollet, 2007; Pedersen & Denollet, 2003). For instance, Kupper and Denollet (2007) explicitly stated that Type D personality is a synergy between NA and SI. Statistically, synergy can best be modeled in terms of a statistical interaction effect between the continuous scores of two constructs (***Chapter 2***). See ***Chapters 6 and 7*** for an empirical application of the continuous method.

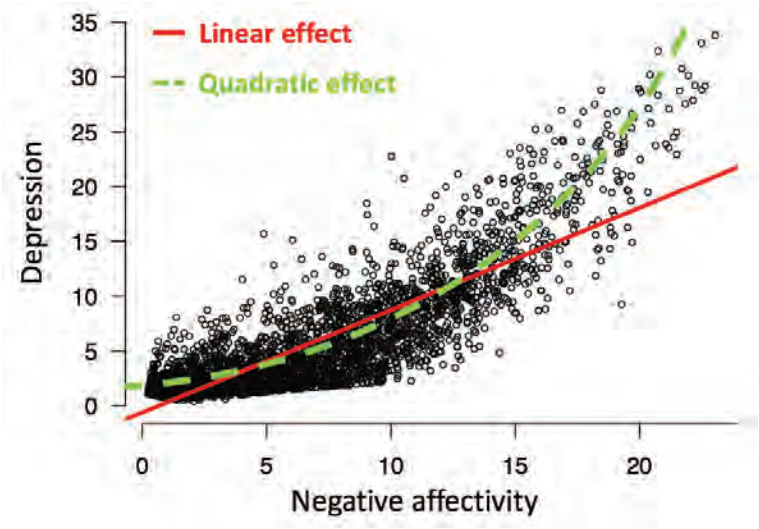
Limitation of continuous method

However, we argue that this proposed continuous method is also not without problems. For this method to perform adequately, it is important that the model is correctly specified. Earlier research has shown that misspecification may result in misleading conclusions about the effects estimated in a statistical model. For instance, if researchers are interested in testing the interaction effect between NA and SI on a dependent measure, then they should also include the first-order NA and SI effects in the model (Aiken & West, 1991). Not doing so may risk spurious interaction effects when only NA or SI is causally related to the outcome. Suppose that in reality only NA is causally related to the dependent measure, then multiplying the NA and SI sum scores will likely result in a significant effect for this interaction term when researchers do not include the first-order effects in the model. This is because the interaction

correlates with NA and not controlling for NA in the model can result in a spurious effect for the interaction between NA and SI. In this chapter, we show that although such spurious interactions can be significantly reduced by mean centering the NA and SI scores before multiplying them, bias is not entirely prevented and we therefore recommend against modeling the Type D effect with only an interaction term. For recent examples of such misspecification in the Type D literature, see Dehghani (2018) or Smith and colleagues (2018).

Another potential misspecification in models testing interaction effects is not considering possible quadratic effects. Various studies have shown that significant interaction effects can masquerade an unmodeled non-linear (e.g., quadratic) effect of one of the constructs involved in the interaction, especially when the two constructs are correlated (Belzak & Bauer, 2019; Bysemeyer & Jones, 1983; MacCallum, Zhang, Preacher, & Rucker, 2002). Suppose that in reality (1) NA and SI are correlated, (2) only NA has a causal quadratic effect on an outcome, and therefore (3) SI is not causally related to the outcome. These three requirements are sufficient in producing spurious interaction effects (i.e., spurious synergistic Type D effects).

Figure 1: Empirical data (from ***Chapter 6***) where a quadratic curve describes the association between the latent NA and Depression estimates better than a linear curve



But how plausible are these assumptions? First, NA and SI typically show a positive correlation around 0.5 (e.g. Ferguson et al., 2009; Grande et al, 2010; Horwood, Anglim & Tooley, 2015). Second, personality traits occasionally show quadratic relationships other variables, regardless of whether the personality traits were modeled as total questionnaire scores or latent variable scores. For instance, quadratic relations have been found between neuroticism and depression (Jorm et al., 2000) and between conscientiousness and job performance (Whetzel, McDaniel, Yost & Kim, 2010). More importantly, several studies have found quadratic effects for NA or SI on various outcome measures (***Chapters 6 and 7***; Kupper, Lodder, Habibovic, Spek & Denollet, preprint). For instance, Figure 1 illustrates that a quadratic curve described the association between NA and depression much better than a linear curve. Interestingly, not adjusting for the quadratic NA and SI effects resulted in a significant interaction between NA and SI (i.e., a synergistic Type D effect). However, this interaction was no longer significant when adjusting for the quadratic effects of NA and SI, suggesting that Type D was not synergistically related to depression and anxiety because this synergy was confounded by the presence of unmodeled quadratic effects. Nevertheless, given that *both* quadratic effects were significant in that study, one could still speak of an *additive quadratic* Type D effect. To sum up, misspecifying the continuous interaction model by either omitting the NA and SI main effects or quadratic effects may cause spurious synergistic Type D effects.

Inconsistent empirical findings

In the empirical literature, several studies have shown diverging results when estimating the Type D effect according to both the 2-group and continuous methods on the same dataset. For instance, Horwood, Anglim & Tooley (2016) used the 2-group method to show that people having a Type D personality have significantly more physical and psychological symptoms than people without a Type D personality. However, the results of the continuous method suggested that these symptoms were only causally related to NA. Neither SI, nor the interaction between NA and SI was statistically significant. Similarly, Bouwens and colleagues (2019) used the 2-group method to indicate that vascular surgery patients with a Type D personality have a significantly lower quality of life than patients without a Type D personality. However, the result of the continuous method suggested that only NA predicted a lower quality of life, not SI or the interaction between NA and SI. These differences indicate that

conclusions regarding whether or not it is the combination of NA and SI that is important in explaining variation in the outcome, depend on the statistical model used to assess the Type D effect. We argue that these methods do not test the same hypotheses and do not detect the same causal mechanisms. Here, we use a computer simulation to illustrate the performance of various statistical models to estimate Type D effects. Although previous simulations have already been published investigating each of the separate methods (***Chapter 2*****;** Belzak & Bauer, 2019), the current simulation will elucidate the performance of all commonly used methods to estimate Type D effects to illustrate how these methods do not detect the same causal mechanisms. Simulation studies are useful tools for discovering whether statistical models perform adequately under a wide variety of circumstances. By simulating datasets that closely match the patterns found in empirical data, simulation studies are not limited to investigating statistical issues (e.g., does a method produce biased effects?), but may also shed light on substantive issues (e.g., are previously reported Type D effects valid?).

The present study

The present study aims to investigate the performance of the 2-group, 4-group and the continuous interaction method in detecting Type D effects and effects of NA or SI only. The goal of this simulation was not so much to precisely assess the performance of each method under a wide variety of underlying effects (see ***Chapter 2*** for that purpose), but more to illustrate the misleading conclusions that could be drawn when assessing Type D personality effects in specific circumstances. The simulation results reported below are therefore limited to several interesting combinations of input parameters. The correlation between NA and SI was of special interest, as this correlation explains why some methods do not perform adequately.

Based on earlier research (e.g., ***Chapter 2***), we expect in our simulations that the 2-group method tends to falsely indicate a significant effect regardless of the causal mechanism relating NA and SI to an outcome measure. Specifically, we expect significant 2-group effects when only one of the underlying personality traits was causally related to the outcome. Similarly, for the continuous interaction method that does not model the first-order NA and SI effects, we expect a tendency towards significant interaction effects when we simulated

that only NA or SI was causally related to the outcome. The bias resulting from such model misspecification is well known, but we aim to illustrate this bias in the context of estimating Type D effects, given that many published Type D studies have used this misspecified model.

We expect the 2-group and misspecified continuous interaction model to produce results that may lead researchers to falsely conclude a synergistic Type D effect, *regardless of the correlation between NA and SI.* Contrarily, we expect the 4-group method to *only* falsely suggest a Type D effect when NA and SI are sufficiently correlated. Lastly, in line with Belzak and Bauer (2019), we expect that a positive correlation between NA and SI causes the continuous interaction model without the quadratic NA and SI effects to falsely detect a synergistic Type D effect (interaction between NA and SI) when only a quadratic NA or SI effect is underlying the data. Readers are encouraged to use the online app specially developed for the purpose of this article, to experiment with simulating different kinds of Type D effects.

METHOD

In this computer simulation, 486000 datasets were generated to test the association between Type D personality and a continuous dependent measure. The simulation R-scripts are available on this project's open science framework page (https://osf.io/9ht35). Readers can experiment with the simulation using a specially developed R-shiny application (https://plodder.shinyapps.io/Type_D_effect_simulation/).

Data generation

In each generated dataset, the two vectors containing latent NA and SI scores were generated using *n* draws from a bivariate normal distribution (M=0; SD=1). The correlation between NA and SI varied across the simulation conditions. The scores on the continuous dependent measure were generated based on the latent NA and SI scores using a linear regression model with six parameters; that is,

$$Y = \beta_0 + \beta_1 NA + \beta_2 SI + \beta_3 (NA * SI) + \beta_4 \mathrm{NA}^2 + \beta_5 \mathrm{SI}^2 + \varepsilon \qquad (1)$$

The model includes an intercept (β_0), the main effects of NA (β_1) and SI (β_2), the interaction between NA and SI (β_3), and the quadratic effects of NA (β_4) and SI (β_5). In the first two simulations, the NA main effect (β_1) and NA quadratic effect (β_4) were varied, while β_0, β_2, β_3, and β_5 were fixed to zero. This means that we simulated data in which only NA was causally related to the outcome. Significant interaction effects in these data sets give a false impression of a synergy between NA and SI. In the third simulation the NA*SI interaction effect (β_3) varied across simulation conditions, while all other effects were restricted to 0.

The latent NA and SI scores were transformed to DS14 item scores using a standard two-factor model with measurement model parameters based on a model fitted to empirical data (factor loadings ranging from 0.68 to 0.82, intercepts from -1.02 to -0.78, and residuals from 0.33 to 0.54). Symmetric threshold parameters were used to transform the generated continuous NA and SI scores to ordinal score on a 0-4 Likert scale (Muthén & Kaplan, 1985). For both NA and SI, the ordinal item scores were summed and the resulting sum scores were used in further analyses. The simulated datasets varied in sample size (n = [100, 300, 500]), and the correlation between NA and SI ($r_{(NA,SI)} = [$-.60, -.45, -.30, -.15, 0, .15, .30, .45, .60]). In the first set of simulations the size of the NA first-order effect ($\beta_1 = [0, 0.5, 1.0, 1.5, 2.0, 2.5]$) was varied, in the second simulation the size of the NA quadratic effect ($\beta_4 = [0, 0.1, 0.2, 0.3, 0.4, 0.5]$), and in the third simulation the size of the NA*SI interaction effect ($\beta_3 = [0, 0.1, 0.2, 0.3, 0.4, 0.5]$), comprising a total of 162+162+162=486 unique simulation conditions with 1000 datasets generated in each condition.

Data analysis

Each simulated dataset was analyzed using seven different linear regression analyses (two personality group analyses and four continuous analyses):

(1) Model 1: the 2-group method,
(2) Model 2: the 4-group method
(3) Model 3: Continuous method: NA*SI (without first mean-centering NA and SI)
(4) Model 4: Continuous method: NA*SI
(5) Model 5: Continuous method: NA*SI + NA + SI
(6) Model 6: Continuous method: NA*SI + NA + SI + NA^2 + SI^2
(7) Model 7: Continuous method: NA + SI + NA^2 + SI^2

For each method, the Type D effect was estimated according to the procedures described in the introduction. In each simulation condition, the 1000 estimated regression coefficients were averaged and the proportion significant effects was determined by dividing the total number of significant effects (at a significance level of 0.05) in a condition by 1000 (replications).

RESULTS

Figures 2 to 7 visualize the simulation results for the seven regression analyses used to estimate a Type D effect. The next sections discuss the results for each method separately.

2-group method

Figure 2 visualizes the simulation results for data where *only* NA was linearly related to an outcome. The size of this underlying NA main effect varied across conditions (separate lines), as did the sample size (figure columns) and the correlation between NA and SI (x-axis). The figure shows the estimated regression coefficients and proportion of significant effects according to the 2-group method and the 4-group method's contrast between the Type D group and the group with high NA scores only. When only one personality trait is simulated to be related to the outcome, there is neither an additive (NA+SI) nor synergistic (NA*SI) Type D personality effect. This implies that if a method should detect *only* such Type D effects, then on average the estimates should be equal to zero and the percentage of significant Type D effects should be equal to 5% (the false positive rate given the chosen significance level).

However, Figure 2 shows that when only NA was related to the dependent measure, the 2-group approach almost always produced statistically significant Type D effects that followed the size of the NA main effect. These false positive rates increased alongside the correlation between NA and SI. The same patterns (not visualized) were observed when we simulated only SI to be related to the outcome. In line with our first expectation, this highlights the problem of the 2-group method for studying synergistic Type D effects: it results in

significant effects even when only one of the two Type D personality traits was related to the outcome.

4-group method

Figure 2 also shows that the 4-group method is not suitable for testing synergistic Type D effects. When only NA and not SI was causally related to the outcome, then a method that adequately detects synergy would indicate that the Type D group does not differ from the NA+SI- group, because these groups should score approximately equal on NA. However, the results indicate that when the correlation between NA and SI differed from zero, the 4-group method suggested a Type D effect (NA+SI+) when in reality only NA *or* SI was causally related to the outcome.

Figure 3 further illustrates the role of the correlation between NA and SI in the bias of the 4-group approach. When this correlation was zero and only NA was causally related to the outcome, then the mean scores in the NA+SI- and Type D groups were equal and differed significantly from the mean scores in the NA-SI+ and NA-SI- groups. However, increasing the correlation between NA and SI resulted in a bias in the mean scores of the Type D group and the NA-SI+ groups.

Continuous interaction model

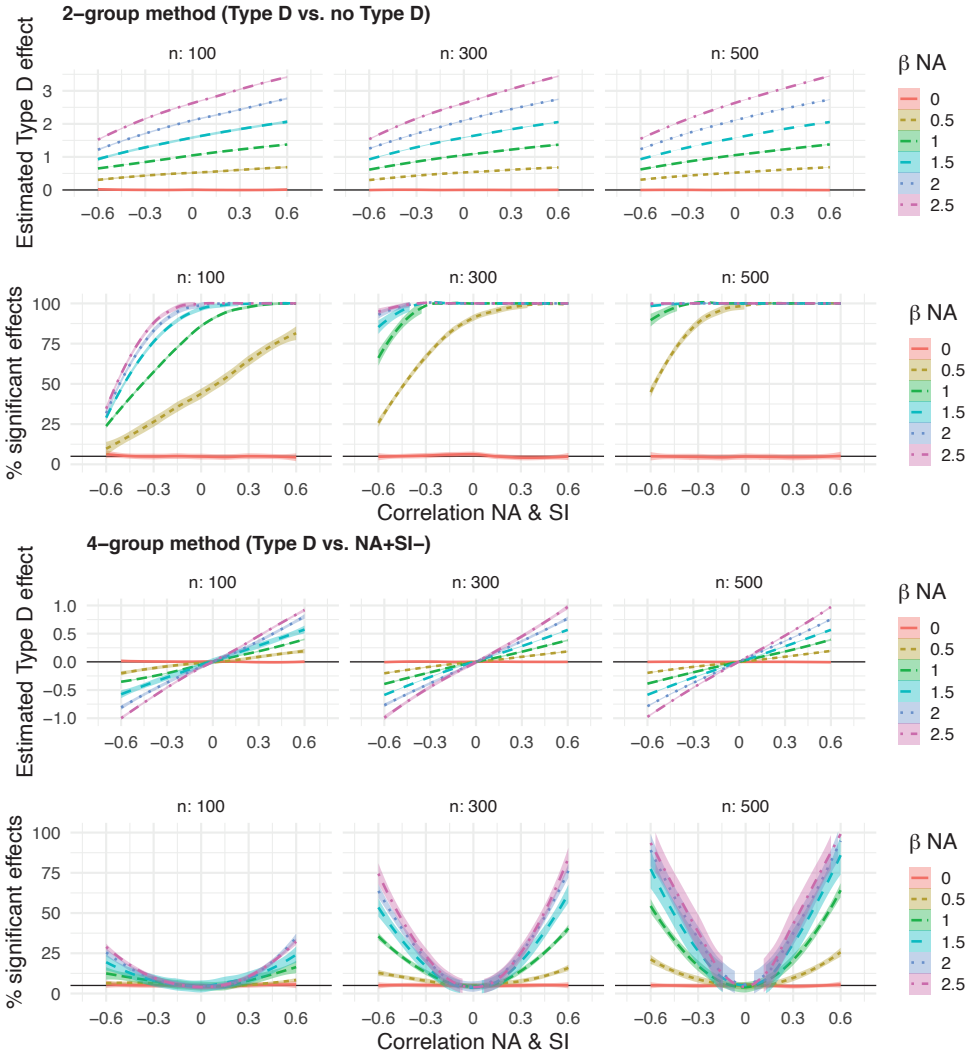
Figure 4 shows the results for simulated datasets in which again *only* NA was linearly related to an outcome. The figure shows the average estimated regression coefficients and detection rates for the continuous interaction models *without* the NA & SI first-order effects (Models 3&4). In line with our expectations, the upper two rows of Figure 4 shows that this model misspecification (model 3) resulted in significant interaction effects (synergistic Type D effect). Although these false positive interaction effects could largely be prevented by first mean-centering the NA and SI sum scores before multiplying them (model 4; bottom two rows), the false positive rate was still higher than the nominal 5% rate whenever the correlation between NA and SI differed from zero. The unadjusted continuous method does therefore not adequately detect a synergistic Type D effect, which is not surprising as statistics textbooks recommend including the first-order effects in interaction models (e.g., Aiken & West, 1991).

Figure 5 shows the mean estimated regression coefficient of the interaction between NA and SI and the proportion of significant effects in simulation conditions where only NA was quadratically related to the outcome and all other effects were fixed to zero. The bottom two rows show estimates for the model including the first-order effects of NA and SI and their interaction (i.e., model 5), while the upper two rows show estimates for the model that also included the NA and SI quadratic effects (i.e., model 6). In line with our expectations, this figure illustrates that not modeling true quadratic effects produced spurious synergistic Type D effect (NA*SI interaction). This bias increased alongside the correlation between NA and SI. The second row of Figure 5 illustrates the importance of including quadratic effects in the model: this kept the false positive rates around 5% when no true NA*SI interaction effect was present in the simulated data.

To show the extent to which the continuous interaction method adequately detects synergistic Type D effects, we also simulated data in which NA and SI interact synergistically (i.e. $\beta_3 > 0$). Figure 6 shows the average bias in the estimated interaction coefficients and proportion of significant effects when a true interaction between NA and SI was driving the outcome. The upper two rows show analyses of a continuous interaction model including NA and SI quadratic effects, while the bottom two rows concern the interaction model excluding quadratic effects.

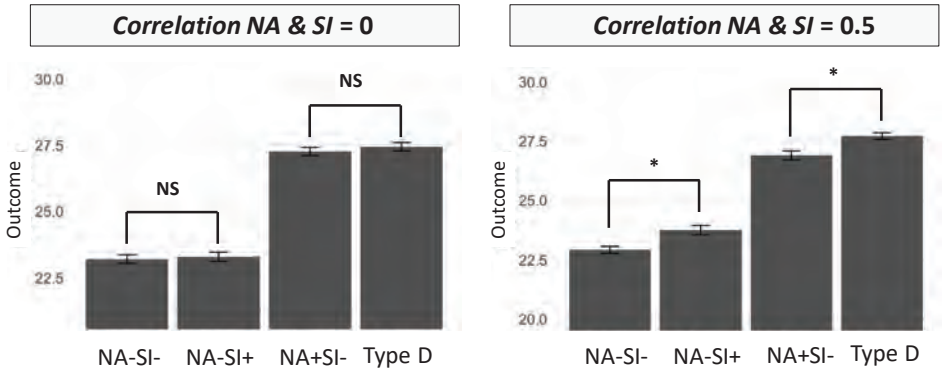
Given that the data were generated from a latent variable model and the simulated scores did not have perfect reliability (i.e., Estimated Cronbach's alpha = 0.88), we expected underestimated regression coefficients. In both models, the interaction effects are slightly underestimated due to such attenuation bias. When NA and SI are uncorrelated, both models perform similarly. However, as the correlation between NA and SI deviates from zero, the power to detect significant interaction effects decreases when quadratic NA and SI terms were included in the model, but this does not happen in the model including only first-order NA and SI effects and their interaction.

Figure 2: The estimated regression coefficients and proportion of significant effects of the 2-group method (upper two rows) and the 4-group method's contrast between the Type D group and the group with high NA scores only (bottom two rows). The simulated datasets varied in sample size (columns), the correlation between NA and SI (x-axis) and the size of the true effect underlying the simulated data (i.e., NA main effect; separate lines). All other effects were fixed to zero.



Lastly, Figure 7 shows the average estimated regression coefficients of the quadratic NA effect and the percentage of significant quadratic NA effects in simulation conditions where only an interaction between NA and SI was causally related to the outcome. These data were analyzed with a model including the first-order and quadratic NA and SI effects, but not their interaction. This figure shows that when a true interaction was causally related to the outcome, but not included in the model, then spurious quadratic NA or SI effect tended to occur. The stronger the positive correlation between NA and SI, the higher the chance on spurious positive quadratic effects, while the reverse pattern was true for negative correlations.

Figure 3: For the 4-group method, the mean scores on a simulated outcome when only NA is causally related to this outcome. The correlation between NA and SI was simulated to be either 0 (left panel) or 0.5 (right panel).



NA-SI- = Reference group; NA-SI+ = Only high SI scores; NA+SI- = only high NA scores; NS = not significant.

** p < .05*

Figure 4: The estimated regression coefficients and proportion of significant effects of the continuous interaction model without the NA & SI first-order effects. When computing the NA*SI interaction term, mean-centering was applied in the bottom two rows but not in the upper two rows. The simulated datasets varied in sample size (columns), the correlation between NA and SI (x-axis) and the true effect underlying the simulated data (NA main effect; separate lines). All other effects were fixed to zero.

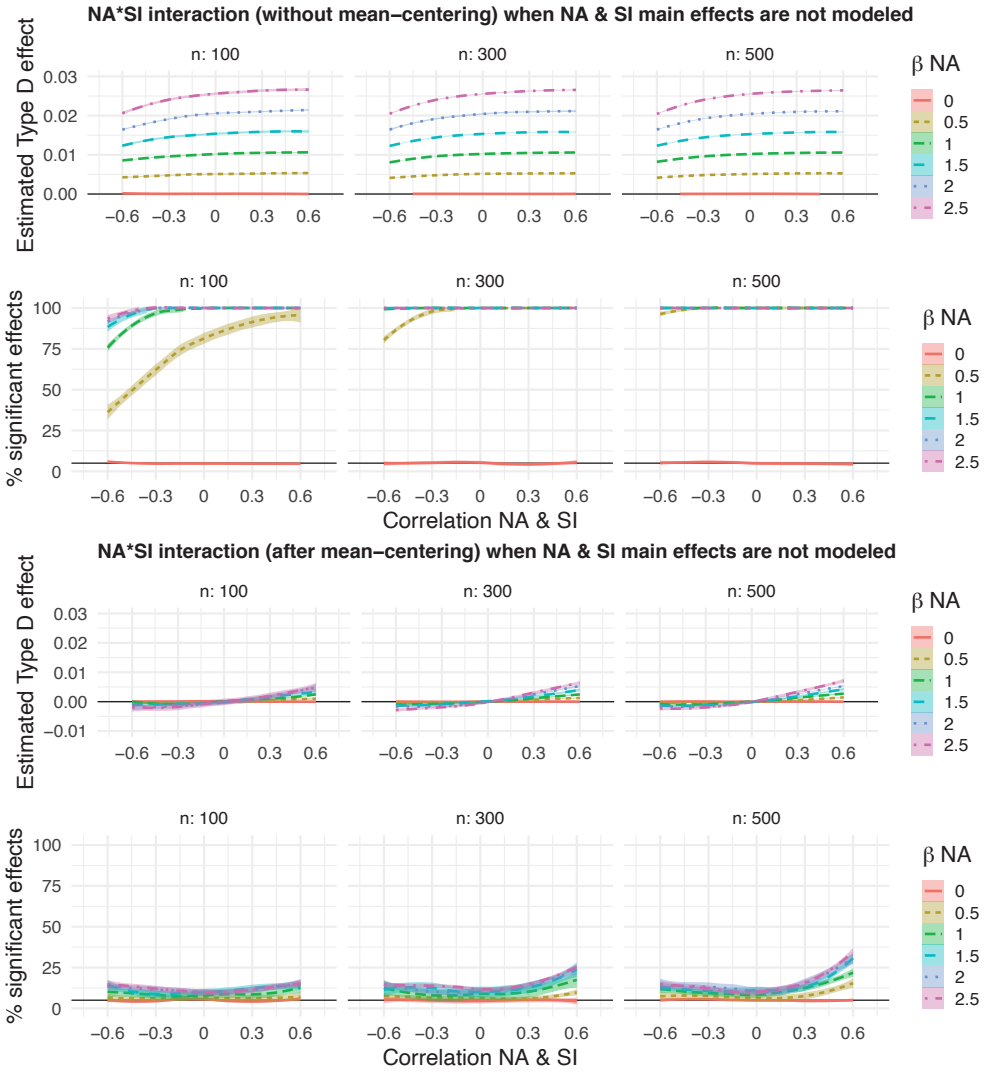


Figure 5: The average regression coefficients for the estimated NA*SI interaction effect and percentage of significant effects. The simulated datasets varied in sample size (columns), the correlation between NA and SI (x-axis), and the true effect underlying the data (only a quadratic NA effect; separate lines). NA & SI quadratic effects were modeled in the upper two rows (i.e., model 6) but not in the bottom two rows (i.e., model 5).

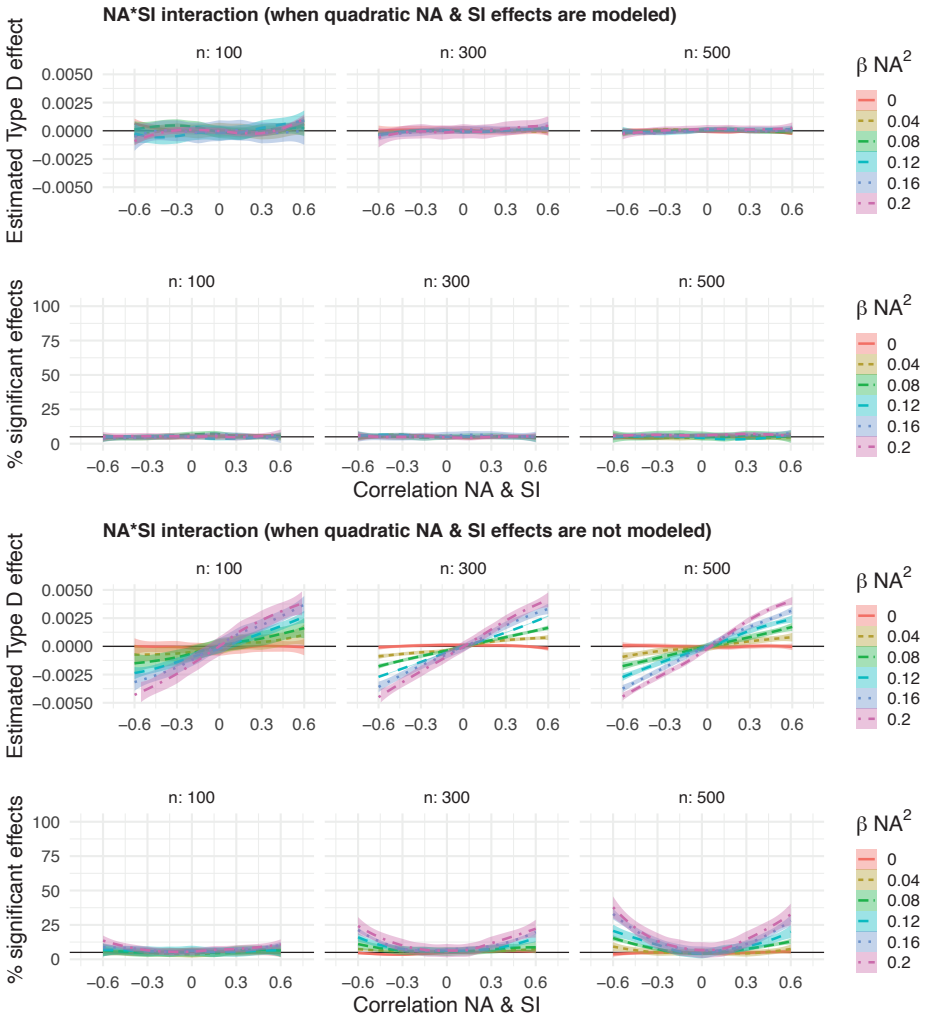


Figure 6: The average bias in the estimated regression coefficients of the NA*SI interaction effect and percentage of significant effects. The simulated datasets varied in sample size (columns), the correlation between NA and SI (x-axis), and the true effect underlying the data (NA*SI interaction effect; separate lines). NA & SI quadratic effects were modeled in the upper two rows (i.e., model 6) but not in the bottom two rows (i.e., model 5).

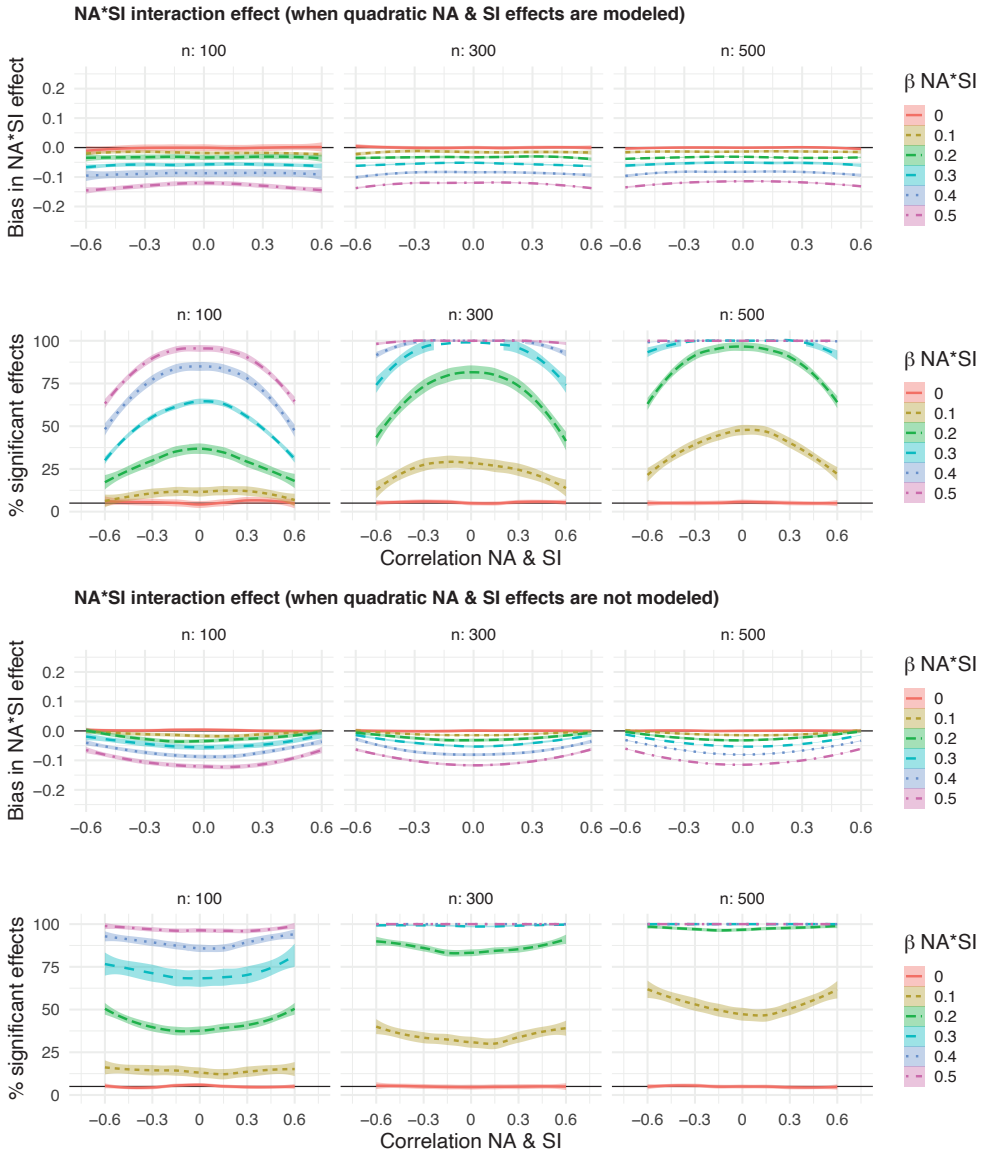
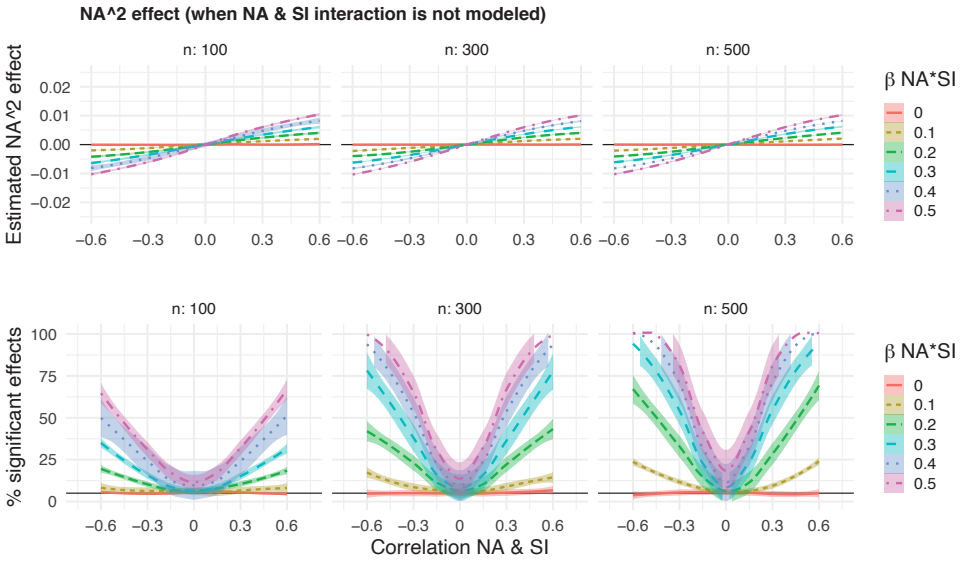


Figure 7: The average estimated regression coefficients of the quadratic NA effect and the percentage of significant effects. The simulated datasets varied across sample size (columns), the correlation between NA and SI (x-axis), and the true effect underlying the data (NA*SI interaction effect; separate lines). The statistical model included the NA & SI first-order and quadratic but not the NA*SI interaction effect (i.e., model 7).



DISCUSSION

This study investigated the performance of various commonly used methods to assess a Type D personality effect. Our results corroborate the earlier finding (***Chapter 2***) that the 2-group method is sensitive to any kind of underlying NA and SI effect. Moreover, our findings indicate inflated false positive rates for the 4-group method and the misspecified continuous methods, given the realistic assumption that there is a positive correlation between NA and SI around 0.5 (e.g., Ferguson et al., 2009; Grande et al, 2010; Horwood, Anglim & Tooley, 2015). When only one of these traits is causally related to an outcome, spurious effects for the other trait may arise due to the positive correlation between the traits, thereby falsely suggesting that both traits are related to the outcome. Although earlier studies have advocated to model the

Type D effect as the continuous NA and SI scores and their interaction (***Chapter 2*****;** Ferguson et al., 2009; Smith, 2011), our study shows that even this approach may result in spurious interaction effects when not including the NA and SI first-order effects in the model. Note that these cited authors did not advocate for constructing such misspecified regression models.

In the simulations focusing on the 2-group and 4-group method, some of the standardized regression coefficients generating the association between NA and the outcome were rather large. An increase of one standard deviation on the latent NA variable was associated with an increase in the outcome measure that ranged between 0 and 2.5 standard deviations. We decided to include these larger effect sizes to clearly illustrate the potential of the 4-group method to indicate a Type D effect when only NA or SI is causally related to an outcome. Although our findings indicate that this issue was less pronounced for the 4-group method than for the 2-group method, we still recommend against using the 4-group method also at smaller effect sizes it produced results that could wrongly be interpreted as a synergistic Type D effect.

Based on the simulation results, we recommend to study Type D effects using a series of models based on the continuous method. In model 1, researchers can include the NA and SI main effects as predictors. In model 2, the researchers can include the first-order and quadratic NA and SI effects. In model 3, both the first-order effects and the interaction between NA and SI can be included to test the synergistic Type D effect. This continuous interaction approach can identify the presence of additive Type D effects (significant NA and SI firs-order effects in model 1), additive quadratic Type D effects (significant NA and SI quadratic effects in model 2) or synergistic Type D effects (significant interaction between NA and SI in model 3). If both quadratic and interaction effects are present then we advise researchers to compare the model fit of models 2 and 3 (Belzak & Bauer, 2019) to find out which of these nonlinear terms shows the best fit to the data. We encourage researchers to preregister the expected kind of Type D effect *before* analyzing the data, in order to prevent capitalizing on chance by conducting several statistical tests of the Type D effect. We also recommend researchers to fit the continuous interaction model using latent variable methods such as structural equation modeling (**Chapters 6 and 7**) to prevent measurement error from attenuating the model estimates, especially those of the quadratic and interaction effects as

those often involve more measurement error than the first-order effects. MacCallum & Mar (1995) argued that when modeling interactions and quadratic effects it is especially important to use latent variable models because the measurement error is considerably larger in quadratic and interaction terms than in individual item scores. Latent variable models can estimate the association between latent variables that are free of measurement error and therefore are not affected by differences in reliability.

Our findings suggest that some of the Type D effects reported in the literature may not be true synergistic Type D effects, but more likely be driven by a causal effect of NA or SI alone, or an additive effect of both. First, all reported synergistic Type D effects may in fact be spurious due to unmodeled quadratic NA and/or SI effects. Second, for all reported synergistic Type D effects that were concluded based on a model without the first-order NA and SI main effects, in reality only NA *or* SI may be causally related to the outcome. Third, for all significant Type D effects using the 2-group and 4-group methods, only NA *or* SI may be causally related to the outcome, even when the 4-group method indicates that the Type D group differs significantly from both the High NA and High SI groups.

It remains unclear how many published studies in the Type D literature have reported a synergistic Type D effect when a more plausible explanation of the outcome would be a single causal effect of NA or SI. It would therefore be interesting to know the percentage of studies where different methods to assess the Type D effect result in different conclusions. Future research could for instance investigate the discrepancy between the 2-group and continuous method, in the subset of studies that reported the results according to both methods (e.g., Bouwens et al., 2019; Horwood, Anglim & Tooley, 2016). Another way to assess the potential bias in the Type D literature would be to conduct individual patient data meta-analyses (Riley, Lambert, & Abo-Zaid, 2010) on commonly investigated dependent measures. This allows for a sufficiently powered test of the Type D effect using a correctly specified continuous method. A first attempt to conduct such an analysis has already been initiated (***Chapter 5***), focusing on the association between Type D personality and adverse (cardiac) events in patients with cardiovascular disease.

In the R-shiny application, the NA and SI scores were not generated from a latent variable model and were assumed to be perfectly reliable, in order to reduce the application's computation time. Consequently, the effects estimated by the app are not attenuated, as typically seen typically seen when analyzing imperfectly reliable measures using their observed score methods rather than latent variable methods (Spearman, 1904). Nevertheless, the simulated datasets in this study were generated according to a latent variable method and the conclusions regarding the bias of the Type D methods are like those resulting from the app.

In sum, this study indicated that all methods commonly used to assess a Type D personality effect produce results that can falsely be interpreted as a synergistic Type D effect. The least biased method to assess the Type D effect (be it additive or synergistic) involves a series of models using the continuous method to assess the NA and SI first-order effects and their interaction, but also the NA and SI quadratic effects. Our findings suggest that some Type D effects reported in the literature are not synergistic Type D effects, but rather other types of NA or SI effects. To shed more light on the extent of this problem, we recommend that earlier published studies investigating a Type D effect should be reanalyzed using the continuous method. Our conclusions are not limited to research on Type D personality, but any field where two continuous measures are transformed in either 2 or 4 groups based on some cutoff runs the risk of falsely concluding a combined causal effect for these measures.

PART II

Reconsidering the Type D personality literature

CHAPTER 4

A systematic review comparing two popular methods to assess a Type D personality effect

P. LODDER
H.M. KUPPER
M. ANTENS
J.M. WICHERTS

General Hospital Psychiatry

ABSTRACT

Introduction: Type D personality, operationalized as high scores on negative affectivity (NA) and social inhibition (SI), has been associated with various medical and psychosocial outcomes. The recent failure to replicate earlier findings could result from the various methods used to assess the Type D effect. Despite recommendations to analyze the continuous NA and SI scores, a popular approach groups people as having Type D personality or not. This method does not adequately detect a Type D effect as it is also sensitive to situations where only NA or SI is causally related to the outcome, suggesting the literature contains studies that falsely conclude that having high scores on both traits is causally important. Here, we systematically assess the extent of this problem.

Method: We conducted a systematic review including 44 published studies assessing a Type D effect with both a continuous and dichotomous operationalization.

Results: The dichotomous method showed poor agreement with the continuous Type D effect. Of the 89 significant dichotomous method effects, 37 (41.6%) were Type D effects according to the continuous method. The remaining 52 (58.4%) are therefore likely not Type D effects based on the continuous method, as 42 (47.2%) were main effects of NA or SI only.

Conclusion: Half of the published Type D effect according to the dichotomous method may be false positives, with only NA or SI driving the outcome.

INTRODUCTION

Type D ("Distressed") personality has been related to various medical and psychosocial outcomes, such as the occurrence of major cardiac events (Denollet et al., 2013; ***Chapter 7***), depression, and anxiety (Nefs et al., 2015; ***Chapter 6***). The construct Type D personality is hypothesized to affect these outcomes through the combined influence of its two subcomponents, the personality traits negative affectivity (NA) and social inhibition (SI). Negative affectivity refers to the tendency of experiencing negative thoughts, feelings and emotions, while socially inhibited people experience difficulty in expressing these emotions and feelings in social situations (Denollet, 2005).

Initially, Type D research mainly focused on how the combined influence of high NA and SI scores affects the prognosis of cardiovascular disease patients (Denollet et al., 1996). These studies mainly involved hard endpoints, such as mortality and various cardiac events. Although some earlier studies on cardiovascular disease patients have not been replicated in subsequent research (Grande et al., 2011), one meta-analysis found support for a Type D effect on adverse events in CAD patients, but not in heart failure patients (Grande, Romppel & Barth, 2012). An explanation for the inconsistent findings involves differences between studies in the sample characteristics and studied endpoints (Kupper & Denollet, 2016). The Type D effect is arguably less pronounced in older patients and mortality endpoints, because at older ages, various medical comorbidities may become more important in explaining mortality than personality effects like Type D. Although such characteristics may in part explain why the Type D effects in studies involving older participants or heart failure patients could not be replicated, here we show that the methodological issue of how to operationalize Type D personality may also play an important role in explain these inconsistencies.

A considerable debate exists on how the NA and SI traits combine in exercising a Type D effect (Ferguson et al., 2009; Smith, 2011; Coyne & de Voogd, 2012). In ***Chapter 2,*** we argued that an *additive* effect would mean that the Type D effect is equal to the sum of the separate NA and SI effects, while a *synergistic* effect would imply that for both NA and SI,

the effect of one trait increases across the entire score range of the other trait. We showed that both additive and synergistic Type D effects can best be modeled by including both the continuous NA and SI scores and their interaction as predictors in a regression analysis. Although several authors have advocated the use of this *continuous method* (Ferguston et al., 2009; Smith, 2011; Coyne & de Voogd, 2012), it has remained more common to assess the Type D effect using a dichotomous operationalization. According to this *dichotomous method*, people are classified as having a Type D personality when they score above a predetermined cut-off on *both* NA and SI. In some studies, the dichotomous method is extended to a *categorical method* with 4 personality groups, which further divides the people without Type D personality in three groups based on the different combinations of scoring above or below the cut-off on the two traits: (1) High NA & Low SI; (2) Low NA & High SI; (3) Low NA & Low SI.

Historically, the development of the dichotomous method was motivated by clinical and empirical considerations. Using cluster analysis, Denollet and colleagues (1996) detected a clinical discontinuity between people with or without high scores on both NA and SI. The prevalence of this Type D personality type was shown to be higher in people suffering from cardiovascular disease than in healthy controls (Denollet et al., 1996). In these earlier studies, NA and SI were assessed with the trait anxiety scale and a social inhibition subscale of the heart patient's psychological questionnaire. Inspired by these measures, Denollet developed the Type-D scale-16 (DS16; Denollet, 1998), measuring each of the NA and SI constructs with eight items. Seven years later the DS16 was revised into the slightly shorter DS14 instrument that provided a more balanced assessment of the various aspects of the NA and SI constructs (Denollet, 2005). From 2005 onwards, the DS14 became the standard instrument used to assess Type D personality.

In the early Type D studies, researchers exclusively used the dichotomous and categorical methods that classified people in two or four personality groups, either based on a predetermined cut-off score of 10 (in case of the DS14) or a median split (in case of other measurement instruments). These methods have been criticized in several studies based on conceptual and empirical arguments (Ferguson et al., 2009; Smith, 2011). Although Whitehead and colleagues (2007) were the first to use a continuous method to estimate a

Type D effect, Ferguson and colleagues (2009) were the first to explicitly argue that Type D personality can better be conceptualized and analyzed as a continuous construct. Several years later, Smith (2011) warned that the dichotomous and categorical methods could produce spurious Type D effects. Recently this was confirmed empirically based on various computer simulations to investigate the adequacy of these methods in estimating the Type D effect (***Chapter 2 and 3***).

These simulations showed that the dichotomous and categorical methods often fail to detect the Type D effect adequately, because they often tend to produce false-positive Type D effect (i.e. Type I errors) when only NA or SI is causally related to the outcome. Although these personality group methods are sensitive to any kind of NA or SI effect, at the same time they are less powerful in detecting a particular significant effect than the continuous method and may therefore also produce more false-negative findings (i.e. Type II errors). Reducing the continuous NA and SI measures to two or four personality types reduces the information about individual differences on these personality traits and this practice is associated with a loss in statistical power up to 60% (Cohen, 1983). The simulation studies also indicated that a correctly specified continuous model does not suffer from this problem and is able to correctly identify the underlying NA and SI effects (***Chapter 2 and 3*)**.

The bias of the dichotomous and categorical approaches is not limited to research on Type D personality, but to any field where two continuous measures are transformed into a variable indicating whether or not someone scores above a cut-off on both measures. Examples are defensive hostility (high scores on both defensiveness and hostility; Helmers et al., 1995), mixed states in bipolar disorder (high scores on both mania and depression; Goldberg, Garno, Leon, Kocsis & Portera, 1998), or androgynous gender schemas (high masculinity and femininity gender scores; Bem, 1981). When the two measures are correlated, the dichotomous and categorical approaches overestimate the presence of Type D effects. When the measures are uncorrelated, only the dichotomous method shows such positive bias, yet the categorical method using 4 groups still results in lower power (***Chapter 2 & 3***).

Given that most of the studies in the Type D literature have used the dichotomous method, the conclusions drawn from significant dichotomous effects may have to be reconsidered. In

these studies, Type D personality may not be responsible for explaining individual differences in the dependent measure, but rather NA or SI only. However, the extent of this bias remains unclear. It would therefore be interesting to know the percentage of studies in which the dichotomous method and continuous method lead to different conclusions. To determine the extent to which the dichotomous and continuous methods have produced different conclusions in the Type D literature, one would ideally like to compare the results of both methods in *all* published empirical studies on Type D personality. Unfortunately, most published studies only used one method to assess the Type D effect, partly because the continuous method has only been argued for in the literature from 2009 onwards (Ferguson et al., 2009). However, it is still possible to investigate the differences in findings between the dichotomous and continuous method, in the subset of studies that reported the results according to both methods. This is the aim of the present study.

Here, we present results of a systematic review of the empirical Type D literature, including any kind of Type D study as long the results were reported according to both the dichotomous and continuous methods. This is no traditional systematic review, as its purpose is not to answer the substantive question of whether Type D personality is related to a particular outcome. Its purpose is rather to assess how often the conclusions drawn from the continuous method and the dichotomous method differ. Based on these comparisons we can estimate the percentage of significant dichotomous effects that do not represent Type D effects (additive or synergistic) according to the continuous method. In line with earlier simulation studies (***Chapter 2 and 3***), we expect that the dichotomous method produce more significant effects than the continuous method. As these simulations also indicated that the dichotomous method is sensitive to main effects of NA or SI only, we expect that some of the significant dichotomous effects will not be Type D effects according to the continuous method, but main effects of NA or SI.

METHOD

Inclusion and exclusion criteria

We included studies that reported the effect of Type D personality on any dependent measure according to both the dichotomous method and the continuous method. Only studies written in the English language were included. We excluded studies using the continuous method if they analyzed the effects of NA and SI in separate univariate analyses (with either NA *or* SI predicting an outcome measure; e.g., zero-order correlations). When in reality only one of these traits is predictive of an outcome, univariate analyses risk finding significant effects for both traits, because of the moderate correlation between NA and SI. If NA is causally related to an outcome, then the correlation with SI may produce a spurious association between SI and the outcome in a univariate analysis. Controlling for NA is important to prevent falsely interpreting the spurious association as a causal effect of SI on the outcome. Studies were also excluded if their continuous model did not include the NA and SI first-order effects in the presence of an interaction, or if the interaction was not modelled at all (i.e., main effects model only).

Search strategy

The literature search was performed on November 4th 2019. The electronic databases Pubmed and PsycINFO were searched for articles containing the term 'Type D personality' in either the title, keywords, or abstract. The hits provided by these databases were filtered to select only empirical studies, resulting in 569 unique studies. Two independent researchers screened the full texts of these studies to determine whether they met the inclusion criteria listed above. This turned out to be the case for 36 studies. Most excluded studies did either not use two methods to assess the Type D effect, or used statistical analyses that did not meet our inclusion criteria (e.g. zero-order correlations). A cited reference search of Type D's measurement instrument, the DS14 questionnaire (Denollet, 2005) was performed using Web of Science and resulted in 665 studies, of which 398 were not yet identified by the earlier search. Of these 398 studies most did not investigate a Type D effect, yet eight not yet identified studies met our inclusion criteria, resulting in a total of 44 included studies. Each of those studies used the DS14 questionnaire to measure NA and SI.

Data extraction

For each study, two researchers independently extracted the following data: [1] First author's name; [2] Journal; [3] Publication year; [4] Sample size; [5] All dependent measures where Type D personality was used as a predictor; [6] For each dependent measure, the reported effect size (e.g. odds ratio (OR), hazard ratio (HR), R^2, Beta) or test statistic (e.g. t, F, Z) and p-value according to the dichotomous method and the [7] continuous method. The effect size was preferred if both an effect size and test statistic was reported. If neither an effect size nor the test statistic was reported, the effect was considered missing and only the p-value was extracted. When articles reported the effect size with a 95% confidence interval instead of a p-value, we calculated the p-value based on the standard error extracted from the confidence interval. If p-values and confidence intervals were not reported, statistical significance was either determined based on whether the authors reported the effect as statistically significant, or else was considered missing. If the interaction effect was not statistically significant, we inspected main effects in models without the interaction effect. Covariate adjusted effects were preferred over unadjusted effects.

Data analysis

For both the dichotomous method and continuous method, the frequency and percentage of significant p-values was calculated. For significant interaction effects, conditional regression slopes were visualized to determine whether the Type D effect was synergistic. P-value distributions were visualized using histograms. Cohen's kappa was used to determine the agreement between the conclusions drawn from both methods. Agreement was also expressed in terms of the percentage of conclusions (i.e., significant or not) that were consistent between the two methods.

We did not conduct standard meta-analytic tests on the extracted data because the purpose of this project was not to aggregate the estimated effect sizes of studies in the Type D literature. Such aggregates would not be very meaningful as the included studies almost never focused on similar outcome measures. Our aim was merely to compare the *conclusions* drawn from two methods commonly used to assess the Type D effect, given the typical sample sizes and statistical power encountered in the Type D literature. For this reason, the known limitations of vote counting (Hedges & Olkin, 1980) are not relevant to

the present study. Vote counting involves counting the statistically significant p-values and this practice can be problematic because studies that are underpowered may produce non-significant effects, even when there are real effects underlying the data (Borenstein, Hedges, Higgins & Rothstein, 2011).

The power to detect significant effects differs between the dichotomous and continuous methods. Computer simulations show that continuous methods in general have more statistical power than methods using dichotomized variables (Cohen, 1983; MacCallum, Zhang, Preacher & Rucker, 2002; ***Chapter 2 and 3***). In ***Appendix C*** we report the results of a small simulation study indicating that if a Type D effect in the form of an interaction between NA and SI is simulated, then the continuous method has always more power to detect a Type D effect than the dichotomous method. This simulation shows that when the dichotomous method results in a significant effect and the continuous method in a non-significant effect, then this difference is likely not explained by a lower statistical power of the continuous method. A more plausible explanation could be that the dichotomous method is sensitive to any kind of NA or SI effect (main/quadratic/interaction), whereas the continuous method adequately detects the presence of each of these different types of effects.

RESULTS

Table D1 in **Appendix D** presents 14 published studies that were excluded from our review because the continuous method was not modeled appropriately. The authors of these studies took seriously the recommendation to analyze Type D personality continuously by investigating the interaction between NA and SI (except for one study where the sum of NA and SI rather than their product was investigated). However, in these excluded studies, these products (or sums) were used in subsequent analyses without adjusting for the continuous NA and SI main effects. As a result, these analyses may suffer from a problem similar to that of the dichotomous approach: they cannot distinguish between four kinds of underlying effects: [1] NA main effect; [2] SI main effect; [3] Additive Type D effect, or [4]

Synergistic Type D effect. Because these studies did not meet our inclusion criteria due to not using the correct continuous method, we have excluded them from further analyses.

Main findings

The flowchart in Figure 1 indicates that of all 967 empirical studies, 44 (7.7%) were included in our review. All included studies were published after 2009, the year when the first recommendation to assess the Type D effect with a continuous method was published. Together, the 44 included studies investigated 158 effects of Type D personality on a dependent measure. Each of those 158 effects was assessed using both the dichotomous and continuous method.

Figure 1: Flowchart of studies included in the systematic review.

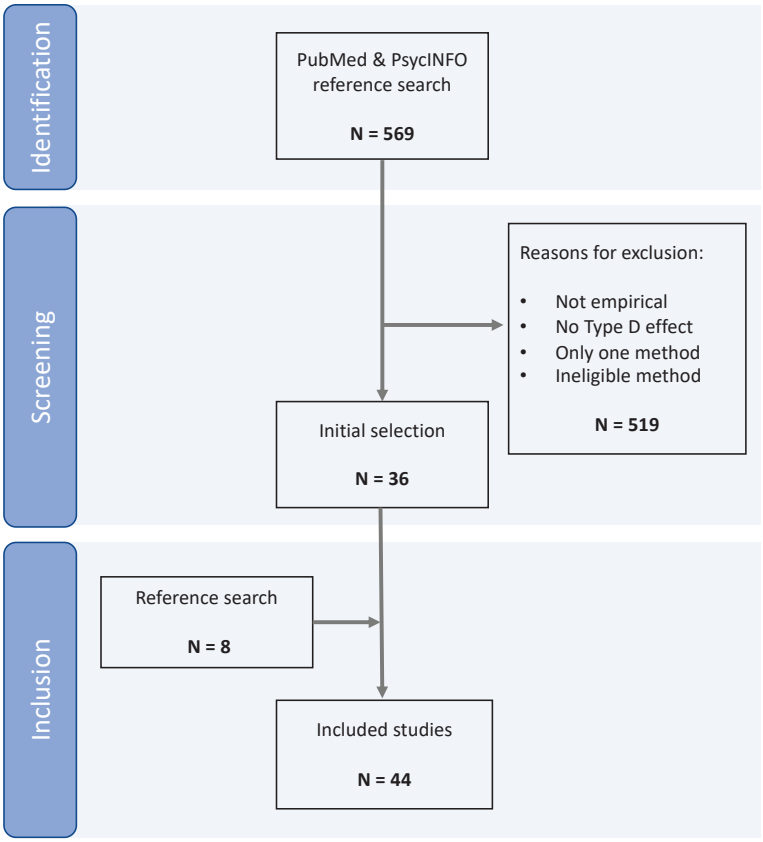


Table 1: For all studies included in our review, the number (%) of statistically significant results according to the continuous (rows) and dichotomous (columns) methods.

Continuous method effect	Significant dichotomous effect		
	Yes	No	Total
No effect	10 (6%)	49 (31%)	59 (37%)
NA main effect	34 (22%)	6 (4%)	40 (26%)
SI main effect	8 (5%)	5 (3%)	13 (8%)
Additive Type D effect (NA+SI)	15 (9%)	1 (1%)	16 (10%)
Synergistic Type D effect (NA*SI)	22 (14%)	8 (5%)	30 (19%)
Total	89 (56%)	69 (44%)	158 (100%)

Table E1 in ***Appendix E*** shows for each study included in our review, the estimated Type D effect and p-value according to the dichotomous and continuous method. A summary of these findings is shown in Table 1, presenting for both the significant and non-significant dichotomous effects the percentage of significant effects according to the continuous method. Of all 158 effects, 89 (56%) were significant according to the dichotomous method, 40 (26%) effects concerned a significant continuous NA effect, 13 (8%) a significant continuous SI effect, 16 (10%) an additive Type D effect (NA+SI), and 30 (19%) a synergistic Type D effect (NA*SI). The direction of all but one (Horwood, Anglim & Tooley, 2016) of the synergistic Type D effects was in the hypothesized direction.

To determine the agreement in statistical significance between the dichotomous and the continuous methods in assessing the Type D effect, both Cohen's kappa and the percentage of agreement were calculated. It turned out that the Type D effect assessed according to the dichotomous method showed poor agreement with both the continuously assessed SI effect (κ = .19; 95%CI = [.08, .31]; agreement = 56.2%), yet reasonable agreement with the NA effect (κ = .55; 95%CI = [.42, .68]; agreement = 77.1%). The dichotomous method showed poor agreement with the additive Type D effect (κ = .14; 95% CI = [.06, .21]; agreement = 51.6%), and even worse agreement with the synergistic Type D effect (κ = .08; 95%CI = [-.03, .18]; agreement = 50.0%). This makes sense, as earlier research has shown that the dichotomous method is not so much sensitive to additive or synergistic Type D effects,but

more to the presence of *any* NA or SI effects. Another explanation for the poor agreement may be the lower statistical power of the dichotomous method relative to the continuous method (***Appendix C***).

Indeed, the dichotomous method showed the best agreement with the detection of *any* continuous effect (i.e. NA or SI main effect or their interaction: $\kappa = .59$; 95%CI =[.46, .72]; agreement = 80.4%). These results indicate that the dichotomous method is very sensitive, but not very specific in detecting the kind of underlying effects. Table F1 in ***Appendix F*** shows the results of sensitivity analyses for different types of outcomes. Cohen's kappa could not be estimated for the mortality outcomes due to low cell counts in the cross table. Nevertheless, regardless of whether researchers studied the cardiometabolic or psychosocial outcomes, the results were similar to those of the overall analysis.

For all studies included in the systematic review, Figure 2 shows the p-value distributions according to both the dichotomous method and the main- and interaction effects resulting from the continuous method. The presence of true effects is indicated by right skewed p-value distributions, with a higher chance on observing lower p-values (e.g., 0.01) than high (e.g., 0.04) p-values. Under the truth of the null hypothesis the p-value distribution is expected to be uniform, with an equal chance on observing any p-value (Simonsohn, Nelson & Simmons, 2014; van Assen, van Aert & Wicherts, 2015). At first sight, the distributions of the dichotomous method and the continuous NA effect look rather similar. These distributions are both very right skewed and therefore indicate a larger evidential value than the distributions of the continuous SI effect and the interaction between NA and SI. Interestingly, the p-value distribution of the interaction effect looks most uniform of all, suggesting the least evidential value for synergistic Type D effects.

Given that the continuous method is much more specific than the dichotomous method in identifying the type of underlying effects, it would be interesting to evaluate the results of the continuous method within the subset of significant dichotomous effects. This would help explain in what way the NA and SI personality traits influence a dependent measure whenever the dichotomous effect is significant.

Figure 2: For all effects included in the systematic review, the distribution of observed p-values resulting from the dichotomous method and the three effects of the continuous method. The black dotted line indicates a significance level of 0.05.

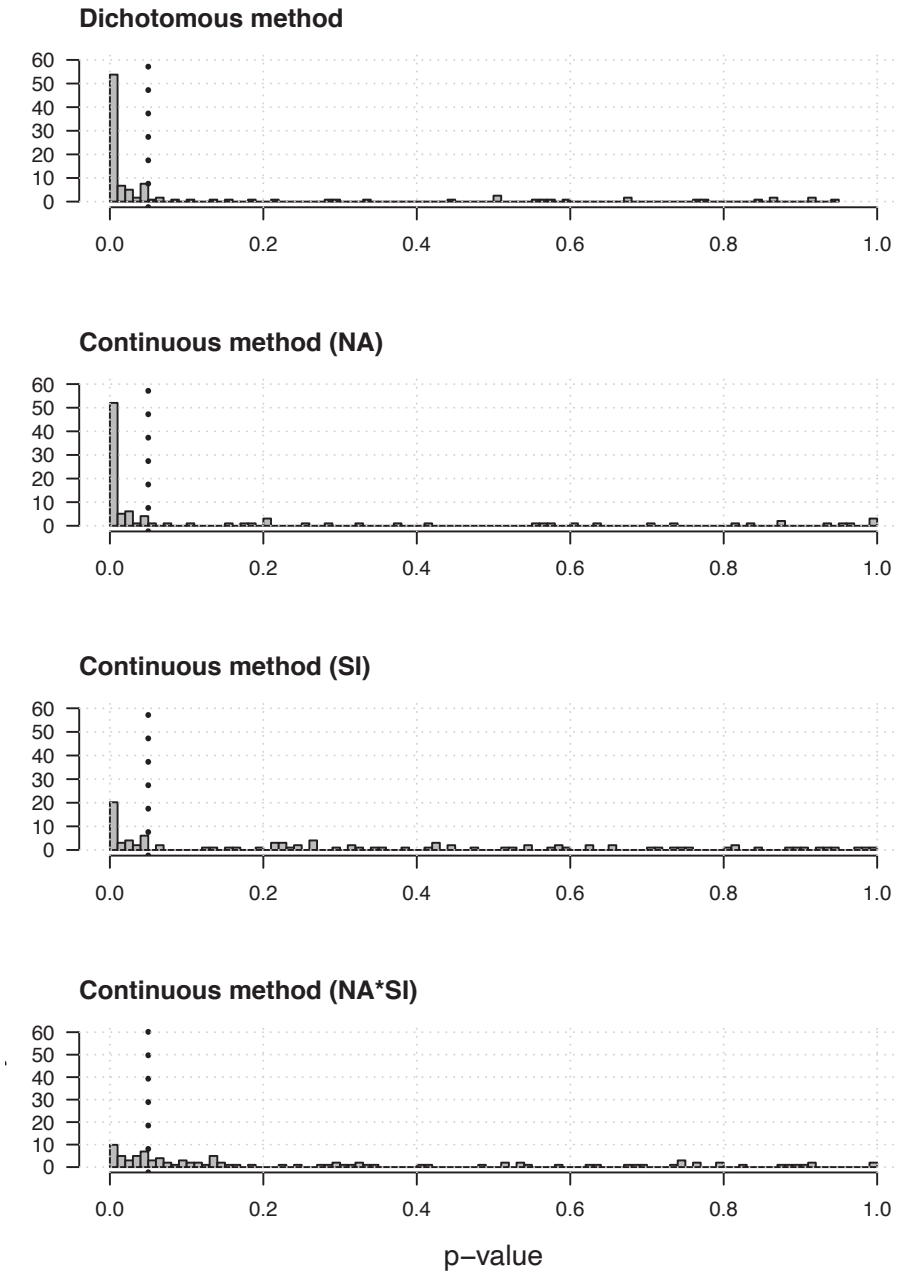
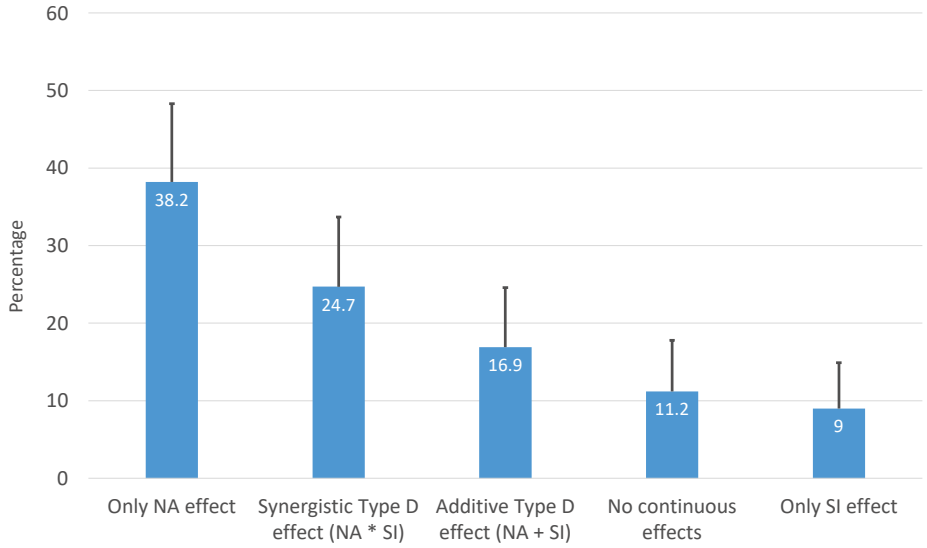


Figure 3 shows for all significant dichotomous effects the percentage of significant effects according to the continuous method. It turned out that of the 89 significant dichotomous effects, 15 (16.9%) were found to be additive Type D effects, and 22 (24.7%) were synergistic Type D effects. Assuming that a Type D effect is either additive or synergistic, 41.6% of the significant dichotomous effects were Type D effects according to the continuous method. The remaining 58.4% significant dichotomous effects method (55 effects) are therefore likely not Type D effects based on the continuous method.

The most frequently observed result in the continuous analyses was a significant effect for NA only in 34 (38.2%) of the significant dichotomous effects. In 8 (9.0%) of the significant dichotomous effects only SI was related to the dependent measure. These results suggest that 47.2% of the significant dichotomous effects are in fact not caused by the combined influence of NA and SI (be it additively or synergistically), but rather by one of these two personality traits only. Lastly, for 10 (11.2%) of the significant dichotomous effects, no significant continuous effect was found.

Figure 3: For all 89 significant Type D effects based to the dichotomous method, the percentage of observed Type D effects according to the continuous method. The black bars represent the upper bound of a 95% confidence interval.



DISCUSSION

The purpose of this study was to determine the discrepancy in the results based on the continuous method and dichotomous method in assessing the Type D effect. Our analyses indicated that the dichotomous method shows poor agreement with both the continuously assessed additive and synergistic Type D effects. For the studies included in our review, the dichotomous method showed reasonable agreement with the NA main effect, suggesting that in many Type D studies only NA is sufficient in explaining variance in the dependent measure.

Earlier research (***Chapter 2 and 3***) indicated that the dichotomous method is not very specific, because it is sensitive to the presence of *any* underlying NA or SI effect, including main effects, quadratic effects and interactions. This suggests that the results of published studies using *only* the dichotomous method should be reconsidered because these studies may have concluded that Type D personality is related to a dependent measure, while in reality the significant dichotomous effect could be caused by NA or SI only. For all dependent measures that are affected by only one of these two personality traits, the Type D personality construct is not necessary in explaining how people vary on the dependent measure (Fiedorowicz, 2020).

The present study showed that 56.3% of the included dichotomous analyses showed statistically significant Type D effects. Of those significant effects, 58.4% were not Type D effects according to the continuous method. Assuming that the studies included in this review are representative (e.g., in terms of sample size and the kind of dependent measures studied) for all studies investigating a Type D effect, it can be concluded that almost 60% of the significant Type D effect reported according to the dichotomous method may be spurious Type D effects caused by the bias of the dichotomous method. Our review suggests that such spurious Type D effects are most likely explained by effects of NA only. These estimates should of course be interpreted with care as their generalizability is conditional on the assumption that the studies included in our review are representative to for all studies investigating a Type D effect. Our review included studies conducted from 2009 onwards,

since at that point Ferguson and colleagues first argued to analyze Type D effects using the continuous method. Differences between included and excluded studies in terms of for instance study population (e.g. cardiac vs. healthy) or dependent measure (e.g. cardiac endpoints vs. mental health questionnaires) may have confounded our estimates. Nevertheless, our results suggest that at least part of the significant dichotomous effects reported in the Type D literature are likely main effects of NA or SI only. This highlights the importance that future research at least takes a closer look at this problem, not only in the context of Type D research, but also in other fields where two correlated continuous measures are dichotomized and transformed into subgroup variables.

Regarding research on Type D personality, a first start would be to re-analyze the earlier published literature using the continuous method. Such analyses should not only investigate the NA and SI main effects, but also their interaction and their quadratic effects (***Chapter 3***). When testing interaction effects, it is important to check whether they are confounded by quadratic effects of the variables involved in the interaction, because not modeling quadratic effects when they are actually present may result in false positive interaction effects (Busemeyer & Jones, 1983; Belzak & Bauer, 2019). Ideally, such re-analyses could be done separately for each type of outcome measure in the form of an individual patient data meta-analysis (Riley, Lambert & Abo-Zaid, 2010). Such meta-analyses combine the raw datasets of earlier published studies focusing on a similar research question. In the context of Type D personality this will allow for a sufficiently powered statistical test using the continuous method, to determine whether the earlier reported dichotomous effects are best explained by NA only, SI only, or the combined Type D effect (additive or synergistic). A first attempt to conduct such an analysis has already been initiated, investigating the Type D effect on adverse (cardiac) events in patients with coronary heart disease (***Chapter 5***).

This study was motivated by earlier findings that the Type D effect can better be analyzed using a continuous approach (***Chapter 2 and 3***). These studies assumed a dimensional conceptualization of the Type D personality construct. A counterargument could be that the true mechanism underlying the Type D effect can better be seen as categorical, with a set of distinct latent personality classes giving rise to the different score patterns on the DS14 questionnaire. However, there appears to be a consensus that personality traits in general

are dimensional in nature, raising the question why NA and SI would be an exception (Ferguson et al., 2009; Suls, 2014). Furthermore, based on a taxometric analysis, Ferguson and colleagues (2009) showed that Type D can better be seen as a dimensional construct than as a categorical construct. Moreover, it is only a small step from pragmatically creating a categorical personality type variable to assigning concrete existence such artificially created categories, a process called reification. Nevertheless, we still believe in the utility of the label Type D personality, but more as a convenient description of a particular NA and SI score pattern, than as having an ontological reality of its own.

Our review also identified a set of 14 studies that did not appropriately use the continuous method to assess the Type D effect. In these studies, the Type D construct was operationalized as the sum or product of NA and SI scores. However, these sums/products were included in subsequent analyses without adjusting for the NA and SI main effects. Therefore, any significant Type D effects may be confounded by the presence of NA and/or SI main effects, making it unclear whether Type D personality is necessary in explaining individual differences in the dependent measure. These effects should be reconsidered in future research using the correctly specified continuous method.

For 11.2% of the significant dichotomous effects, no significant continuous effects were found. This finding is likely explained in terms of differences between the approaches in statistical power. Assuming that both NA and SI have a very small negative effect on a dependent measure and that the power to detect such effects is too low, resulting in non-significant continuous effects. Although dichotomizing continuous variables results in lower statistical power (Cohen, 1983), the dichotomous effect may not necessarily have lower power than the continuous method, because it combines information from *two* dichotomized variables. Since the dichotomous method is sensitive to *any* underlying NA or SI effect, it may pick up explained variance from both the NA and SI main effects, producing effects large enough to be detected with sufficient power, even when the continuous tests of single personality traits are underpowered.

Strengths and limitations

A strength of this study is that is the first to review the results of all published studies that have used both the continuous and dichotomous method to estimate a Type D effect. A comparison of the findings of those methods was necessary, as earlier simulations indicated that the dichotomous method could cause researchers to falsely interpret the presence of Type D effects. Our study is the first to show that a major part of the significant dichotomous method effects published Type D literature is not a Type D effect according to the continuous method, but an effect of only one of its underlying personality traits.

A limitation of this review is that we could only include a small percentage of the Type D literature, because most studies did not use several methods to assess the Type D effect. A second limitation is that we only included published studies. Therefore, generalization of the current results to the entire Type D literature is conditional on the similarity between the included and excluded studies.

A third limitation is that in some included studies the Type D effect was assessed on more than one dependent measure. The Type D effects within these studies may be correlated, either because of they are based on the same sample of participants or because the investigated dependent measures are similar (e.g., the subscale scores of a multidimensional questionnaire). This could mean that the data used in our Cohen's kappa analyses were not independent, possibly resulting in biased estimates of the agreement between the dichotomous and continuous methods. The estimated percentage of significant Type D effects in the literature is likely not affected by this limitation, assuming that this violation of independence is similar in the set of excluded studies.

A fourth limitation of this review is that the continuous methods did not investigate the confounding influence of quadratic NA or SI effects. Research shows that when such quadratic effects have a causal influence but are not included in the statistical analysis model, then spurious interaction effects may arise when the two traits involved in the interaction are correlated, which is the case for NA and SI (***Chapter 3***), but also for many other kinds of studies, such as those testing interaction effects between anxiety and depression (Rutledge et al., 2009). For Type D research, this implies that every reported

significant interaction between NA and SI could be a quadratic effect of NA or SI only. It was not possible to investigate this issue in the current review, because, to our knowledge, only two published studies have investigated whether the synergistic Type D effect is confounded by the quadratic NA and SI effects (***Chapter 6 and 7***). Future research should investigate the extent to which in the published literature unmodeled quadratic effects may have resulted in spurious synergistic Type D effects.

To conclude, this study showed that the majority of the significant dichotomous effects in the Type D literature are likely not additive or synergistic Type D effects, but rather main effects of NA or SI only. This stresses the importance of reconsidering all earlier studies on Type D personality that have used only the dichotomous method to assess the Type D effect. Although this paper focused on Type D personality, we hope our findings also motivate reanalysis of earlier results in other fields involving the combined effects of two correlated variables. These fields may contain many spurious findings if analyses were conducted using subgroup variables based on the dichotomization of two continuous variables. Reconsidering such earlier studies may shed light on why, in general, so many published findings are difficult to replicate. The use of adequate statistical methods minimizes the chance on incorrect conclusions (i.e., Type I & II errors) and is one way to increase the replicability of psychological science (Asendorpf et al., 2015).

CHAPTER 5

Type D personality as a risk factor for adverse outcome in patients with cardiovascular disease: An individual patient data meta-analysis

P. LODDER
J.M. WICHERTS
M. ANTENS
C. ALBUS
I.S. BESSONOV
E. CONDÉN
K. DULFER
S. GOSTOLI
G. GRANDE
P. HEDBERG
C. HERRMANN-LINGEN
T. JAARSMA
M. KOO
P. LIN
T. LIN
T. MEYER
G. PUSHKAREV
G. RAFANELLI
O.I. RAYKH
A. SCHAAN DE QUADROS
M.M. SCHMIDT
A.N. SUMIN
E.M.W.J. UTENS
D.J. VAN VELDHUISEN
J. WANG
H.M. KUPPER

Psychosomatic Medicine

ABSTRACT

Objective: Type D personality, a joint tendency toward negative affectivity (NA) and social inhibition (SI), has been linked to adverse events in patients with heart disease, though with inconsistent findings. Here, we apply an individual patient-data meta-analysis to data from 19 prospective cohort studies (N=11151), to investigate the prediction of adverse outcomes by Type D personality in acquired cardiovascular disease (CVD) patients.

Method: For each outcome (all-cause mortality, cardiac mortality, myocardial infarction (MI), coronary artery bypass grafting, percutaneous coronary intervention, major adverse cardiac event (MACE), any adverse event), we estimated Type D's prognostic influence and the moderation by age, sex, and disease type.

Results: In CVD patients, evidence for a Type D effect in terms of the Bayes factor (BF) was strong for MACE (BF=42.5; OR = 1.14) and any adverse event (BF=129.4; OR = 1.15). Evidence for the null hypothesis was found for all-cause mortality (BF=45.9; OR = 1.03), cardiac mortality (BF=23.7; OR = 0.99) and MI (BF=16.9; OR = 1.12), suggesting Type D had no effect on these outcomes. This evidence was similar in the subset of coronary artery disease (CAD) patients, but inconclusive for heart failure (HF) patients. Positive effects were found for NA on cardiac- and all-cause mortality, the latter being more pronounced in males than females.

Conclusion: Across 19 prospective cohort studies, Type D predicts adverse events in CAD patients, while evidence in HF patients was inconclusive. In both CAD and HF patients, we found evidence for a null effect of Type D on cardiac- and all-cause mortality.

INTRODUCTION

Type D ("distressed") personality is defined as the joint tendency toward negative affectivity (NA) and social inhibition (SI). Individuals with high NA tend to experience negative emotions across time and situations, while those with high SI tend to feel inhibited and insecure during social interactions (Denollet, 2005). Both Type D personality traits are associated with other well-known personality traits. For instance, neuroticism correlates positively with both NA (r =. 68) and SI (r = 0.43), while extraversion correlates negatively with SI (r = -0.65; Denollet, 2005). NA also correlates strongly with trait anxiety (r = .81; Denollet 2000) and the trait anxiety scale of the HPPQ questionnaire has been used to measure negative affectivity before the existence of dedicated Type D personality scales such as the DS14 (Denollet, 2005) and DS16 (Denollet, 1998) .

Although SI is associated with introversion, it is also a distinct construct because introversion does not necessarily involve a distressed experience, while high social inhibition also implies high emotionality and personal distress (Denollet, 2000). SI further expresses how people cope with negative emotions, yet it differs from emotional coping styles such as repression and defensiveness as those involve low distress and unconscious exclusion of negative emotions, while SI is characterized by high interpersonal distress and conscious suppression of emotions. Indeed, the correlation between SI and defensiveness is small (r = -.06) (Denollet, 1998).

Type D personality has been linked to various medical and psychological outcomes (Grande, Romppel & Barth, 2012; O'Dell, Masters, Spielmans, & Maisto, 2011; Versteeg, Spek, Pedersen & Denollet, 2012). The cornerstone of Type D research is the prognostic risk this distressed personality type is thought to pose to cardiovascular disease (CVD) patients. Previous research has found that individuals who inhibit emotional states are at increased risk of cardiovascular dysregulation and complications, such as decreased heart rate variability (Horsten et al., 1999) and cardiovascular recovery (Brosschot & Thayer, 1998) and atherosclerosis (Matthews et al., 1998). Moreover, high SI individuals report that they perceive less social support and are less likely to seek help (Parker et al., 2005). Individuals

with high NA and high SI persistently experience negative emotional states and inhibit the expression of these emotions in social situations, thereby increasing their risk on adverse cardiovascular events for which they are not likely to seek help.

Several meta-analyses have indicated that Type D personality is associated with an increased risk of adverse events in patients with coronary artery disease (CAD), while this has not been found for other types of CVD (Grande, Romppel & Barth, 2012; O'Dell, Masters, Spielmans, & Maisto, 2011). Some have argued that the effect sizes expressing the prognostic risk posed by Type D personality have declined over the years, based on the observation that the earlier studies with smaller sample sizes showed larger effects than more recent and larger studies (Coyne & de Voogd, 2012). However, others have stated that the difficulty in replicating some of the earlier studies can be explained in terms of differences across studies in endpoints and patient characteristics such as age and cardiac diagnosis (Kupper & Denollet, 2016). For instance, a meta-analysis concluded an increased mortality risk of Type D patients with CAD, but no increased mortality risk in patients with heart failure (Grande, Romppel & Barth, 2012). Furthermore, a re-analysis of four earlier published studies indicated that in CAD patients, Type D personality was not predictive of all-cause mortality, but it did show an increased risk on cardiac events, primarily in adult patients younger than 70 years old (Kupper & Denollet, 2016).

Estimating a Type D personality effect

Two constructs synergistically affect another when the conditional effect of each construct on the outcome increases with higher scores on the other construct. Various scholars have argued that a Type D effect involves a synergy between its subcomponents NA and SI (Denollet, Sys, & Brutsaert, 1995; Pedersen & Denollet, 2003; Kupper & Denollet, 2007). Most earlier studies aimed to capture this synergistic effect by classifying people in a Type D group when they score high on *both* the NA and SI total scores (Denollet, 2005). Various researchers have criticized this *2-group method*, not only for resulting in less statistical power, but also for risking spurious Type D effect (Ferguson et al., 2009; Smith, 2011). A *4-group method* was commonly applied to solve this issue by also including groups for people with high scores on only one of the two NA or SI traits. However, two recent simulation studies showed that not only the 2-group method may suggest a Type D effect when in

reality only NA *or* SI was driving the effect, but that the 4-group method to a lesser extent suffers from a similar problem due the correlation between NA and SI (***Chapter 2 and 3***). In some of this simulated data, only one personality trait (e.g., only NA) was causally related to an outcome. However, analyzing such data with the 2-group and 4-group methods often produced statistically significant effects of the Type D group compared to the other groups. In empirical studies, researchers have often interpreted such effects as indicating that it is Type D personality that was driving the effect, while re-analysis with the continuous method indicated that only NA influenced the outcome (***Chapter 4***). This implies that methods that estimate the Type D effect based on two or four personality groups cannot distinguish a causal effect of Type D personality from an effect of only one of the underlying personality traits NA or SI

In line with earlier recommendations (Ferguson et al., 2009; Smith, 2011), these simulation studies concluded that a *continuous method* that does not create personality groups but rather uses the original scale of NA and SI scores, is least biased. This method models the effect of both continuous variables NA, SI, as well as their quadratic effects and interaction. A quadratic effect for NA or SI would imply that the risk this personality trait poses on adverse events is not constant but increases with higher trait scores. Detecting that *both* NA and SI independently predict an outcome would point to an additive Type D effect because the effect of both NA and SI remains constant across the entire score range of these traits. However, if there is an interaction effect between NA and SI on the outcome, then the effect of these traits is not constant, but the effect of one trait changes across scores on the other trait. If the interaction effect is positive, then the effect of one trait on the outcome increases for higher scores on the other trait. We consider such an interaction to reflect a synergy between NA and SI, because higher scores on both traits result in increasingly higher predicted values on the outcome measure. Negative interaction effects would not represent a synergistic effect, because then the effect of one personality trait on the outcome decreases with higher scores on the other trait. Various researchers have argued that the Type D effect involves a synergy between NA and SI (Kupper & Denollet, 2007; Pedersen & Denollet, 2003; Denollet, Sys, & Brutsaert, 1995) and that such synergistic Type D effects can be adequately tested with an interaction effect between two continuous variables (Smith, 2011; ***Chapters 2 and 3***).

Reconsidering the published Type D literature

Although these simulations indicate that the 2-group and 4-group methods may lead researchers to erroneously conclude a Type D effect when only NA *or* SI explains variation in the outcome, the extent of this problem in the Type D literature remains unclear. A recent systematic review of all published studies in the Type D literature included all studies that have estimated a Type D effect according to both the 2-group and continuous method. It turned out that most of the significant 2-group effects were not Type D effects according to the continuous method, but effects of NA *or* SI only (***Chapter 4***). This suggests a major inconsistency in the conclusions drawn from these two methods, questioning the validity of the conclusions drawn from earlier published studies using only the 2-group method. The conclusions of earlier published meta-analyses are equally affected, as those were invariably based on 2-group method effects (Grande, Romppel & Barth, 2012; O'Dell et al., 2011).

The continuous method, however, is also often not adequately applied. According to earlier simulation studies (Belzak & Bauer, 2019; ***Chapter 3***), the continuous method should not only include both the NA and SI main effects and their interaction, but also check whether this interaction is confounded by NA and SI *quadratic* effects. Most published studies using a continuous method did not model these quadratic NA and SI effects. To the best of our knowledge only two earlier published studies have done this (***Chapter 6 and 7***). This suggests that for the remaining literature it stays unclear whether a significant NA*SI interaction indicates a Type D effect, or merely a main- or quadratic effect of NA or SI. This highlights the importance of reconsidering the published Type D literature.

A first reanalysis of Type D's prognostic effect in CAD patients modeled the Type D effect according to both the 2-group and continuous approaches (Kupper & Denollet, 2016). Both approaches showed that Type D increased the risk on cardiac event in CAD patients. A follow-up analysis revealed that this effect was only found for patients younger than 70 years old and did not apply to older patients, possibly because patients who have reached a higher age may experience less environmental (work) pressure and may therefore be less susceptible to stress related cardiac events (Kupper & Denollet, 2016). Nevertheless, it remains unclear why Type D personality does not seem to be a risk factor for cardiac events in older individuals with CAD.

A limitation of the reanalysis by Kupper and Denollet (2016) is that the quadratic NA and SI effects were not included and that only the dichotomous method was used to show that the Type D effect was less pronounced at older ages, making it unclear whether age moderated the Type D effect, or whether it moderated a NA or SI effect only. A second limitation is a possible selection bias because the included data originated from four subsequent cohorts from the same university hospital. Individual patient meta-analysis on data from a diverse set of research groups is essential to achieve a more representative sample of studies.

Here, we present the results of an individual patient meta-analysis focusing on Type D's prognostic effect in cardiovascular disease patients. Individual patient meta-analysis enables an efficient re-analysis of large collections of studies designed to answer a similar research question (Ioannidis, 2012). This results in high statistical power to detect small effects that are hard to detect in each of the included studies individually. Whereas traditional meta-analyses are only able to estimate moderator effects at the study level, individual patient meta-analyses can test moderator effects at the individual level, resulting in more power to detect moderators of Type D's prognostic effect.

Our first aim is to aggregate the data of earlier published prospective cohort studies and test the association between Type D personality and the occurrence of adverse events during follow-up in patients with cardiovascular disease. Another aim is to determine whether this Type D effect depends on age, sex and cardiac diagnosis. Previous research has found that males with Type D personality show a more elevated heart rate response to social tasks than females with Type D personality (Riordan, Howard, & Gallagher, 2019). Studies have also shown that Type D is more predictive of MACE in younger ages than older ages (Kupper & Denollet, 2016) and a meta-analysis concluded an increased mortality risk of Type D patients with CAD, but no such risk in patients with heart failure (Grande et al., 2012). Although our final conclusions will be based on the continuous method, a secondary aim is to estimate the Type D effect according to the 2-group, 4-group and continuous methods to illustrate the difference in their results. In line with earlier research, we expected (1) that Type D personality is a risk factor for cardiac events but not for all-cause mortality and (2) that Type D effect to be more pronounced in younger than in older individuals (Kupper & Denollet, 2016).

METHOD

Inclusion criteria

We only included prospective cohort studies involving patients who at baseline were diagnosed with cardiovascular disease, coronary artery disease, heart failure or ventricular arrhythmia, who were measured on Type D's subtraits NA and SI using the DS16 (Denollet, 1998) or DS14 (Denollet, 2005) or any other validated instrument designed to measure these personality traits. A further requirement was that the occurrence of adverse events was recorded prospectively across the study's follow-up time. We excluded case-control, cross-sectional studies, imaging studies, case series and case reports. When several studies had been published on the same cohort, we included the study with the largest sample size and/or longest follow-up time. Of each included study we contacted the corresponding author (or other authors in case of non-response) and requested the raw data listed below. Participation was rewarded with a maximum of two co-authorships on the current article. Included studies at least had to provide data on Type D personality (individual item scores or total scores for NA and SI) and adverse outcomes (at least one of the following: all-cause mortality; cardiac mortality; myocardial infarction (MI); coronary artery bypass grafting (CABG); percutaneous coronary intervention (PCI)). Additionally, we requested data on clinical characteristics (type of cardiovascular disease), demographic characteristics (age; sex) and study characteristics (date of baseline measurement; follow-up duration).

Search strategy

We conducted a literature search on January 4th 2020, using the electronic databases PubMed, Web of Science and PsycINFO. We updated this literature search on April 1st 2022. We searched for the terms 'Type D personality' AND ['cardiovascular disease' OR 'coronary artery disease' OR 'coronary heart disease' OR 'heart failure' OR 'ventricular arrhythmia'] AND ['adverse event' OR 'myocardial infarction' OR 'mortality' OR 'cardiac death' OR 'cardiac event' OR 'MACE']. Furthermore, we performed hand searches, selecting articles included in earlier systematic reviews and meta-analyses. We limited our search to a period between 1996 and January 2020, because the first publication on Type D personality was in 1996. Two authors (PL & MA), a third reviewer (NK) was consulted. We have used the QUIPS tool to

assess the quality of the prospective cohort studies included in our meta-analysis (Hayden et al., 2013). During the quality assessment we have not evaluated the statistical analysis and inclusion of confounders, because we are responsible for those analysis choices in our individual patient data meta-analysis. ***Appendix H*** presents the results of this quality assessment.

Data extraction

The participating researchers were requested to share their data in either an Excel or SPSS file. Because the shared data already contained all information required to conduct the individual patient data meta-analysis, it was not necessary to further extract data from the included articles. If raw DS-14 item scores were shared, then we checked the calculation of the NA and SI total scores to prevent errors in calculating the total scores (e.g., reverse coding). For each study, the NA and SI scores were standardized within studies to accommodate for the fact that three of the included studies did not use the DS14 questionnaire but other instruments to measure NA and SI that preceded the DS14. Within cluster (i.e., within study) standardization is recommended in multilevel studies when effects of person level predictors (e.g., personality traits) are of primary interest (Enders & Tofighi, 2007).

Operationalizing Type D personality

We operationalized Type D personality according to the continuous interaction method. ***Appendix G*** presents the methods and results for analyses based on the 2-group and 4-group methods. The continuous method models both the continuous NA and SI main effects, as well as their interaction (***Chapter 3***). The method further investigates whether the interaction is confounded by quadratic NA or SI effects. The quadratic and interaction effects are calculated by multiplying the mean-centered or standardized NA and SI scores. In models investigating interactions between correlated variables, it is important to investigate the presence of quadratic effects for the variables involved in the interaction. When two variables are correlated and one of them has a quadratic causal influence on the outcome, then an interaction model without quadratic terms often falsely shows an interaction effect between the two variables (***Chapter 3***; Belzak & Bauer, 2019). Including both the interaction and quadratic effects in the same model could reduce the power to detect each of those

effects (Belzak & Bauer, 2019). Therefore, we estimated the NA*SI interaction effects both in models including and excluding the NA and SI quadratic effects. When no quadratic NA or SI effects are found, the interaction effect in the model without the quadratic effects was used to represent the Type D effect.

Endpoints

We investigated seven endpoints, including five observed endpoints (all-cause mortality, cardiac mortality, MI, CABG, and PCI) and two composite endpoints (MACE and any adverse event). MACE was defined as the occurrence of cardiac mortality, MI, CABG, or PCI during follow-up. *Any adverse event* was defined as the occurrence of MACE or all-cause mortality during follow-up. If the effect of a composite endpoint is only driven by one of the observed endpoints included in the composite, then a significant composite endpoint could wrongly raise the impression that the other observed endpoints are also affected (Ferreira-González et al., 2007). Therefore, we did not limit our analyses to these composite endpoints, but also present the findings for each of the directly observed outcomes. The included studies differed in the number of recorded endpoints. When computing the MACE and *any adverse event* endpoints, only studies were included that recorded each of the endpoints included in these composites. For instance, if a study only recorded cardiac mortality, then this study could not be used in analyses of the MACE or any adverse events endpoint because it was unknown whether these patients had an MI or underwent CABG or PCI.

Statistical analysis

We conducted our primary individual patient data meta-analysis according to a one-stage approach (Burke, Ensor & Riley, 2017). This approach aggregates the data across the included studies and uses a multilevel approach to allow for variation in the estimated regression coefficients across studies. We used a Bayesian estimation procedure to determine the evidence in favor of both the null and the alternative hypothesis. Bayesian multilevel logistic regression models were fitted using the R-package brms (Bürkner, 2017). All regression coefficients (intercept + predictor coefficients) were modeled as random parameters to capture the dependency between scores of participants included in the same study. Parameters were estimated using Markov Chain Monte Carlo (MCMC) sampling with three chains and 3000 iterations, including 1000 burn in iterations. This number of iterations

resulted in an effective sample size of at least 400 for each estimated parameter. Trace plots were inspected to assess convergence. Age and sex were both included as covariates and as potential moderators of the Type D effects on each endpoint. The effects of Type D personality on each endpoint were estimated according to each of the 2-group, 4-group, and continuous approaches. Final conclusions were based on the continuous method, because this approach is the least biased according to earlier simulation results (***Chapter 2 and 3***). Age and sex were both included as covariates and as potential moderators of the Type D effects on each endpoint. Moderation models were estimated separately for age, sex, and disease type, each model including the interaction effect between age/sex/disease on the one hand, and the personality trait variables on the other hand (NA, SI, NA^2, SI^2, NA*SI).

For all models, effects were expressed in terms of odds ratios, including 95% Bayesian credible intervals. In line with earlier research (Van Zwet, 2019), we assumed the priors of the regression coefficients to be normally distributed $N(\mu=0, \sigma=2)$. As a sensitivity analysis we also investigated the same prior but with smaller or larger standard deviation ($\sigma=1$ and $\sigma=4$). For each method, the evidence for a Type D effect in terms of the Bayes factor was quantified as the evidence ratio of the posterior probability of a hypothesis against its alternative (i.e., complement). For example, the evidential value for a Type D effect according to the continuous method was computed as the ratio of the posterior probability that the regression coefficient of the NA*SI interaction was larger than 0, against to the posterior probability that this coefficient was 0 or smaller. To quantify the evidence in favor of the null hypothesis of no Type D effect (regression coefficient of NA*SI interaction = 0), Bayes factors were estimated according to the Savage-Dickey density ratio method (Verdinelli & Wasserman, 1995). Bayes factors can be used to quantify the support of one model compared to another model. In contrast to frequentist statistics, this allows us to quantify evidence in favor of a hypothesis (e.g., evidence in favor of the null hypothesis of no Type D effect). Bayes factors were interpreted according to guidelines by Kass & Raftery (1996) (BFs 1 to 3.2 = “Anecdotal”; BFs 3.2 to 10 = “Substantial”; BFs 10 to 100 = “Strong”; BFs 100 or larger = “Decisive”).

As a sensitivity analysis, we also conducted two-step meta-analyses to investigate whether the results of our one-step analysis are robust against the selection of a different meta-

analytic approach (Riley, Lambert & Abo-Zaid, 2010). In the first step, logistic regression analyses were conducted separately for each included study, to estimate for each endpoint the association with Type D personality according to the continuous method. In the second step, a fixed effects meta-analysis (Rice & Higgins, 2018) was conducted for each endpoint on the log odds ratios and standard errors estimated in step 1. The exponentiated (odds ratio) results of those analyses were visualized in forest plots. All analyses were conducted using R (Team R, 2017) and the script is available on this project's open science framework page together with the preregistration of the data collection and analysis plan: https://osf.io/czmhs/.

RESULTS

Our initial literature search resulted in 367 unique studies. The flowchart in Figure 1 shows that after reviewing the titles and abstracts, 330 studies were excluded because they either did not use a prospective cohort design or did not involve patients with cardiovascular disease. Of the resulting 37 studies an additional 12 were excluded for similar reasons after examining the full text. We emailed the corresponding authors of the remaining 25 eligible studies. In case of no response, we first sent two reminders before e-mailing other authors. Researchers of 20 studies responded to our emails and 18 were willing to participate in this project by sharing their data. The authors of the remaining studies did either not respond or indicated that the data could not be shared because projects involving that dataset were still in progress. After updating the literature search during the review process, we included one additional study in our analysis, resulting in 19 included prospective cohort studies.

Table 1 shows the characteristics of these 19 studies, featuring a total of 11151 patients with CVD who were followed for a median follow-up time of 47.3 months (IQR = 18.0 to 65.4). The included studies differed in cardiac diagnosis, age, and sex of patients, but on average patients were 62.5 years old (SD = 11.3), the majority were male (75.6%) and most were diagnosed with CAD (N_{CAD} = 8096; N_{HF} = 2027; N_{VA} = 638; N_{CVD} = 390).

Figure 1: Flow chart of the systematic literature review

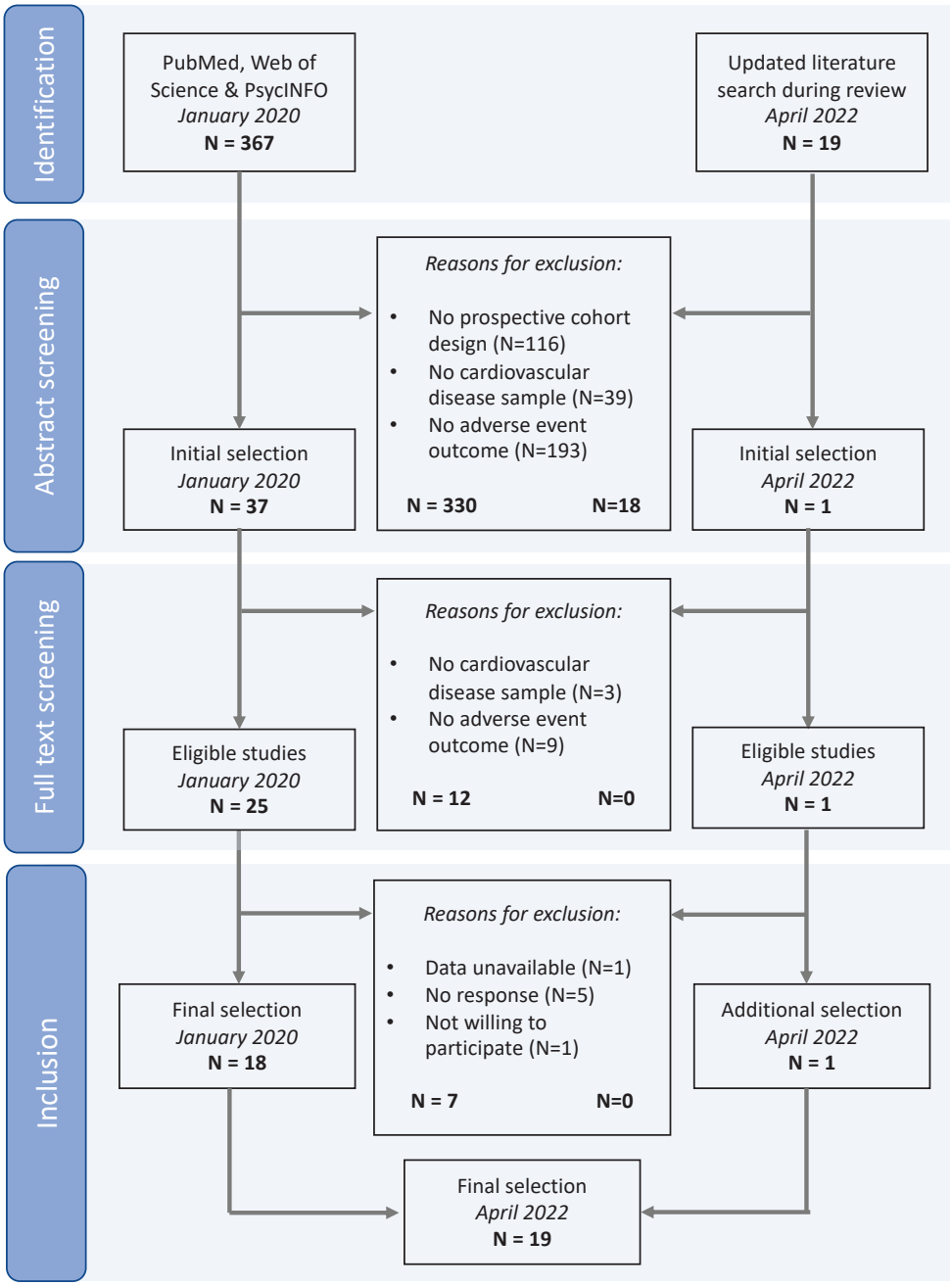


Table 1: Characteristics of studies included in the individual patient-data meta-analysis

Study	N	Diagnosis	Country	Follow-up (months)	Age (M)	Men (%)	Type D Measurement	NA (M, SD)	SI (M, SD)
Denollet et al. (1996)	378	CAD	Belgium	95	55.6	88.6%	STAI & HPPQ	9.8 (6.7)	10.5 (6.6)
Denollet et al. (2000)	364	CAD	Belgium	60	56.5	91.8%	DS16	9.7 (6.6)	14 (6.6)
Denollet et al. (2006)	326	CAD	Netherlands	20	56.8	87.1%	DS16	9.2 (6.6)	13.3 (6.3)
Martens et al. (2010)	466	CAD	Netherlands	22	59.3	78.5%	DS14	7.3 (6.2)	9.1 (6.5)
Pelle et al. (2010)	641	HF	Netherlands	37	66.4	74.3%	DS14	7.1 (6.4)	9.1 (6.5)
Schmidt et al. (2011)	137	CAD	Brazil	12	60.2	63.5%	DS14	10.6 (6.7)	10.3 (7.4)
Coyne et al. (2011)	1047	HF	Netherlands	18	70.9	62.6%	DS14	6.3 (6.0)	7.8 (6.9)
Herrmann-Lingen et al. (2016)	569	CAD	Germany	18	59.2	78.9%	DS14	15.8 (4.8)	11.8 (5.9)
Grande et al. (2011)	1091	MIX	Germany	71	62.7	74.8%	DS14	10.1 (5.7)	8.3 (5.2)
Denollet et al. (2013a)	638	VA	Netherlands	38	62.9	80.6%	DS14	7.5 (6.4)	9.0 (6.3)
Denollet et al. (2013b)	541	CAD	Belgium	60	58.7	87.4%	DS14	9.0 (6.3)	9.8 (6.3)
Meyer et al. (2014)	470	CAD	Germany	60	63.7	76.8%	DS14	10.6 (5.7)	9.2 (5.7)
Sumin et al. (2015)	682	CAD	Russia	12	58.5	81.8%	DS14	9.1 (4.1)	9.3 (3.5)
Dulfer et al. (2015)	1190	CAD	Netherlands	120	62.3	72.6%	DS14	9.4 (6.8)	9.1 (6.5)
Gostoli et al. (2016)	117	VA	Italy	24	63.1	74.4%	DS14	8.1 (6.7)	7.4 (6.5)
Pushkarev et al. (2017)	939	CAD	Russia	12	58.7	75.3%	DS14	10.4 (5.8)	9.7 (5.5)
Conden et al. (2017)	941	CAD	Sweden	76	70.5	66.7%	DS14	6.6 (5.6)	7.9 (5.8)
Lin et al. (2019)	222	HF	Taiwan	18	60.4	66.2%	DS14	6.5 (5.1)	6.0 (5.7)
Lv et al. (2020)	392	CAD	China	12	61.6	68.9%	DS14	11.4 (4.7)	10.9 (4.9)

CAD = coronary artery disease; HF = heart failure; VA = ventricular arrythmia; MIX = mix of various cardiovascular disease diagnoses.
HPPQ = heart patients psychological questionnaire; STAI = state trait anxiety inventory

Figure 2: For each included study, a scatterplot of the NA and SI sum scores. The dot size represents the frequency of a NA and SI score combination.

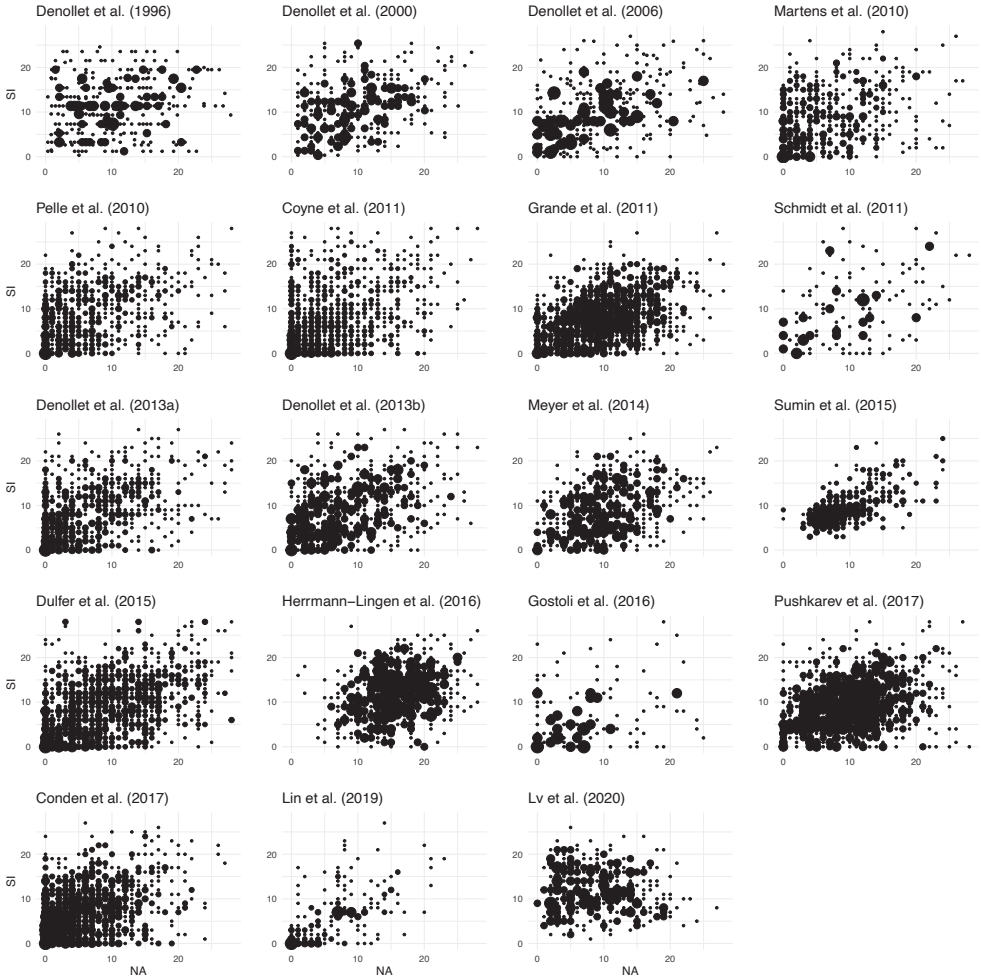


Table 2: For each endpoint, the estimated odds ratios [95% Bayesian credible interval] of demographic predictors, the Type D effects according to the continuous method. The 95%CI of bold cells does not include an odds ratio of one.

Outcome	All-cause mortality	Cardiac mortality	Myocardial infarction	CABG	PCI	MACE	Any adverse event
Sample size	N = 10647	N = 6166	N = 6269	N = 2832	N = 2840	N = 4315	N = 6013
Predictor	OR [95%CI}	OR [95%CI}	OR [95%CI}	OR [95%CI}	OR [95%CI}	OR [95%CI}	OR [95%CI}
Demographics [m3]							
Age (standardized)	**1.933 [1.559, 2.355]**	**1.56 [1.238, 1.991]**	1.107 [0.785, 1.615]	0.945 [0.561, 1.65]	0.835 [0.674, 1.032]	1.087 [0.833, 1.436]	1.14 [0.918, 1.456]
Men	**1.27 [1.012, 1.597]**	**1.681 [1.096, 2.648]**	1.089 [0.735, 1.676]	0.832 [0.299, 2.224]	0.725 [0.406, 1.212]	1.069 [0.762, 1.461]	1.021 [0.748, 1.369]
Continuous method							
NA [m1]	**1.156 [1.045, 1.296]**	**1.284 [1.088, 1.51]**	1.118 [0.96, 1.325]	1.296 [0.829, 1.861]	**1.205 [1.002, 1.435]**	**1.283 [1.146, 1.44]**	**1.269 [1.139, 1.425]**
SI [m1]	1.011 [0.922, 1.136]	1.061 [0.873, 1.341]	1.09 [0.947, 1.277]	0.977 [0.69, 1.384]	1.045 [0.874, 1.247]	1.049 [0.945, 1.166]	1.05 [0.952, 1.165]
NA2 [m2]	1.038 [0.975, 1.107]	1.054 [0.918, 1.179]	1.074 [0.977, 1.181]	1.026 [0.757, 1.298]	1.004 [0.842, 1.144]	1.023 [0.948, 1.101]	1.019 [0.952, 1.09]
SI2 [m2]	1.005 [0.916, 1.084]	0.927 [0.809, 1.059]	1.076 [0.946, 1.26]	1.088 [0.82, 1.422]	1.041 [0.9, 1.194]	1.058 [0.967, 1.157]	1.041 [0.962, 1.126]
NA * SI [m3]	0.996 [0.918, 1.092]	0.996 [0.851, 1.191]	1.069 [0.937, 1.232]	1.171 [0.844, 1.666]	1.112 [0.896, 1.351]	**1.14 [1.001, 1.286]**	**1.167 [1.033, 1.313]**
NA * SI [m4]	1.02 [0.953, 1.116]	0.994 [0.870, 1.142]	1.120 [0.995, 1.282]	1.165 [0.891, 1.507]	1.099 [0.884, 1.308]	1.126 [0.99, 1.268]	**1.135 [1.029, 1.253]**

CABG = coronary artery bypass grafting; MACE = major adverse cardiac event; OR = Odds ratio; PCI = percutaneous coronary intervention

m1: Model = Age + Men + NA + SI

m2: Model = Age + Men + NA + SI + NA2 + SI2

*m3: Model = Age + Men + NA + SI + NA2 + SI2 + NA*SI*

*m4: Model = Age + Men + NA + SI + NA*SI*

Table 3: For each endpoint, Bayes factor (BF) estimates and evidential value for the presence (main hypothesis) or absence (null hypothesis) of a Type D effect according to the continuous methods. Bold faced cells indicate strong or decisive evidential value.

	All-cause mortality		Cardiac mortality		Myocardial infarction		CABG		PCI		MACE		Any adverse event	
Type D effect	BF	Evidence	BF	Evidence	BF	Evidence	BF	Evidence	BF	Evidence	BF	Evidence	BF	Evidence
Complete sample (N=11151)														
Main hypothesis: NA*SI > 0	0.82	Anecdotal	0.86	Anecdotal	5.45	Substantial	4.87	Substantial	6.37	Substantial	**40.1**	**Strong**	**99.0**	**Decisive**
Null hypothesis: NA*SI = 0	**47.18**	**Strong**	**23.34**	**Strong**	**19.29**	**Strong**	8.05	Substantial	9.87	Substantial	3.83	Substantial	1.57	Anecdotal
CAD patients (N=8096)														
Main hypothesis: NA*SI > 0	4.96	Anecdotal	4.04	Anecdotal	3.89	Substantial	5.24	Substantial	6.65	Substantial	**39.0**	**Strong**	**175.47**	**Decisive**
Null hypothesis: NA*SI = 0	**17.38**	**Strong**	**10.98**	**Strong**	**22.35**	**Strong**	8.19	Substantial	9.57	Substantial	3.54	Substantial	1.04	Anecdotal
HF patients (N=2027) *														
Main hypothesis: NA*SI > 0	0.55	Anecdotal	0.81	Anecdotal		-		-		-	1.15	Anecdotal	1.24	Anecdotal
Null hypothesis: NA*SI = 0	**10.14**	**Strong**	4.90	Substantial		-		-		-	2.52	Anecdotal	2.88	Anecdotal

BF = Bayes factor; CABG = coronary artery bypass grafting; CAD = coronary artery disease; HF = heart failure; MACE = major adverse cardiac event; PCI = percutaneous coronary intervention

** Empty cells indicate that insufficient data was available to estimate the Type D effect on a particular endpoint for this patient sample*

Table 4: For each endpoint, the estimated odds ratios [95% Bayesian credibility interval] of the moderating influence of age, sex, and disease on the Type D effects according to the continuous method. The 95%CI of bold cells does not include an odds ratio of one.

Outcome	*All-cause mortality*	*Cardiac mortality*	*Myocardial infarction*	*CABG*	*PCI*	*MACE*	*Any adverse event*
Predictor	*OR [95%CI}*	*OR [95%CI}*	*OR [95%CI}*	*OR [95%CI}*	*OR [95%CI}*	*OR [95%CI}*	*OR [95%CI}*
Moderating effect of sex ^{m1}							
Men * NA	1.05 [0.861, 1.269]	0.943 [0.571, 1.518]	0.969 [0.696, 1.338]	1.301 [0.502, 3.174]	0.968 [0.622, 1.466]	1.008 [0.755, 1.344]	0.964 [0.727, 1.257]
Men * SI	0.948 [0.78, 1.155]	1.025 [0.66, 1.613]	1.006 [0.727, 1.395]	0.755 [0.298, 1.87]	1.165 [0.782, 1.753]	1.088 [0.839, 1.427]	1.023 [0.787, 1.326]
Men * NA2	**1.184 [1.026, 1.353]**	1.115 [0.842, 1.525]	1.109 [0.875, 1.39]	0.596 [0.289, 1.091]	0.857 [0.63, 1.144]	0.971 [0.801, 1.179]	0.985 [0.825, 1.188]
Men * SI^2	1.005 [0.852, 1.179]	0.811 [0.599, 1.099]	1.075 [0.826, 1.396]	0.673 [0.31, 1.453]	0.852 [0.619, 1.185]	0.925 [0.758, 1.138]	0.962 [0.787, 1.186]
Men * NA * SI	0.987 [0.837, 1.156]	0.984 [0.666, 1.426]	1.008 [0.756, 1.338]	1.707 [0.825, 3.538]	0.864 [0.578, 1.272]	0.923 [0.706, 1.191]	0.916 [0.702, 1.177]
Moderating effect of age ^{m2}							
Age * NA	1.02 [0.902, 1.15]	0.977 [0.772, 1.243]	0.873 [0.747, 1.027]	1.088 [0.702, 1.649]	0.944 [0.705, 1.334]	0.97 [0.833, 1.13]	0.993 [0.839, 1.175]
Age * SI	0.955 [0.849, 1.059]	0.991 [0.799, 1.233]	0.997 [0.848, 1.196]	0.964 [0.637, 1.46]	1.077 [0.87, 1.342]	0.968 [0.852, 1.115]	0.913 [0.799, 1.041]
Age * NA^2	1.028 [0.953, 1.112]	1.113 [0.97, 1.278]	1.077 [0.943, 1.234]	0.99 [0.729, 1.37]	1.052 [0.896, 1.241]	1.053 [0.946, 1.173]	1.039 [0.936, 1.151]
Age * SI^2	1.011 [0.935, 1.099]	1.105 [0.919, 1.336]	0.988 [0.864, 1.141]	1.063 [0.712, 1.605]	0.984 [0.811, 1.205]	1.035 [0.927, 1.164]	1.053 [0.937, 1.188]
Age * NA * SI	0.967 [0.878, 1.056]	0.952 [0.785, 1.171]	1.007 [0.86, 1.181]	1.001 [0.663, 1.532]	0.841 [0.679, 1.042]	0.96 [0.833, 1.112]	0.953 [0.812, 1.099]
Moderating effect of disease ^{m3}							
Disease * NA	1.001 [0.769, 1.319]	0.97 [0.605, 1.568]	1.036 [0.066, 15.269]	1.085 [0.059, 19.496]	1.105 [0.075, 18.703]	0.937 [0.575, 1.599]	0.959 [0.584, 1.522]
Disease * SI	1.127 [0.852, 1.534]	1.492 [0.82, 2.46]	1.062 [0.077, 16.481]	0.995 [0.064, 16.002]	1.002 [0.064, 15.805]	0.751 [0.449, 1.271]	0.759 [0.462, 1.213]
Disease * NA^2	0.916 [0.774, 1.075]	0.908 [0.6, 1.363]	1.037 [0.07, 15.358]	0.979 [0.052, 15.393]	0.959 [0.065, 13.495]	0.94 [0.671, 1.323]	0.912 [0.644, 1.259]
Disease * SI^2	0.924 [0.707, 1.132]	0.791 [0.532, 1.162]	1.039 [0.066, 15.071]	1.003 [0.062, 15.562]	1.028 [0.064, 16.064]	1.353 [0.952, 1.946]	1.278 [0.949, 1.739]
Disease * NA * SI	1.145 [0.909, 1.408]	1.18 [0.737, 1.873]	1.018 [0.061, 16.898]	1.005 [0.061, 16.547]	1.07 [0.07, 18.78]	1.079 [0.62, 1.825]	1.152 [0.724, 1.851]

*m1 = Age + Men + NA + SI + NA2 + SI2 + NA*SI + Men*NA + Men*SI + Men*NA2 + Men*SI2 + Men*NA*SI*

*m2 = Age + Men + NA + SI + NA2 + SI2 + NA*SI + Age*NA + Age*SI + Age*NA2 + Age*SI2 + Age*NA*SI*

*m3 = Age + Men + Disease + NA + SI + NA2 + SI2 + NA*SI + Disease*NA + Disease *SI + Disease*NA2 + Disease*SI2 + Disease*NA*SI*

Figure 2 visualizes the bivariate distribution of the NA and SI scores in each study. Across all studies, NA and SI were positively correlated (*r* = .373). ***Appendix H*** report the quality assessment of each included study. Although some studies were potentially more biased than others, most were at low risk of bias and none of the included studies showed a high risk of bias.

We used a Bayesian multilevel logistic regression analysis to estimate the Type D effects. The number of iterations of the MCMC procedure was sufficient to reach an effective sample size of at least 500 in the estimation of each model parameter. The R-hat value of each estimated regression coefficient was smaller than 1.05, indicating proper convergence (Kruschke, 2014). Table 2 shows for each endpoint the estimated odds ratios (including 95% Bayesian credible interval) of age, sex, the Type D effects according to the three operationalizations. Older age and male sex predicted the occurrence of all-cause mortality and cardiac mortality, but none of the other endpoints. Based on the continuous method, NA and SI showed a synergistic Type D effect on the occurrence of any adverse event during follow-up (OR=1.135, 95%CI=1.029, 1.253). Though the interaction model including quadratic effects also showed a synergistic Type D effect on MACE, when excluding the non-significant quadratic NA and SI effects from the continuous interaction model the 95% Bayesian credible interval contained an odds ratio of one, suggesting no effect (OR=1.126, 95%CI=0.99, 1.286).

In ***Appendix I***, Table I1 shows for each endpoint the standard deviation (including 95% Bayesian credible interval) of all random predictor effects according to the continuous method. The fact that many of these credible intervals did not include a standard deviation of zero suggests that these effects differ across studies, supporting our choice to model these parameters as random effects.

Table 3 presents the Bayes factor (BF) estimates according to the continuous method, expressing the evidential value for the presence or absence of a Type D effect on each endpoint for the complete sample and for CAD and HF patients separately. Evidence for a Type D effect in the complete sample was strong for the endpoint MACE (BF=40.1) and decisive for any adverse event (BF=99.0). Strong evidence *against* a Type D effect was found

for all-cause mortality (BF=47.18), cardiac mortality (BF=23.34) and myocardial infarction (BF=19.29). The evidence for a Type D effect on CABG and PCI was inconclusive, showing substantial evidential value both in favor and against a Type D effect. When limiting the sample to CAD patients, similar evidential values were found. For patients with HF, however, we found substantial to strong evidence against a Type D effect on all-cause mortality (BF=10.1), while for the other endpoints the evidence was either inconclusive or could not be estimated due to sparse data.

The results in Table 4 indicate that age and sex did not moderate the Type D effects in terms of the interaction between NA and SI on any of the studied endpoints. However, sex turned out to moderate the quadratic NA effect, indicating that increasingly high NA scores were associated with a higher odds on all-cause mortality and this effect was more pronounced for male than for female patients (OR=1.184, 95%CI = 1.026, 1.353). A Bayes factor of 89.9 indicated very strong evidence that the population odds ratio of this effect is larger than 1.

Figure 3 visualizes the Type D effects on each endpoint according to the continuous method estimates for the model including the NA and SI main effects and their interaction. For various standardized NA and SI scores, the figure shows the predicted posterior probability on the occurrence of each endpoint. The colored shades represent the 95% prediction intervals for each level of SI scores. The figure indicates the positive interaction effect between NA and SI on both MACE and any adverse events. The probability on the occurrence of these events during follow-up increased for higher NA scores and these positive effects became more pronounced for larger scores on SI. Similarly shaped curves, but smaller effects were found for CABG or PCI, though statistical evidence for these Type D effects was inconclusive. To facilitate interpretation of these figures, across the included datasets patients on averaged scored 9.02 on the NA (SD=6.33) and 9.20 on the SI (SD=6.01) measurements of the DS14. Based on these statistics, Figure 3 indicates that the probability on any adverse event during follow-up is 0.14 for patients with average NA and SI scores. For patients scoring two standard deviations above the average on NA (21.7), this risk increases to 0.20. For Type D patients, such as those who score two standard deviations above the average on both NA (21.7) and SI (21.3), the risk of an adverse event increases even further to 0.30. To facilitate the significant interaction effects between NA and SI on

any adverse events in cardiovascular disease patients, ***Appendix J*** reports for both NA and SI the simple slope analysis. The effect of SI on adverse events increases across higher NA scores and the 95%CI of the simple slopes starts to exclude a slope of zero (no effect) at NA scores of 13.8 or higher. The effect of NA on adverse events increases across higher SI scores and the 95%CI of the simple slope starts to exclude a slope of zero at SI scores of 6.2 or higher.

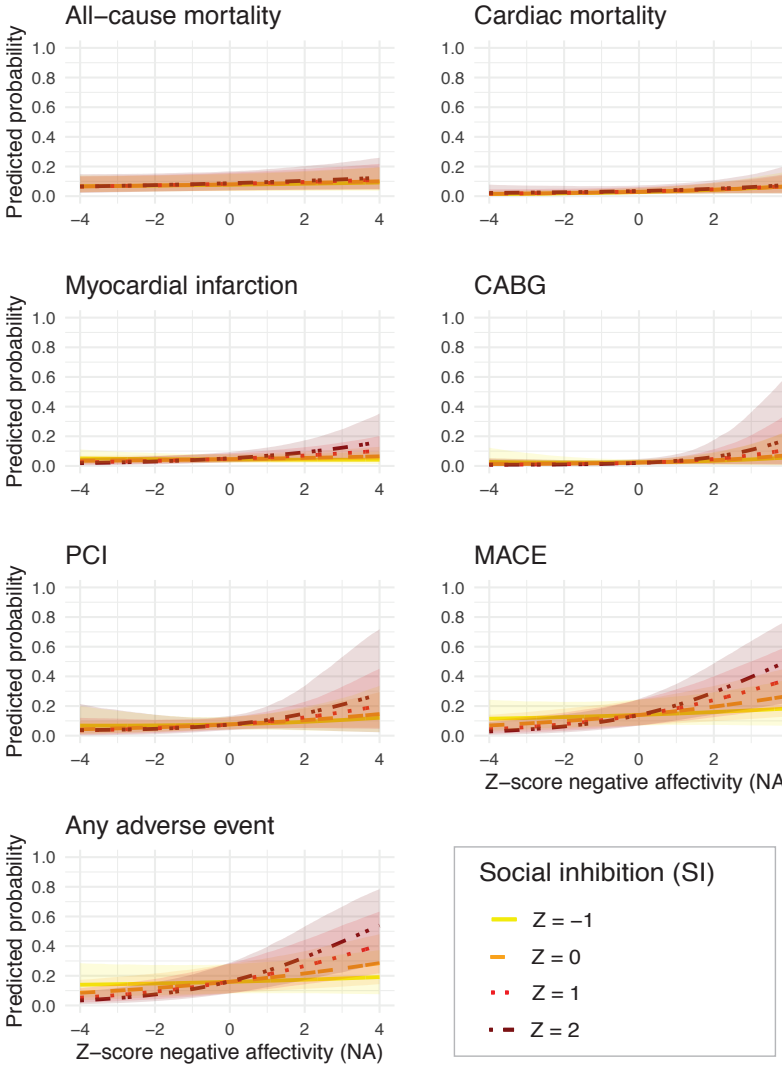
To facilitate interpretation of our model estimates, we have created an online tool (https://anonymousresearcher.shinyapps.io/AdverseEvent_Prediction_TypeD_CVD/) that uses the age, sex, negative affectivity and social inhibition scores and type of cardiovascular disease to calculate according to our model estimates the predicted probability on a particular outcome within the average follow-up time of our meta-analysis. For instance, for a 60-year-old male cardiovascular disease patient with a high NA score (20), the probability of having an adverse event within 48 months is 40.72% when the SI score is average (10), while the probability increases with 4% to 44.85% when the SI score is high (20).

As a sensitivity analysis, ***Appendix K*** shows for each endpoint a forest plot presenting the results of the two-step meta-analyses. These results are like those of the one-step meta-analysis, suggesting that Type D personality (operationalized according to the continuous method) was significantly associated with MACE and any adverse event, but not with any of the other endpoints. Table L1 in ***Appendix L*** presents the results of leave-one-out sensitivity analyses, that repeats the meta-analysis multiple times, each time with a different study left out. This sensitivity analysis shows that our findings were generally not driven by a single study, except that excluding one of the studies (Denollet et al., 2013b) attenuated the Type D effect on MACE, resulting in a Bayesian 95% credible interval that included the value of no effect (OR=1) and suggesting that the MACE effect is largely driven by that study.

Table L2 in ***Appendix L*** estimated the impact of prior distribution specification for the regression coefficients of the Type D effect according to the continuous method. The results show similar conclusions for each endpoint except MACE, with different prior distributions resulting in similarly sized Type D effects, yet slightly wider 95% credible intervals including an odds ratio of no effect, suggesting uncertainty regarding Type D's effect on MACE. Lastly,

Table L3 in ***Appendix L*** presents for each method to estimate the Type D effect a brier score, expressing the accuracy of predicting the observed endpoint based on the model estimates. For each method and outcome, the brier scores are close to zero, indicating high predictive accuracy.

Figure 3: predicted posterior probability on the occurrence of several endpoints during follow-up, given various scores on the standardized NA and SI scores.



DISCUSSION

We conducted an individual patient meta-analyses across 19 published prospective cohort studies investigating the prognostic effect of Type D personality in cardiovascular disease patients. We estimated the Type D effect according to the continuous interaction method, which performed best in several simulation studies. Bayes factors indicated strong evidence for the hypothesis that Type D predicts the occurrence of adverse events in patients with coronary artery disease. Simple slope analysis indicated that the influence of both NA and SI on any adverse event increased across higher scores on the other personality trait. Although Bayes factors indicated strong evidence for the Type D effect on MACE, various sensitivity analyses produced 95% credible intervals containing an odds ratio of one, suggesting that we should entertain the possibility of no Type D effect on MACE.

Evidence for a null effect was found for the outcomes all-cause mortality and cardiac mortality. The risk on those mortality endpoints increased with older age, male sex, and higher NA scores. A moderation of sex on a quadratic NA effect suggested that the higher NA scores increasingly resulted in a higher risk of all-cause mortality and this pattern was more pronounced for men in comparison to women. In the subset of patients with HF, there was slightly more evidence against a Type D effect on each studied endpoint, yet generally evidence for Type D's prognostic influence in HF patients remains inconclusive. Future research could investigate potential moderators of Type D's prognostic influence on adverse events in HF patients, for instance by comparing different etiologies (e.g., valvular or ischemic HF).

When interpreting the Type D effect on MACE and any adverse event, it is useful to inspect the effects on each of the MACE components. The Type D effects on CABG, PCI and MI are slightly smaller than the effects on MACE and based on both the Bayes factors and the 95% credible intervals we cannot exclude the possibility of a null effect. Nevertheless, the Type D effects on any of these individual outcomes point in the same direction and they may have become more noticeable when combined in a composite endpoint such as MACE or any adverse event. One could argue that endpoints such as the risk on MACE or any adverse

event are more interesting to patients than individual endpoints such as PCI or CABG, as those endpoints reflect a similar disease pathway while their occurrence also depends on more arbitrary factors such as healthcare availability or the location of atherosclerosis.

Our finding that Type D predicts adverse events in patients with CAD is in line with the conclusions drawn from earlier meta-analyses (Grande, Romppel & Barth, 2012; O'Dell et al., 2011) and a reanalysis of four of the earlier studies on this topic using the continuous method (Kupper & Denollet, 2016). However, our multilevel model indicated significant differences between studies in the estimate of this Type D effect. Our two-step meta-analysis reported in Appendix I reveals the studies that primarily drive this effect. The analysis indicated that all but two of the included studies showed positive estimates of the Type D effect on MACE, yet the effect appears to be predominantly driven by three studies (Denollet et al., 2006; 2013b; Martens et al., 2010). Indeed, our leave-one-out meta-analysis reported in Supplemental Table 5 showed that the Type D effect on MACE was no longer statistically significant when excluding one of those studies from the meta-analysis (Denollet et al., 2013b). This study involved a sample of 541 relatively young (M=58.7) and mostly male (87%) patients with CAD. According to the quality assessment there was no reason to exclude this study from our analysis. Nevertheless, our finding that the Type D effect on MACE depends primarily on this particular study raises doubt on the robustness of this effect. This uncertainty is corroborated by two other observations in our statistical analysis. First, the continuous interaction model excluding the quadratic NA and SI effects no longer showed a significant interaction between NA and SI on MACE. Second, even when including those quadratic effects in the model, the 95% credible interval for the interaction between NA and SI on MACE contained one when using a flat instead of normally distributed prior for the regression coefficients. Altogether, these observations suggest that there is still uncertainty regarding the effect of Type D on MACE. Nevertheless, our various sensitivity analyses all suggest an association between Type D personality and adverse events in cardiovascular disease patients.

Our finding that not Type D personality, but only NA was associated with both all-cause and cardiac mortality contrasts with the conclusion of an earlier published meta-analysis [1]. This discrepancy is likely explained by the fact that this previous meta-analysis included Type D

effects estimated according to the 2-group method. As this method is not able to distinguish Type D effects from effects of NA or SI only (***Chapters 2 and 3***), meta-analyses including such effects have the same limitation. Previous research estimated that approximately half of all published Type D effects according to the 2-group method were effects of NA or SI only according to the continuous method (***Chapter 4***). Supplemental Figures 1 and 2 show that only one of the currently included studies showed a statistically significant Type D effect on all-cause and cardiac mortality according to the continuous method, while the earlier published meta-analysis included many studies with significant effects according to the 2-group method (Grande et al., 2012). The current study suggests that many of these earlier studies showing a link between Type D personality and mortality endpoints were in fact effects of NA only. Indeed, studies using the continuous method to estimate the Type D effect have shown that only NA was associated with various outcomes, such as in-stent neoatherosclerosis (Lee et al., 2019), coronary lipid plaque (Wang et al., 2016), and medication adherence (Wu & Moser, 2014). Future research should use individual patient-data meta-analyses to test whether these findings are confirmed when aggregating across multiple studies.

The absence of a moderation by age of the Type D effect on MACE contrasts with a previous analysis of several published studies showing that Type D only predicted MACE in CAD patients if they were younger than 70 years old (Kupper & Denollet, 2016). Our moderation analysis also found no evidence that the Type D effect on any outcome differs across the type of cardiovascular disease. However, the confidence intervals for these moderations by disease were very wide, suggesting considerable uncertainty in these estimates. Indeed, the subgroup analyses reported in Table 3 show that the Type D effects in CAD patients are similar to those in the full sample, yet much uncertainty remains regarding the effects in HF patients. Sex did not moderate the Type D effect on any outcome, yet moderated a quadratic NA effect on all-cause mortality, suggesting that this quadratic effect differs between the sexes. The prediction model in our shiny app reveals the risk on all-cause mortality increases quadratically with higher NA scores for male CVD patients, while females do not show such an NA effect. This finding resonates with earlier research showing that negative mood episodes such as depression increase the mortality risk more in males than females (Gilman et al., 2017).

Our data only allowed adjusting the Type D effects for age and sex. It therefore remains unclear whether the Type D effect on adverse events is confounded by other risk factors, such as lifestyle or depressive symptoms. Alternatively, these risk factors may also signify increased disease progression, and therefore not confound but rather mediate or explain the association between Type D personality and adverse events. Given the high correlation between NA and depressive symptoms, depression may have confounded or mediated the Type D effects found in our study. Similarly, we were also not able to control for other potential physical or mental morbidities that could produce both an increase in for instance both NA and the risk on adverse events. For these reasons, our findings do not support a *causal* influence of Type D personality on adverse events. On the other hand, the studies included in our analysis that showed the largest effects of Type D on adverse events (Denollet et al., 2006; Denollet et al., 2013b) did adjust their analyses for confounders such as decreased systolic function / LVEF, exercise tolerance, and psychological stress. Nevertheless, future research could perform a highly powered preregistered investigation into the added predictive value of Type D personality on adverse events in cardiovascular disease patients above and beyond the effect of depression and other clinical risk factors, while modeling Type D personality according to the continuous interaction method.

Should such a high-powered preregistered analysis detect a Type D effect on adverse events, then subsequent research could shed more light on the biological pathways underlying this association. Although in earlier work Type D personality has been associated with impaired endothelial function (Denollet et al., 2018), subclinical inflammation (van Dooren et al., 2016) and various inflammatory biomarkers (Conraads et al., 2005; Denollet et al., 2009), these analyses were based on the biased personality group methods. Future work should therefore reanalyze these studies using the continuous method to find out whether these effects were truly driven by Type D personality, or by an effect of NA or SI only. Recent work using the continuous method showed that Type D is associated with higher levels of coronary artery calcification, after adjusting for many known CAD risk factors such as depression, smoking, diabetes and hypertension (Raykh et al., 2020). Coronary artery calcification is itself related to an increased risk of adverse cardiac events, and an unhealthy lifestyle could explain why some individuals develop high coronary artery calcification levels (Liu et al., 2015). Type D personality has been associated with less regular physical exercise

(Bunevicius et al., 2014), a less healthy diet (Booth & Williams, 2015), and poor self-management (Kessing et al., 2017). Therefore, future research could focus on testing the role of an unhealthy lifestyle as a possible behavioral pathway mediating Type D's effect on coronary artery calcification and other indicators of heart disease (Kupper & Denollet, 2018).

One clinical implication of our finding is that interventions to reduce mortality risk in cardiovascular disease patients should mainly target NA, because elevated SI does not confer additional risk. Given the close relation between NA and other negative mood episodes such as depression, it may therefore be worthwhile to treat these CVD patients with interventions that are effective in reducing depressive symptoms. Although a randomized controlled trial found no benefit of stepwise psychotherapy in reducing depressive symptoms in CAD patients, a subgroup analysis revealed the intervention was more effective in those with Type D personality than in those without Type D personality (Hermann-Lingen et al., 2016). For preventing adverse events in CVD patients, it may be worthwhile to besides NA additionally intervene on SI. Acceptance and commitment therapy (ACT) could allow those with high SI to improve their emotion regulation skills (Forman et al., 2007). Although SI is generally considered a temporally stable personality trait, when individuals show increased SI due to traumatic interpersonal experiences, then targeting such experiences may potentially reduce SI and thereby its increased risk on adverse events in those with high NA.

Strengths and limitations

Strengths of the current research are the large sample size (N=11151), the Bayesian estimation approach (allowing quantification of the evidential value for both the null and alternative hypotheses), the sensitivity analysis (one-step vs. two-step individual patient data meta-analysis), and the various contrasted Type D operationalizations (2-group vs. 4-group vs. continuous method) confirming previous work that the 2-group and 4-group methods cannot distinguish synergistic Type D effects, from effects of NA or SI only (***Chapters 2 & 3***).

Despite these strengths, our study also has several limitations. First, the cardiac mortality endpoint may be unreliable because identifying the cause of mortality can be difficult,

particularly in elderly multimorbid patients. Second, we did not have sufficient data to adjust our estimate of the Type D effect for earlier received treatments or non-cardiac somatic and psychiatric diagnoses. This raises the question of whether baseline NA or SI measurements were influenced by disease or treatment related factors. Nevertheless, some of the studies included in this meta-analysis found significant Type D effects after controlling for a history of cardiac events such as CABG, PCI, or MI (Denollet et al., 2006; 2013b; Martens et al., 2010).

Third, 7 of the 25 identified eligible studies could not be included either due to non-response or the reluctance of sharing the raw data. This resulted in excluding the potential data of 1457 patients with HF and 1035 patients with CAD. Although our analyses still involved 2027 patients with HF and 8096 patients with CAD, it was not possible to estimate a Type D effect for some endpoints in patients with HF due to sparse data. As a result, it remains unclear whether Type D is associated with an increased risk on MI, CABG and PCI in patients with HF. Of the seven excluded studies, two out of three studies in heart failure patients showed a significant association between Type D personality and mortality using the 2-group method (Bundgaard et al., 2019; Denollet et al., 2007; Volz et al., 2011). The four remaining studies focused on CAD patients, three of which used the 2-group method to show that Type D personality was associated with MACE (Du et al., 2016; Imbalzano et al, 2018; Leu et al., 2019), while one study indicated that a cluster with CAD patients scoring high on Type D personality had an increased risk of all-cause mortality during follow-up than other patient clusters (Modica et al., 2012). None of these seven studies used the continuous method to estimate Type D effects, leaving it unclear whether Type D personality was driving these effects. This is likely only true for some of these studies, given that approximately half of the studies with significant Type D effects based on the 2-group method are effects of NA or SI only according to the continuous method (***Chapter 4***).

Another limitation is that we did not include unpublished studies. Although one earlier meta-analysis did not find evidence of publication bias in the sample of studies investigating the MACE endpoint (O'Dell et al., 2011), another indicated that studies with smaller sample sizes showed larger Type D effects than studies with larger sample sizes, possibly hinting at publication bias (Grande et al., 2012). Should it be the case that there exist unpublished

studies investigating the risk of Type D on adverse events in cardiovascular disease patients, and that those studies differ from published studies in their effect sizes, then publication bias may have affected our conclusions.

Our meta-analysis was applied to total NA and SI scores because individual item scores were no longer available for various studies included in our analysis. Therefore, we were not able to conduct item-level analyses, testing whether specific combinations of NA and SI items interact in predicting adverse events. We recommend researchers in future studies on Type D personality to test item-level interaction effects to investigate which items primarily drive a potential significant interaction effect between NA and SI. Due to the lack of individual item scores, we were not able to conduct an IRT-based measurement harmonization to link the differently sized DS14 and DS16 scales. As a workaround we standardized the NA and SI total scores to the same z-score metric. We were also not able to determine whether the measurement instruments showed signs of differential item functioning across the included studies. Nevertheless, previous research using item response theory has shown that the DS14 instrument provides fairly comparable measurements across the general and clinical populations (Emons, Meijer & Denollet, 2007). Future research could investigate this measurement invariance across other factors such as age, sex, or type of cardiovascular disease.

Conclusion

In light of recent findings that a major part of the published Type D effects may be false positives masquerading for effects of NA or SI only (***Chapters 2, 3, and 4***) our study is a first endeavor at a large scale reanalysis of the published Type D literature. Using the continuous method, our reanalysis suggests that some of the earlier published Type D effects on all-cause and cardiac mortality (Denollet, Sys & Brutsaert, 1995; Denollet et al., 1996) are likely effects of NA only. Nevertheless, based on this individual patient data meta-analysis of 19 published prospective cohort studies, Type D personality poses an increased risk on the occurrence of adverse events in patients suffering from coronary artery disease.

PART III

A latent variable model of Type D personality

CHAPTER 6

Modeling interactions between latent variables in research on Type D personality: A Monte Carlo simulation and clinical study of depression and anxiety

P. LODDER
J. DENOLLET
W.H.M. EMONS
G. NEFS
F. POUWER
J. SPEIGHT
J.M. WICHERTS

Multivariate Behavioral Research

ABSTRACT

Introduction: Several approaches exist to model interactions between latent variables. However, it is unclear how these perform when item scores are skewed and ordinal. Research on Type D personality serves as a good case study for that matter.

Methods: In Study 1, we fitted a multivariate interaction model to predict depression and anxiety with Type D personality, operationalized as an interaction between its two subcomponents negative affectivity (NA) and social inhibition (SI). We constructed this interaction according to four approaches: (1) Sum score product; (2) Single product indicator; (3) Matched product indicators; (4) Latent Moderated Structural equations (LMS). In Study 2, we compared these interaction models in a simulation study by assessing for each method the bias and precision of the estimated interaction effect under varying conditions.

Results: In Study 1, all methods showed a significant Type D effect on both depression and anxiety, although this effect diminished after including the NA and SI quadratic effects. Study 2 showed that the LMS approach performed best with respect to minimizing bias and maximizing power, even when item scores were ordinal and skewed. However, when latent traits were skewed LMS resulted in more false positive conclusions, while the Matched PI approach adequately controlled the false positive rate.

Conclusion: LMS was the least biased method to estimate interactions between latent variables. Using this method, we found a latent interaction between NA and SI on depression and anxiety.

INTRODUCTION

In the social and behavioral sciences, researchers commonly investigate the effect of an interaction between two predictors on an outcome variable. Traditionally, such interaction effects have been analyzed by including the product of the sum scores of two interaction constructs in a standard regression analysis. However, the presence of measurement error in the predictor variables can lead to biased estimates of the regression coefficients, especially for interactions between constructs both measured with error (Busemeyer & Jones, 1983; Cole & Preacher, 2014; Embretson, 1996; Kang & Waller, 2005; MacCallum, Zhang, Preacher, & Rucker, 2002). Although latent variable modeling can be used to take into account this measurement error, there is no consensus on how to best model interactions in this context, especially when the item scores are of an ordinal nature and not normally distributed. In this article, we investigate this issue based on a Monte Carlo simulation study and an empirical application.

The construct of Type D personality (Denollet, 2005) serves as a great case study for this matter, because according to some authors (Smith, 2011), the Type D effect is hypothesized to constitute a statistical interaction between the construct's two subcomponents, which are both measured by items on an ordinal scale with skewed response distributions. In this article, we will first give an overview of research on Type D personality and of the various methods used to handle non-normal ordinal data and to investigate interaction effects. As an illustration we will study the relation between Type D personality, depression, and anxiety according to several statistical interaction models. As will become clear, such complex statistical modeling involves making choices on several matters, including the estimation methods, the interaction model, and the techniques used to handle non-normal and ordinal data. Therefore, in line with recommendations by Muthén and Muthén (2002) and Steiger (2006a; 2006b), the second part of this article presents the results of a Monte Carlo simulation study where we examine the influence of these modeling choices on the bias and accuracy of estimated interaction effects.

Type D Personality

People with a Type D personality have a tendency to: (1) experience negative emotions (i.e., negative affectivity) and (2) inhibit the expression of their behavior and emotions in social interactions (i.e., social inhibition). Type D personality has been associated with a poor prognosis of cardiovascular disease (for two meta-analyses see: Grande, Romppel & Barth, 2012; O'Dell, Masters, Spielmans & Maisto, 2011) as well as with emotional factors such as anxiety and depression (Pedersen, van Domburg, Theuns, Jordaens, & Erdman, 2004; Schiffer et al., 2005; Nefs et al., 2015). People experiencing depression or anxiety symptoms also show an increased risk of developing cardiovascular diseases (Frasure-Smith & Lespérance, 2008; Kubzansky, Cole, Kawachi, Vokonas & Sparrow, 2006; Roest, Martens, de Jonge & Denollet, 2010; Wulsin & Singal, 2003). The Type D subcomponent Social inhibition (SI) was proposed to moderate the effect of the other subcomponent negative affectivity (NA) on health problems (Kupper & Denollet, 2007; Kupper & Denollet, 2016). With respect to cardiovascular disease, this implies that the negative association between NA and cardiovascular health is stronger for people who score highly on SI. According to Smith (2011) and other authors, this implies that the association between Type D and health seems to constitute a statistical interaction between subcomponents NA and SI, rather than the separate additive effects of NA and SI.

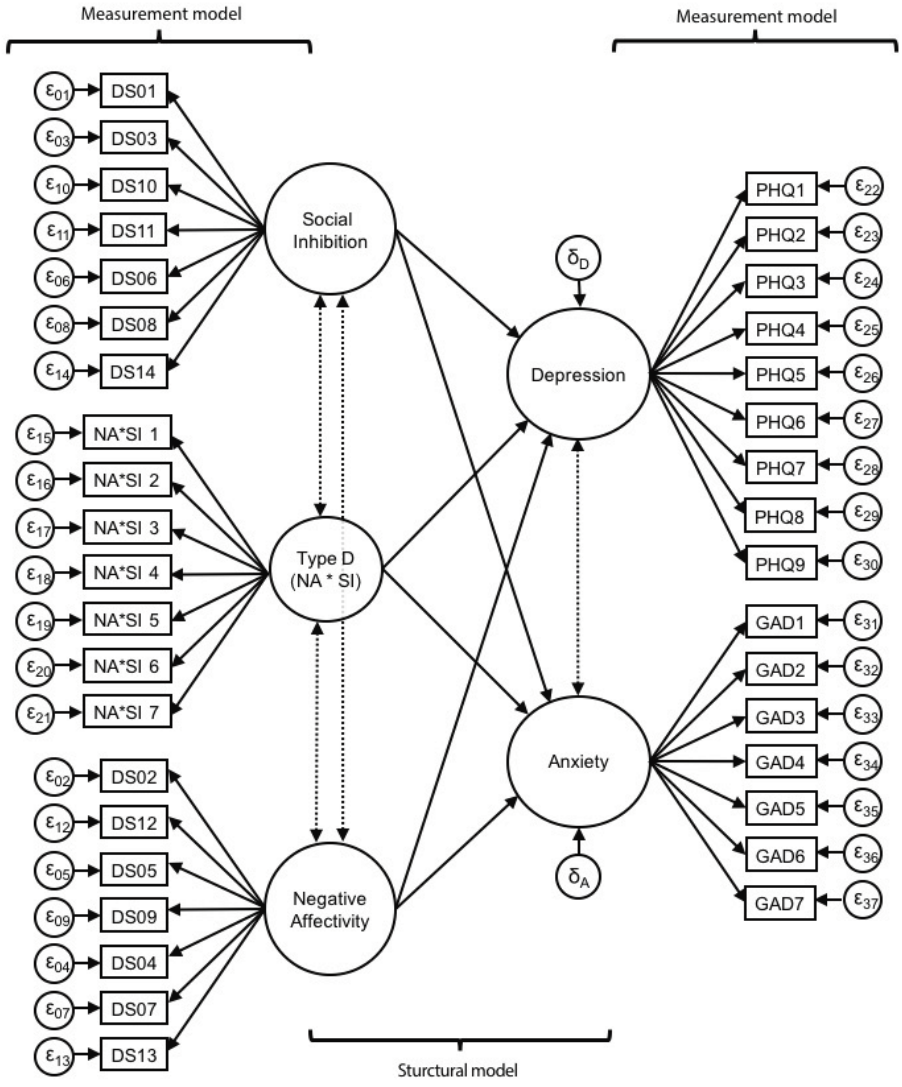
While some studies testing the interaction between NA and SI on health outcomes demonstrated a significant interaction (Denollet, Pedersen, Vrints, & Conraads, 2013 [small effect: Odds Ratio (OR)=1.36]; Kupper, Denollet, Widdershoven, & Kop, 2013 [small effect: partial h^2=.04]; Whitehead, Perkins-Porras, Strike, Magid, & Steptoe, 2007 [small to medium effect: r = .30]), others failed to demonstrate such an effect (Coyne et al., 2011 [trivial effect; Hazard Ratio (HR)=.90]; Grande et al., 2011 [trivial effect; HR=1.01]; Pelle et al., 2010 [trivial effect; HR=1.16]). In these studies, researchers assessed the interaction effect by including both the product of the NA and SI sum scores as well as the separate NA and SI sum scores as predictors in linear or logistic regression analysis. Throughout this paper, we define a sum score as the unweighted sum of the scores on individual items that measure a particular construct. We note that the unweighted sum score is a linear transformation of the mean score of a set of items.

Although linear regression using sum scores is predominantly used in practice to test interactions, it has several disadvantages (Busemeyer & Jones, 1983). First, spurious interactions can arise when the relation between a latent variable and its observed variables is non-linear, for instance when observed variables are measured on an ordinal scale (Busemeyer & Jones, 1983; Embretson, 1996; Kang & Waller, 2005; MacCallum, Zhang, Preacher, & Rucker, 2002). A second disadvantage of regression using sum scores is that the model does not account for measurement errors. Responses to questionnaire items contain measurement errors and in subsequent analyses these may attenuate associations with other variables (Spearman, 1904). This attenuation bias is particularly problematic in studies that investigate an interaction between two variables that are both subjected to random measurement errors. Indeed, when multiplying two sum scores having measurement errors the resulting product variable contains even more measurement error than the two separate sum scores, because the parts of the sum scores that contain measurement error also are multiplied. Thus, measurements of interaction effects are less reliable than those of main effects. As a result, the true strength of the interaction is underestimated. Conversely, ignoring measurement error not only increases the risk of attenuation bias, in some complex models it can also result in overestimated associations (Cole & Preacher, 2014).

An alternative approach for testing interactions between imperfectly measured psychological attributes is latent variable modeling (e.g., Skrondall & Rabe-Hesketh, 2004). A latent variable is not observed, but rather is believed to underlie a set of observed variables that are indicators of the construct of interest (e.g., depression; Borsboom, Mellenbergh & Van Heerden, 2003). For example, in a latent variable model for depression, the latent construct ‘depression’ is assumed to be reflected by the observed item scores of a depression questionnaire. According to one interpretation of latent variable theory, (co)variation in the item scores is caused by variation in people’s position on the latent variable (Borsboom et al., 2003). This interpretation is closely connected to the local independence assumption, which states that the observed item scores are conditionally independent upon the latent variables scores (Hambleton, Swaminathan & Rogers, 1991). Regarding the latent construct *depression*, this implies that when holding the latent depression scores constant, the item scores on the depression questionnaire become statistically independent. It is possible to relax this assumption in confirmatory factor

analysis models, by freely estimating the covariance between the error terms in the measurement model.

Figure 1: Latent prediction model of Type D personality, depression, and anxiety according to the matched product indicator approach. Circles correspond to latent variables; rectangles to observed variables; e to measurement error; d to prediction error; dashed lines to latent covariances and curved lines to residual covariances.



In the present study, we adopted the following terminology regarding the various aspects of latent variable models. In a latent variable model, the *measurement model* specifies the relation between a latent variable and a set of observed item scores intended to measure that latent variable. In the *structural model,* the relations between latent variables are modelled. Finally, the *latent prediction model* is the complete model of interest, including both the structural model and the measurement models for each latent variable. Figure 1 visualizes a latent prediction model of Type D personality, depression, and anxiety. In this figure, circles correspond to latent variables, rectangles to observed item scores, e to the measurement error of an item, and d to the residual error of a latent prediction effect. Dashed lines refer to covariances between latent variables and curved lines to residual covariances.

To our knowledge, the relation between Type D personality, depression, and anxiety has never been investigated while taking into account measurement error and non-normally distributed item scores. Therefore, the aim of the present study was to tackle these problems in two separate studies. In Study 1, we aimed to provide less biased estimates of the association between Type D, depression and anxiety by applying a latent prediction model to a dataset of people with diabetes. In Study 2, we aimed to investigate the accuracy and stability of the estimated Type D effect in Study 1, by conducting a simulation study. Building such a latent prediction model, however, requires modeling the latent interaction. Therefore, in the next paragraph we discuss some issues related to building latent interaction models based on skewed and ordinal data.

Assessing Latent Variable Interactions

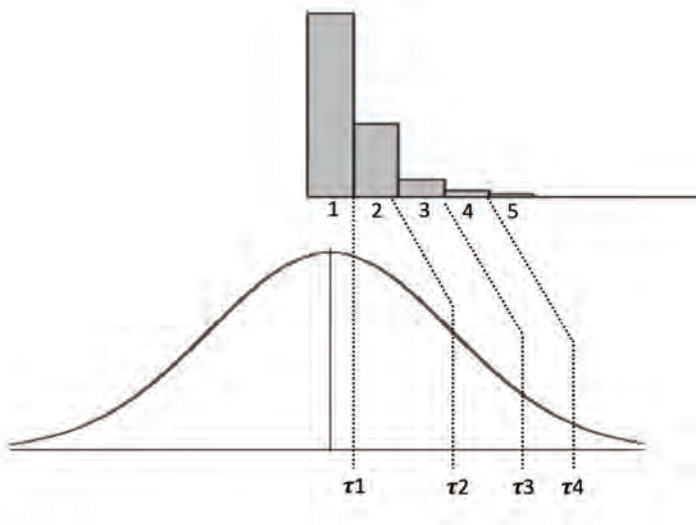
How to handle skewed ordinal item scores?

The Type D personality, depression, and anxiety questionnaires contain items that are all measured on an ordinal scale. The item score distributions also show substantial positive skewness, as is common for clinical data (Reise & Waller, 2009). For instance, most clinical questionnaires ask people to indicate how often they experience a symptom, with response indicated on a five-point Likert scale. Especially in the general population, such questionnaires usually show an overabundance of low scores indicating that people do not

frequently experience these symptoms. Positively skewed distributions pose a problem to estimation methods that assume the data to be normally distributed.

By default, most statistical packages use maximum likelihood (ML) to estimate the parameters in a latent prediction model. A standard assumption of ML estimation applied to factor analysis is that the observed item scores are continuous and normally distributed (Bollen, 1989, pp. 131–134). However, in practice, factor analysis is frequently used in conjunction with ML estimation to model ordinal questionnaire data (ten Holt, van Duijn, & Boomsma, 2010). Treating the distributions of items as normal and continuous while they are in fact skewed and ordinal can lead to underestimated standard errors of the parameter estimates and hence to more false positive conclusions on the significance of these estimates, especially if there are five of fewer response categories (Dolan, 1994; Muthén & Kaplan, 1985; Kline, 2010; Rhemtulla, Brosseau-Liard & Savalei, 2012). Moreover, if item response variables have an ordinal measurement level, then those items are likely not linearly associated, even when they are fairly normally distributed. Such linearity is assumed by traditional factor analytic models that are based on product-moment correlations (Flora, LaBrish & Chalmers, 2012).

Figure 2: Ordinal item score distribution (top) and the assumed underlying continuous distribution (bottom), connected by a set of polychoric threshold parameters (t_k).



Multiple alternative methods have been proposed to handle ordinal item scores in latent variable models. For example, researchers can use item response models that do not assume that the item scores are normally distributed (Rasch, 1960; Birnbaum, 1968; Samejima, 1997). If it is important to stay within a structural equation modeling framework, then research can for instance use ML estimation with robust standard errors and a robust test statistic. This robustness may result from bootstrapping or jack-knife techniques (Bollen & Stine, 1993), a Satorra-Bentler correction (Satorra & Bentler, 1988) or a sandwich estimator (Yuan & Bentler, 2000).

Another method that explicitly models the ordinal item scores makes use of polychoric correlations. Such correlations are estimates of the relation between two ordinal variables. This method assumes that each ordinal variable has an underlying continuous latent variable and estimates the correlation between those underlying latent variables. The two latent variables are assumed to show a bivariate normal distribution. Figure 2 illustrates the relation between such an ordinal variable and its assumed underlying continuous dimension. For each item, the boundary between each ordinal response category is connected to a latent continuous distribution through a set of threshold parameters (t_k, where *k* equals the number of response categories minus one). Each threshold marks a point on a latent indicator scale where values above and below the threshold correspond to different responses on the ordinal item. Latent scores between minus infinity and the value of the first threshold correspond to the lowest ordinal score. Continuous scores between the first and second threshold correspond to the second lowest ordinal score, etcetera.

The estimated threshold parameters are subsequently combined with the information in the bivariate contingency table of two ordinal variables to estimate (using maximum likelihood) the correlation between the two underlying latent variables when they would have been observed directly (Flora & Curran, 2004). Applying this method to handle the ordinal item scores in latent prediction models first requires the matrix of polychoric correlations between all items in the model. Next, the latent prediction model is fitted to this polychoric correlation matrix. Flora and Curran (2004) suggested to use a weighted least squares (WLS) estimation procecure (Browne, 1984) to estimate the model parameters, because using ML estimation results in biased test statistics and standard errors (Babakus, Ferguson &

Jöreskog, 1987; Dolan, 1994). This WLS method uses the asymptotic variances and covariances of these polychoric correlations to estimate a weight matrix. This matrix is subsequently used in the WLS fit function to weigh the squared difference between the sample statistics and the model-implied population parameters (Muthen, 1984). Given that this method takes into account the skewness and kurtosis of the raw data, it is not necessary to specify distributional assumptions for the observed variables, making WLS an asymptotically distribution free estimator (Browne, 1984). A disadvantage of this method is that this asymptotic covariance matrix quickly gets larger as the number of observed variables increases, which can result in computational problems. Furthermore, it is well established that WLS requires very large samples and under small sample sizes this method might produce inflated chi-square statistics (Dolan, 1994) and negatively biased standard errors (Potthast, 1993). In such a scenario, it is recommended to use unweighted least-squares (ULS) or Diagonally weighted least-squares (DWLS or Robust WLS) because these methods do not suffer from these limitations (Flora & Curran, 2004; Flora, LaBrish & Chalmers, 2012).

An assumption of the polychoric correlation method is that the observed bivariate distributions between ordinal indicators can be explained from underlying bivariate normally distributed continuous variables (Olsson, 1979). This shifts the distributional assumption from the observed scores to the latent indicator level. Even then, this bivariate normality assumption is under some circumstances not necessary. Indeed, WLS estimation of a CFA based on the polychoric correlation matrix is robust to moderate violations of this normality assumption (Flora & Curran, 2004). Furthermore, Muthén and Kaplan (1985) showed that WLS estimation in general is robust to non-normality when sample size is larger than 1000. A final advantage of this WLS method based on polychoric correlations is that it is much more efficient than full information ML estimation, especially when modeling multiple correlated traits (Forero & Maydeu-Olivares, 2009). Given these advantages, we aim to test the fit of our latent prediction model to the polychoric correlation matrix and estimate our parameters with DWLS estimation. This method results in a less parsimonious model because it requires estimation of additional threshold parameters for each item. We will therefore also investigate whether ML estimation with the Satorra-Bentler correction for robust standard errors (MLR estimation) works equally well.

How to construct a latent interaction model?

To investigate the effect of Type D personality on depression and anxiety within a latent prediction model, one has to decide on how to extract information from the observed scores to draw inferences about the interaction between the NA and SI traits at a latent level. In methods based on sum scores, the interaction variable results from a multiplication of the two sum scores. An alternative is to multiply – rather than sum – item scores, resulting in one or more multiplied item pairs serving as indicators of a latent interaction construct.

We can test latent interaction effects according to multiple methods. When the *indicant product* approach is utilized, all items of the first construct are multiplied with all possible combinations of items of the second construct (Kenny & Judd, 1984). With an increasing number of items that measure each construct, this method quickly results in a very large number of indicators of the latent interaction variable. For example, both NA and SI are measured with seven items by the 14-item Type D Scale (DS14; Denollet, 2005), resulting in a total of 49 (7 multiplied by 7) additional indicators. These new indicator variables share a lot of information because they are all based on the same collection of observed variables. This large number of overlapping indicators might result in convergence problems and is computationally very demanding (Ping, 1995). The method also requires adding complex parameter constraints to the model. Therefore, it is preferable to use methods that both require a smaller number of indicator variables and that do not require complex parameter constraints.

Marsh, Wen & Hau (2004) proposed an unconstrained method to model latent interaction effects using less indicator variables than the indicant product approach. Two examples of such unconstrained approaches are the *single product indicator (Single PI)* approach and the *matched product indicator (Matched PI)* approach. These methods differ with respect to the number of new indicators loading on the latent interaction construct. The Single PI approach uses a Single PI, while the Matched PI uses two or more new product indicators. Each indicator is the result of a multiplication of two items – one of each construct. Items sets can either be chosen at random or based on the ranking of the standardized factor loadings. For example, first the items are ranked according to their reliability based on the standardized factor loadings, and then the product of items with a similar standardized factor loading

ranking is computed (Marsh, Wen & Hau, 2004). In a simulation study, Marsh and colleagues (2004) showed that the Matched PI approach results in the same amount of bias in the estimated interaction as the indicant product approach, while the Single PI approach is more biased than both other approaches. Of the two approaches with the least amount of bias, the Matched PI approach is preferable over the indicant product approach, because it is easier to implement, does not require complex constraints and is computationally less demanding.

Another way to model latent interactions is the Latent Moderated Structural Equations approach (LMS; Klein and Moosbrugger, 2000; see also Maslowsky, Jager & Hemken, 2015). Compared to the Single PI and Matched PI approaches, LMS does not multiply item scores, nor does it model the interaction term as a latent variable. Instead, LMS directly takes into account the non-normality of non-linear effects by representing the joint distribution of all indicator variables as a mixture of normal distributions. Interactions are inherently non-linear and tend to have non-normal distributions, even when the latent variables constituting the interaction are themselves normally distributed. LMS assumes that the indicators of the exogenous latent variables conform to a multivariate normal distribution. Given the non-normally distributed item scores of the Type D questionnaire, we expect in line with earlier research (Kelava & Nagengast, 2012; Kelava, Nagengast & Brandt, 2014), that the LMS approach shows bias in the estimation of the latent interaction. Furthermore, we expect the Single PI and Matched PI approach to show less bias than the LMS method because earlier research showed that these approaches, especially the Matched PI approach, were less biased than LMS when the data are not normally distributed (Marsh et al, 2004; Cham, West, Ma & Aiken, 2012).

Study overview

In Study 1, we used a latent prediction model to investigate the association of Type D personality with depression and anxiety. Type D personality was modeled as an interaction between its components negative affectivity and social inhibition (Smith, 2011). All constructs are positively skewed and measured with multiple items measured on an ordinal scale. Given that earlier simulation studies on latent interaction modeling did not investigate the performance of the Single PI, Matched PI and LMS approaches under these specific

circumstances, we applied all three approaches and compared their performance in a simulation study. Therefore, in Study 2, we performed a Monte Carlo simulation to investigate to what extent the models used in Study 1 provided accurate and stable parameter estimates given the specific characteristics of Study 1 (i.e., large sample size, large positive skewness, and ordinal item scores).

STUDY 1: LATENT PREDICTION MODEL IN ADULTS WITH DIABETES

METHOD

Participants

Data were used from the Diabetes MILES study (Nefs, Bot, Browne, Speight & Pouwer, 2012), containing a sample of 3,314 Dutch adults with type 1 or 2 diabetes. The psychological research ethics committee of Tilburg University approved of the study protocol (EC-2011 5). All participants gave informed consent.

Measures

DS14

The traits underlying Type D personality (NA and SI) were measured using the DS14 questionnaire (Denollet, 2005). Each trait was measured with a scale consisting of seven questions with five ordinal response categories ranging from “false” (0) to “true” (4). The DS14 has been validated in several populations (Denollet, 2005). In our sample the coefficient alpha of NA and SI were 0.89 and 0.90, respectively.

PHQ-9

Depressive symptoms were measured using the nine-item Patient Health Questionnaire (PHQ-9; Kroenke, Spitzer & Williams, 2001), with each item having four ordinal response categories ranging from “not at all” (0) to “nearly every day” (3). The PHQ-9 has been validated in both the general- and the diabetes population (Martin, Rief, Klaiberg & Braehle,

2006; van Steenbergen-Weijenburg et al., 2010). In our sample, the coefficient alpha of the PHQ-9 was 0.86.

GAD-7

Anxiety symptoms were measured using the seven-item General Anxiety Disorder questionnaire (GAD-7; Spitzer, Kroenke, Williams & Löwe, 2006), with each item having four ordinal response categories, ranging from "not at all" (0) to "nearly every day" (3). The GAD-7 has been validated in the general population (Löwe et al., 2008). In our sample the coefficient alpha of the GAD-7 was 0.90. The item correlation matrices and skewness and kurtosis estimates are reported in ***Appendix M*** (DS14) and ***Appendix N*** (PHQ-9 and GAD-7).

Software

We performed all analyses in the freely available R programming software (Version 3.2.3; R development Core Team, 2008) and used Mplus software (Version 8; Muthén & Muthén, 1998-2010) to build our latent interaction models. The Mplus syntax files are available at this project's open science framework page: https://osf.io/kf6d5/.

Model Building

To identify our model, we fixed the first factor loading of each latent trait to a value of one. The latent prediction model was built in several steps. We first created separate measurement models for Type D personality, depression, and anxiety. These measurement models were then used to evaluate whether the data fitted a predetermined factor structure. Both the depression and anxiety questionnaires should exhibit a one-factor structure (Kroenke, Spitzer & Williams, 2001; Spitzer et al., 2006), whereas the Type D questionnaire should show a two-factor structure.

In the next step, we connected these measurement models by including a structural model. We took a multivariate approach by first regressing depression and anxiety on both NA and SI, thereby investigating the main effects of Type D on depression and anxiety. Subsequently, we added the latent interaction between NA and SI, thereby testing whether the interaction between NA and SI explains any additional variance in depression or anxiety above and beyond the variance explained by the additive effects of NA and SI.

Our model selection procedure involved the comparison of three nested models. The first model is a baseline model where all regressions between latent constructs were fixed to zero. This model served as a reference model against which we compared the fit of our second model. In Model 2, we estimated the main effects of NA and SI on both depression and anxiety. Finally, in Model 3 we included the interaction between NA and SI on both depression and anxiety. Lastly, we tested the quadratic effects of both NA and SI to determine whether a possible interaction effect was merely picking up an unmodeled quadratic effect rather than a true interaction (MacCallum & Mar, 1995).

Interaction models

We assessed the interaction between NA and SI according to six different methods: (1) Regression of sum scores, (2) LMS using robust maximum likelihood (MLR) estimation, (3) Matched PI using MLR estimation, (4) Single PI using MLR estimation, (5) Matched PI based on the polychoric correlation matrix using diagonally weighted least squares (DWLS) estimation, and (6) Single PI based on the polychoric correlation matrix using DWLS estimation. For the matched and Single PI approaches the pairing of the NA and SI indicators (constituting the latent interaction term) was based on the ranking of the estimated standardized factor loadings of the Type D measurement model. For the Matched PI approach we used seven indicator pairs and for the Single PI approach one indicator pair.

Model Fit

Model fit was assessed by inspecting the χ^2 test, the Bayesian Information Criterion (BIC; Schwarz, 1978), as well as the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI), and Root Mean Square Error of Approximation (RMSEA). For both the TLI and CFI we assumed values above 0.95 to indicate adequate fit (Hu & Bentler, 1999). For the RMSEA we considered values above 0.10 as unacceptable and values below 0.06 as indicating good model fit (Chen et al., 2008). For some of the investigated methods not all of these fit indices were available. For instance, because the BIC is based on the log likelihood, Mplus only reports this fit index for models involving maximum likelihood estimation. Consequently, this fit index is not reported for our models involving weighted least squares estimation. Furthermore, when numerical integration is required (e.g. with the LMS method), means, variances, and covariances are not sufficient statistics for model estimation and chi-square

and related fit statistics are not available. As a result, Mplus does for the LMS method not report the chi-square statistics and related indices (RMSEA, CFI, TLI). In these situations all available fit indices were reported. Nested models were compared using a chi-square or log likelihood difference test, depending on the type of interaction model. For the PI MLR methods we used the Satorra & Bentler (2010) scaled χ^2 difference between two models against their difference in degrees of freedom (df). For the PI DWLS we used the T3 test (Asparouhov & Muthén, 2010a), a nested model comparison test developed specifically for ordinal non-normal data. Lastly, for the LMS method we used the chi-square difference test based on log likelihood values and scaling correction factors obtained with MLR estimation.

RESULTS

Of all 3,314 participants, 3 (0.1%) did not complete the PHQ-9 questionnaire and 9 (0.27%) did not complete the GAD-7 questionnaire. We excluded these participants from our latent prediction model. Tables 1 through 3 show the fit statistics and parameter estimates for several tested models according to the six different ways to model the interaction effect between NA and SI on depression and anxiety. Table 1 focuses on the Sum score approach, while Tables 2 and 3 focus on the latent variable approaches.

Sum score approach

According to the Sum score approach the models including the interaction between NA and SI fitted the data better than the models without the interaction term. There was a significant interaction between NA and SI on both depression (B=0.561, t(3308)=9.807, p<.001, β=0.126) and anxiety (B=0.329, t(3302)=7.481, p<.001, β=0.093). However, first including the NA and SI quadratic resulted in significant quadratic effects and rendered the interaction between the two constructs non-significant, both with respect to depression (B=0.048, t(3308)=0.678, p=0.498, β =0.012) and anxiety (B=-0.052, t(3302)=-0.95, p=0.342, β=-0.017).

Table 1: Association between Type D personality, depression, and anxiety using linear hierarchical regression analyses.

Dependent Measure	Model	Model fit (F-test)	Nested model difference test (Δ*F*)	NA		SI		SI x NA		NA^2		SI^2	
				B (SE)	*β*	B (SE)	*β*	B (SE)	*β*	B (SE)	*β*	B (SE)	*β*
Depression	1	F(2,3308) = 1346.0*	-	3.08 (.07)*	.66	.06 (.07)	.01	-	-	-	-	-	-
	2a	F(3,3307) = 955.3*	F(1,3307) = 96.2*	3.02 (.07)*	.68	.003 (.07)	.001	.56 (.06)*	.13	-	-	-	-
	2b	F(4,3306) = 802.4*	F(1,3306) = 143.0*	2.69 (.07)*	.70	.03 (.07)	.01	-	-	.79 (.05)*	.21	.21 (.05)*	.06
	3	F(5,3305) = 641.9*	F(1,3305) = .46	2.69 (.07)*	.69	.03 (.07)	.01	.05 (.07)	.01	.78 (.06)*	.20	.19 (.06)*	.05
Anxiety	1	F(2,3302) = 1561.0*	-	2.64 (.05)*	.72	-.18 (.05)*	-.05	-	-	-	-	-	-
	2a	F(3,3301) = 1077.0*	F(1,3301) = 56.0*	2.60 (.05)*	.74	-.21 (.05)	-.06	.33 (.04)*	.09	-	-	-	-
	2b	F(4,3300) = 902.7*	F(1,3300) = 125.9*	2.33 (.05)*	.77	-.17 (.05)*	-.06	-	-	.60 (.04)*	.20	.08 (.04)*	.03
	3	F(5,3299) = 722.3*	F(1,3299) = .90	2.34 (.05)*	.75	-.17 (.05)*	-.05	-.05 (.05)	-.02	.62 (.05)*	.20	.10 (.05)*	.03

Model 1: NA & SI main effects

Model 2a: NA & SI main + interaction effects

Model 2b: NA & SI main + quadratic effects

Model 3: NA & SI main + quadratic + interaction effects

β = standardized regression coefficient; B = unstandardized regression coefficient

** p < .05*

Table 2: Model fit for three models and five methods to model latent interaction effects.

Method	Model	Parameters (df)	Model fit indices					
			χ^2 / LL	χ^2 or LL difference test (df)†	BIC	RMSEA (95%CI)	CFI	TLI
Single PI MLR	1	96 (400)	6915.1*	-	216231.9	.067 [.066 .069]	0.843	0.831
	2	100 (396)	4935.6*	2437.6 (4)*	213474.9	.057 [.055 .058]	0.891	0.881
	3	102 (394)	4852.3*	61.0(2)*	**213353.8**	**.056 [.055 .058]**	**0.893**	**0.883**
Single PI DWLS	1	140 (356)	59500.9*	-	NA‡	.204 [.202 .205]	0.584	0.551
	2	144 (352)	7983.4*	6376.5 (4)*	NA‡	.073 [.072 .075]	0.947	0.942
	3	146 (350)	7831.4*	144.7 (2)*	NA‡	**.073 [.071 .074]**	**0.948**	**0.943**
Matched PI MLR	1	115 (588)	7464.9*	-	282889.9	.058 [.056, .059]	0.844	0.833
	2	119 (584)	5549.2*	2479.6 (4)*	280149.3	.049 [.048 .050]	0.887	0.879
	3	121 (582)	5479.3*	111.6 (2)*	**280053.3**	**.049 [.048 .050]**	**0.889**	**0.880**
Matched PI DWLS	1	159 (544)	60679.9*	-	NA‡	.171 [.169 .172]	0.602	0.576
	2	163 (540)	8410.1*	6209.4 (4)*	NA‡	.062 [.060 .063]	0.948	0.945
	3	165 (538)	8164.7*	156.1 (2)*	NA‡	**.061 [.060 .062]**	**0.950**	**0.946**
LMS MLR	1	92 (373)	-101940.5	-	204626.5	NA§	NA§	NA§
	2	96 (369)	-100563.7	2553.6 (4)*	201905.3	NA§	NA§	NA§
	3	98 (367)	-100445.3	88.4(2)*	**201684.7**	NA§	NA§	NA§

Model 1: Covariates

Model 2: Covariates + NA + SI

*Model 3: Covariates + NA + SI + NA*SI*

Numbers printed in bold indicate better fit of one model relative to the other.

** p < .05*

† This column shows for the PI MLR methods the Satorra & Bentler (2010) scaled χ^2 difference between two models against their difference in degrees of freedom (df). For the PI DWLS methods the column shows the T3 test (Asparouhov & Muthén, 2010a), a nested model comparison test developed for ordinal non-normal data. For the LMS method this column shows the likelihood ratio test based on log likelihood values and scaling correction factors obtained with MLR estimation.

‡ As the BIC is based on the log likelihood which is not computed in the DWLS methods, Mplus does not report this fit index for models involving weighted least squares estimation.

§ When numerical integration is required (e.g. with the LMS method), means, variances, and covariances are not sufficient statistics for model estimation and chi-square and related fit statistics are not available. Therefore, Mplus does for report the chi-square statistics and related indices (RMSEA, CFI, TLI) for the LMS method.

Table 3: Model fit indices for three models and five methods to model latent interaction and quadratic effects for NA and SI on the latent variables depression and anxiety.

Method	Model	Parameters (df)	Model fit indices					
			χ^2 / LL	χ^2 or LL difference test (df)†	BIC	RMSEA (95%CI)	CFI	TLI
Single PI MLR	1	115 (381)	5401.8*	-	237632.6	**.056 [.054 .057]**	**0.886**	**0.874**
	2	117 (379)	5390.9*	12.32 (2)*	**237629.7**	**.056 [.055 .057]**	**0.886**	**0.874**
Single PI DWLS	1	159 (337)	9334.1*	-	NA‡	.075 [.073 .076]	0.947	**0.942**
	2	161 (335)	9225.5*	128.8 (2)*	NA‡	.075 [.073 .076]	**0.948**	**0.942**
Matched PI MLR	1	172 (531)	15356.2*	-	**437610.0**	**.060 [.059 .060]**	**0.780**	**0.767**
	2	174 (529)	15362.4*	2.64 (2)	437621.5	.060 [.059 .061]	**0.780**	**0.767**
Matched PI DWLS	1	230 (473)	18482.5*	-	NA‡	**.093 [.092 .094]**	**0.738**	**0.723**
	2	232 (471)	18477.7*	38.0 (2)*	NA‡	.093 [.092 .094]	**0.738**	**0.723**
LMS MLR	1	100 (365)	-100012.1	-	**200834.4**	NA§	NA§	NA§
	2	102 (363)	-100201.0	12.55 (2)*	201228.4	NA§	NA§	NA§

Model 1: NA + SI + NA^2 + SI^2

*Model 2: NA + SI + NA^2 + SI^2 + NA*SI*

** p < .05*

† This column shows for the PI MLR methods the Satorra & Bentler (2010) scaled χ^2 difference between two models against their difference in degrees of freedom (df). For the PI DWLS methods the column shows the T3 test (Asparouhov & Muthén, 2010a), a nested model comparison test developed specifically for ordinal non-normal data. For the LMS method this column shows Wald test on the parameter constraints that the estimates of the interaction effect on both depression and anxiety are equal to zero.

‡ As the BIC is based on the log likelihood which is not computed in the DWLS methods, Mplus does not report this fit index for models involving weighted least squares estimation.

§ When numerical integration is required (e.g. with the LMS method), means, variances, and covariances are not sufficient statistics for model estimation and chi-square and related fit statistics are not available. Therefore, Mplus does for report the chi-square statistics and related indices (RMSEA, CFI, TLI) for the LMS method.

Table 4: Regression coefficients for five methods to model latent interaction effects between the personality traits NA and SI on the latent variables depression and anxiety.

Method	Model	Latent regression effects on depression				Latent regression effects on anxiety			
		SI	NA	SI x NA	β	SI	NA	SI x NA	β
Single PI MLR	1	-	-	-	-	-	-	-	-
	2	-.05(.01)*	.74(.04)*	-	-	-.08(.01)*	.73(.03)*	-	-
	3	-.05 (.01)*	.66 (.04)*	.13(.02)*	0.24	-.09 (.01)*	.67 (.03)*	.09(.02)*	0.16
Single PI DWLS	1	-	-	-	-	-	-	-	-
	2	-.11 (.02)*	1.37 (.04)*	-	-	-.20 (.02)*	1.49 (.04)*	-	-
	3	-.12 (.02)*	1.28 (.04)*	.14 (.01)*	0.16	-.20 (.02)*	1.43 (.04)*	.09 (.01)*	0.11
Matched PI MLR	1	-	-	-	-	-	-	-	-
	2	-.04 (.01)*	.74 (.04)*	-	-	-.08 (.01)*	.72 (.03)*	-	-
	3	-.05 (.01)*	.70 (.04)*	.09 (.01)*	0.16	-.09 (.01)*	.71 (.03)*	.05 (.01)*	0.09
Matched PI DWLS	1	-	-	-	-	-	-	-	-
	2	-.12 (.02)*	1.41(.04)*	-	-	-.21 (.02)*	1.51(.04)*	-	-
	3	-.12 (.02)*	1.32 (.04)*	.10(.01)*	0.13	-.20 (.02)*	1.46 (.04)*	.06(.01)*	0.07
LMS MLR	1	-	-	-	-	-	-	-	-
	2	-.04(.01)*	.76(.04)*	-	-	-.08(.01)*	.75(.03)*	-	-
	3	-.04(.01)*	.69(.04)*	.21(.02)*	0.21	-.08(.01)*	.70(.03)*	.13(.02)*	0.14

*All effects are reported in terms of unstandardized regression coefficients and their corresponding standard errors. β denotes the standardized regression coefficient of the latent interaction. *p < .001*

Table 5: Regression coefficients for five methods to model latent interaction and quadratic effects for the personality traits NA and SI on the latent variables depression and anxiety.

Method	Model	Latent regression effects on depression				Latent regression effects on anxiety			
		SI^2	NA^2	SI x NA	β	SI^2	NA^2	SI x NA	β
Single PI MLR	1	.35 (.11)*	-.07 (.06)	0	0	.22 (.07)*	0 (.04)	0	0
	2	.09 (.05)	.01 (.02)	.10 (.03)*	0.19	.07 (.04)	.04 (.02)*	.05(.02)*	0.10
Single PI DWLS	1	-.29 (.06)*	.22 (.02)*	0	0	-.30 (.06)*	.22 (.02)*	0	0
	2	-.23 (.05)*	.10 (.02)*	.18 (.02)*	0.21	-.30 (.05)*	.15 (.03)*	.14 (.02)*	0.16
Matched PI MLR	1	.04 (.02)*	.35 (.06)*	0	0	.05 (.02)*	.19 (.05)*	0	0
	2	.04 (.02)*	.35 (.06)*	.00 (.02)	0.00	.06 (.02)*	.22 (.05)*	-.02 (.02)	-0.04
Matched PI DWLS	1	2.54 (1.50)	-1.54 (.86)	0	0	3.64 (2.12)	-2.06 (1.21)	0	0
	2	.45 (.18)*	-.32 (.10)*	-.02(.01)	-0.05	.42 (.27)*	-.18 (.14)*	-.04(.02)*	-0.07
LMS MLR	1	0 (.01)	.49 (.04)*	0	0	-.02 (.01)*	.42 (.04)*	0	0
	2	.05(.02)*	.22(.01)*	-.09(.03)*	-0.15	.04(.01)*	.20(.01)*	-.11(.03)*	-0.19

*All effects are reported in terms of unstandardized regression coefficients and their corresponding standard errors. β denotes the standardized regression coefficient of the latent interaction. *p < .001*

Moreover, for the sum score approach the model including the interaction term did not significantly explain additional variance in both depression and anxiety, above and beyond the NA and SI main and quadratic effects. This suggests that the Type D effect might be confounded by quadratic NA and SI effects.

Single PI (MLR) approach

Table 2 shows that for the Single PI approach with MLR estimation, all three nested models resulted in a significant χ^2 value, indicating that the model-implied covariance matrix deviated from the observed covariance matrix, a sign of poor exact model fit. The BIC favored the model including the interaction between NA and SI (Model 3) above the model without interaction effects (Model 3). The RMSEA suggested that both Models 2 and 3 showed good fit, because the upper bound of the 95% confidence interval was below 0.06. According to the CFI and TLI, none of the models showed acceptable fit, although the model including interaction effects showed better fit than the other two models. The chi-square difference tests indicated an improved fit of each model over its predecessor. Of particular interest is the significant difference in fit between Models 2 and 3, indicating that the model including the interaction effect significantly explained additional variation in anxiety and depression scores above and beyond the model with the NA and SI main effects only. Table 4 indicates that the Single PI MLR approach showed significant regression coefficients for the interaction between NA and SI for both depression (B=0.13, z=6.29, p<0.001, β=0.24) and anxiety (B=0.09, z=4.70, p<.001, β=0.16).

To investigate whether the interaction effect merely reflected unmodeled quadratic effect, Table 3 shows the comparison of a model including the NA and SI main- and quadratic effects with a model also including the interaction effect. As the RMSEA, CFI and TLI almost always resulted in similar fit measures for both models, we decided to only report these results in Table 3. It turned out that based on the BIC and chi-square differences test, the model including main, quadratic *and* interaction effect fitted the data significantly better than the model with main and quadratic effects only, suggesting that the interaction effect did not merely reflect quadratic NA and SI effects. Table 5 shows that after including the quadratic effects, the estimated interaction effects remained statistically significant for both depression (B=0.10, z=4.01, p<0.001, β=0.19) and anxiety (B=0.05, z=2.35, p=0.019, β=0.10).

Single PI (DWLS) approach

Table 2 shows the results for the Single PI approach with WLS estimation based on the polychoric correlation matrix. Again, all three nested models resulted in a significant χ^2 value, a sign of poor exact model fit. The RMSEA suggested that both Models 2 and 3 showed reasonable fit with the upper bound of the 95% confidence interval below 0.08. For both models 2 and 3, the CFI and TLI were just below 0.95, suggesting good fit. The two nested model comparison tests indicated an improved fit of Model 2 over Model 1 and of Model 3 over Model 2, suggesting that both the NA and SI main effects, as well as their interaction is of added value to the model. Table 4 indicates that the Single PI DWLS approach resulted in significant regression coefficients for the interaction between NA and SI for both depression (B=0.14, z=11.34, p<0.001, β=0.16) and anxiety (B=0.09, z=7.21, p<0.001, β=0.11).

Table 3 shows that based on the nested model comparison test, the model including main, quadratic and interaction effects fitted the data significantly better than the model including main and quadratic effects only. Table 5 reveals that after including the quadratic effects, the estimated interaction effects remain statistically significant for both depression (B=0.18, z=10.07, p<0.001, β=0.21) and anxiety (B=0.14, z=6.77, p<0.001, β=0.16).

Matched PI (MLR) approach

Table 2 shows the results of the Matched PI approach with MLR estimation. All three models resulted in a significant χ^2 value, a sign of poor exact model fit. However, the BIC favored the model with the interactions (Model 3) above the model without the interactions (model 2). According to the RMSEA, CFI and TLI, the model including the interactions fitted the data best. The RMSEA of both model 2 and 3 showed an upper bound of the 95% confidence interval of 0.05. According to both the CFI and TLI, all models showed suboptimal fit with values of approximately 0.88. The chi-square difference tests indicated an improved fit of each model over its predecessor. Hence, including the interactions (Model 3) resulted in significantly better fit than fixing the interactions to zero (Model 2). Table 4 indicates that for Model 3 the Matched PI (MLR) approach resulted in a significant interaction between NA and SI for both depression (B=0.09, z=7.77, p<0.001, β=0.16) and anxiety (B=0.05, z=4.59, p<.001, β=0.09).

Table 3 shows that based on the BIC and the nested model comparison test, the model including main, quadratic and interaction effects did *not* significantly fit the data better than the model including main and quadratic effects only. Table 5 reveals that after including the quadratic effects, the estimated interaction effects did no longer reach statistical significance for both depression (B=0.00, z=-0.14, p=0.892, β=0.00) and anxiety (B=-0.02, z=-1.40, p=0.163, β=-0.04).

Matched PI (DWLS) approach

Table 2 shows the results of the Matched PI approach with DWLS estimation based on the polychoric correlation matrix. All three models resulted in a significant χ^2 value, a sign of poor exact model fit. According to the RMSEA, CFI and TLI, the model including the interactions fitted the data best. Both models 2 and 3 showed RMSEA 95% confidence intervals around 0.06. According to both the CFI and TLI, models 2 and 3 showed good fit with values of approximately 0.95. The nested model comparison tests indicated an improved fit of each model over its predecessor. Hence, including the interactions (Model 3) resulted in significantly better fit than fixing the interactions to zero (Model 2). Table 4 indicates that for Model 3 the Matched PI (DWLS) approach resulted in a significant interaction between NA and SI for both depression (B=.10, z=12.00, p<.001, β=.13) and anxiety (B=.06, z=6.95, p<.001, β=.07).

Table 3 shows that based on the nested model comparison test, the model including main, quadratic and interaction effects fitted the data significantly better than the model including main and quadratic effects only. Table 5 reveals that after including the quadratic effects, the estimated interaction effect no longer reached statistical significance for depression (B=-0.02, z=-1.53, p=0.126, β =-0.05) yet remained significant for anxiety (B=-0.04, z=-2.17, p=0.03, β=-0.07).

LMS approach

Table 2 shows the results of the LMS approach using MLR estimation. According to the BIC, the model with the interaction term (Model 3) fitted the data better than the model without the interaction term (Model 2) for both outcome variables. The likelihood ratio test too indicated that inclusion of the interaction terms resulted in a significantly improvement in

model fit. Table 4 indicates that for Model 3 the LMS approach resulsted in a significant interaction between NA and SI for both depression (B=0.21, z=8.93, p<0.001, β=0.21) and anxiety (B=0.13, z=6.02, p<0.001, β=0.14).

Table 3 shows that based on the nested model comparison test, the model including main, quadratic and interaction effects fitted the data significantly better than the model including main and quadratic effects only (Wald(2) = 12.55, p = 0.002)[1]. Lastly, Table 5 reveals that after including the quadratic effects, the estimated interaction effects remained statistically significant, but switched signs both for depression (B=-0.09, z=-2.7, p=0.007, β =-0.15) and anxiety (B=-0.11, z=-3.54, p<0.001, β =-0.19).

Factor scores

Figure 3 shows in four separate plots for both NA and SI the association with depression and anxiety. Each plot shows the effect of NA or SI for three different levels of the other construct (Low < -1 SD < Average < +1 SD < High). All axes show factor scores estimated with the Maximum A Priori method in Mplus. Visual inspection of these plots suggest two competing interpretations: either the effects of NA or SI on depression and anxiety get stronger at higher levels of the other construct, or NA and SI show quadratic effects on depression and anxiety, where the association gets stronger at higher levels of each construct. The estimated factor scores show good factor indeterminacy values (NA = 0.964; SI = 0.952; DEP = 0.959; ANX = 0.965).

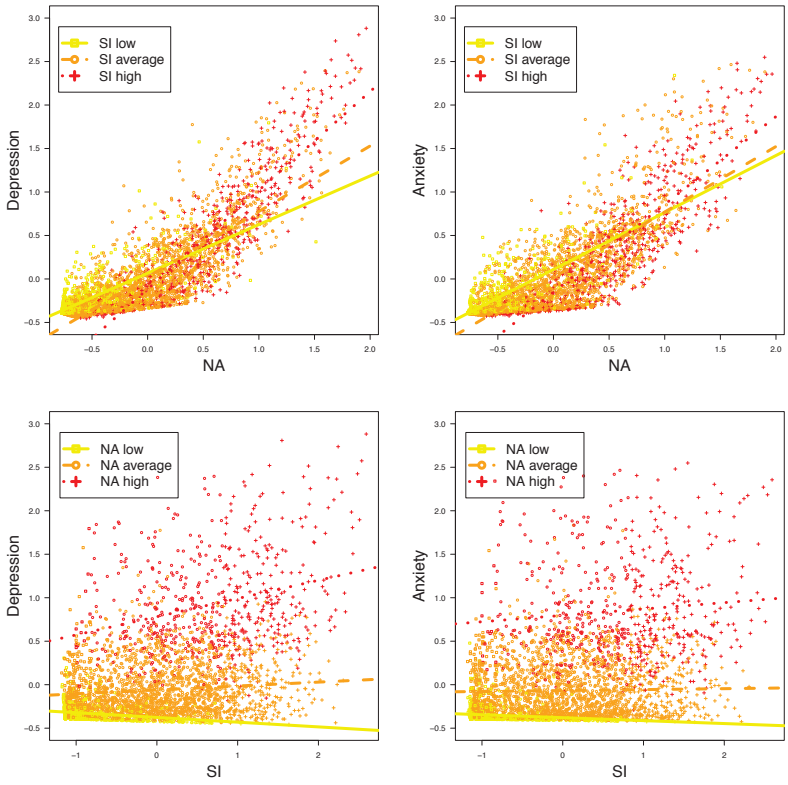
Summary

To summarize, according to all six methods, the interaction between NA and SI was significantly associated with both depression and anxiety, although the size of this interaction varied across the tested methods. Moreover, when first adding the quadratic effects to the model, all methods produced smaller estimates of the interaction effects. For both Single PI methods the interaction effects remained significant, while for other methods

[1] Originally, we compared these two models with a log likelihood difference test. However, the resulting chi-square difference value was negative, and thus the test is invalid as it should result in a positive chi-square difference. Negative chi-square values can arise when using MLR estimation. As an alternative, we used the Wald test, which is available in MPLUS, to test the constraint that both interaction effects are equal to zero.

the interaction reduced to zero (Sum score; Matched PI MLR; Matched PI DWLS), or even changed the effect in the opposite direction (LMS). Because the estimates of both the quadratic and interaction effects vary not only across the method used to model the non-linear effects, but also across the estimation method (e.g. MLR vs. DWLS) we doubt the robustness of the models including both quadratic and interaction terms and we advise readers to interpret the results of these models with care. Given these conflicting results we need additional evidence to support the robustness of the findings in Study 1. Therefore, in Study 2, we conducted a simulation to compare these six methods to model interactions on the bias and precision of the estimated interaction effect.

Figure 3: Plots showing the association between NA and Depression (upper left), NA and anxiety (upper right), SI and depression (lower left), and SI and Anxiety (lower right). Each plot shows the effect of NA or SI for three levels of the other construct (-1 SD, Mean, +1 SD). All axes show factor scores estimated with the Maximum A Priori method in Mplus.



STUDY 2: SIMULATION STUDY

METHOD

Procedure

In our simulation study, we varied four different design parameters: scale (continuous, ordinal), item skewness (0, 2, 3), the size of the interaction (0, .10, .20, .40) and sample size (250, 500, 3000). All possible combinations of these four design parameters resulted in a factorial design with 2x3x4x3=72 conditions. The anxiety (GAD-7) and depression (PHQ-9) questionnaires showed similar psychometric properties. For reasons of clarity, we decided to focus our simulation on estimating the interaction effect of NA and SI on depression.

Design Parameter 1: Scale Level

The first design parameter was the scale level of the simulated DS14 and PHQ-9 items. We either simulated continuous or ordinal item scores. For ordinal scores, the number of response categories depended on the type of questionnaire (PHQ-9 = 0-3 Likert scale; DS14 = 0-4 Likert scale).

Design Parameter 2: Skewness

The second design parameter was the amount of skewness in the distribution of the latent traits NA and SI. We used the method of Vale and Maurelli (1983) as implemented in the R-package *fungible* (version 1.5; Waller, 2016), to simulate a multivariate distribution of NA and SI. We varied across three skewness values (0, 2 and 3; with corresponding kurtosis values 0, 7 and 21), while retaining the product moment correlation between NA and SI (Study 1 estimate of .553),

Design Parameter 3: Size of Interaction

The third design parameter indicated the strength of the interaction between NA and SI on depression. We based the true size of the interaction on the standardized regression coefficient of the estimated interaction effect in Study 1 according to the LMS approach

(β=0.20). In our simulation, we allowed the interaction to be either absent (0), half the size of the Study 1 interaction (0.10), the exact size of the Study 1 interaction (0.20), or twice the size of the Study 1 interaction (0.40).

Design Parameter 4: Sample size

The fourth design parameter indicated the sample size of the simulated dataset. We varied between a small, medium and large sample size condition. In the *large sample* condition, we used a sample size of 3,000 participants, resembling the sample size of the dataset used in Study 1. In the *medium sample* condition we simulated data for 500 participants, representing sample sizes commonly encountered when analyzing latent variable models. Lastly, in the *small sample* condition we simulated a sample size of 250 participants.

Data Simulation

For each of the 72 conditions, we simulated 500 datasets containing scores on items measuring the constructs depression (9 items), NA (7 items) and SI (7 items). We generated data using the parameter estimates (i.e., factor loadings, latent (co)variances, regression coefficients, thresholds and error variances) of the latent prediction model in Study 1[2]. First, we randomly sampled vectors of NA and SI latent trait scores according to the multivariate skew distribution, given the NA and SI (co)variance(s) from Study 1 and given the skewness design parameter. Second, we used Equation 1 to calculate the continuous item scores for each individual (*i*) and for each item (*j*) measuring the traits (*t)* NA or SI:

$$Y_{ij} = \lambda_{jt}\, \eta_{ti} + \varepsilon_{ij} \qquad (1)$$

In this equation λ_{jt} denotes the factor loading of item *j* loading on trait *t*. η_{ti} represents the score of individual *i* on latent trait *t,* and ε_{ij} the residual error of individual *i* on item *j*. When generating the continuous item scores we used as input a matrix with individual NA and SI trait scores ($\boldsymbol{\Psi}$), the factor loading matrix retrieved from Study 1 ($\boldsymbol{\Lambda}$), and a residual error matrix ($\boldsymbol{\Theta}$) based on a multivariate normal distribution with a mean vector of zeroes and a

[2] We used the parameter estimates resulting from the LMS method, because this is the default method in Mplus to model interaction effects.

diagonal covariance matrix with variances retrieved from the output in Study 1. In line with earlier research (Flora & Curran, 2004), for ordinal scenarios we transformed these continuous item scores into ordinal scores using the case 1 thresholds proposed by Muthén and Kaplan (1985).

To simulate PHQ-9 (depression) item scores, we had to take into account that the scores on the depression measure depended on the scores of the NA and SI traits, their interaction, and prediction error. Therefore, we used Equation 2 to first compute the latent depression scores:

$$\psi_{Di} = \beta_{SI}\psi_{SIi} + \beta_{NA}\psi_{NAi} + \beta_{NA*SI}\psi_{NAi}\psi_{SIi} + \delta_i \qquad (2)$$

We then used Equation 1 to compute the continuous depression item scores and if appropriate we used the case 1 thresholds (Muthén & Kaplan, 1985) to transform them into ordinal item scores. In Equation 2, ψ_{Di} denotes the latent depression score of individual *i*, the three $\beta's$ represent the regression coefficients of the latent regression of depression on NA (ψ_{NAi}), SI (ψ_{SIi}) and the interaction between NA and SI ($\psi_{NAi} * \psi_{SIi}$). Lastly, δ_i denotes the prediction residual of individual *i,* based on normal distribution with mean zero and a variance retrieved from the output of Study 1.

Estimating interaction effects

After simulating 500 datasets in each of the 72 conditions, we analyzed each dataset according to the same methods used in Study 1: (1) Regression of sum scores; (2) LMS with MLR estimation; (3) Single PI with MLR estimation; (4) Matched PI with MLR estimation; (5) Single PI with DWLS estimation based on the polychoric correlation matrix; (6) Matched PI with DWLS estimation based on the polychoric correlation matrix. Note that we only applied methods 5 and 6 to datasets with ordinal item scores, because the polychoric thresholds could not be estimated for data with continuous measurement levels. We implemented all latent interaction models in Mplus and conducted the simulation using the R-package MplusAutomation (Hallquist & Wiley, 2011). The R-script of this simulation study is available at this project's open science framework page: https://osf.io/kf6d5/.

Outcome Measures

We aggregated the parameter estimates over 500 replications and used the mean and standard deviation to compute a 95% confidence interval for the parameters of interest. Our main outcome was the bias and precision of the parameter estimates. The amount of bias was computed as the difference between the mean of a parameter estimate and the true value (i.e., the β values of used to generate the data; see Equation 2). We also assessed the mean squared error, defined as the squared distance between the estimated value of the interaction effect and the true value of the interaction, averaged across 500 replications. We used the width of the 95% confidence interval as a measure of precision in the parameter estimate.

Expectations

With respect to simulation conditions with continuous item scores, we expected in line with earlier research (Kelava & Nagengast, 2012; Kelava, Nagengast & Brandt, 2014), that LMS would perform best when the latent traits are normally distributed and that it would get more biased as skewness increased. Furthermore, we expected the PI approaches to show less bias in skewed conditions than the LMS method because earlier research showed that these approaches, especially the Matched PI approach, were less biased than LMS when the data is not normally distributed (Marsh et al, 2004; Cham, West, Ma & Aiken, 2012). With respect to simulation conditions with ordinal item scores, we expected the WLS approach that used the polychoric correlation matrix to outperform the MLR approach based on the product moment correlation matrix, because the former method directly models skewness by estimating threshold parameters for each item. Finally, we expected the Sum score approach to underestimate the interaction effects because the presence of measurement error attenuates the true association between the latent constructs.

RESULTS

All methods except LMS showed a 100% convergence rate. In conditions without skewness LMS also reached 100% convergence, but as skewness increased the convergence rate decreased to approximately 90% for N=500, and 85% for N=250 conditions. Non-converged solutions have been excluded from further analyses.

For all 72 conditions in our stimulation study, Tables O1 and O2 in ***Appendix O*** show respectively for continuous and ordinal item score conditions the mean parameter estimates of the interaction (including 95% confidence interval). Table O3 presents for all simulation conditions the bias in the estimated interaction effect in terms of the squared distance between the estimated interaction and the true interaction, averaged across all 500 replications (mean squared error). Table 4 summarizes these statistics by reporting the total mean squared error for each level of every design factor used in our simulation study. Lastly, Table 5 shows for each scenario the percentage of significant interaction effects across all 500 replications, given a significance level of 0.05.

Figures 4 and 5 illustrate respectively for continuous and ordinal item score conditions the mean bias in the estimate of the interaction between NA and SI for each of the 72 scenarios in our simulation study. Each figure shows nine plots, divided over three rows and three columns. The rows represent different sample sizes and the columns different amounts of skewness. Within each plot, the *x*-axis shows varying sizes of the true interaction between NA and SI. The y-axis shows the bias in the estimate of the standardized regression coefficients of the interaction between NA and SI. In each plot the colors and shape of the data points correspond to different methods to model the interaction effect. Each dot corresponds to the bias in the parameter estimate averaged over 500 replications and the error bars represent the 95% confidence interval of the mean.

Continuous item scores

When comparing the four methods used on continuous item scores, Table 6 shows that the LMS MLR performed best across almost all conditions with respect to minimizing bias,

followed by the Sum score approach, and Matched PI MLR. For all methods, the least biased conditions were those with the largest sample size, a zero interaction effect, or no skewness. Overall, the bias tended to increase as the sample size decreased and as the skewness increased. Against our expectations, when skewness was present in the latent traits LMS performed best, while Single PI performed poorly.

Figure 4 illustrates that for all methods the average bias was mostly positive, implying that the interaction effect was overestimated in conditions with continuous item scores. For the LMS and Sum score methods, the size of the interaction did not have a large impact on the bias in the estimates. However, both PI methods overestimated the larger interaction effects, especially when skewness was present. Of all methods, LMS had the highest precision because it showed the least variability in the estimated interaction effects, as indicated by the narrowest confidence intervals in Figure 4. As expected, for all methods larger sample sizes reduced the variability in the estimated interaction effects across the 500 replications. An interesting finding is that higher skewness resulted in more variability in the estimated effects for all methods, but especially so for the Single PI MLR method.

Table 7 shows that the power to detect a significant interaction effect was above 0.80 for all methods in the continuous item score conditions, except for the Single PI MLR method. This method was especially underpowered as skewness increases, even at larger sample sizes, possibly because it resulted in widely varying estimates of the interaction effect, as indicated by the broad confidence intervals around the mean estimate of the interaction effect in conditions with skewness (see Figure 4).

Table 7 further indicates for both the Single PI MLR and Matched PI MLR approaches were able to retain a false positive rate of approximately 5% across all simulation conditions. Both the LMS and Sum score method also showed acceptable false positive rates, but only when the latent traits were not skewed. As skewness increased, more false positive findings emerged, and this effect was magnified with larger sample sizes.

Table 6: For all methods, the total mean squared error across conditions involving specific design factors.

Design factor	Continuous item scores				Ordinal item scores					
	Sum score	*LMS*	*Single PI*	*Matched PI*	*Sum score*	*LMS*	*Single PI*		*Matched PI*	
	regression	*MLR*	*MLR*	*MLR*	*regression*	*MLR*	*MLR*	*DWLS*	*MLR*	*DWLS*
All simulation conditions	0.145	0.058	5.723	0.291	0.110	0.085	0.530	0.494	0.117	0.116
N=250	0.059	0.028	2.018	0.107	0.048	0.041	0.185	0.175	0.059	0.057
N=500	0.048	0.019	2.261	0.095	0.036	0.028	0.176	0.164	0.037	0.037
N=3000	0.039	0.012	1.445	0.089	0.026	0.017	0.169	0.155	0.021	0.021
Skewness = 0	0.012	0.012	0.022	0.045	0.014	0.013	0.083	0.09	0.038	0.031
Skewness = 2	0.043	0.023	0.881	0.083	0.044	0.034	0.206	0.186	0.036	0.04
Skewness = 3	0.091	0.023	4.820	0.163	0.052	0.038	0.241	0.219	0.043	0.045
β = 0	0.011	0.008	0.114	0.013	0.018	0.021	0.007	0.008	0.019	0.022
β = .1035	0.032	0.015	0.972	0.050	0.019	0.023	0.016	0.015	0.029	0.033
β = .207	0.057	0.023	2.095	0.109	0.010	0.016	0.073	0.067	0.029	0.03
β = .414	0.045	0.012	2.542	0.119	0.064	0.026	0.434	0.405	0.04	0.03

β denotes the standardized regression coefficient of the true interaction effect

Figure 4: Comparison of four methods to estimate interaction effects between NA and SI on depression in scenarios with continuous items**,** varying over the true interaction size (x axis), amount of skewness (columns) and sample size (rows)**.** Each data point shows the mean bias (including 95% confidence interval) in the standardized regression coefficient of the estimated interaction effect.

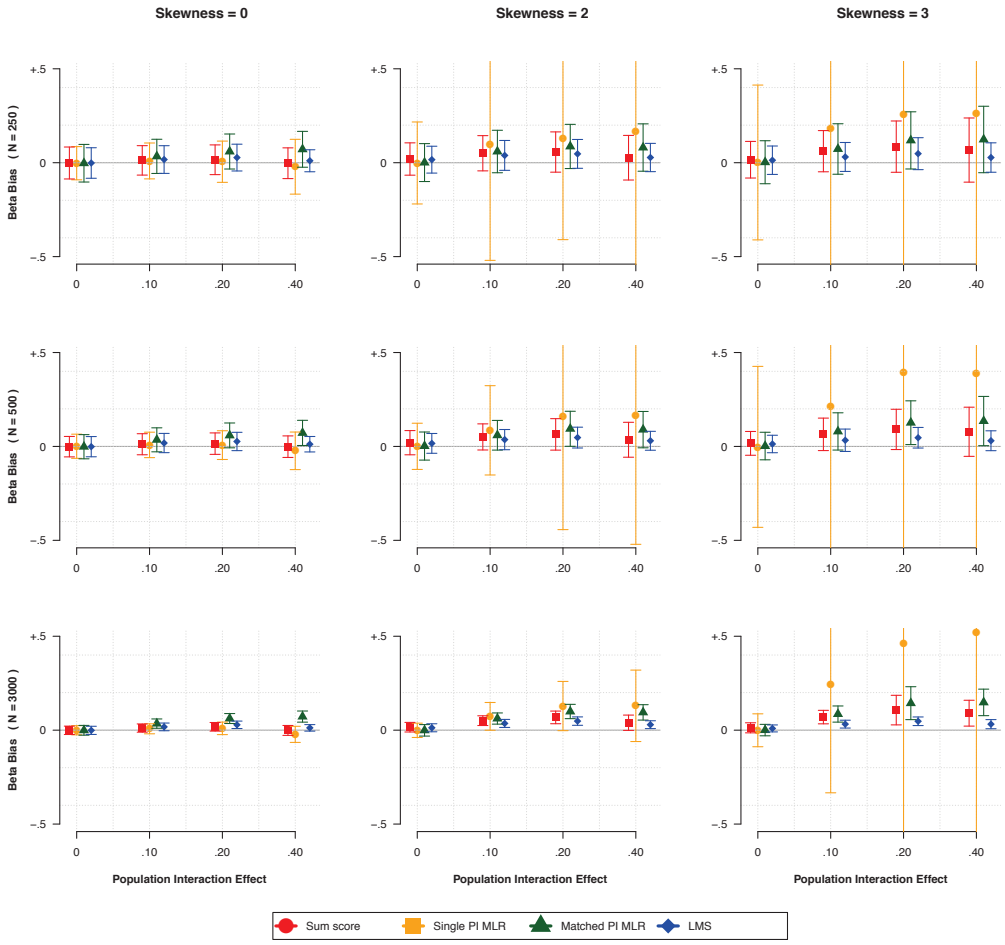


Figure 5: Comparison of four methods to estimate interaction effects between NA and SI on depression in scenarios with ordinal items**,** varying over the true interaction size (x axis), amount of skewness (columns) and sample size (rows)**.** Each data point shows the mean bias (including 95% confidence interval) in the standardized regression coefficient of the estimated interaction effect.

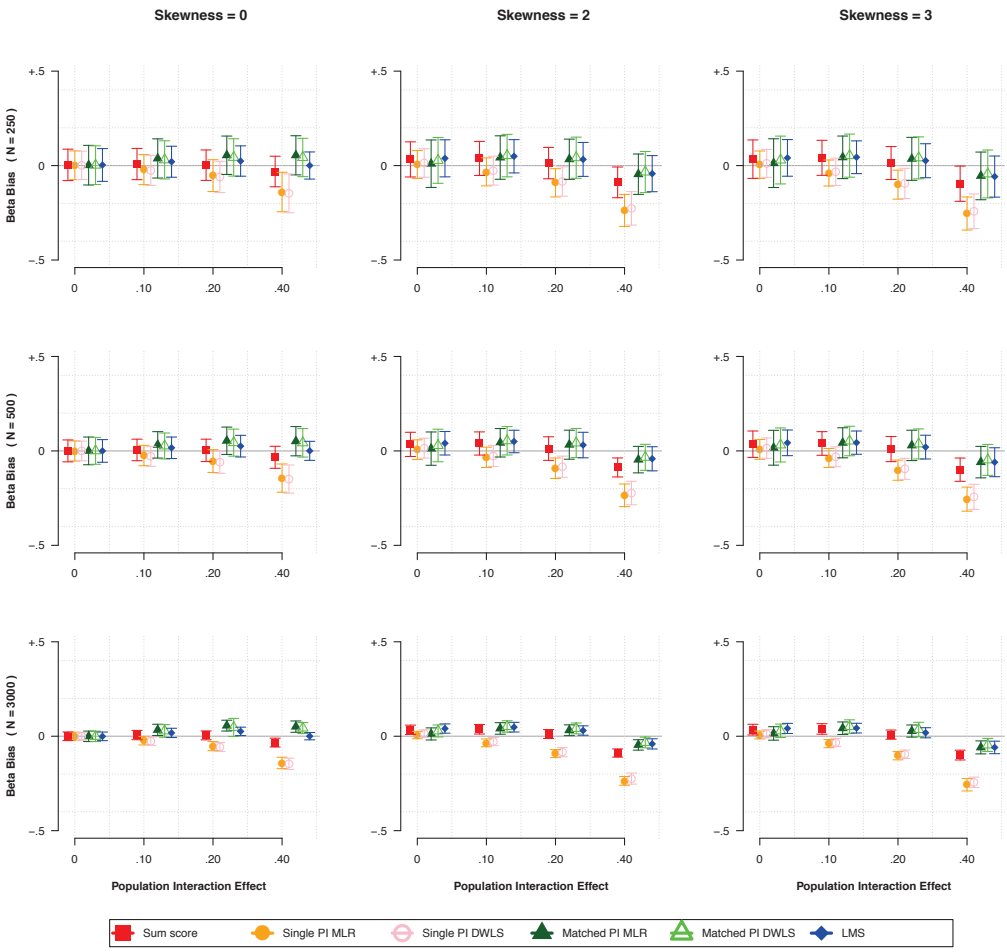


Table 7: The percentage of statistically significant ($p < .05$) interaction effects across conditions for methods estimating interaction effects while modeling continuous item scores.

N	Skewness	True size Interaction	Continuous item scores			
			Sum score regression	*LMS MLR*	*Single PI MLR*	*Matched PI MLR*
250	0	0	7.6	8.4	5.4	8.2
		.1035	81.6	86.8	67.4	82.6
		.207	100	100	98	100
		.414	100	100	100	100
	2	0	7.0	7.04	2.2	4.4
		.1035	93.4	94.5	28.8	85.0
		.207	100	100	50.8	99.8
		.414	100	100	63.6	100
	3	0	6.2	8.81	3.6	5.6
		.1035	93.8	92.5	21.8	82.6
		.207	100	100	38.4	100
		.414	100	100	48.4	99.8
500	0	0	4.2	5.2	5.2	6.0
		.1035	98.6	99.8	90.4	99.2
		.207	100	100	100	100
		.414	100	100	100	100
	2	0	9.8	9.82	2.2	6.8
		.1035	99.6	100	52.4	98.4
		.207	100	100	65.4	100
		.414	100	100	75.8	100
	3	0	6.6	6.9	1.8	5.6
		.1035	99.0	98.9	33.2	98.2
		.207	100	100	42.2	100
		.414	100	100	49.6	100
3000	0	0	3.4	5.0	3.8	6.0
		.1035	100	100	100	100
		.207	100	100	100	100
		.414	100	100	100	100
	2	0	22.2	22.9	3.6	6.0
		.1035	100	100	98.4	100
		.207	100	100	98.2	100
		.414	100	100	99.4	100
	3	0	16.4	14.2	1.0	5.2
		.1035	100	99.8	61.8	100
		.207	100	100	66.2	100
		.414	100	100	74	100

Table 8: The percentage of statistically significant ($p < .05$) interaction effects across conditions for methods estimating interaction effects while modeling ordinal item scores.

N	Skewness	True size Interaction	Ordinal item scores					
			Sum score regression	*LMS*	*Single PI*		*Matched PI*	
				MLR	*MLR*	*DWLS*	*MLR*	*DWLS*
250	0	0	6.4	5.8	5.4	13.0	5.6	12.8
		.1035	75.4	82.0	50.2	69.2	75.0	86.0
		.207	100	100	94.8	98.0	100	100
		.414	100	100	100	100	100	100
	2	0	13.4	15	4.0	15.4	7.0	21.8
		.1035	89	92.4	46.8	71.8	69.2	91.8
		.207	100	100	86.4	96.0	98.6	99.8
		.414	100	100	99.4	100	100	100
	3	0	15.6	15.8	4.2	20.0	5.8	27.0
		.1035	87	90.6	42.4	71.6	72	90.4
		.207	99.2	99.6	78.2	96	97.8	99.6
		.414	100	100	97.6	99.4	100	100
500	0	0	5.0	7.0	4.2	11.8	5.6	10.4
		.1035	95.6	97.8	79.2	89.2	95.8	97.6
		.207	100	100	100	100	100	100
		.414	100	100	100	100	100	100
	2	0	18.6	22.6	5	17.8	7.4	23.2
		.1035	99.8	100	72.2	90.4	94.0	99.8
		.207	100	100	99.4	100	100	100
		.414	100	100	100	100	100	100
	3	0	20.4	25.2	6.2	23.0	9.0	28
		.1035	99.6	99.8	72.4	94.0	94.4	99.6
		.207	100	100	98.6	99.6	100	100
		.414	100	100	100	100	100	100
3000	0	0	4.2	3.64	3.8	11.0	4.0	7.4
		.1035	100	100	100	100	100	100
		.207	100	100	100	100	100	100
		.414	100	100	100	100	100	100
	2	0	74.6	88.2	8.2	36.0	8.8	54.6
		.1035	100	100	100	100	100	100
		.207	100	100	100	100	100	100
		.414	100	100	100	100	100	100
	3	0	72.8	87.6	10.8	43.0	12.6	53.8
		.1035	100	100	99.8	100	100	100
		.207	100	100	100	100	100	100
		.414	100	100	100	100	100	100

Ordinal item scores

When comparing the six methods used on ordinal item scores, Table 6 shows that the LMS MLR performed best across almost all conditions with respect to minimizing bias, followed by the Sum score, Matched PI DWLS and Matched PI MLR approaches. For all methods, the least biased conditions were those with the largest sample size and no skewness, and bias increased as the sample size decreased and as skewness increased. Both Single PI MLR and Single PI DWLS performed poorly with respect to the mean squared error, especially when skewness was present and as the true size of the interaction became larger. Contrary to our expectations, for skewed latent traits Matched PI MLR was slightly less biased than Matched PI DWLS. However, the Single PI DWLS did perform slightly better than Single PI MLR when skewness was present.

Figure 5 illustrates that for both Single PI methods the average bias was mostly negative, implying that these methods underestimated the interaction effect when item scores were ordinal. The average bias for the other methods depended on specific design factors. The bias of the LMS MLR and both Matched PI methods was largely similar, though without skewness LMS MLR was less biased than both Matched PI methods. The width of the confidence intervals in Figure 5A indicates there were no large differences with respect to the precision of the six methods. Interestingly, the Single PI methods extremely variable estimates in skewed and continuous conditions did not apply to the ordinal conditions. As expected, for all methods larger sample sizes reduced the variability in the estimated interaction effects across the 500 replications. Higher skewness resulted in slightly more variability in the estimated effects for all methods.

Table 8 shows that for ordinal conditions, LMS outperformed the other methods with respect to maximizing the power to detect a significant interaction effect, followed closely by the Sum score method. As expected, for all methods the power increased with larger sample sizes and for N=250 the Single PI MLR, Single PI DWLS and Matched PI MLR methods were underpowered to detect the smallest simulated effect. Increasing the sample size to N=500 resulted in adequate power for all methods except the Single PI MLR method, still underpowered to detect the smallest effect.

With respect to minimizing the percentage of false positive findings, Table 8 shows that the Single PI MLR method outperformed all other methods, but was closely followed by the Matched PI MLR method. These methods managed to keep the false positive rate close to 5% in most conditions. Although they produced somewhat poorer false positive rates when skewness and sample size were large, they still outperformed the other methods in those simulation conditions. While both the Single- and Matched PI DWLS methods never showed acceptable false positive rates, the LMS and Sum score method only showed acceptable false positive rates without skewness. Both the LMS and Sum score method showed an increase in false positive rates as skewness increased, and this effect was most pronounced with larger sample sizes.

DISCUSSION

In this study, we investigated the relation between Type D personality, depression, and anxiety using a latent prediction model. To our knowledge, the association between these constructs has not been analyzed previously with latent variable models that take into account the measurement error present in the scales that measure these constructs. These modeling approaches allowed us to prevent the kind of bias that is likely to occur when analyzing such data with regular regression analysis based on sum score variables (Busemeyer & Jones, 1983; Embretson, 1996; Kang & Waller, 2005; MacCallum, Zhang, Preacher, & Rucker, 2002).

In Study 1 we applied a latent prediction model to existing data of 3,314 persons with Type 1 or 2 diabetes. Results according six methods to model interaction effects suggested a small but significant effect of Type D personality (viz. an interaction of its subcomponents NA and SI) on depression and anxiety, implying that the association of negative affectivity with both depression and anxiety tends to get stronger as people show more social inhibition. These findings are consistent with earlier research that did not use latent variable modeling (Nefs et al., 2015). However, the inclusion of quadratic NA and SI effects to the models reduced the size of the interaction effect. Indeed, both Single PI methods produced smaller but still

significant interaction effects, while for other methods the interaction effect either reduced to zero (Sum score; Matched PI MLR; Matched PI DWLS), or even changed in the opposite direction (LMS). Given these conflicting results we conducted a Monte Carlo simulation study to investigate the bias and precision of each of the six methods used to model the interaction effect.

In our simulation study, we found that the six methods to model interaction effects differ in the accuracy and precision of the estimated interaction effects. In general, the LMS approach performed best with respect to minimizing bias in the estimated interaction effect. This conclusion applies both to continuous as well as ordinal item scores. Unexpectedly, when skewness was present in the latent traits LMS was the least biased method, although it still overestimated the true value of the interaction on these occasions. This finding contrasts with earlier research showing the Matched PI approach to be less biased than LMS when skewness was present (Marsh et al., 2004; Kelava, Nagengast & Brandt, 2014).

In general, the Single PI approach did not perform well in terms of minimizing bias. Especially when skewness was high and item scores were continuous this approach resulted in widely varying interaction effect estimates. This finding also applies to the Matched PI approach, yet to a much lesser extent. The worse performance of the Single PI method compared to the Matched PI method is not surprising, because the Single PI method has to rely on a single product indicator when estimating the interaction effect. In light of these findings we strongly recommend against using the Single PI approach when modeling latent interaction effects.

With respect to maximizing the power to detect a non-zero interaction effect, LMS outperformed the other methods, both for continuous as well as for ordinal item scores. However, the Sum score approach and Matched PI approaches were also adequately powered in most simulated conditions. Regarding the percentage of false positives, the Single PI MLR and Matched PI MLR methods outperformed the other methods in most simulation conditions, keeping the false positive rate around 5%. Both LMS and the sum score method were only able to control the false positive rate without skewness in the latent traits. As the latent traits got more skewed, these methods produced more false

positives. This is no surprise, given that these conditions fail to meet LMS's assumption that the item scores of the exogenous latent variables show a multivariate normal distribution (Klein & Moosbrugger, 2000). These high power and the inflated false positive of the LMS approach are in line with findings from other simulation studies examining the LMS approach (Marsh et al., 2004; Kelava & Nagengast, 2012). Although the Single PI MLR and Matched PI MLR approach were able to better control the false positive rate than LMS in simulation conditions with skewness, both PI approaches still performed worse than than LMS in terms of bias and the power to detect a non-zero interaction effect.

Earlier research showed that for ordinal data, fitting a confirmatory factor model to the polychoric correlation matrix and estimating the parameters with (D)WLS estimation is fairly robust to moderate violations of the assumption that the underlying latent traits are normally distributed (Flora & Curran, 2004). Interestingly, our results indicated that for ordinal data, fitting the model to the polychoric correlation matrix (rather than the product moment correlation matrix) and using DWLS (rather than MLR estimation) was not beneficial with respect to minimizing the bias and false positive conclusions, but did show slightly better power. Our results highlight that if one decides to use the Single- or Matched PI approaches, it is not necessary to use the categorical DWLS approach as the continuous MLR strategy performs at least as well, and is more parsimonious.

In terms of minimizing bias, the Sum score approach also performed reasonably well in general, being the second least biased method, after LMS. Without skewness the Sum score approach even performed equally well as LMS, independent of whether the item scores were continuous or ordinal. We expected that the Sum score approach would show attenuated interaction effect estimates because this method does not take into account measurement error. However, the results of our simulation indicated that the Sum score approach overestimated the interaction effect. A possible explanation for this finding could be that ignoring measurement error can under some circumstances lead to overestimated associations, especially as models become more complex (Cole and Preacher, 2014). The Sum score approach resulted in spurious interactions only when skewness was high and/or sample size was large and this effect was more pronounced for ordinal than for continuous

items. These results align with earlier research showing that the Sum score method can indeed lead to spurious interactions (Embretson, 1996; Schwabe & Van den Berg, 2014).

To the best of our knowledge, this is the first simulation study comparing the performance of the Sum score, LMS and PI approaches when modeling interactions between latent variables based on *ordinal* data. Based on our findings, we would advise researchers to use latent variable modeling when testing for interactions between variables that are measured with error. This echoes similar statements made over the years that highlight the importance of latent variable models in isolating a construct from its measurement error (Rasch, 1960; Birnbaum, 1968; Embretson, 1996; Kang & Waller, 2005; Schwabe & Van den Berg, 2014). When comparing continuous- with ordinal simulation conditions LMS appears to perform slightly better when item scores are continuous, while the other methods perform better when item scores are ordinal. Nevertheless, if one aims to minimize bias in estimating the latent interaction effect, then we recommend to use the LMS approach, as this approach is the least biased across all simulation conditions, even when latent traits are skewed and item scores are ordinal. If one aims at maximizing the power to detect a non-zero interaction effect, then LMS is also the method of choice. However, LMS did show increased false positive rates as the skewness of the latent traits increased. Both PI MLR approaches adequately kept the false positive rate close to 5%. Of those two, the Matched PI MLR approach appears much less biased than the Single PI MLR approach. Therefore, if one aims at minimizing the chance on false positive findings, one could consider using the Matched PI MLR approach. This conclusion resonates with earlier research showing the benefits of this particular method relative to other methods to construct latent interactions (Marsh, Wen & Hau, 2004; Cham, West, Ma & Aiken, 2012).

Although our simulation showed LMS to be the least biased method, it still overestimated the interaction effect and has an inflated false positive rate when the latent traits were skewed, especially when item scores were ordinal. This aligns with earlier research indicating that LMS shows biased parameter estimates when skewness is introduced (Kelava & Nagenast, 2012; Kelava, Nagengast & Brandt, 2014; Cham, West, Ma & Aiken, 2012). Other promising methods that fell beyond the scope of the present study use mixture modeling to model skewed exogenous latent variables (Dolan & Van der Maas, 1998). For instance, the

recently developed Nonlinear Structural Equation Mixture Modeling (Kelava, Nagengast & Brandt, 2014) and the Bayesian finite mixture model (Kelava & Nagengast, 2012) both show promise in modeling latent interaction and quadratic effects when the latent exogenous variables are not normally distributed. However, it is unclear how these methods perform when item scores are ordinal rather than continuous. This would be an interesting avenue for future research.

The primary motivation for our simulation study was to place the results of our empirical study in context, assessing the bias and precision of the six methods used to model interaction effects. Of all simulation conditions, the one with a sample size of 3000, ordinal item scores, skewed latent traits, and an interaction effect of .207 most resembles the circumstances of our empirical study. Interestingly, our simulation study indicated that the Sum score approach is least biased in that specific scenario. This method showed that the interaction between NA and SI is no longer statistically significant after adding the quadratic NA and SI effects to the model. However, in the simulations, the Sum score approach exhibited many false positives when latent traits were skewed and item scores were ordinal. To lower the possibility that the significant quadratic NA and SI effects reflect false positive findings, we can inspect the results of the Matched PI MLR approach, which showed nominal false positive rates close to 5%. The empirical results of the Matched PI MLR approach are also similar to those of the Sum score method: adding the significant quadratic NA and SI effects to the model rendered the interaction between NA and SI statistically insignificant.

Combined, our findings fail to support our main hypothesis that the association between negative affectivity and both depression and anxiety gets stronger at higher levels of social inhibition. Rather, our results suggest that the effect of NA on both depression and anxiety might get stronger at higher levels of NA, and that there exist a similar but smaller quadratic effect for SI on both depression and anxiety. Effects similar to these have been reported in earlier research, where the personality traits neuroticism (correlation with NA: $r = 0.68$; De Fruyt & Denollet, 2002) and introversion (correlation with SI: $r = 0.52$; De Fruyt & Denollet, 2002) did not show a significant interaction on depression and anxiety, yet both did show significant quadratic effects (Jorm et al., 2000). Quadratic effects are known to be well approximated by interaction effects, especially as the constructs involved in the interaction

correlate highly (Kang & Waller, 2005). Therefore, in line with other researchers we stress the importance of always first adding the quadratic effects to models that test for the presence of interaction effects (Lubinski & Humphreys, 2000).

Our study investigated the association between Type D personality and anxiety and depression. Apart from these clinical psychological constructs, there exists a large body of research investigating whether Type D personality is related to worse health-related outcomes in the general population (Mols & Denollet, 2010) or in people with cardiovascular disease (Denollet et al., 2010; Denollet et al., 2018). Because most of these studies did not apply latent variable modeling and did not test for quadratic NA and SI effects, an interesting avenue for future research would be to do so in preregistered, highly powered direct replication projects. Such studies will not only highlight the importance of latent variable modeling, but also answer recent critiques (e.g. Smith, 2011; Grande, Romppel & Barth, 2012) questioning the replicability of research on Type D personality (for a more detailed discussion of this issue, see Denollet et al., 2013).

Finally, we would like to note that the methods used in this study are not limited to research on Type D personality. The investigated approaches to model latent interactions can be readily applied to future studies aimed at accounting for measurement error while analyzing interactions between psychological constructs.

CHAPTER 7

Latent logistic interaction modeling: A simulation and empirical illustration of Type D personality

P. LODDER
J. DENOLLET
W.H.M. EMONS
J.M. WICHERTS

Structural Equation Modeling

ABSTRACT

Introduction: This study focuses on three popular methods to model interactions between two constructs containing measurement error in predicting an observed binary outcome: logistic regression using (1) observed scores, (2) factor scores, and (3) Structural Equation Modeling (SEM). It is still unclear how they compare with respect to bias and precision in the estimated interaction when item scores underlying the interaction constructs are skewed and ordinal.

Method: We investigated this issue using both a Monte Carlo simulation and an empirical illustration of the effect of Type D personality on cardiac events.

Results: Our results indicated that the logistic regression using SEM performed best in terms of bias and confidence interval coverage, especially at sample sizes of 500 or larger.

Conclusion: Although for most methods bias increased when item scores were skewed and ordinal, SEM remained relatively unbiased interaction effect estimates when items were modeled as ordered categorical.

.

INTRODUCTION

In the medical and behavioral sciences, researchers often investigate the effect of predictor variables on a binary outcome variable. Occasionally, researchers are not only interested in the main effects of single predictors, but also in how multiple predictors influence each other's effects on the outcome variable. When one predictor moderates the effect of another predictor, one speaks of an interaction effect. Although there are several ways to assess the interaction between two variables on a binary outcome measure, researchers typically use a logistic regression analysis. In the logistic regression model, the interaction effect is assessed by multiplying the observed scores of two (or more) constructs involved in the interaction, and including the resulting product variable as a predictor (e.g. see Field, 2010; Tabachnick & Fidell, 2007).

In psychological research, the two interacting constructs are commonly unobserved (latent) and measured with questionnaires containing items measured on an ordinal scale. The scores on these items are typically summed and the resulting sum score is assumed to represent the construct of interest. One disadvantage of using such observed scores in regression analyses is that the presence of measurement error in these scores can either attenuate the regression coefficients or result in spurious effects (Busemeyer & Jones, 1983; Embretson, 1996; Kang & Waller, 2005; MacCallum, Zhang, Preacher, & Rucker, 2002). We will argue that this is especially true for interaction effects.

A second approach to assess such interactions is structural equation modeling (SEM). This approach takes into account the measurement error in the observed item scores by specifying a measurement model that shows the association between a latent construct and the items used to measure the construct. This measurement model allows for separating the variance in the item scores caused by variation in the latent construct from the variance caused by residual factors (i.e. measurement error). However, SEM often requires a large sample size, especially as models become more complex (Schumacker & Lomax, 2004; Kline, 2010).

A third approach to assessing interaction effects is by using factor analysis to first estimate the factor scores of the constructs, and to subsequently regress the binary outcome measure on the product of these estimated factor scores in a logistic regression (Lu, Kwan, Thomas & Cedzynski, 2012; Devlieger & Rosseel, 2017). The key difference between this factor score interaction approach and the regular SEM approach is that the latter estimates the parameters of the measurement and structural model in a single step, whereas according the former approach those parameters are estimated in two separate steps.

Although all three approaches are commonly used by applied researchers, it is still unclear how these three approaches perform with respect to bias and precision when estimating the interaction effect on a binary outcome measure when sample size is small and the observed item scores are ordinal and non-normally distributed. In this study, we aim to answer this question based on both a Monte Carlo simulation and an empirical illustration.

Type D personality

The Type D ("distressed") personality construct (Denollet, 2005) serves as a good case study for modeling interaction effects on binary outcome measures. First, some researchers (e.g. ***Chapter 2***; Smith, 2011) argue that the effect of Type D personality is best modeled as an interaction between its two subcomponents negative affectivity (NA) and social inhibition (SI). People with a Type D (distressed) personality tend to both experience negative emotions (i.e. NA) and are inhibited in expressing their emotions and behavior in social situations (i.e. SI). The two subcomponents NA and SI are hypothesized to show a combined synergistic effect that is more than the sum of their parts (Smith, 2011), suggesting the Type D effect is more than the additive NA and SI effects. A second reason why Type D personality serves as a good case study is that the medical outcomes associated with Type D personality are often measured on a binary scale. For instance, a systematic review showed that patients with cardiovascular disease who have a Type D personality show an increased risk on adverse cardiac events (Denollet, Schiffer & Spek, 2010). These findings were corroborated by a meta-analysis (O'Dell, Masters, Spielmans & Maisto, 2011). Another meta-analysis indicated that having a Type D personality imposes an increased mortality risk in people with coronary artery disease (Grande, Romppel & Barth, 2012).

Most of the studies included in these meta-analysis did not operationalize Type D personality as an interaction between its two subcomponents NA and SI, but classified people as having a Type D personality when they showed a sum score of 10 or higher on both the NA and SI construct. This approach has been criticized (e.g. Smith, 2011) as it involves the dichotomization of continuous variables, a practice known to reduce the power and effect size in statistical analyses and that may even give rise to spurious main- or interaction effects (MacCallum, Zhang, Preacher, & Rucker, 2002).

To answer this criticism, Denollet, Pedersen, Vrints and Conraads (2013) showed that Type D personality, operationalized as the interaction between its two continuous subcomponents NA and SI, significantly predicted the later occurrence of major cardiac events. This study serves as a perfect case study for the present article, as the authors used a logistic regression analysis to assess the interaction between two continuous predictor variables on a binary manifest outcome. Therefore, in the first part of the current study we will reanalyze the data of Denollet et al. (2013) and not only use the original logistic regression analysis, but also a logistic regression on factor scores and a logistic regression using SEM. As it not yet clear how these models perform with respect to the bias and precision of the estimated interaction effect when the item scores are ordinal and non-normally distributed, the second part of this article present the results of a Monte Carlo simulation study, assessing for each of those three interaction models the bias and precision in estimating the interaction effect under various conditions.

Modeling logistic interaction effects

Several methods exist to model an interaction between two continuous variables on a binary manifest outcome variable. These methods all operate within a logistic regression framework, to model the effect of continuous predictor variables on a binary outcome variable. They model the interaction effect between two variables on a binary outcome variable by including in the regression both the main effects of the two variables in the model, as well as their interaction. Let ξ_1 and ξ_2 be the latent predictors, respectively, and $\xi_1\xi_2$ their product representing their interaction. Furthermore, let $p(\mathbf{x}) = P(\text{Cardiac event}|\mathbf{x} = (\xi_1, \xi_2)$ be the probability of a cardiac event, given a set of predictors $\boldsymbol{\xi}$. The logistic model then equals:

$$\ln\left(\frac{p(\xi)}{1-p(\xi)}\right) = \beta_0 + \beta_1\xi_1 + \beta_2\xi_2 + \beta_3\xi_1\xi_2 \qquad (1)$$

In Equation 1, the natural logarithm of the odds of an event (i.e. the log odds or logit) depends on an intercept (β_0), an effect (β_1) of the first predictor (ξ_1), an effect (β_2) of the second predictor (ξ_2), and the interaction effect between those two predictors (β_3).

We will now discuss three general methods used to model interaction effects when the outcome is both manifest and binary: (1) logistic regression on observed scores; (2) logistic regression using SEM; and (3) logistic regression on factor scores. These methods differ with respect to the theoretical meaning of the predictors ξ_1 and ξ_2 and in how they handle the measurement error in the item scores.

Logistic regression on observed scores

The approach most often encountered in introductory statistics textbooks involves modeling the interaction using logistic regression where the predictors are observed scores. According to this method, the terms ξ_1 and ξ_2 in Equation 1 can typically be seen as an unweighted sum of all questionnaire item scores measuring a construct. The interaction term is typically constructed by multiplying the mean-centered sum scores of the two constructs constituting the interaction (e.g. see Field, 2010, p. 279; or Tabachnick & Fidell, 2007, p. 442). Using observed scores in a regression analysis tacitly assumes that these scores are a perfectly reliable measure of the construct that is supposed to be measured. By assuming this, researchers ignore the measurement error that is often present in questionnaire scores. Ignoring this measurement error leads to reduced standardized associations in subsequent analyses, a phenomenon known as attenuation bias (Spearman, 1904). Such attenuation bias is especially problematic when modeling interactions between two observed scores containing random measurement error. When multiplying unreliable observed scores to construct the interaction term, the measurement error of the resulting product variable is larger than the sum of the two parts, because the measurement error present in each of the two scores also gets multiplied rather than summed.

Logistic regression using SEM

A second approach to model interaction effects fits a structural equation modeling framework, combining a latent variable measurement model to model errors with a structural model expressing the relations between these latent variables. According to this SEM approach, the terms ξ_1 and ξ_2 in Equation 1 can be seen as the unidimensional latent variables representing the construct of interest. Within a SEM framework, for each latent variable (ξ) a separate measurement model relates the vector of observed item scores ($\boldsymbol{x}$) to the latent construct based on a vector of factor loadings ($\boldsymbol{\Lambda}_x$) and a vector denoting measurement error ($\boldsymbol{\delta}$):

$$\boldsymbol{x} = \boldsymbol{\Lambda}_x \xi + \boldsymbol{\delta}. \qquad (2)$$

Equations (1) and (2) together define the structural equation model, where (1) describes the structural model and (2) the measurement model. All parameters in a SEM model are simultaneously estimated. Note that within a SEM context a binary observed outcome can also be modeled as a latent variable with a single binary indicator by fixing the item's factor loading and residual variance to specific values (Hayduk & Littvay, 2012). However, in the present study we used the observed binary outcomes in our SEM, similar to how the observed binary outcomes are modeled using the sum score and factor score regression methods.

There are different approaches to model interactions within a SEM framework. In its basic form, interaction effects are modeled by means of the product of two latent variables. However, within a latent variable modeling framework, interaction (and other non-linear) effects can also be modeled using various other techniques. First, the *indicant product* approach (Kenny & Judd, 1984) multiplies the item scores of items loading on the first construct involved in the interaction, with the item scores of all possible combinations of items of the other construct involved in the interaction. These multiplications constitute new items that are modeled to load on a new latent interaction variable. For example, when two latent constructs are each measured with three items, nine different multiplications are possible between the item scores of these two sets of three items. Hence, nine new observed variables are created that will load on the new latent interaction variable.

The structural model then includes both the effects of the two original latent variables as well the new interaction variable ξ_{int}, thus replacing the $\xi_1\xi_2$ by ξ_{int} in Equation 1. Two main disadvantages of this method are (1) that the number of items loading on the new latent interaction variable can quickly get very large, and (2) that the method additionally requires the specification of a set of complex parameter constraints. To solve these problems, Marsh, Wen & Hau (2004) proposed an *unconstrained approach* that no longer needed the complex parameter constraints and that modeled the latent interaction variable based on the same number of items as each latent variable involved in the interaction. In a simulation study, the authors showed that their unconstrained approach performed better than the traditional constrained approach.

Another popular method to model latent interaction effects is the Latent Moderated Structural equations (LMS) approach (Klein & Moosbrugger, 2000). This approach does not require new items loading on a latent interaction variable, but makes use of mixture modeling to express the non-normality of an interaction effect. Due to their multiplicative nature, the product scores representing interaction effects often show non-normal distributions, even when the latent variables involved in the interaction are themselves normally distributed. LMS models interaction effects by representing the joint distribution of the indicator variables as a mixture of normal distributions. The method assumes that the indicators of the latent predictor variables show a multivariate normal distribution. Earlier simulation studies showed LMS to perform very well when item scores are normally distributed, yet a violation of this assumption caused LMS to show more bias than for instance the *unconstrained approach* (Marsh et al, 2004; Cham, West, Ma & Aiken, 2012).

Some other recently developed latent variable methods make use of mixture modeling to handle non-normally distributed latent variables. Examples include the Nonlinear Structural Equation Mixture Modeling approach (NSEMM; Kelava, Nagengast & Brandt, 2014) and an approach making use of Bayesian finite mixture models (Kelava & Nagengast, 2012). Although both approaches appear to be useful when modeling latent interaction effects, in the current study we decided to use the LMS approach, as it is commonly used by applied researchers, and is the default method implemented in the Mplus software (Muthén & Muthén, 1998-2010). Moreover, in a previous simulation study (***Chapter 6***) we found that

LMS outperformed two *unconstrained* approaches with respect to minimizing bias and maximizing power when estimating the interaction between two continuous latent variables on a continuous outcome variable, even when the latent traits were non-normally distributed.

Logistic regression on factor scores

Although latent variable modeling is often considered the preferred choice when analyzing associations between variables containing measurement error, latent variable approaches do have some disadvantages. First, they often require a large sample size, especially when the models become more complex. Furthermore, because latent variables typically estimate all the parameters in a single step, increasingly complex models can produce unstable parameter estimates (Devlieger, Talloen & Rosseel, 2019). To overcome these problems, researchers may use a two-step approach, called factor score regression (Lu, Kwan, Thomas & Cedzynski, 2012).

This approach is similar to the logistic regression on observed scores, but the observed scores of the predictor variables are replaced with factor scores estimated in a separate latent variable model. Accordingly, the terms ξ_1 and ξ_2 in Equation 1 can be seen as the estimated factor scores of the constructs of interest. Factor score logistic regression differs from latent variable modeling approach in that it first estimates for each person the factor scores from the items scores (Step 1), and then uses those estimated factor scores in a logistic regression (Step 2). Factor score regression requires a smaller sample size than structural equation models. In a recent simulation study, Devlieger and Rosseel (2017) showed that factor score regression can be a suitable alternative to structural equation modeling when estimating the effect of continuous latent predictor variables on continuous latent outcome variables.

Within a factor score regression framework there exist several techniques to estimate the factor scores that can be used in a subsequent regression analysis. Devlieger, Mayer and Rosseel (2016) showed that some of these techniques are inherently biased depending on whether factor scores or observed scores are used for the predictor or outcome variables.

The authors compared four approaches to estimate factor scores: the (1) *Regression* method (Thurstone, 1935), (2) *Bartlett* method (Bartlett, 1937), (3) *Bias avoiding* method (Skrondal & Laake, 2001), and (4) *Bias correcting* method (Croon, 2002). Devlieger, Mayer and Rosseel (2016) analytically showed that the latter two approaches should be unbiased, independent of whether factor scores or observed scores are used for the predictor or outcome variables. Their analysis also indicated that the *Bartlett* method is expected to only show bias when factor scores are used for the predictor variable(s). The reverse is true for the *Regression* method, which is expected to only show bias when factor scores are used for the outcome variable(s). The authors confirmed these predictions in a Monte Carlo simulation study where factor scores were used for both the predictor and outcome variables.

In the current study, the outcome variables concern different types of cardiac events that are directly observed. The predictor variables NA and SI could be considered unobserved latent variables. This means that the *Bartlett* approach is expected to show bias, while the remaining three approaches to estimate factor scores should be unbiased. Of those three approaches we decided to use the *Regression* approach in our study, because this is both one of the best known approaches and one of the easiest to implement by applied researchers.

$$\boldsymbol{A}_{\xi} = var(\xi)\boldsymbol{\Lambda}_{x}\boldsymbol{\Sigma}_{x}^{-1} \qquad (3)$$

Equation 3 shows how this *Regression* approach estimates the factor scores of a latent variable ($\boldsymbol{A}_{\xi}$) based on the variance of the latent factor $var(\xi)$, the vector with factor loadings ($\boldsymbol{\Lambda}_{x}$) and the inverse of the model implied covariance matrix of the indicator variables ($\boldsymbol{\Sigma}_{x}^{-1}$). The last step of the factor score regression method is to use these estimated factor scores (and their interaction effect) as predictors in a logistic regression analysis.

Modeling skewed ordinal item scores

In clinical research, questionnaire item scores often show positively skewed distributions (Reise & Waller, 2009), with lower response categories being more often endorsed than higher response categories. Such skewed distributions can be problematic if the statistical method used to analyze the data assumes these data to be normally distributed. Although linear regression assumes the prediction residuals to be normally distributed, logistic regression does not impose such distributional assumptions. Nevertheless, the continuous predictors in a logistic regression are assumed to have a linear relation with the logit of the binary outcome variable (Field, 2010).

In latent variable modeling, an assumption of maximum likelihood (ML) estimation applied to factor analysis is that the observed item scores are continuous and normally distributed (Bollen, 1989, pp. 131-134). Analyzing skewed ordinal item scores as if they are continuous and normally distributed can lead to biased parameter estimates and underestimated standard errors, increasing the chance of false positive conclusions about the significance of these estimates, especially if there are five or fewer response categories (Dolan, 1994; Muthén & Kaplan, 1985; Kline, 2010; Rhemtulla, Brosseau-Liard & Savalei, 2012). Researchers could avoid this bias by using modeling techniques that do not assume the item scores to be normally distributed, such as (1) item response models (Rasch, 1960; Birnbaum, 1968; Samejima, 1997) or (2) factor models that estimate the parameters using weighted least squares (WLS) estimation and fit the model to the polychoric correlation matrix rather than the product moment correlation matrix.

Another popular method to handle skewed ordinal item scores is (3) using ML estimation with robust standard errors and a robust statistic, such as the Satorra-Bentler correction (also known as MLR estimation; Satorra & Bentler, 1988). In a previous Monte Carlo simulation study (Lodder et al., 2019), we found that when item scores are ordinal and non-normally distributed, using MLR estimation adequately controls the false positive rate when estimating the interaction between two latent variables on a continuous latent outcome variable, while WLS estimation resulted in an inflated false positive rate. Therefore, in the present study we chose to estimate the parameters of our latent variable models using MLR estimation. In earlier work (Lodder et al., 2019), we also showed that a linear regression on

sum scores resulted in negatively biased interaction effects as well as inflated false positive rates. It is not yet clear, however, whether these results also apply to regression models with a binary and manifest outcome, especially because logistic regression, as opposed to linear regression, does not assume the prediction residuals to be normally distributed.

According to an earlier simulation study, latent variable as well as factor score regression methods showed a negative bias in the standard errors of the structural regression coefficients when the item scores were non-normally distributed (Devlieger et al., 2016). However, this simulation study only investigated the main effect of continuous latent predictor on a continuous latent outcome, raising the issue of whether skewed item scores also result in biased standard errors when estimating the interaction effect of two continuous latent variables on a binary and manifest outcome.

The main goal of the present study is to compare several popular methods to model interactions between two continuous latent variables on a binary and manifest outcome variable. It is not yet known how these methods perform in terms of bias, accuracy and precision, especially when the item scores underlying the two continuous constructs are non-normally distributed and measured on an ordinal scale. We aim to shed more light on this issue using both a Monte Carlo simulation as well as an empirical illustration.

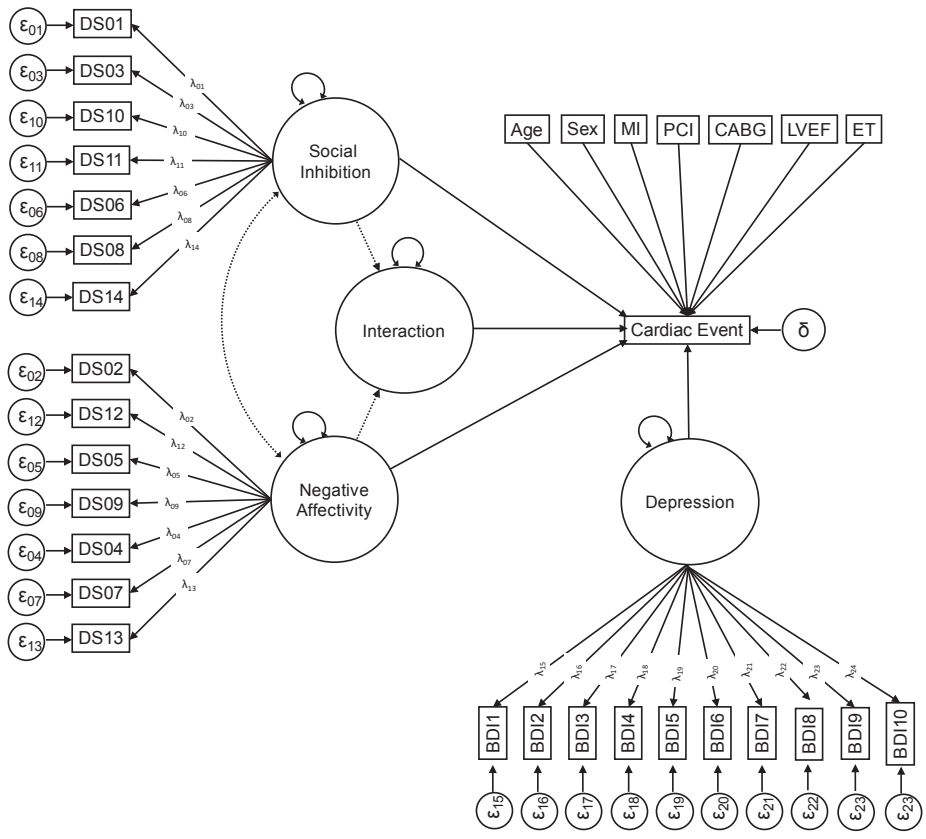
Study overview

In Study 1, our empirical illustration, we investigate the effect of Type D personality on the two endpoints *major cardiac event* (MACE) and *myocardial infarction* (MI) separately, and after adjusting for theoretically important confounding factors (see Denollet et al. 2013). Figure 1 shows the structural equation model of the relation between Type D Personality (modeled as the interaction between NA and SI) and Cardiac Events, while controlling for the covariates depression, age, gender, MI at baseline, percutaneous coronary intervention (PCI) at baseline, coronary artery bypass grafting (CABG) at baseline, left ventricular ejection fracture (LVEF), and poor exercise tolerance (ET).

As the Type D personality effect is hypothesized to reflect an interaction between its components negative affect and social inhibition (Smith, 2011), we studied four methods to

model this interaction effect: (1) logistic regression, modeling the interaction as a multiplication of sum scores, (2) logistic regression, modeling the interaction as a multiplication of factor scores, (3) latent logistic regression, modeling the interaction using the LMS approach in a structural equation model treating the ordinal item scores as continuous, and (4) latent logistic regression using LMS, but modeling the item scores at their appropriate measurement level (ordered categorical).

Figure 1: Structural Equation Model of the Relation between Type D Personality (modeled as the interaction between NA and SI) and Cardiac Events, while controlling for the covariates Depression, Age, Sex, MI at baseline, Percutaneous Coronary Intervention (PCI) at baseline, Coronary Artery Bypass Grafting (CABG) at baseline, Left Ventricular Ejection Fracture (LVEF), and Poor Exercise Tolerance (ET).



NA and SI, the two constructs involved in the interaction, are measured with multiple items measured on an ordinal and occasionally positively skewed scale. It is still unclear how these interaction models perform with respect to bias and precision when the observed outcome is binary and the item scores of the exogenous latent variables are ordinal and non-normally distributed. Therefore, in Study 2, we aimed to answer this question by conducting a Monte Carlo simulation investigating to what extent the models used in Study 1 provided accurate and stable parameter estimates, while varying across (1) sample size, (2) the reliability of the NA and SI scales (3) the size of the interaction effect, (4) skewness in the latent NA and SI traits, and (5) skewness in the item scores (while the latent NA and SI traits are normally distributed). This simulation should provide valuable information on the bias and accuracy of these popular methods to study interaction effects on dichotomous outcomes.

STUDY 1: EMPIRICAL ILLUSTRATION

Method

Participants

Secondary data were used of a study by Denollet et al. (2013), containing a sample of 541 patients with cardiovascular disease. The mean age of these patients was 58.7 years (SD = 10.5) and 87% were men. The patients were included from Antwerp University Hospital Belgium between January 1998 and December 2005. The Medical Ethics Committee of the hospital approved of the study protocol (5/48/193). Patients filled out the psychological questionnaires at baseline and after 5 years each patient (or family) was contacted to determine the end points (see below). All participants gave informed consent.

Measures

End points

We investigated two different but related endpoints marking the occurrence of a cardiac event: The *major cardiac event* (MACE) and *cardiac death or myocardial infarction* (MI). The endpoints are related in such a way that every MI is a MACE, but not every MACE is an MI.

Time-to-event data were not available for these endpoints, explaining our choice for logistic regression rather than cox regression.

Type D Personality

The traits underlying Type D personality (NA and SI) were measured using the DS14 questionnaire (Denollet, 2005). Each trait was measured with a scale consisting of seven questions with five ordinal response categories ranging from "false" (0) to "true" (4). The DS14 has been validated in several populations (Denollet, 2005) and several studies showed a two-factor structure to best fit the data (Nefs, Pouwer & Denollet, 2012; Romppel, Herrmann-Lingen, Vesper & Grande, 2012). In our sample both NA and SI showed a coefficient alpha of .87. See ***Appendix P*** for more information on the DS14 item scores.

Depression

Depressive symptoms were measured using the ten-item version of the Beck Depression Inventory (BDI10; Denollet, Martens, Smit & Burg, 2009), with each item having three ordinal response categories ranging from 0 through 2. Because depressive symptoms may confound the relation between Type D personality and cardiac events, we included the BDI10 score as a covariate in our models. The BDI10 has been validated in both the general- and a post-MI population (Denollet et al., 2009). In our sample, the coefficient alpha of the BDI10 was .83. In our latent variable analyses we also added a measurement model for the BDI10 to adjust the Type D personality effect for the *latent* depression score.

Model building

For the logistic regression on sum scores method, the interaction variable was constructed by multiplying the mean-centered NA and SI sum scores. To account for uncertainty in the factor scores when estimating the standard errors, the R-package boot (version 1.3-24; Canty & Ripley, 2019) was used to bootstrap the standard errors using 1000 bootstrap samples. To identify our latent variable models, we fixed the first factor loading of each latent trait to a value of one. For each of the two cardiac endpoints separately, we fitted four nested models where we regressed the endpoints on the covariates and on Type D personality. In Model 0 we only included the covariates age, sex, depression, ET, LVEF, MI at baseline, CABG at baseline and PCI at baseline. Subsequently, in Model 1 we added the main

effects of NA and SI. To make sure the interaction effect between NA and SI would not merely reflect an unmodeled quadratic effect (MacCallum & Mar, 1995; ***Chapter 3***), we added the quadratic effects of NA and SI in Model 2. Finally, in Model 3 we included the interaction between NA and SI. We assessed the interaction between NA and SI according to four different methods: (1) Logistic regression of sum scores using maximum likelihood (ML) estimation, (2) Logistic regression of cardiac events on factor scores using robust maximum likelihood (MLR) estimation to estimate the parameters in the NA and SI measurement models, (3) latent logistic regression using LMS with MLR estimation (treating ordinal items as continuous), and (4) latent logistic regression using LMS with MLR estimation, modeling the ordered categorical items using polychoric threshold parameters. Although earlier research has advocated to use weighted least squares estimation (Flora & Curran, 2004) for ordered categorical SEM, the software used to estimate latent interaction effects according the LMS method only allowed for MLR estimation.

Model fit

For our latent variable models, model fit was assessed by inspecting the Akaike (1974) Information Criterion (AIC; Akaike, 1974), the Bayesian Information Criterion (BIC; Schwarz, 1978), the sample size adjusted BIC (SABIC). When numerical integration is required (e.g. when modeling latent interactions with the LMS method), means, variances, and covariances are not sufficient statistics for model estimation and chi-square and related fit statistics are not available. As a result, for the LMS method the Mplus software does not report the chi-square statistics and related indices (RMSEA, CFI, TLI). We therefore decided to report these fit indices only to assess the fit of the correlated measurement models. With respect to the fit of the interaction model, we compared the fit of the nested models by conducting a difference test based on the log likelihood values of the two compared models. For LMS these log likelihood values are adjusted based on scaling correction factors obtained with MLR estimation. P-values smaller than .05 were considered statistically significant.

Software

We conducted the hierarchical logistic regression analysis in IBM SPSS Statistics 23. We used the Mplus software (Version 8; Muthén & Muthén, 1998-2010) to analyze our latent interaction models. All other analyses were conducted using the freely available R

programming software (Version 3.2.3; R development Core Team, 2008), including the R-package Lavaan (Version 0.6.1; Rosseel, 2012) to model the factor score logistic regression and the R-package Amelia II (Version 1.7.5; Honaker, King & Blackwell, 2011) to handle any missing data. Based on guidelines by Lodder (2013), we used multiple imputation to impute any missing values in the DS-14 or BDI10 item scores and we imputed 10 datasets. Parameter estimates and fit indices were pooled across imputed datasets. SPSS reported pooled results for the sum score regression models. The pooling of the factor score regression models was done using the R-package semTools (Jorgensen, Pornprasertmanit, Schoemann, & Rosseel, 2019). Lastly, Mplus provided the pooled results for the latent interaction models (Asparouhov & Muthen, 2010b). As only a few observations were missing, the test statistics and fit indices were similar across All syntax files are available at this project's open science framework page: https://osf.io/yhvnp/.

Hypotheses

We first expected to reproduce the findings of Denollet et al. (2013), who originally analyzed this dataset using hierarchical logistic regression analysis on the NA and SI sum scores. However, Denollet et al. (2013) did not adjust for quadratic NA and SI effects and ignored the missing values when computing the NA, SI and BDI10 sum scores. We did not expect this to have a major impact on the results. Our second expectation was that the four methods to model the interaction between NA and SI would produce approximately similar results. However, we expected the logistic regression method based on observed sum scores to show smaller effects compared to the latent variable and factor score logistic regressions because it does not take into account the measurement error present in the observed item scores.

RESULTS

Of all 514 participants, 113 patients (20.89%) experienced a MACE. Of those 113 MACE patients, 47 (8.69%) experienced MI. On 6 items of the DS14, 8 patients showed a total of 9 missing values. On 10 items of the BDI, 13 patients showed a total of 31 missing values. We used multiple imputation to impute these missing values before running our main analyses. Little's (1988) MCAR test suggested that the missing values are likely missing completely at random, $\chi^2(255) = 289.62$, $p = .067$. For the MACE endpoint, Tables 1 and 2 show the results of the sum score and factor score methods to model interaction effects for the MACE and MI endpoints, while Tables 3 and 4 shows those results for the two SEM LMS methods. The latent variables NA and SI both showed adequate factor indeterminacy values (NA = .948; SI = .940). To estimate the factor scores required for the factor score regression, a correlated three factor model (NA, SI, depression) was estimated and factor scores were computed for further analyses. The CFA showed a reasonable yet suboptimal fit to the data ($\chi^2(239) = 637.775$, $p < .001$; RMSEA = .056, 95% CI = [.050, .061]; CFI = .907; TLI = .892).

Major Cardiac Events (MACE)

Logistic regression on sum scores

According to the logistic regression using the NA and SI sum scores, the -2 log likelihood difference test indicated that after adjusting for covariates, the model including the interaction between NA and SI on MACE fitted the data better than the models without the interaction term, $c^2(1) = 6.489$, $p = .011$. The main- and quadratic effects of NA and SI failed to reach significance in all tested models. In Model 3, the interaction between NA and SI on MACE was statistically significant (OR = 1.411, 95% CI = [1.063, 1.873]). This effect is reasonably similar to the effect found in the original study (OR = 1.36, 95% CI = [1.11, 1.67]).

Logistic regression on factor scores

First, the correlated three factor model (NA, SI, depression) showed a less than optimal fit to the data, ($\chi^2(249) = 779.347$, $p < .001$; RMSEA = .063, 95% CI = [.058, .067]; CFI = .867; TLI = .852). The factor scores were saved and subsequently used in the logistic regression to predict MACE. The residual deviance difference test indicated that after adjusting for

covariates, the model including the interaction between NA and SI on MACE yielded a better fit than the models without the interaction term ($\chi^2(1) = 6.070$, $p = .014$). The main- and quadratic effects of NA and SI failed to reach significance in all tested models. In Model 3, the interaction between NA and SI on MACE was statistically significant (OR = 1.484, 95% CI = [1.030, 2.140]).

Logistic regression using SEM LMS (continuous)

When using SEM to estimate the latent interaction using the LMS approach, treating ordinal items as continuous, the -2 log likelihood (-2LL) difference tests preferred the model with covariates only (Model 0) to all other models (including the main-, quadratic-, and interaction effects of NA and SI). Based on the AIC we would choose the model including the interaction term, yet the BIC preferred the model with covariates only, while the SABIC preferred both models equally. Inspection of the regression coefficients of the latent interaction models showed non-significant main- and quadratic effects for NA and SI, yet their interaction was significant (OR = 1.582, 95% CI = [1.042, 2.402]). Although the model including the interaction did not show the best model fit, the estimated interaction effect points in a similar direction as the regular logistic regression effect reported above.

Logistic regression using SEM LMS (categorical)

When modeling the ordered categorical items not as continuous, but at their appropriate ordinal measurement level, both the -2LL difference test as well as the AIC, BIC and SABIC indicated that the model including the latent interaction between NA and SI best fitted the data. The estimated regression coefficients revealed non-significant main- and quadratic effects for NA and SI, but a significant interaction (OR = 1.85, 95% CI = [1.11, 3.08]).

Myocardial Infarction (MI)

Logistic regression on sum scores

According to the logistic regression using the NA and SI sum scores, the -2 log likelihood difference test indicated that after adjusting for covariates, the model including the interaction between NA and SI on MI fitted the data better than the models without the interaction term, $\chi^2(1) = 11.12$, $p = .001$. The main effects of NA and SI failed to reach significance in all tested models. However, in Model 3 there was a negative significant

Table 1: Association between Type D personality and Major Adverse Cardiac Events (MACE) and Myocardial Infarction (MI), modeling the logistic interaction between NA and SI as the product of sum scores.

	Sum score logistic regression		
	Model 1	Model 2	Model 3
MACE			
-2 log likelihood	509.13	507.24	500.75
Difference test ‡	$\chi^2(2)=2.58$	$\chi^2(2)=1.88$	$\chi^2(1)=6.49$*
Type D Personality			
NA	0.84 (0.63, 1.12)	0.81 (0.60, 1.09)	0.79 (0.58, 1.08)
SI	1.17 (0.93, 1.48)	1.12 (0.88, 1.43)	1.14 (0.88, 1.48)
NA^2	-	1.08 (0.89, 1.31)	0.95 (0.75, 1.19)
SI^2	-	1.10 (0.91, 1.33)	0.97 (0.78, 1.22)
NA x SI	-	-	1.41 (1.06, 1.87)*
MI			
-2 log likelihood	271.86	268.97	257.85
Difference test ‡	$\chi^2(2)=0.11$	$\chi^2(2)=2.88$	$\chi^2(1)=11.12$*
Type D Personality			
NA	0.94 (0.76, 1.16)	0.84 (0.61, 1.17)	0.88 (0.74, 1.06)
SI	1.04 (0.87, 1.24)	1.09 (0.90, 1.33)	1.04 (0.81, 1.33)
NA^2	-	1.24 (1.12, 1.38)*	0.91 (0.75, 1.10)
SI^2	-	0.87 (0.64, 1.19)	0.53 (0.31, .90)*
NA x SI	-	-	2.47 (1.27, 4.81)*

All effects were adjusted for the covariates Age, Sex, MI at baseline, Percutaneous Coronary Intervention at baseline, Coronary Artery Bypass Grafting at baseline, Left Ventricular Ejection Fracture, Poor Exercise Tolerance and Depression. All effects are reported as exponentiated regression coefficients, interpretable as odds ratios. A regression coefficient is statistically significant if the confidence interval does not include an odds ratio of 1.

* $p < .05$.

‡ This row shows the likelihood ratio test between a model and the nested previous model. The difference in residual deviance is chi-square distributed.

Table 2: Association between Type D personality and Major Adverse Cardiac Events (MACE) and Myocardial Infarction (MI), modeling the logistic interaction between NA and SI as the product of factor scores.

	Factor score logistic regression		
	Model 1	Model 2	Model 3
MACE			
-2 log likelihood	512.19	508.29	502.22
Difference test ‡	$\chi^2(2)=1.99$	$\chi^2(2)=3.89$	$\chi^2(1)=6.07$*
Type D Personality			
NA	0.77 (0.49, 1.20)	0.70 (0.43, 1.12)	0.68 (0.41, 1.11)
SI	1.18 (0.90, 1.53)	1.13 (0.85, 1.50)	1.16 (0.85, 1.58)
NA^2	-	1.15 (0.93, 1.41)	0.97 (0.74, 1.26)
SI^2	-	1.11 (0.91, 1.36)	0.92 (0.69, 1.21)
NA x SI	-	-	1.48 (1.03, 2.14)*
MI			
-2 log likelihood	272.87	269.22	259.80
Difference test ‡	$\chi^2(2)=0.42$	$\chi^2(2)=3.65$	$\chi^2(1)=9.42$*
Type D Personality			
NA	0.82 (0.42, 1.58)	0.70 (0.34, 1.39)	0.75 (0.35, 1.62)
SI	1.03 (0.69, 1.55)	1.10 (0.70, 1.73)	1.03 (0.57, 1.86)
NA^2	-	1.31 (0.96, 1.77)	0.90 (0.56, 1.43)
SI^2	-	0.86 (0.60, 1.23)	0.48 (0.24, .94)*
NA x SI	-	-	2.65 (1.11, 6.28)*

All effects were adjusted for the covariates Age, Sex, MI at baseline, Percutaneous Coronary Intervention at baseline, Coronary Artery Bypass Grafting at baseline, Left Ventricular Ejection Fracture, Poor Exercise Tolerance and Depression. All effects are reported as exponentiated regression coefficients, interpretable as odds ratios. A regression coefficient is statistically significant if the confidence interval does not include an odds ratio of 1.

* $p < .05$.

‡ This row shows the likelihood ratio test between a model and the nested previous model. The difference in residual deviance is chi-square distributed.

Table 3: Association between Type D personality and Major Cardiac Events (**MACE**) and Myocardial Infarction (**MI**), modeling the interaction with SEM and the LMS approach, modeling the ordinal items as continuous.

	SEM LMS (continuous)		
	Model 1	Model 2	Model 3
MACE			
-2 log likelihood	-13749.54	-13747.61	-13744.86
Difference test ‡	$\chi^2(2)=1.04$	$\chi^2(2)=1.94$	$\chi^2(1)=2.74$
AIC	27671.08	27671.21	27667.73
BIC	28040.32	28049.03	28049.84
SABIC	27767.32	27769.69	27767.32
Type D Personality			
NA	0.76 (0.47, 1.23)	0.69 (0.42, 1.11)	0.66 (0.39, 1.1)
SI	1.19 (0.92, 1.54)	1.14 (0.86, 1.51)	1.19 (0.85, 1.66)
NA^2	-	1.16 (0.94, 1.43)	0.95 (0.71, 1.28)
SI^2	-	1.12 (0.91, 1.37)	0.9 (0.66, 1.21)
NA x SI	-	-	1.58 (1.04, 2.4)*
MI			
-2 log likelihood	-13629.62	-13630.12	-13623.19
Difference test ‡	$\chi^2(2)=0.21$	$\chi^2(2)=0.50$	$\chi^2(1)=6.93$*
AIC	27431.23	27436.24	27424.38
BIC	27800.47	27814.06	27806.50
SABIC	27527.47	27534.72	27523.98
Type D Personality			
NA	0.81 (0.41, 1.61)	0.7 (0.34, 1.42)	0.75 (0.34, 1.68)
SI	1.03 (0.75, 1.42)	1.1 (0.73, 1.66)	1.03 (0.51, 2.08)
NA^2	-	1.32 (1.01, 1.72)*	0.82 (0.5, 1.34)
SI^2	-	0.85 (0.6, 1.19)	0.36 (0.15, 0.85)
NA x SI	-	-	3.71 (1.25, 11.02)*

All effects were adjusted for the covariates Age, Sex, MI at baseline, Percutaneous Coronary Intervention at baseline, Coronary Artery Bypass Grafting at baseline, Left Ventricular Ejection Fracture, Poor Exercise Tolerance and Depression. All effects are reported as exponentiated regression coefficients, interpretable as odds ratios. A regression coefficient is statistically significant if the confidence interval does not include an odds ratio of 1.

* $p < .05$.

‡ This row shows the likelihood ratio test between a model and the nested previous model. The difference in -2LL is chi-square distributed.

Table 4: Association between Type D personality and Major Cardiac Events (**MACE**) and Myocardial Infarction (**MI**), modeling the interaction with SEM and the LMS approach, modeling the ordinal items as ordered categorical.

	SEM LMS (categorical)		
	Model 1	Model 2	Model 3
MACE			
-2 log likelihood	-12493.70	-12491.78	-12487.36
Difference test ‡	$\chi^2(2)=0.98$	$\chi^2(2)=1.92$	$\chi^2(1)=4.42$*
AIC	25215.40	25215.56	25208.72
BIC	25704.85	25713.60	25711.04
SABIC	25342.97	25345.37	25339.65
Type D Personality			
NA	0.8 (0.51, 1.24)	0.81 (0.53, 1.24)	0.76 (0.47, 1.24)
SI	1.15 (0.88, 1.51)	1.15 (0.89, 1.48)	1.17 (0.84, 1.63)
NA^2	-	1.12 (0.94, 1.34)	0.85 (0.62, 1.17)
SI^2	-	1.13 (0.94, 1.34)	0.85 (0.62, 1.17)
NA x SI	-	-	1.85 (1.11, 3.08)*
MI			
-2 log likelihood	-12372.21	-12370.93	-12367.70
Difference test ‡	$\chi^2(2)=0.12$	$\chi^2(2)=1.28$	$\chi^2(1)=3.23$
AIC	24972.42	24973.86	24969.40
BIC	25461.87	25471.85	25471.33
SABIC	25100.00	25103.67	25100.33
Type D Personality			
NA	0.89 (0.47, 1.68)	0.9 (0.5, 1.62)	0.92 (0.44, 1.93)
SI	0.99 (0.71, 1.38)	0.99 (0.67, 1.45)	0.9 (0.44, 1.83)
NA^2	-	1.24 (1.01, 1.52)*	0.72 (0.44, 1.18)
SI^2	-	0.87 (0.65, 1.17)	0.36 (0.13, 1.02)
NA x SI	-	-	4.36 (1.2, 15.83)*

All effects were adjusted for the covariates Age, Sex, MI at baseline, Percutaneous Coronary Intervention at baseline, Coronary Artery Bypass Grafting at baseline, Left Ventricular Ejection Fracture, Poor Exercise Tolerance and Depression. All effects are reported as exponentiated regression coefficients, interpretable as odds ratios. A regression coefficient is statistically significant if the confidence interval does not include an odds ratio of 1.

* $p < .05$.

‡ This row shows the likelihood ratio test between a model and the nested previous model. The difference in -2LL is chi-square distributed.

quadratic effect of SI (OR = 0.532, 95% CI = [0.314, 0.900]). This model also showed a significant interaction between NA and SI on MI (OR = 2.469, 95% CI = [1.269, 4.805]), indicating that the odds of having a cardiac death or myocardial infarction was 2.469 times larger for people with score X+1 on the interaction term compared to people who score X on the interaction term. This effect is substantially larger than the effect found in the original study (OR = 1.48, 95% CI = [1.11, 1.96]). Further re-analysis of the original data showed that this larger effect results from first including the quadratic effects in our model.

Logistic regression on factor scores

The residual deviance difference test indicated that after adjusting for covariates, the model including the interaction between NA and SI on MI fitted the data better than the models without the interaction term, $\chi^2(1) = 9.421$, $p = .002$. The main- effects of NA and SI failed to reach significance in all tested models. In Model 3, both the interaction between NA and SI on MI (OR = 2.647, 95% CI = [1.110, 6.280]) and the negative quadratic effect of SI (OR = 0.476, 95% CI = [0.239, 0.940]) were statistically significant.

Logistic regression using SEM LMS (continuous)

According to the logistic regression conducted within a SEM framework and using the LMS approach to model the interaction between the latent variables NA and SI, the -2 log likelihood difference test indicated that after adjusting for covariates, the model including the interaction between NA and SI on MI fitted the data better than the models without the interaction term, $\chi^2(1) = 6.929$, $p = .008$. However, although based on the AIC and SABIC we would choose the model including the interaction term, the BIC preferred the model with covariates only. Inspection of the regression coefficients of the latent interaction models showed non-significant main effects for NA and SI, yet their interaction was significant (OR = 3.706, 95% CI = [1.246, 11.021]). There also turned out to be a significant negative quadratic effect for SI (OR = 0.359, 95% CI = [0.151, 0.853]), suggesting that higher SI scores are associated with a lower chance on MI and that this effect gets stronger at higher levels of SI. Although the model including the interaction did not show the best model fit according to the BIC and SABIC, the estimated interaction effect points in a similar direction as the regular logistic regression effect reported above.

Logistic regression using SEM LMS (categorical)

When modeling the ordered categorical items not as continuous, but at their appropriate ordinal measurement level, both the -2LL difference test as well as the BIC and SABIC indicated that the model with covariates only (without any of the Type D effects) best fitted the data. The AIC preferred model 3, including the interaction effect between NA and SI. This model showed a significant latent interaction effect between NA and SI (OR = 4.36, 95% CI = [1.20, 15.83]), yet the confidence interval was very broad, likely due to the small number of observed MI events.

Synthesis

The current findings support the hypothesis that Type D personality, operationalized as the interaction between NA and SI, predicts the occurrence of a major cardiac events, and even more strongly predicts the occurrence of a cardiac death or myocardial infarction. Figure 2 shows the factor score distributions of NA, SI & NA x SI, separately for people who did (red curve) or did not (green curve) have a cardiac event. The top row shows the results for the MACE endpoint and the bottom row for the MI endpoint. The top of each plot shows the result of the two sample Kolmogorov-Smirnoff test for the equality of the two distributions.

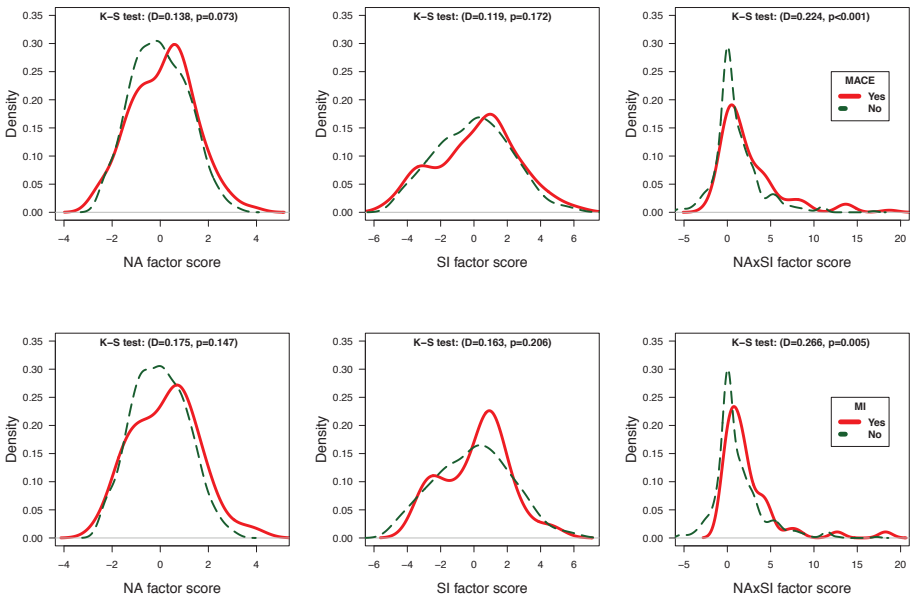
As the main effects of NA and SI were not associated with both MACE and MI, we expect the factor score distributions of people with- and without a cardiac event to be samples from the same population distribution. Conversely, because the interaction between NA and SI significantly predicted the occurrence of both MACE and MI, we expect the factor score distributions of people with- and without a cardiac event to differ. The results of the two sample Kolmogorov-Smirnoff tests confirmed these expectations. The null hypothesis of equal factor score distributions for people with- and without an event could not be rejected for the NA and SI factor scores, but was rejected for the NA x SI interaction factor scores for both MACE (D=.224, $p < .001$) and MI (D=.266, $p=.005$).

In general, the four methods to model the interaction effect agreed on the direction and the statistical significance of the interaction effect, but differed in their estimated size of the interaction. As expected, the effects of the sum score logistic regression analyses were smaller than those of the other two methods, likely because this approach did not take into

account the measurement error in the item scores. However, the interaction effects estimated by the two latent logistic regression approaches using LMS were substantially larger than the effects estimated by the factor score logistic regression.

An earlier study comparing four factor score regression methods (Devlieger, Mayer & Rosseel, 2016) analytically and numerically showed that the *Regression* method should not be biased when factor scores are only used for the predictors and the outcome is observed. However, their simulation study involved *continuous* outcome variables and did not include interaction effects, making it difficult to generalize these findings to the models applied in the current study. In our second study we aimed to shed more light on this issue by conducting a Monte Carlo simulation study, comparing the bias, power and false positives of four methods used to model interaction effects on binary observed outcomes.

Figure 2: Factor score distributions of NA, SI, & NA x SI, separately for people who did (red/straight curve) or did not (green/broken curve) have cardiac events. The top row shows the results for the MACE endpoint and the bottom row for the MI endpoint. Each plot shows the result of the two sample Kolmogorov-Smirnoff test for the equality of the distributions.



STUDY 2: MONTE CARLO SIMULATION

METHOD

Design

In our simulation study, we varied five different design parameters: scale (continuous, ordinal), scale reliability (0.60, 0.87), skewness (no skewness, moderate latent skewness, high latent skewness, moderate item skewness, large item skewness), the size of the interaction (0, 0.154, 0.308, 0.616) and sample size (250, 500, 1000). This resulted in a fully-crossed factorial design with 2x2x5x4x3=240 conditions. We simulated 500 datasets in each of those 240 conditions and we analyzed each of those datasets with three different methods to model interaction effects.

Design Parameter 1: Scale measurement level

The first design parameter was the scale level of the simulated DS14 items. We either simulated continuous item scores or ordinal item scores with five response categories (0-4 Likert scale).

Design Parameter 2: Scale Reliability

The second design parameter was the reliability of the NA and SI scales. We simulated both scales to either have a reliability observed in Study 1 (0.87), or a substantially lower reliability of 0.60. This lower reliability was achieved by multiplying all estimated Study 1 factor loadings of the NA and SI measurement models with 0.7.

Design Parameter 3: Latent skewness

The third design parameter was the amount of skewness in the distribution of the latent traits NA and SI. We used the method of Vale and Maurelli (1983) as implemented in the R-package *fungible* (version 1.5; Waller, 2016), to simulate a multivariate distribution of NA and SI. We varied across three latent skewness values (0, 2 and 3; with corresponding kurtosis values 0, 7 and 21), while retaining the product moment correlation between NA

and SI (Study 1 estimate of .428). Besides generating skewness on the latent level, in some ordinal item scenarios skewness was also generated on the item level while keeping the underlying latent traits normally distributed (see Data simulation paragraph).

Design Parameter 4: Size of Interaction

The fourth design parameter indicated the strength of the interaction between NA and SI on depression. We based the true size of the interaction on the standardized regression coefficient of the estimated interaction effect in Study 1 according to the LMS approach (β=0.308). In our simulation, we allowed the interaction to be either absent (0), half the size of the Study 1 interaction (0.154), the exact size of the Study 1 interaction (0.308), or twice the size of the Study 1 interaction (0.616).

Design Parameter 5: Sample size

The fifth design parameter indicated the sample size of the simulated dataset. We varied across small (n=250), medium (n=500) and large (n=1000) sample size conditions. The number of items loading on a construct was not manipulated in this simulation. As this may also affect the power to detect interaction effects, readers are advised to interpret our findings using the number of cases per variable (n/p ratio) rather than the raw sample size. For the sample sizes included in our simulation (250, 500, 1000) the n/p ratios are approximately 18, 36 and 72.

Data simulation

For each of the 240 conditions, we simulated 500 datasets containing scores on items measuring the constructs NA (7 items) and SI (7 items). We generated data using the parameter estimates (i.e., factor loadings, latent (co)variances, regression coefficients, thresholds and error variances) of the latent interaction model in Study 1[3]. First, we randomly sampled vectors of NA and SI latent trait scores according to the multivariate skew

[3] We used the parameter estimates resulting from the LMS method, because this is the default method in Mplus to model interaction effects. We focused on the MACE endpoint, because the MI endpoint shows a relatively low proportion of participants with an event compared to the MACE endpoint. For reasons of simplicity, we did not study the effects of covariates in this Monte Carlo simulation. We therefore fitted the Study 1 LMS model without covariates and used those parameter estimates as input for our simulation study.

distribution, given the NA and SI (co)variance(s) from Study 1 and given the skewness design parameter. Second, continuous item scores for each individual (*i*) and for each item (*j*) measuring the traits (*t)* NA or SI were obtained as follows:

$$Y_{ij} = \lambda_{jt}\, \xi_{ti} + \epsilon_{ij} \qquad (4)$$

As input we used a matrix with individual NA and SI trait scores (**Ξ**), the factor loading matrix retrieved from Study 1 (**Λ**), and residual error matrix (**Θ**) based on a multivariate normal distribution with a mean vector of zeroes and a diagonal covariance matrix with variances retrieved from the output in Study 1. In line with earlier research (Flora & Curran, 2004), for ordinal scenarios with latent skewness we transformed these continuous item scores into ordinal scores using the symmetric Case 1 thresholds (-1.645, -0.643, 0.643, 1.645) proposed by Muthén and Kaplan (1985)[4]. In scenarios where skewness was generated on the item level while keeping the latent NA and SI traits normally distributed, we used the Case 2 (-1.645, -0.643, 0.643, 1.645) and Case 3 (-1.645, -0.643, 0.643, 1.645) thresholds to transform the normally distributed continuous item scores into skewed ordinal item scores.

To simulate the cardiac event scores, we had to take into account that these scores depended on the scores of both the latent NA and SI traits as well as the interaction between NA and SI. Therefore, we used Equation 5 to compute the event probabilities based on a logistic regression model:

$$\rho_{Ci} = \frac{1}{1+e^{-(\beta_0+\beta_{SI}\xi_{SIi}+\beta_{NA}\xi_{NAi}+\beta_{NA*SI}\xi_{NAi}\xi_{SIi})}} \qquad (5)$$

In Equation 5, ρ_{Ci} denotes the probability of a cardiac event score of individual *i*. β_0 represents the intercept and the three other $\beta's$ denote the standardized regression coefficients of the structural logistic regression of MACE on NA (ξ_{NAi}), SI (ξ_{SIi}) and the interaction between NA and SI ($\xi_{NAi} * \xi_{SIi}$). Lastly, *e* denotes the base of natural logarithms.

[4] The average estimated skewness of the simulated skewed ordinal item scores was .16 and .20 when the generated latent skewness was equal to 2 and 3, respectively.

Data analysis

After simulating 500 datasets in each of the 240 conditions, we analyzed each dataset according to the same methods used in Study 1: (1) Logistic regression of sum scores using maximum likelihood (ML) estimation, (2) Logistic regression of factor scores based on the Regression method and using MLR to estimate the parameters in the NA and SI measurement models, and (3) Latent variable interaction using LMS with MLR estimation while treating the ordinal items as continuous. In scenarios involving ordinal item scores we also estimated the latent interaction effect using (4) LMS with MLR estimation while modeling the items as ordered categories using polychoric threshold parameters. We implemented all latent interaction models in Mplus and conducted the simulation using the R-package MplusAutomation (Hallquist & Wiley, 2011). The R-script of this simulation study is available at this project's open science framework page: https://osf.io/yhvnp/.

Outcome measures

Our main outcomes were bias, precision and accuracy of the estimated interaction effects. The bias was computed as the difference between the mean of the parameter estimate across 500 replications and the true value; that is, the β values used to generate the data. We used the standard deviation and corresponding 95% variability interval of the parameters estimates across the 500 replications as a measure of precision. We also assessed the mean squared error (MSE) as a measure of accuracy, where MSE is defined as the squared distance between the estimated value of the interaction effect and the true value of the interaction effect, averaged across 500 replications. Additionally, we computed for each condition, a 95% confidence interval coverage rate, as the percentage of replications where the 95% confidence interval of a single estimated interaction effect contained the true value of the interaction effect. Lastly, we determined in each condition the percentage of replication with a significant estimated interaction effects, in order to shed light on the power to detect non-zero interaction effects and the percentage of false positives when the true interaction effect was equal to zero.

Expectations

We expected the sum score method to underestimate the true size of the interaction because this approach does not take into account the measurement error inherent in the

item scores and this may attenuate the true association between the latent construct and the manifest binary outcome. Though factor score methods account for measurement error when estimating parameters in the measurement model, the coefficients in the factor score regression may still be contaminated because the sample moments of the factor score deviate from the true moments. For linear factor score regression the biased moments can cancel out, but it is unclear whether this also generalizes to a logistic regression analysis. We therefore investigated this empirically in our simulation. Lastly, we did not expect LMS to suffer from attenuated regression coefficients, because it takes into account the measurement error of the item scores when estimating the interaction between the latent variables. In line with earlier simulation studies (Kelava & Nagengast, 2012; Kelava, Nagengast & Brandt, 2014) we expected LMS to perform well with large sample sizes (500 or higher) and no skewness in the latent traits. We also expected the factor score approach to perform equally to the LMS method, because earlier research showed this to be the case for linear models involving main effects rather than interactions (Devlieger & Rosseel, 2017).

RESULTS

For both the sum score and factor score methods, the convergence rate was 100% in all simulation conditions. Although for the LMS approach most conditions also showed 100% convergence rates, conditions with large skewness sometimes resulted in non-convergence (convergence rates were 98.4% or higher). No non-positive definite covariance matrices were encountered. We removed the non-converged solutions from further analyses.

Tables 5, 6 and 7 show for all 240 simulation conditions the relative bias in the estimated interaction effect averaged across all 500 replications. Table 8, 9, and 10 report for all simulation conditions the percentage of replications in which the 95% confidence interval of a single estimated interaction effect contained the true value of the interaction effect. ***Appendix Q*** contains six supplemental Tables. Tables Q1, Q2 and Q3 present the bias in the estimated interaction effect in terms of the mean squared error (the squared difference between the true interaction effect and the estimated interaction, averaged across 500

replications). Tables Q4, Q5, and Q6 show the power of detecting the interaction effect (the percentage of significant interactions in case of a non-zero true interaction) and false positive rates (the percentage of significant interaction effects when a true interaction effect was absent).

For each method, the average bias in the estimated interaction effect is visualized in Figures 3 (continuous item scores), 4 (ordinal item scores with latent skewness), and 5 (ordinal item scores with skewness generated on the item level). ***Appendix Q*** contains similar figures for both the NA and SI main effects. Each figure contains 9 plots, where the rows represent the different sample size conditions and the columns the different skewness conditions. In each of the 9 plots the x-axis shows the size of the true interaction effect and the y-axis the bias in the estimated standardized regression coefficient. The three different methods used to model the interaction effect are visualized with varying colors and shapes of the data points. Filled shapes indicate high reliability conditions, while open shapes represent conditions with poor reliability. Each data point corresponds to the bias of a particular method in the estimated interaction effect, averaged across 500 replications. The error bars represent the 95% confidence interval of the mean estimated bias.

Absolute and relative bias

Inspection of Figures 3, 4 and 5 (absolute bias) and Tables 5, 6, and 7 (relative bias) reveals some interesting patterns. As expected, the sum score method underestimated the interaction effect in every simulation condition and this effect became more pronounced as the reliability decreased, as the skewness increased, as the size of the true interaction increased, or when item scores were ordinal. Interestingly, a similar pattern was observed for the factor score method, suggesting that this method did not appropriately adjust for measurement error when estimating interaction effects in a generalized linear modeling context.

Figure 3: Comparison of three methods to estimate interaction effects between NA and SI on a binary outcome in scenarios with **continuous items**, varying over the true interaction size (x axis), reliability of the NA and SI scales, amount of **latent skewness** (columns) and sample size (rows). Each data point shows the mean bias (including 95% confidence interval) in the standardized regression coefficient of the estimated interaction effect.

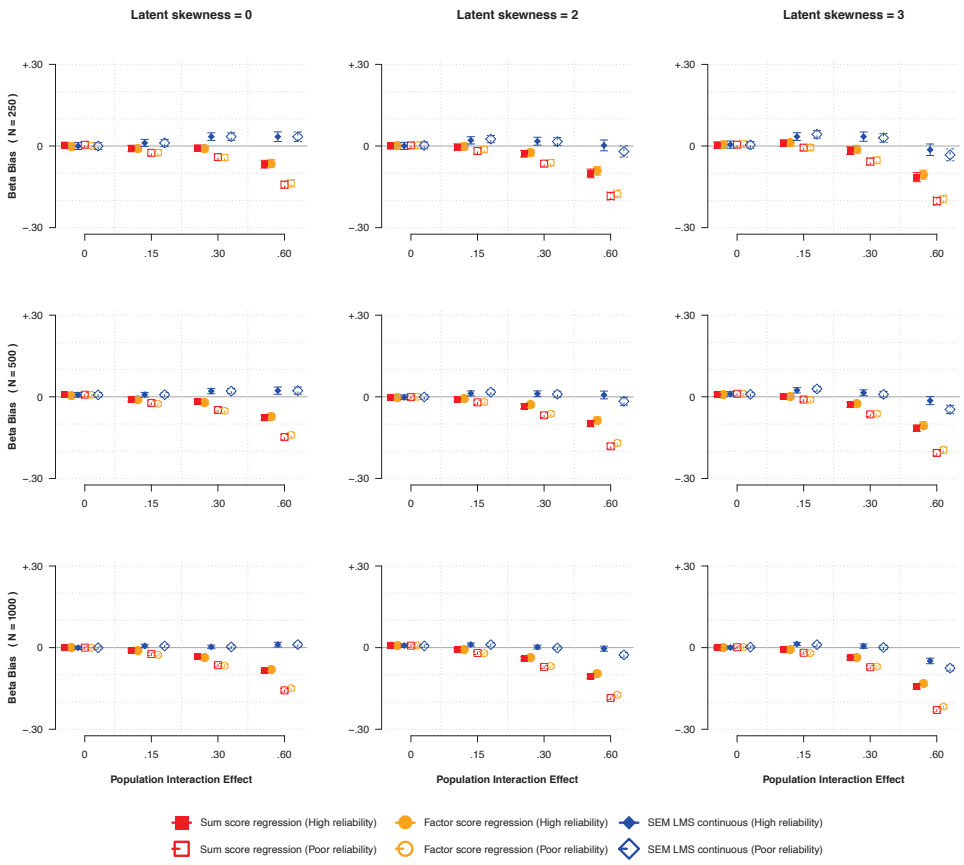


Figure 4: Comparison of four methods to estimate interaction effects between NA and SI on a binary outcome in scenarios with **ordinal items**, varying over the true interaction size (x axis), reliability of the NA and SI scales, amount of **latent skewness** (columns) and sample size (rows). Each data point shows the mean bias (including 95% confidence interval) in the standardized regression coefficient of the estimated interaction effect.

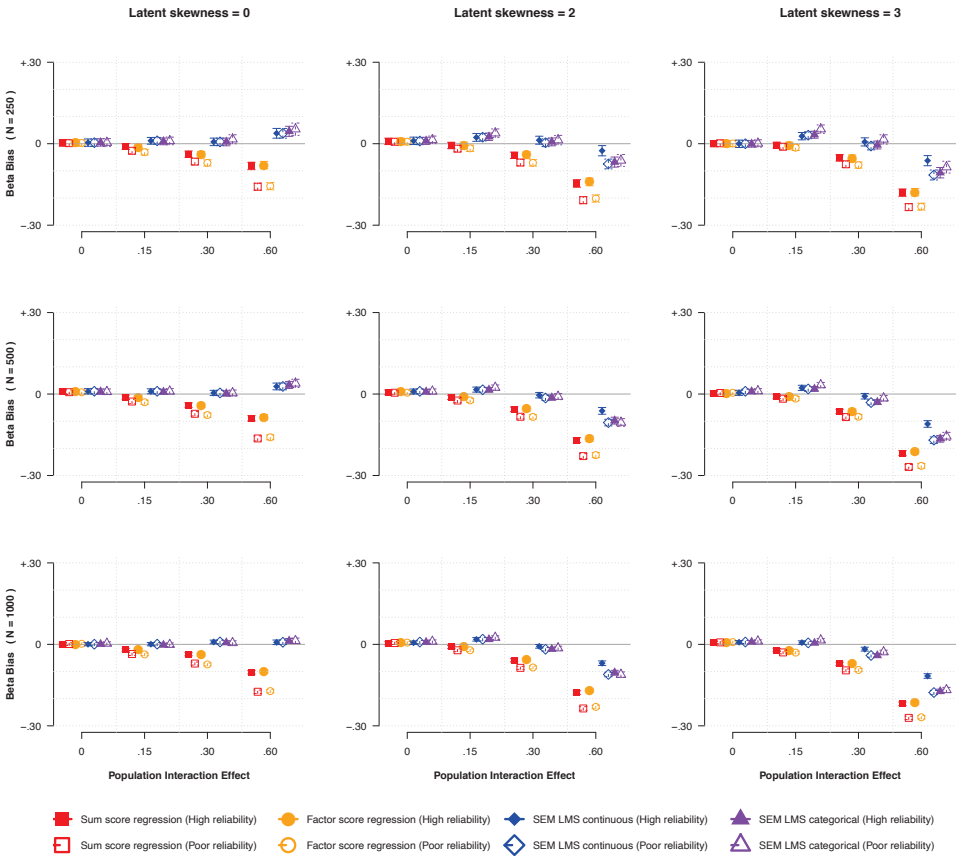


Figure 5: Comparison of four methods to estimate interaction effects between NA and SI on a binary outcome in scenarios with **ordinal items**, varying over the true interaction size (x axis), reliability of the NA and SI scales, amount **item skewness** (columns) and sample size (rows). Each data point shows the mean bias (including 95% confidence interval) in the standardized regression coefficient of the estimated interaction effect.

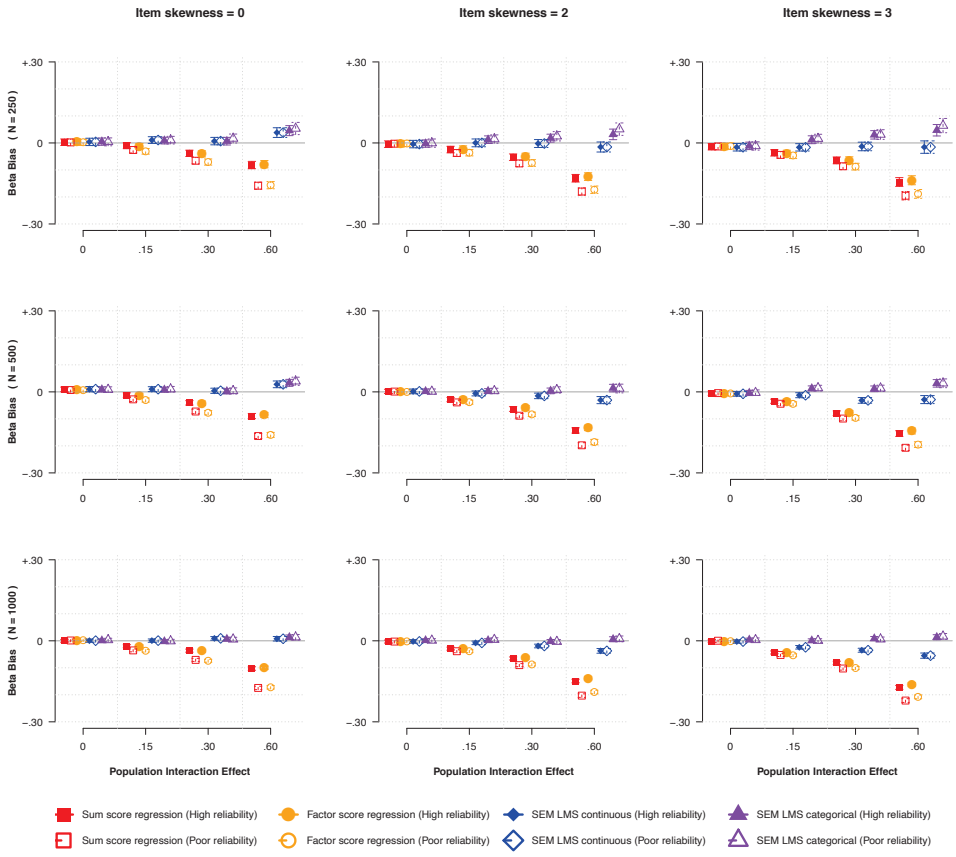


Table 5: Relative bias in the estimated interaction effects for continuous item scores and skewness generated at the latent level. Bold faced cells indicate acceptable bias (<10%).

N	N/p ratio	Skewness	Interaction	Sum score regression		Factor score regression		SEM LMS continuous	
				α =.87	α =.60	α =.87	α =.60	α =.87	α =.60
250	18	0	0.15	**-5.91**	-16.19	**-5.47**	-15.87	**7.41**	**7.41**
			0.31	**-2.57**	-13.24	**-2.97**	-13.77	11.21	11.21
			0.62	-10.86	-23.08	-10.29	-22.21	**5.55**	**5.55**
		2	0.15	**-2.4**	-11.9	**-0.32**	**-8.63**	13.55	16.49
			0.31	**-9.23**	-21.01	**-7.98**	-19.87	**5.74**	**5.27**
			0.62	-16.34	-29.82	-14.85	-28.52	**0.36**	**-3.25**
		3	0.15	**6.91**	**-3.87**	**7.29**	**-3.59**	22.41	27.88
			0.31	**-5.56**	-18.56	**-3.74**	-17.03	11.18	**9.75**
			0.62	-18.58	-32.86	-17.08	-31.67	**-2.23**	**-5.27**
500	36	0	0.15	**-5.56**	-14.63	**-7.14**	-16.58	**5.03**	**5.03**
			0.31	**-5.55**	-15.66	**-6.52**	-16.66	**6.85**	**6.85**
			0.62	-12.54	-24.05	-11.69	-22.79	**3.69**	**3.69**
		2	0.15	**-5.51**	-13.25	**-4.68**	-12.4	**8.19**	11.22
			0.31	-11.23	-22	**-9.47**	-20.14	**3.65**	**3.11**
			0.62	-15.98	-29.54	-13.87	-27.54	**1.08**	**-2.66**
		3	0.15	**1.84**	**-6.47**	**1.62**	**-6.96**	15.53	18.78
			0.31	**-9.44**	-20.94	**-8.57**	-20.18	**4.85**	**2.92**
			0.62	-18.68	-33.48	-16.85	-31.73	**-2.28**	**-7.55**
100	72	0	0.15	**-6.71**	-15.44	**-7.94**	-16.95	**3.96**	**3.96**
			0.31	-10.94	-20.91	-11.54	-21.55	**0.87**	**0.87**
			0.62	-13.9	-25.51	-13.08	-24.23	**1.81**	**1.81**
		2	0.15	**-5**	-12.92	**-5.2**	-13.24	**7.34**	**7.09**
			0.31	-12.95	-23.01	-11.98	-22.16	**0.63**	**-0.44**
			0.62	-17.15	-29.99	-15.2	-28.12	**-0.63**	**-4.33**
		3	0.15	**-3.76**	-12.28	**-4.13**	-13	**8.76**	**7.05**
			0.31	-12.37	-23.36	-11.48	-22.58	**1.76**	**0.22**
			0.62	-23.41	-37.23	-21.35	-35.14	**-7.95**	-12.23

Table 6: Relative bias in the estimated interaction effects for ordinal item scores and skewness generated at the latent level. Bold faced cells indicate acceptable bias (<10%).

N	N/p ratio	Skewness	Interaction	Sum score regression		Factor score regression		SEM LMS continuous		SEM LMS categorical	
				a =.87	a =.60	a =.87	a =.60	a =.87	a =.60	a =.87	a =.60
250	18	0	0.15	**-6.66**	-16.93	**-8.06**	-20.18	**6.95**	**6.95**	**5.58**	**6.68**
			0.31	-12.46	-21.4	-13.15	-22.94	**2.21**	**2.21**	**2.12**	**5.32**
			0.62	-13.29	-25.7	-12.77	-25.39	**6.22**	**6.22**	**7.41**	**8.68**
		2	0.15	**-4.23**	-11.85	**-3.88**	-10.96	15.02	15.54	16.66	25.01
			0.31	-13.83	-22.51	-13.5	-22.9	**4.21**	**1.17**	**2.2**	**4.29**
			0.62	-23.75	-33.75	-22.84	-32.74	**-4.19**	-12.11	-10.98	-10
		3	0.15	**-3.28**	**-7.19**	**-3.42**	**-9.39**	18.02	20.31	21.57	34.73
			0.31	-16.98	-24.61	-17.24	-25.54	**2.17**	**-2.73**	**-1.53**	**5.11**
			0.62	-29.25	-37.9	-28.96	-37.62	-10.13	-18.7	-17.33	-14.01
500	36	0	0.15	**-7.68**	-17.93	**-8.23**	-19.62	**6.5**	**6.5**	**5.08**	**6.35**
			0.31	-13.19	-23.6	-13.6	-25.1	**1.31**	**1.31**	**0.67**	**1.15**
			0.62	-14.53	-26.55	-13.88	-25.85	**4.53**	**4.53**	**5.43**	**6.37**
		2	0.15	**-7.75**	-15.54	**-6.83**	-15.1	10.57	10.15	10.21	15.42
			0.31	-18.23	-27.16	-17.75	-27.34	**-1.37**	**-4.73**	**-4.22**	**-3.07**
			0.62	-27.59	-37.1	-26.73	-36.52	-10.07	-17.06	-15.89	-17.03
		3	0.15	**-5.65**	-11.18	**-5.24**	-10.89	14.77	12.2	12.27	21.37
			0.31	-20.8	-27.53	-20.4	-27.1	**-2.62**	-10.16	**-9.67**	**-5.09**
			0.62	-35.27	-43.57	-34.41	-42.93	-17.89	-27.56	-26.74	-25.3
100	72	0	0.15	-12.96	-23.04	-13.13	-23.88	**0.48**	**0.48**	**-1.55**	**-0.78**
			0.31	-11.82	-22.93	-12.05	-24.01	**2.91**	**2.91**	**2.1**	**1.75**
			0.62	-16.65	-28.33	-16.18	-27.95	**1.24**	**1.24**	**2.12**	**2.14**
		2	0.15	**-6.19**	-14.15	**-5.4**	-14.05	11.96	12.32	11.89	16.14
			0.31	-19	-28.07	-18.28	-27.62	**-2.36**	**-5.87**	**-5.69**	**-4.59**
			0.62	-28.81	-38.31	-27.69	-37.38	-11.26	-17.99	-16.91	-18.09
		3	0.15	-13.57	-19.24	-13.64	-19.64	**4.2**	**3.51**	**3.22**	**9.76**
			0.31	-23.28	-31.1	-22.7	-30.61	**-5.73**	-13.29	-13.3	**-9.28**
			0.62	-35.48	-43.95	-34.99	-43.57	-18.91	-28.79	-28.1	-27.26

Table 7: Relative bias in the estimated interaction effects for ordinal item scores and skewness generated at the item level. Bold faced cells indicate acceptable bias (<10%).

N	N/p ratio	Skewness	Interaction	Sum score regression		Factor score regression		SEM LMS continuous		SEM LMS categorical	
				α =.87	α =.60	α =.87	α =.60	α =.87	α =.60	α =.87	α =.60
250	18	0	0.15	**-6.66**	-16.93	**-8.06**	-20.18	**6.95**	10.5	**5.58**	**6.68**
			0.31	-12.46	-21.4	-13.15	-22.94	**2.21**	**9.96**	**2.12**	**5.32**
			0.62	-13.29	-25.7	-12.77	-25.39	**6.22**	**8.17**	**7.41**	**8.68**
		2	0.15	-15.9	-24.19	-14.94	-23.55	**0.17**	**-0.05**	**7.83**	**9.34**
			0.31	-17.1	-24.6	-16.22	-24.11	**-0.68**	**6.10**	**5.77**	**8.04**
			0.62	-21.22	-29.21	-20.24	-28.11	**-2.43**	**8.00**	**5.15**	**8.28**
		3	0.15	-23.9	-28.72	-24.89	-30.16	-10.6	**-7.12**	**7.7**	10.2
			0.31	-20.97	-28.04	-20.8	-28.49	**-4.22**	**2.72**	**9.74**	10.1
			0.62	-23.48	-31.73	-22.42	-30.62	**-2.5**	**7.20**	**7.63**	10.43
500	36	0	0.15	**-7.68**	-17.93	**-8.23**	-19.62	**6.5**	**3.37**	**5.08**	**6.35**
			0.31	-13.19	-23.6	-13.6	-25.1	**1.31**	**2.95**	**0.67**	**1.15**
			0.62	-14.53	-26.55	-13.88	-25.85	**4.53**	**5.98**	**5.43**	**6.37**
		2	0.15	-18.43	-25.3	-17.58	-24.73	**-3.74**	**2.46**	**1.79**	**2.2**
			0.31	-20.61	-28.66	-19.29	-27.03	**-4.81**	**-3.11**	**1.17**	**2.35**
			0.62	-23.19	-32.1	-21.65	-30.17	**-4.95**	**1.08**	**2.28**	**2.02**
		3	0.15	-22.43	-28.52	-22.36	-28.81	**-8.04**	**-9.01**	**8.33**	**9.08**
			0.31	-25.67	-32.07	-25.09	-31.26	-10.24	**-3.96**	**3.71**	**4.25**
			0.62	-25.05	-33.61	-23.52	-31.7	**-4.64**	**-1.45**	**5.12**	**5.15**
100	72	0	0.15	-12.96	-23.04	-13.13	-23.88	**0.48**	**2.50**	**-1.55**	**-0.78**
			0.31	-11.82	-22.93	-12.05	-24.01	**2.91**	**1.74**	**2.1**	**1.75**
			0.62	-16.65	-28.33	-16.18	-27.95	**1.24**	**2.04**	**2.12**	**2.14**
		2	0.15	-18.86	-25.06	-18.42	-24.95	**-4.97**	**-4.67**	**0.93**	**2.66**
			0.31	-21.23	-29.23	-20.31	-28.38	**-6.22**	**-1.52**	**-0.66**	**-1.04**
			0.62	-24.17	-32.94	-22.43	-30.77	**-6.05**	**-0.33**	**0.93**	**1.16**
		3	0.15	-28.04	-34.01	-28.36	-34.83	-15.57	-10.7	**0.44**	**0.29**
			0.31	-26.48	-33.29	-25.8	-32.68	-11.33	**-8.80**	**2.7**	**1.84**
			0.62	-28	-35.91	-26.32	-33.7	**-8.92**	**-3.12**	**2.15**	**2.64**

Table 8: 95% confidence interval coverage rates of the estimated interaction effects for continuous item scores with skewness generated at the latent level. Bold faced cells indicate acceptable coverage.

N	N/p	Skewness	Interaction	Sum scores		Factor scores		LMS	
				α =.87	α =.60	α =.87	α =.60	α =.87	α =.60
250	18	0	0	**94.6**	**95**	**95**	**95**	**94.6**	**94.6**
			0.15	**94.8**	**94**	**95.2**	**96**	**95.4**	**95.4**
			0.31	**95.6**	**94.8**	**96.8**	**95.2**	**96.6**	**96.6**
			0.62	91.6	80	91.4	82.8	**96.4**	**96.4**
		2	0	**95.4**	**96**	**96.8**	97.8	**96.6**	**96.6**
			0.15	**96.8**	**95.8**	97.8	98	**96.4**	**95.8**
			0.31	**95**	**93.2**	**96.8**	**95.2**	**96**	**96.4**
			0.62	85.8	74.2	90.8	75.6	**94.8**	**93.4**
		3	0	**95.4**	**95.2**	**97**	**97.2**	**96**	**96.8**
			0.15	**94.4**	**95**	**97**	**96.6**	**95.4**	**94.2**
			0.31	**94**	90.4	**97.2**	**94.4**	**96.2**	**93.2**
			0.62	84.6	70.2	89.4	74	**92.8**	91.2
500	36	0	0	**95.4**	**95.4**	**96**	**96.8**	**96.2**	**96.2**
			0.15	**95.2**	**93**	**96.2**	**95.2**	**95.8**	**95.8**
			0.31	**94**	90.4	**93**	90.8	**95.2**	**95.2**
			0.62	86.6	68.4	88.2	70.8	**94.4**	**94.4**
		2	0	**94**	**94**	**94.8**	**95**	**94.6**	**93.8**
			0.15	**94.6**	**93.6**	**95.8**	**94.8**	**95.6**	**94.8**
			0.31	**92.8**	87.8	**93.8**	88.6	**95.2**	**94.8**
			0.62	84	60.2	86.8	66.4	**95.8**	**93.4**
		3	0	**96**	**95.8**	**97.4**	**96.8**	**96.4**	**95.58**
			0.15	**94.8**	**94.6**	**96**	**94.6**	**94.6**	**93.4**
			0.31	**93.4**	88.6	**93.2**	90.4	**94.79**	**95.99**
			0.62	83	57.2	86.6	61.4	**93.2**	90.78
100	72	0	0	**94**	**94.2**	**94.6**	**95.2**	**94.8**	**94.8**
			0.15	**93.8**	91.8	**93.2**	92	**93.8**	**93.8**
			0.31	91.8	82.6	91.4	81.2	**96.2**	**96.2**
			0.62	76.8	46	79.6	47.8	**95.6**	**95.6**
		2	0	**93.8**	**94.2**	**93.8**	**94.2**	**93.4**	**94**
			0.15	**95**	**94.4**	**95.8**	**94.4**	**95.6**	**94.6**
			0.31	91.6	82.4	**92.8**	81.4	**95.19**	**95.6**
			0.62	75	34.6	81.2	42.2	**96.4**	**96**
		3	0	**94.8**	**96.2**	**95.6**	**96.6**	**94.35**	**95.92**
			0.15	**94.6**	**93.6**	**95.2**	**93.6**	**95.56**	**94.25**
			0.31	89.8	79.2	90.8	82	**94.72**	**94.46**
			0.62	62.4	22.4	67.2	29	91.94	85.77

Table 9: 95% confidence interval coverage rates of the estimated interaction effects for ordinal item scores with skewness generated at the latent level. Bold faced cells indicate acceptable coverage.

N	N/p ratio	Skew-ness	Inter-actio	Sum scores		Factor scores		LMS		LMS	
				α =.87	α =.60	α =.87	α =.60	α =.87	α =.60	α =.87	α =.60
250	18	0	0	**94.6**	**94.2**	**95.8**	**96.6**	**95.4**	**95.4**	**95.6**	**95.8**
			0.15	**95.6**	**95.8**	**97.2**	**97**	**96.8**	**96.8**	**97.2**	98.2
			0.31	**94.4**	90.6	**95.2**	92.2	**96**	**96**	**95.8**	**95.6**
			0.62	90	79.2	89.8	78.6	**96**	**96**	**95.6**	**96.8**
		2	0	**94.4**	**94.8**	**94.8**	**95.6**	**94.2**	**94.6**	**94.4**	**97**
			0.15	**95**	**94.6**	**96**	**95.6**	**96.4**	**97**	97.6	97.6
			0.31	**93.6**	90.6	**95.2**	92.4	**97**	**96.2**	**96.2**	**96**
			0.62	81.6	68.6	84	72	**93.4**	91	91.6	90.2
		3	0	**95.4**	**96**	**96.8**	**97.2**	**96.6**	**97.4**	**96.6**	**97.2**
			0.15	**97.2**	**96.4**	**97.4**	**97.4**	97.6	97.8	98.2	98.8
			0.31	**93**	90.8	**95.4**	**94**	**96.2**	**94.2**	**94**	**95.4**
			0.62	76	62.2	78.8	66.6	90	83.6	84.8	86.6
500	36	0	0	**93.6**	**93.4**	**94**	**92.6**	**93.4**	**93.4**	**93**	**94**
			0.15	**94.6**	**93.6**	**95.4**	**92.8**	**95.4**	**95.4**	**95.8**	**94.8**
			0.31	**93.2**	87.2	**93**	86	**95.4**	**95.4**	**94.8**	**95.2**
			0.62	84.2	65.2	85.4	64.4	**96.2**	**96.2**	**96.2**	**96.8**
		2	0	**95.2**	**95.6**	**96**	**96.2**	**96.2**	**96.4**	**96.4**	**96**
			0.15	**94.8**	**94.2**	**95**	**94.4**	**95.6**	**96.2**	**96.6**	**97**
			0.31	90	83.8	91.6	84.2	**94.8**	**94.2**	**94.4**	**94.8**
			0.62	60.8	41	65	43.8	90.4	82.4	84.4	85.2
		3	0	**94**	**95.8**	**94.6**	**96.2**	**94.8**	**95.8**	**95.4**	**96**
			0.15	**95.4**	**93.2**	**96.4**	**95.2**	**95.4**	**95.8**	**95.2**	**96.6**
			0.31	88.6	82.4	88.8	85	**95.4**	**92.8**	92	**94**
			0.62	45.4	27.8	49.6	30.6	81.8	64.4	66.6	71.8
100	72	0	0	**95.6**	**95**	**95.4**	**96**	**95.8**	**95.8**	**95.6**	**95.8**
			0.15	**93.6**	92	**94.4**	91	**95.4**	**95.4**	**95.4**	**96.2**
			0.31	91	79.6	92.4	77.4	**97.2**	**97.2**	**97.4**	**95.8**
			0.62	72.2	34.8	75	37.2	**96**	**96**	**96**	**95.6**
		2	0	**95.4**	**95.4**	**95**	**95.8**	**94.6**	**93.8**	**93.8**	**95.6**
			0.15	**94.2**	**94.2**	**95.2**	**93.4**	**95.4**	**95**	**95.2**	**96**
			0.31	85.2	75.8	86.6	73.6	**95.8**	**94.8**	**95**	**95.4**
			0.62	33	14	38.8	14.4	86.6	74	77.6	79.4
		3	0	**96**	**95.8**	**96.2**	**96.2**	**96**	**94.4**	**95.6**	**95.6**
			0.15	**92.6**	91.4	**93.8**	92.2	**94.6**	**95.4**	**95.4**	**95.2**
			0.31	80	69.4	82	69.2	**92.8**	88.2	87.2	91.6
			0.62	18.6	5.4	19.4	6.2	74.8	47.4	49.6	55.2

Table 10: 95% confidence interval coverage rates of the estimated interaction effects for ordinal item scores with skewness generated at the item score level. Bold faced cells indicate acceptable coverage.

N	N/p ratio	Skew-ness	Inter-actio	Sum scores		Factor scores		LMS		LMS	
				α =.87	α =.60	α =.87	α =.60	α =.87	α =.60	α =.87	α =.60
250	18	0	0	**94.6**	**94.2**	**95.8**	**96.6**	**95.4**	**96.2**	**95.6**	**95.8**
			0.15	**95.6**	**95.8**	**97.2**	**97**	**96.8**	97.8	**97.2**	98.2
			0.31	**94.4**	90.6	**95.2**	92.2	**96**	**96**	**95.8**	**95.6**
			0.62	90	79.2	89.8	78.6	**96**	**95.8**	**95.6**	**96.8**
		2	0	**93.4**	**92.6**	**94.4**	**95.2**	**94.8**	**96.8**	**96.4**	**97.4**
			0.15	**94.2**	**92.6**	**95.8**	**93.4**	**95.4**	**96**	**96.4**	**97**
			0.31	**92.6**	89.4	**94.4**	89.6	**95.6**	**95.6**	**96.6**	**96.4**
			0.62	84	73.6	83.8	75.6	**93.2**	**95.2**	**95**	**96**
		3	0	**94**	**94.2**	**95.8**	**95.4**	**95.2**	**95.8**	**96**	**97.4**
			0.15	**95.4**	**93.4**	**95.6**	**94.8**	**97.4**	**97**	**96.4**	**97.4**
			0.31	91.4	89.4	**93**	89.6	**96**	**95.2**	**96.2**	**96.4**
			0.62	76.2	66.2	80.2	69.8	**93.4**	**96**	**95**	**96.6**
500	36	0	0	**93.6**	**93.4**	**94**	**92.6**	**93.4**	**95.6**	**93**	**94**
			0.15	**94.6**	**93.6**	**95.4**	**92.8**	**95.4**	**95.8**	**95.8**	**94.8**
			0.31	**93.2**	87.2	**93**	86	**95.4**	**96**	**94.8**	**95.2**
			0.62	84.2	65.2	85.4	64.4	**96.2**	**94**	**96.2**	**96.8**
		2	0	**96**	**96.6**	**96**	**97.4**	**96.2**	**96**	**97.2**	**96.8**
			0.15	**93**	**93**	**94.6**	**93**	**97**	**96.8**	**95.6**	**97**
			0.31	87.6	81.2	89.6	83.4	**94.6**	**95.6**	**95.2**	**95.4**
			0.62	65.2	47.2	70	52.2	**95**	**93.8**	**95.4**	**94.6**
		3	0	**95.8**	**96**	**95.6**	**96.6**	**95.2**	**96.6**	**93.8**	**96.4**
			0.15	**93.4**	**93**	**94.4**	**92.8**	**95.4**	**96.2**	**96.2**	**97**
			0.31	84.2	79	84.6	80	91.8	**93.8**	**95.8**	**95.6**
			0.62	65.2	50.6	69.8	55.2	91.8	**93.6**	**96**	**96.2**
100	72	0	0	**95.6**	**95**	**95.4**	**96**	**95.8**	**97**	**95.6**	**95.8**
			0.15	**93.6**	92	**94.4**	91	**95.4**	**95.4**	**95.4**	**96.2**
			0.31	91	79.6	92.4	77.4	**97.2**	**96.8**	**97.4**	**95.8**
			0.62	72.2	34.8	75	37.2	**96**	**97.4**	**96**	**95.6**
		2	0	**95.2**	**94.8**	**96.2**	**95.2**	**95.8**	**96.6**	**97**	**96.4**
			0.15	91.2	89.6	91.8	90.2	**94.6**	**96,4**	**95.6**	**95.6**
			0.31	83	71.4	85.2	70.6	**93.8**	**96.2**	**96.2**	**97**
			0.62	47.8	22.2	52.6	27.4	**92.6**	**94.6**	**95.2**	**95**
		3	0	**95.2**	**95.6**	**95.4**	**95.6**	**95.4**	**95.2**	**95.4**	**96**
			0.15	87.6	83.2	87.6	85	92.4	**94**	**94.8**	**95.8**
			0.31	72.4	61.6	75.6	64.2	90.8	**93.8**	**94.6**	**95.4**
			0.62	39.4	19.4	46.2	25.2	88	**93.6**	**95.4**	**95.6**

When item scores were normally distributed, both LMS methods tended to slightly overestimate the interaction effects, especially when the sample size was small (n/p ratio = 18). In general, LMS outperformed the factor score and sum score methods both in terms of average absolute and relative bias in the estimated interaction. As expected, LMS performed especially good at a sample size of 1000 (n/p ratio = 72) and when item scores were normally distributed. As expected, when latent skewness was introduced, both LMS methods underestimated the larger interaction effects, though this effect was still much more apparent for both the sum score and factor score methods. When item scores were ordinal, continuous LMS performed slightly better than categorical LMS when skewness was generated at the latent level. Interestingly, when the latent variables were normally distributed and skewness was introduced at the item level, only the categorical LMS method remained relatively unbiased. It showed similar bias regardless of whether the ordinal item scores were skewed. This pattern was not observed for the LMS method that treated the ordinal item scores as continuous. That method still resulted in underestimated interaction effects when the ordinal item scores were skewed, yet this effect became less pronounced with lower scale reliability.

Confidence interval coverage

All methods showed acceptable coverage probabilities for the smallest interaction effects. The largest interactions resulted in lower coverage for both the sum score and factor score regression methods, especially at larger sample sizes. A possible explanation is that these methods produce biased estimates for the larger interaction effects, and these biased estimates get increasingly narrow confidence intervals due to the larger sample size, resulting in lower coverage rates. Both LMS methods performed best in terms of confidence interval coverage, with acceptable coverage probabilities in almost all simulation conditions, except when skewness was introduced at the latent level, and either the sample size or the true interaction effect was large. When skewness was introduced at the item level rather than at the latent level, the coverage rates of the continuous LMS methods remained suboptimal, while those of categorical LMS were adequate.

Mean squared error

In **Appendix Q**, Tables Q1, Q2, and Q3 present the simulation results with respect to the mean squared error in the estimated interaction effects. For all methods, as the sample size became larger, the mean squared error became smaller, likely because of less variable estimates. At a sample size of 250 (n/p ratio = 18), LMS showed larger mean squared error than the factor score and sum score methods, likely due to the larger variability in the estimated interaction effects. Lowering the reliability of the questionnaires from .87 to .60 only slightly affected the relative bias and mean squared error for both LMS methods, but it largely affected both the sum score and factor score methods, with lower reliability resulting in more bias, especially for larger interaction effects. When item scores were ordinal and skewness was generated at the item level, categorical LMS outperformed continuous LMS, though it performed slightly worse when skewness was generated at the latent level.

Power and false positives

In **Appendix Q**, Tables Q4, Q5, and Q6 present the simulation results with respect to the percentage of statistically significant interaction effects. As expected, for all methods larger sample sizes resulted in more statistical power to detect true interaction effects. With a sample size of 250 (n/p ratio = 18), all methods were only able to detect the largest interaction effects (0.616) with a power of at least 0.80. At a sample size of 500 (n/p ratio = 36), all methods also became able to detect medium interactions (.308) with sufficient power. Having ordinal item scores resulted in slightly less power for all methods compared to using continuous item scores. As the reliability of the questionnaires decreased, the sum score and factor score methods showed less statistical power to detect the interaction effect, while this only affected both LMS methods to a minor extent. For each method the power decreased as the skewness of the item scores increased, regardless of whether the skewness was generated on the item level or on the latent variable level. An exception was the categorical LMS approach, showing only minor power reductions when skewness was introduced at the item level. All methods showed acceptable false positive rates close to the nominal level of 5% in each of the simulation conditions (ranging between 2.6% and 6.2%).

Synthesis

Inspection of the simulation condition that most resembled the circumstances of our empirical study (500 participants, moderately skewed ordinal item scores (skewness = 2), estimated reliability of 0.87 and an interaction effect of β=.308), suggested that SEM according to both LMS methods performed best in terms of power, coverage probability, relative bias and mean squared error in the estimated interaction. Both the sum score method as well as the factor score method showed slightly lower power, lower coverage probabilities and underestimated the true size of the interaction effect. Whether continuous or categorical LMS performed best depends on whether the skewness originated at the latent or item level. If the latent NA and SI traits are skewed, then categorical LMS performed slightly worse than continuous LMS in terms of absolute bias and power, while the reverse was true when the latent traits were normally distributed and the skewness originated at item score level. Given that the latent NA and SI distributions in Figure 2 are not very skewed, we assume the skewed item scores to have originated at the item level. In light of this we consider the categorical LMS approach to perform best in the circumstances of our empirical study and we will therefore base our discussion on these Study 1 results.

DISCUSSION

The goal of this paper was twofold. Our starting point was the empirical question of whether Type D Personality, operationalized as the interaction between its two subcomponents NA and SI, predicts the occurrence of cardiac events in a population of patients suffering from coronary artery disease. We used three different methods to model the interaction effect. Our second goal was to compare the bias, precision and accuracy of these four methods in a Monte Carlo simulation study, in order to shed more light on the inconsistent estimates resulting from the three interaction models used in our empirical study.

As expected, our empirical study showed that Type D personality was associated with the occurrence of major cardiac events, and even more strongly associated with the occurrence of a cardiac death or myocardial infarction. In general, the four methods used to model the

interaction agreed on the direction and the statistical significance of the interaction effect, but differed in their estimated size of the interaction. As expected, the effects of the sum score logistic regression analyses were smaller than the effects estimated by the factor score logistic regression and latent logistic regression methods, likely because the sum score approach did not take into account the measurement error in the item scores. Although as expected the estimates resulting from the latent logistic regression were larger than those of the sum score regression, they were unexpectedly also larger than the estimates produced by the factor score logistic regression. This finding motivated our Monte Carlo simulation study.

In our simulation SEM using LMS to model the interaction effect outperformed all other methods in terms of relative bias when the sample size was large and there was no skewness. This result aligns with our expectations, because LMS assumes that the indicators are multivariate normally distributed (Klein & Moosbrugger, 2000). Indeed, earlier research showed LMS to be biased when the scores of the items loading on the latent exogenous variables are skewed (Kelava & Nagenast, 2012; Kelava, Nagengast & Brandt, 2014; Cham, West, Ma & Aiken, 2012). We replicated this findings by showing that LMS underestimated the larger interaction effects when item scores were skewed due to skewness at the latent variable level. However, when skewness was introduced at the item level rather than at the latent variable level, categorical LMS produced acceptable estimates of the interaction effects at a sample size of 1000. This findings suggests that when sample size is large enough, categorical LMS becomes robust to violations of the assumption of multivariate normally distributed indicators, as long as the underlying latent traits are normally distributed. This robustness does not apply to the continuous LMS method that treats the ordinal item scores as continuous.

As expected, the sum score interaction method in general produced more biased estimates than those of the latent interaction methods. This corroborates earlier research showing that using sum scores may attenuate the estimates of the regression coefficients because using sum scores includes random measurement errors (Busemeyer & Jones, 1983; Embretson, 1996; Kang & Waller, 2005; MacCallum, Zhang, Preacher, & Rucker, 2002).

Because of this finding we recommend researchers not to use the sum score method when analyzing the interaction between two continuous variables on a manifest binary outcome.

Interestingly, the estimates of the factor score interaction more closely resembled those of the sum score method than those of the latent interaction method. Both the factor score method and the SEM LMS account for measurement error when estimating the parameters in the measurement model. However, the factor score regression's two step method may result in contaminated structural regression coefficients if the factor score sample moments are different from the true moments. Though in the context of linear factor score regression this bias cancels out (see for instance Devlieger & Rosseel, 2017), our simulation shows this is not the case in a *generalized* linear modeling context (e.g. with an observed binary outcome). Showing this analytically would be an interesting avenue for future research.

Recommendations

We have a number of recommendations to researchers planning to model the interaction between two continuous variables on an observed binary outcome. First, whenever possible use items measured on a continuous scale, as these typically result in less bias than items measured on an ordinal scale. Second, in case of a large sample size (e.g. $N \geq 500$ or n/p ratio ≥ 36) consider SEM using LMS to estimate a latent interaction model, as these models are the least biased when sample size is large, especially in the absence of skewness. In that situation researchers could either use categorical or continuous LMS, because in line with earlier research (Dolan, 1994; Rhemtulla, Brosseau-Liard & Savalei, 2012) our simulation suggests that normally distributed ordinal item scores with five categories can be considered continuous. Third, when the estimated reliability of the constructs involved in the interaction is not sufficient, SEM results in less biased estimates than the sum score or factor score regression methods. Fourth, latent skewness introduces negative bias in the estimated interactions for most methods. When the skewed item scores are continuous, LMS is the least biased method, but researchers should take into account that this method may underestimate the larger interaction effects. When the skewed item scores are ordinal, consider using categorical LMS when the underlying latent variables are still normally distributed. Although none of the investigate methods produce unbiased interaction effects when the underlying latent variables are skewed, researchers should consider using

continuous LMS in these circumstances, as this method slightly outperforms categorical LMS and the other methods in terms of minimizing bias. However, researchers may also consider modeling the interaction between non-normally distributed latent variables using a Bayesian approach. Although the present study did not focus on this approach, it produced unbiased interaction effects in another simulation study (Kelava & Nagengast, 2012).

Limitations

In our simulation we focused both on items with continuous scales and items with five category ordinal scales. The differences between those scale types was primarily a matter of degree rather than kind. Overall, using ordinal items resulted in less precise and slightly more biased estimates, especially when skewness was high. This is in line with earlier simulation studies showing that when skewness is absent, ordinal items with at least five categories can be treated as continuous items in subsequent analyses (Dolan, 1994; Rhemtulla, Brosseau-Liard & Savalei, 2012). As these studies did not focus on interaction modeling, future research could investigate whether interaction models perform well when the interaction constructs are based on items with a smaller number of ordinal response categories (e.g., 2, 3 and 4).

Another limitation of our simulation is that we did not include covariates in the model for reasons of simplicity. Given that covariates are often part of a statistical model in the medical and behavioral sciences, future simulation studies could assess whether the inclusion of covariates affects the performance of methods used to model interactions.

A further limitation of our study is that we only focused on the LMS method to model the latent interaction effect, while there exist many other methods to model latent interactions, such as the product indicator approach (Kenny & Judd, 1984; Jöreskog & Yang, 1996; Marsh, Hau & Wen, 2004), the two stage least squares approach (Bollen & Paxton, 1998), or mixture modeling (Kelava & Nagenast, 2012; Kelava, Nagengast & Brandt, 2014). In earlier work (Lodder et al., 2019) we compared LMS with two different product indicator approaches in a Monte Carlo simulation study and found that LMS was the least biased method when modeling the interaction between two latent variables on a continuous latent outcome

variable. Future research could aim at extending the results of the present study to other latent interaction models applied within a logistic regression context.

In our study Type D personality was operationalized as the interaction effect between its continuous subcomponents NA and SI. According to Denollet and colleagues (2013), the effect of Type D personality on cardiac outcomes in theory implies that it is the *combination* of having both high NA and high SI that is most detrimental to cardiac health. Smith (2011) interpreted this as saying that the Type D effect is more than the sum of its parts NA and SI, a classic example of synergy. This synergistic effect would imply that it is not the separate NA and SI effects that are essential to Type D personality, but the combined NA and SI effect that is still present above and beyond the sum of their additive effects. Smith (2011) argued that this synergy is modeled statistically by testing an interaction effect while also including the main effects of the variables constituting the interaction. Although in our study we followed Smith (2011) by using a *variable-centered* approach and modeling Type D personality as an interaction between NA and SI, another commonly used method to operationalize Type D personality is a *person-centered* approach that classifies people in subgroups based on whether they have crossed a particular cutoff score for both NA and SI (e.g. Denollet et al., 2013; Denollet et al., 2018; Hillen, 2017). Person-centered approaches are often useful when the data shows substantial amounts of heterogeneity and the effect of interest is present for some people and not for others. However, classifying people in groups based on cutoff values has been criticized because such dichotomization produces less sensitive statistical tests and may result in spurious findings that are not robust against using other cut-off values and do therefore likely not reflect real differences between the groups (MacCallum, Zhang, Preacher & Rucker, 2002; Royston, Altman & Sauerbrei, 2006). However, it is not necessary to classify people based on arbitrary cut-off values, because within a latent variable framework it is possible to use the individual item scores to classify people in a set of distinct latent classes and subsequently use class membership to predict the scores on an outcome variable. Therefore, future research could investigate whether such a person-centered approach is more beneficial when studying Type D personality, than the variable-centered interaction model we used in our study.

Conclusions

When seeing the estimates in our empirical study in the light of the results of our simulation study, we can draw several conclusions. First, given the characteristics and outcomes of our empirical study (500 participants, moderately skewed ordinal item scores and an interaction effect of *b*=.308), the latent interaction model using categorical LMS performed best with respect to minimizing the average bias. It performed slightly better than the continuous LMS method and much better than both the sum score and factor score methods. Those latter two methods produced similar estimates that were both lower than those of the two LMS methods. It is interesting to note that this exact pattern was also found in our empirical study. If we follow the results of the latent interaction model, then we can conclude that Type D personality is a significant predictor of both major cardiac events and an even stronger predictor of cardiac death or myocardial infarction.

In this article we showed that Type D personality is an important risk factor in the occurrence of cardiac events, in line with earlier research on this issue (Denollet et al., 2013; Du et al., 2016; Kupper & Denollet, 2016). We used several statistical interaction models to assess this association, resulting in varying estimates, yet similar conclusions. To the best of our knowledge this study includes the first Monte Carlo simulation comparing the performance of these methods when estimating interaction effects between two continuous variables on an observed binary outcome variable. To our knowledge this is also the first simulation study to show that the mechanism causing the skewed item scores determines what method should be used to model the interaction effects. Although our simulation study was motivated by an issue we encountered in our empirical study, the results are not limited to research on Type D personality. Because our simulation varied over a wide range of design factors, we consider these results to be generalizable to many other research areas involving interactions between continuous latent variables on binary manifest outcomes.

CHAPTER 8

Assessing the temporal stability of psychological constructs: An illustration of Type D personality in relation to anxiety and depression

P. LODDER
H.M. KUPPER
F. MOLS
W.H.M. EMONS
J.M. WICHERTS

Journal of Research in Personality

ABSTRACT

Introduction: Various methods exist to assess the temporal stability of psychological constructs. In this paper we discuss common methods based on a review of the Type D personality literature. Most of these methods ignore the measurement error in the questionnaire scores and most studies do not test the longitudinal measurement invariance assumption. We illustrate how to handle these issues using three longitudinal latent variable models, investigating the temporal stability of Type D personality in relation to depression and anxiety.

Method: We used data from 2625 cancer survivors involving four yearly measurements of depression, anxiety, and the Type D personality traits negative affectivity (NA) and social inhibition (SI). Longitudinal measurement invariance was tested before fitting univariate and multivariate second-order latent growth curve models and latent trait-state-occasion models. These models were designed to handle the skewed and ordinal item scores.

Results: All questionnaires showed longitudinal measurement invariance. Although the univariate growth models suggested temporal stability on group level, there were significant individual differences in the change of NA, depression and anxiety over time. In the multivariate growth models, individual changes in NA correlated with changes in depression and anxiety. The trait-state-occasion models revealed that SI was most trait-like, while NA was least trait-like.

Conclusion: Latent variable models can be used to assess the temporal stability of psychological constructs while handling measurement error and non-normal score distributions. Individual changes in NA covaried with changes in depression and anxiety, suggesting that NA is not purely a stable trait and may be affected by changes in psychological states.

INTRODUCTION

Personality traits are considered relatively enduring sets of behaviors, feelings and thoughts that characterize individuals (Roberts & Mroczek, 2008). These behavioral, emotional, and cognitive patterns develop from an interplay between biological and environmental influences, and were typically thought to remain stable after reaching adulthood (McCrea & Costa, 1994). However, more recent evidence suggests that personality traits may continue to change throughout adulthood and even into old age (Mroczek & Spiro III, 2003; Mroczek, Graham, Turiano, & Aro-Lambo, 2021). Such change in personality can either be normative or non-normative. Normative change is defined as the generalizable patterns of personality development typically seen in most individuals, whereas non-normative change reflects all individual deviation from the normative developmental trajectories (Roberts, Walton & Viechtbauer, 2006). An example of normative change is that people generally become more socially mature (calm, responsible, confident) as they grow older (Roberts & Wood, 2006). However, (epi)genetic influences or environmental factors, such as major life events and work experiences, may alter the course of such normative development in directions unique to how each individual interacts with his or her environment (Roberts, Wood & Caspi, 2008; Leszko, Elleman, Bastarache, Graham & Mroczek, 2016). Although personality can change into adulthood, several studies suggests that both the genetic and environmental influences on personality increase in stability with age (Briley & Tucker-Drob, 2014; Li-Gao et al., 2021).

Throughout the years, several statistical methods have been used to assess the temporal stability of personality traits, or rather the stability of psychological constructs in general (for an overview see De Fruyt et al., 2006). A first distinction can be made between methods assessing absolute change vs. methods that focus on relative change (also known as differential change, Caspi, Roberts & Shiner, 2005). The absolute change perspective involves determining whether an individual or aggregate score on one time point differs from the score at one or more other time points. Absolute change can be assessed for separate individuals (*individual-level absolute stability*) or for groups of individuals such as the entire sample (*mean-level absolute stability*). From a relative change perspective, it is by definition not possible to assess the change of a single individual, because relative change is

defined as change relative to others. Therefore, *relative stability* methods typically assess whether the ranking of people's scores on a measured construct changes over time. In Study 1, we focus on the literature on the temporal stability of Type D personality to review current practices and to discuss methods designed to detect various types of temporal stability.

Personality and other psychological characteristics are often assessed using scores on multi-item questionnaires. Scores on such psychological questionnaires are known to contain measurement error, where someone's item score does not perfectly reflect this person's score on the latent psychological construct. As a result, measurement error may obscure the true association between constructs, leading to attenuated effect sizes, a phenomenon known as attenuation bias (Spearman, 1904). Moreover, as statistical models get more complex, ignoring measurement error may even result in overestimated associations between constructs (Cole & Preacher, 2014). This highlights the importance of using statistical models that can handle measurement error when assessing the stability of psychological constructs. In Study 2, we use state of the art psychometric (latent variable) modelling approaches to investigate the temporal stability of the two Type D personality traits in relation to depression and anxiety.

Type D personality

Type D personality is most prominently studied in the field of psychosomatic medicine, where it is seen as a risk factor of cardiac events in patients suffering from cardiovascular disease (Grande, Romppel & Barth, 2012; Piepoli et al., 2016; Kupper & Denollet, 2018). Research on the temporal stability of Type D personality illustrates many statistical and psychometric issues in the study of the temporal stability of psychological constructs. Type D is measured with a multi-item questionnaire (DS14; Denollet, 2005) involving ordinal and often skewed item scores. Individuals with a Type D (Distressed) personality are considered to score high on the two personality traits negative affectivity (NA) and social inhibition (SI). NA concerns the tendency of people to experience negative thoughts and emotions, while SI concerns the difficulty in expressing such thoughts and emotions, especially in social interactions.

These is a strong association between Type D's two 'distressed' personality traits and the negative emotional states depression and anxiety (Lodder et al., 2019). This association is especially pronounced for NA and depression, with correlations between scale scores ranging between 0.4 and 0.7 (Spindler, Kruse, Zwisler, & Pedersen, 2009; Ossola, De Panfilis, Tonna, Ardissino, & Marchesi, 2015). These correlations point to a significant statistical overlap between the *trait* NA and the more *episodic* depression, leading some scholars to question whether NA is really a personality trait, or whether depression has *trait*-like characteristics (Ossola et al., 2015).

A limitation of these studies is that they failed to take into account the measurement error in the questionnaire scores. In Study 2, we illustrate how to assess temporal stability of psychological constructs. Our illustration will focus on the personality traits NA and SI and how their temporal stability relates to that of depression and anxiety. We will use latent variable models that not only take into account measurement error in the questionnaire scores, but can also appropriately model the non-normally distributed ordinal item scores typically encountered in psychological research. Treating such ordinal scores as continuous and normally distributed may result in biased parameter estimates (Rhemtulla, Brosseau-Liard & Savalei, 2012; Lodder, Emons, Denollet & Wicherts, 2021) and may therefore result in misleading conclusions regarding the stability of psychological constructs.

Study aims and overview

The aims of the present chapter are twofold. First, in Study 1, we systematically review the methods typically used by researchers to assess the stability of psychological constructs, and use Type D personality (Denollet, 2005) as an example. We discuss how these common methods risk incorrect conclusions regarding the temporal stability of the personality traits NA and SI by ignoring the presence of measurement error in the item scores and by not testing the often-ignored assumption of longitudinal measurement invariance. This review does not only shed light on the earlier research on this issue, but also provides an ideal opportunity to introduce the statistical methods applied researchers typically use to assess temporal stability. Second, in Study 2, we illustrate how to handle these issues using a series of three longitudinal latent variable models used to investigate and compare the temporal stability of Type D personality, anxiety and depression. We will discuss how to handle the

often-overlooked problem that the questionnaire item scores are ordinal and non-normally distributed when building the latent variable models. We subsequently illustrate how to test the often overlooked—yet crucial—assumption underlying the longitudinal analysis of questionnaire data, namely that the properties of your instrument's measurement model (e.g. item factor loadings) are invariant across all measurement occasions (Liu et al., 2017). The relative temporal stability and autoregressive effects of the psychological constructs can also be inferred from these models. Next, we show how latent growth curve models (Hertzog, Lindenberger, Ghisletta & von Oertzen, 2006) can be used to investigate the mean and individual level absolute stability of psychological constructs, while taking into account measurement error in the item scores. Multivariate latent growth curve models also allow for estimating how intra-individual change in for instance depression correlates with intra-individual change in negative affectivity. Lastly, we illustrate the benefit of a latent trait-state-occasion model (Cole, Martin & Steiger, 2005) to estimate what part of a construct can be considered a stable trait and what part a changeable state.

We hypothesized that NA and SI both show absolute and relative temporal stability over time, and that both constructs correspond more to a stable trait than to a changeable state. In line with earlier research (Ossola et al., 2015), we further hypothesized that any individual changes in the personality trait NA would correlate with individual changes in depression and anxiety, but that SI would not show this association.

STUDY 1: SYSTEMATIC REVIEW

The goal of this systematic review is to review the temporal stability studies conducted in the context of research on Type D personality, to document common practices used to analyze stability in this literature, and to discuss the limitations of these methods. Our review included all studies that assessed Type D personality on at least two measurement occasions and used a statistical analysis to determine whether Type D personality, NA, or SI showed temporal stability.

METHOD

On November 4th 2019, the electronic databases Pubmed and Psycinfo were used to search the full text of empirical articles for the terms '("Type D personality" OR "negative affectivity" OR "social inhibition") AND ("stability" OR "test-retest")'. The search resulted in 142 unique studies. After screening the full texts, 24 studies met the inclusion criteria of our review. In total, those 24 studies reported 75 tests for the temporal stability of either Type D personality or its subcomponents NA or SI. We subsequently divided the studies by statistical approach(es) taken to analyze the stability of NA and SI.

RESULTS

For each of the 75 tests included in the review, Table 1 reports the sample characteristics, the statistical method used to assess stability, the personality construct studied, the longest follow-up time, and the results of the stability assessment. The following sections will discuss the findings of these stability tests separately for each of the investigated stability types.

Relative stability

Of all 24 studies included in the review, 22 (91.7%) investigated relative stability, making this the most popular approach to study temporal stability. Relative stability measures assess whether the relative ranking of individual scores remains stable over time. In their basic form, statistical models assessing relative stability estimate whether scores on T_X covary with scores on T_Y, where X and Y denote two distinct measurement occasions.

Table 1: For all 24 studies included in the systematic review, the type of investigated stability, the sample characteristics, the statistical method used, the personality construct studied, the longest follow-up time in months, and the results of the stability assessment.

Study	Sample	Statistical method	Construct	FU (months)	Results
Relative stability					
Dannemann (2010)	126 German cardiac patients	Test-retest correlation	NA	6	Rxx' = 0.61
Romppel (2012)	679 German cardiac patients	Test-retest reliability	NA	72	Rxx' = 0.61
Gremigni (2005)	30 Italian cardiac patients	Test-retest correlation	NA	1	Rxx' = 0.62
Bunevicius (2013)	49 Lithuanian CHD patients	Test-retest reliability	NA	0.5	Rxx' = 0.69
Denollet (2005)	121 Cardiac rehabilitation patients)	Test-retest correlation	NA	3	Rxx' = 0.72
Zohar (2016)	285 Israeli community volunteers	Test-retest correlation	NA	72	Rxx' = 0.72
Aluja (2019)	65 Spanish university students	Test-retest reliability	NA	2	Rxx' = 0.77
Denollet (1998)	60 Belgian CHD patients	Test-retest reliability	NA	3	Rxx' = 0.78
Spindler (2009)	117 Danish cardiac patients	Test-retest reliability	NA	3	Rxx' = 0.78
Alcelik (2012)	100 Turkish hemodialysis patients	Test-retest reliability	NA	1	Rxx' = 0.84
Pedersen (2009)	57 Healthy Ukrainians	Test-retest reliability	NA	1	Rxx' = 0.85
Bagherian (2011)	71 Iranians (MI + healthy)	Test-retest correlation	NA	2	Rxx' = 0.86
Montero (2017)	253 Spaniards (MI + cancer + healthy)	Test-retest reliability	NA	6	Rxx' = 0.88
Dannemann (2010)	126 German cardiac patients	Test-retest correlation	SI	6	Rxx' = 0.59
Romppel (2012)	679 German cardiac patients	Test-retest reliability	SI	72	Rxx' = 0.60
Pedersen (2009)	57 Healthy Ukrainians	Test-retest reliability	SI	1	Rxx' = 0.63
Bagherian (2011)	71 Iranians (MI + healthy)	Test-retest correlation	SI	2	Rxx' = 0.77
Alcelik (2012)	100 Turkish hemodialysis patients	Test-retest reliability	SI	1	Rxx' = 0.78
Spindler (2009)	117 Danish cardiac patients	Test-retest reliability	SI	3	Rxx' = 0.79

Gremigni (2005)	30 Italian cardiac patients	Test-retest correlation	SI	1	Rxx' = 0.81
Bunevicius (2013)	49 Lithuanian CAD patients	Test-retest reliability	SI	0.5	Rxx' = 0.81
Aluja (2019)	65 Spanish university students	Test-retest reliability	SI	2	Rxx' = 0.82
Denollet (2005)	121 Cardiac rehabilitation patients)	Test-retest correlation	SI	3	Rxx' = 0.82
Zohar (2016)	285 Israeli community volunteers	Test-retest correlation	SI	72	Rxx' = 0.82
Denollet (1998)	60 Belgian CHD patients	Test-retest reliability	SI	3	Rxx' = 0.87
Montero (2017)	253 Spaniards (MI + cancer + healthy)	Test-retest reliability	SI	6	Rxx' = 0.89
Zohar (2016)	285 Israeli community volunteers	Test-retest reliability	Type D (NA*SI)	72	Rxx' = 0.78
Ossola (2015)	304 Italian CHD patients	ICC	NA	12	ICC = 0.48
Nefs (2012)	1012 Dutch primary care patients	ICC (2-way; consistency; average)	NA	12	ICC = 0.64 (men); 0.63 (women)
Conden (2014)	313 Swedish acute MI patients	ICC (2-way)	NA	12	ICC = 0.71
Bouwens (2019)	294 Dutch vascular surgery patients	ICC	NA	12	ICC = 0.71
Loosman (2017)	249 Dutch dialysis patients	ICC	NA	6	ICC = 0.72
Lim (2011)	111 Korean CHD patients	ICC	NA	2	ICC = 0.76
Yu (2010)	100 Chinese CHD patients	ICC	NA	3	ICC = 0.76
Kupper (2011)	730 Dutch twins	ICC	NA	108	ICC = 0.78 (twin A); 0.72 (twin B)
Spindler (2009)	117 Danish cardiac patients	ICC (2-way; consistency; average)	NA	3	ICC = 0.87
Loosman (2017)	249 Dutch dialysis patients	ICC	SI	6	ICC = 0.69
Ossola (2015)	304 Italian CHD patients	ICC	SI	12	ICC = 0.70
Nefs (2012)	1012 Dutch primary care patients	ICC (2-way; consistency; average)	SI	12	ICC = 0.73 (men); 0.65 (women)
Yu (2010)	100 Chinese CHD patients	ICC	SI	3	ICC = 0.74
Lim (2011)	111 Korean CHD patients	ICC	SI	2	ICC = 0.77
Conden (2014)	313 Swedish acute MI patients	ICC (2-way)	SI	12	ICC = 0.80
Bouwens (2019)	294 Dutch vascular surgery patients	ICC	SI	12	ICC = 0.80

Kupper (2011)	730 Dutch twins	ICC	SI	108	ICC = 0.83 (twin A); 0.82 (twin B)
Spindler (2009)	117 Danish cardiac patients	ICC (2-way; consistency; average)	SI	3	ICC = 0.88
Kupper (2011)	730 Dutch twins	ICC	Type D (2-groups)	108	ICC = 0.62 (twin A); 0.58 (twin B)
Ossola (2015)	304 Italian CHD patients	ICC	Type D (NA*SI)	12	ICC = 0.52
Bouwens (2019)	294 Dutch vascular surgery patients	ICC	Type D (NA*SI)	12	ICC = 0.72
Conden (2014)	313 Swedish acute MI patients	ICC	Type D (NA*SI)	12	ICC = 0.76
Mean absolute stability					
Pedersen (2009)	57 Healthy Ukrainians	Paired t-test	NA	1	d = 0.004, NS *
Pedersen (2009)	57 Healthy Ukrainians	Paired t-test	SI	1	d = 0.10, NS *
Romppel (2012)	679 German cardiac patients	Cohen's d	NA	72	d = 0.08, NS
Romppel (2012)	679 German cardiac patients	Cohen's d	SI	72	d = 0.01, NS
Dannemann (2010)	126 German cardiac patients	RM ANOVA	NA	6	d = 0.04, NS *
Dannemann (2010)	126 German cardiac patients	RM ANOVA	SI	6	d = 0.12, p<.05 *
Individual absolute stability					
Romppel (2012)	679 German cardiac patients	RCI	NA	72	Significant change: 26.4%
Romppel (2012)	679 German cardiac patients	RCI	SI	72	Significant change: 22.7%
Ipsative stability					
Pelle (2008)	386 Dutch CAD patients	% caseness	Type D (2-groups)	3	Stable classification: 81%
Zohar (2016)	285 Israeli community volunteers	% caseness	Type D (2-groups)	72	Stable classification: 82%
Nefs (2012)	1012 Dutch primary care patients	% caseness	Type D (2-groups)	12	Stable classification: 85%
Martens (2007)	475 Dutch acute MI patients	Logistic regression	Type D (2-groups)	18	$\chi^2(2) = 1.6$, $p = 0.45$

Zohar (2016)	285 Israeli community volunteers	Chi-square test	Type D (2-groups)	72	κ = 0.50 *
Bouwens (2019)	294 Dutch vascular surgery patients	Cohen's Kappa	Type D (2-groups)	12	κ = 0.32
Conden (2014)	313 Swedish acute MI patients	Cohen's Kappa	Type D (2-groups)	12	κ = 0.40
Romppel (2012)	679 German cardiac patients	Cohen's Kappa	Type D (2-groups)	72	κ = 0.42
Ossola (2015)	304 Italian CHD patients	Cohen's Kappa	Type D (2-groups)	12	κ = 0.49
Loosman (2017)	249 Dutch dialysis patients	Cohen's Kappa	Type D (2-groups)	6	κ = 0.52
Dannemann (2010)	126 German cardiac patients	RM ANOVA	Type D (2-groups)	6	κ = 0.26
Conden (2014)	313 Swedish acute MI patients	Cohen's Kappa	NA (dichotomized)	12	κ = 0.48
Bouwens (2019)	294 Dutch vascular surgery patients	Cohen's Kappa	NA (dichotomized)	12	κ = 0.49
Conden (2014)	313 Swedish acute MI patients	Cohen's Kappa	SI (dichotomized)	12	κ = 0.53
Bouwens (2019)	294 Dutch vascular surgery patients	Cohen's Kappa	SI (dichotomized)	12	κ = 0.54
Genetic stability					
Kupper (2011)	730 Dutch twins	ACE model	NA	108	Non-genetic variance 55-60%
Kupper (2011)	730 Dutch twins	ACE model	SI	108	Non-genetic variance 51-58%
Kupper (2011)	730 Dutch twins	ACE model	Type D (2-groups)	108	Non-genetic variance 51-66%
Longitudinal measurement invariance					
Romppel (2012)	679 German cardiac patients	SEM	NA & SI	72	Stable factor loadings
Conden (2014)	313 Swedish acute MI patients	SEM	NA & SI	12	Stable factor loadings

ACE = statistical model used to analyze the results of twin studies; CHD = coronary heart disease; FU = follow-up; ICC = intraclass correlation; MI = myocardial infarction; NA = negative affectivity; RCI = reliable change index; SEM = structural equation modeling; SI = social inhibition

** effect size calculated based on statistics reported in the published study*

Test-retest correlation

In the reviewed literature, the test-retest correlation (in some included studies referred to as test-retest reliability) was the most popular method to study relative stability of Type D personality traits, arguably because it simply involves computing the Pearson correlation coefficient between the T_X and T_Y scores. The Pearson correlation coefficient can be used when the association between two measurements is linear, while Spearman's rho is useful for estimating non-linear but monotonically increasing associations.

The Pearson correlation coefficients reported in Table 1 ranged from 0.61 to 0.88 (median *r* = .77) for NA and from 0.59 to 0.89 (median *r* = 0.81) for SI. One way to operationalize Type D personality is by multiplying the scores of the NA and SI variables (***Chapter 2*;** Ferguson et al., 2009). One study (Zohar, 2016) reported a correlation of .78 between the repeated measurements of this NA*SI product score. Taken together, these findings suggest that both Type D as well as NA and SI generally showed acceptable relative stability based on the Pearson correlation coefficients.

Drawbacks of the Pearson correlation coefficient are that it is limited to correlations between two measurements and that it ignores the measurement error if no proper correction for attenuation (Muchinsky, 1996) is applied. We instead recommend using a latent variable model to directly estimate the correlation between the latent variable scores at two time points (see Study 2).

Intraclass correlation

A popular method to assess relative stability for two or more repeated measurements is the intraclass correlation coefficient (ICC; Bartko, 1976; McGraw & Wong, 1996; Weir, 2005). Historically it was developed as an index of reliability (e.g. test-retest or interrater reliability), but several other purposes exist. There exist various types of intraclass correlation models, but when the goal is to assess temporal stability of repeated measurements, then the 2-way mixed-effects model is the method of choice (Koo & Li, 2016). Researchers also need to decide whether the ICC is calculated based on single item scores or on an average (or sum) of multiple item scores. Use of average measurements is the preferred option when psychological constructs are measured using multi-item

questionnaires. Lastly, researchers should decide whether they are interested in consistency or absolute agreement.

Similar to the correlation coefficient, the consistency ICC is sensitive to the relative ranking of individuals, but a key difference between them is that the correlation expresses the degree to which variables Y and X are associated through a linear transformation (Y=aX + b), while the consistency ICC measures the extent to which Y and X are associated by adding a constant (Y = X + b). The model used to estimate ICCs assumes equal variance at the repeated measurements. Violating this assumption often results in attenuated ICC estimates relative to the correlation coefficient (McGraw & Wong, 1996). If all individuals change to the same degree, then the consistency ICC will be equal to one. The absolute agreement ICC, on the other hand, also considers absolute changes over time and will therefore be lower than one if there are individual differences in the intra-individual change.

Table 1 shows that for NA, the eleven included ICCs ranged from 0.48 to 0.87 (median ICC = 0.72), while for SI the eleven ICCs ranged from 0.65 to 0.88 (median ICC = 0.77). For Type D personality (the NA*SI product score), the three included ICCs ranged from 0.52 to 0.76 (median ICC = 0.72). For most ICCs included in our systematic review researchers did not specify the investigated type of ICC. Given that decisions regarding ICC type (single vs. average ratings & consistency vs. absolute agreement) may considerably influence the estimated ICC, it is difficult to determine whether Type D, NA and SI are temporally stable based on the ICCs reported in these studies.

Neither of the two studies that reported the chosen ICC method used an absolute agreement definition. The ICC estimated using a consistency definition is always equal to or larger than the ICC estimated according to the absolute agreement definition. Consequently, the results of studies that have investigated temporal stability using the consistency ICC (Nefs et al., 2012; Spindler et al., 2009), may appear to be more temporally stable than they really are than if they would have also taken into account absolute stability.

Mean-level absolute stability

Of all 24 studies included in the review, three (12.5%) investigated mean-level absolute stability, each using a different statistical method, including the paired t-test, the repeated measures ANOVA and the standardized mean difference.

Paired t-test

The paired (or dependent) t-test assesses absolute difference in the mean scores of two repeated or dependent measurements. Commonly, the null hypothesis of a paired t-test is that the mean scores on the two measurements are equal. This null hypothesis is rejected when the difference becomes large enough in relation to its standard error to be statistically significant, with the standard error being a function of the sample size, standard deviation and correlation between the two repeated measurements. When assessing absolute stability, researchers typically conclude absolute stability when the difference between two measurements is *not* statistically significant, which is a statistically invalid conclusion. Table 1 shows that one study (Pedersen et al., 2009) included in the review assessed the absolute stability of NA and SI using a paired t-test. Absolute stability was concluded based on both tests because they were not statistically significant with *t*-values of 0.064 (NA) and 0.7 (SI).

Standardized mean difference

Absolute stability can also be assessed by computing the standardized mean difference of the scores on two measurements using the Cohen's *d* effect size for repeated measures. ***Appendix R*** shows the formula used to compute Cohen's *d* for paired data. This method assumes homogeneous variances across repeated measurements (Cohen, 1988). It determines the mean-level absolute stability and can therefore not assess individual-level absolute stability. Table 1 shows that one included study (Romppel, Herrmann-Lingen, Vesper & Grande, 2012) assessed the mean-level absolute stability of NA and SI using the standardized mean difference. This study concluded absolute stability for both personality traits based on non-significant Cohen's *d* estimates of 0.08 (NA) and 0.01 (SI).

Repeated measures ANOVA

The mean-level absolute stability of two or more repeated measurements can also be assessed using a repeated measures (RM) ANOVA. When there are only two repeated

measurements, the RM ANOVA is equivalent to a paired t-test. Typically, researchers conclude absolute stability when the within-subjects effect (e.g. Time or Measurement) does not reach statistical significance. Table 1 indicates that one study (Dannemann et al., 2010) used an RM ANOVA to test the absolute stability of NA and SI based on the within subjects Time effect. It turned out that absolute stability was concluded for NA, but not for SI. Note that this non-significant mean difference does not necessarily indicate the absence of temporal stability in the population because the p-value of this test is also influenced by the sample size.

Individual-level absolute stability

Whereas the absolute stability methods discussed in the previous section all assessed stability at the group level (e.g., the full sample), the reliable change index (RCI) is a method developed to determine for each individual separately whether there is significant absolute change over time (Jacobson & Truax, 1992). Whether this individual change is statistically significant depends in part on the amount of measurement error in the questionnaire scores. As many psychological questionnaires are not perfectly reliable, the RCI can be used to assess whether the observed individual change is larger than the change that may occur due to measurement error. Although change scores have often been criticized for having low reliability, recent psychometric advanced suggest that this is not necessarily the case when modeling change scores from a multilevel perspective, distinguishing individual change on the within-subjects level from group change on the between-subjects level (Gu, Emons, & Sijtsma, 2019). ***Appendix R*** present the mathematical details behind computing the RCI. Note that when calculating the RCI from reliability estimates within a classical test theory perceptive, then researchers make the strong assumption that the variance of the two measurements are equal, as well as the error variances (implying equal reliability coefficients across measurements; Maassen, Bossema & Brand, 2009).

In our review, one included study used the RCI to assess individual stability of NA and SI (Romppel et al., 2012). Although this study did not find absolute change averaged across all participants, significant *individual* change was observed for 26.4% of the participants on NA for 22.7% on SI. Of these changes, the proportion of significant change involving either increased or decreased scores was approximately equal.

Ipsative stability

Of all 24 studies included in the review, nine (37.5%) investigated the ipsative stability of Type D personality. Ipsative stability refers to the continuity or temporal stability of a trait pattern within individuals (De Fruyt et al., 2006). This trait pattern typically involves two or more traits, but in some instances, researchers assess the continuity of having high scores on a single trait. The temporal stability of trait patterns can be assessed using latent variable models such as latent transition analysis or repeated measures latent class analysis (Collins & Lanza, 2009). In the Type D literature, researchers have assessed ipsative stability by investigating temporal changes in the classification of individuals in personality groups. The classification in personality groups is based on whether or not individuals score above a cutoff on NA and/or SI. This either results in two (Type D & no Type D) or four (Type D, NA+SI-, NA-SI+, NA-SI-) personality groups and researchers subsequently calculate the percentage of individuals that change group membership across time. Some studies have assessed the ipsative stability of NA and SI separately by classifying participants in High vs. Low NA groups, and High vs. Low SI groups. A major disadvantage of this approach is that the initial classification in personality groups ignores valuable information on individual differences in these personality traits. For NA and SI, ipsative stability methods cannot detect changes happening within each of the 0-9 or 10-28 ranges, because changes do not affect classification. We therefore argue that ipsative stability methods should only be used when the main interest is the stability in the classification, rather than stability in the underlying personality traits.

The studies included in our review utilized several statistical methods to assess stability in classification, including descriptive statistics, logistic regression, chi-square tests and Cohen's Kappa. Table 1 indicates that three studies (Pelle et al. 2008; Nefs, et al., 2012; Zohar, 2016) used descriptive statistics to assess the ipsative stability of Type D personality. Note that for one of these studies (Pelle et al., 2008), the goal was not to assess temporal stability, but rather to investigate whether patients receiving cardiac rehabilitation would change in their Type D classification. According to the three studies, between 81% and 85% of the participants did not change in their Type D or no Type D classification over time. As these analyses did not involve inferential statistics it is not possible to generalize these findings beyond the studied samples. To solve that problem, another study (Martens et al. 2007)

used a logistic regression to show that measurement occasion did not predict the Type D classification, suggesting that it is stable over time. Another study (Zohar, 2016) used a chi-square test to reject the null hypothesis that the classifications at two measurement occasions are independent. However, the null hypothesis of no dependency is unrealistic, as at least some dependency in classification is expected when the same participants are measured twice. Lastly, five studies used Cohen's Kappa to assess the agreement in classification at two measurement occasions. The Kappa estimates ranged from 0.32 to 0.52 (median = 0.42), suggesting fair to moderate agreement between in Type D classifications over time. Two of these studies also used Cohen's Kappa to study the ipsative stability separately for NA and SI, indicating moderate agreement with Kappa estimates of 0.48 and 0.49 for NA, and 0.53 and 0.54 for SI.

Genetic stability

In behavior genetics, ACE models are frequently applied to twin data to estimate for various psychological traits the proportion of variance explained by either additive genetic (A), shared (C), or non-shared (E) environmental influences (Rijsdijk & Sham, 2002). Such studies typically use data from both identical and fraternal twins to determine the relative contribution of these latent genetic and environmental components in explaining variation in a psychological trait of interest. When longitudinal data is available, then researchers can investigate whether the relative importance of these components remains stable over time.

Traditional genetic stability models do not assess the mean structure of the psychological construct and can therefore not assess absolute stability (neither on the individual level nor on the sample level; for exceptions see McArdle, 1982; Neale & McArdle, 2000; Nivard et al., 2015). Whereas relative and absolute stability methods determine *whether* and *to what extent* the construct shows temporal stability, genetic stability methods intend to elucidate *why* individual differences on a construct show temporal stability or not (Figueredo, de Baca, & Black, 2014). Genetic stability methods estimate the proportion of variance in traits that is attributable to either genetic or environmental influences and determines whether this variance decomposition is stable across time by assessing whether later time points contain genetic or environmental influences not shared with the first point. As this provides valuable information regarding the etiology of the psychological traits, genetic stability methods

could complement relative and absolute stability methods when assessing the temporal stability of psychological traits.

Table 1 indicates that one study included in our review used an ACE model to assess the genetic stability of both Type D personality and its subcomponents NA and SI (Kupper et al., 2011). This study used structural equation modeling to fit longitudinal ACE model on the aggregate NA and SI scores. The results showed that across nine years, the heritability of NA was stable and varied only slightly (between 40 and 45%). Similar genetic stability over time was found for SI, with heritability estimates varying between 42 and 49%.

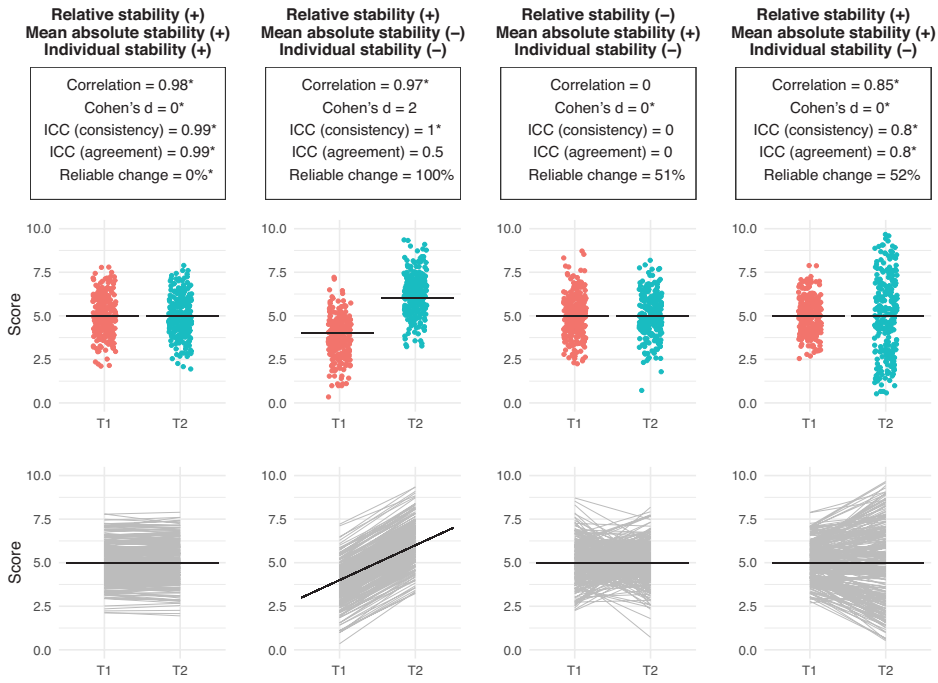
A limitation of this study is that modeling the aggregate NA and SI scores rather than raw item scores fail to consider measurement error in the item scores. It also inhibits testing the assumption of longitudinal measurement invariance (Liu et al., 2015). Simulation studies have indicated that modeling aggregated rather than raw item scores result in underestimated heritability estimates and component correlations across time, which may bias conclusions regarding the stability of the genetic and environmental components (van den Berg, Glas & Boomsma, 2007; Schwabe, Gu, Tijmstra, Hatemi & Pohl, 2019). Future research could prevent this problem by specifying a measurement model for the latent NA and SI constructs when testing the genetic stability of these traits with an ACE model.

A small simulation study

The methods reviewed above differ in the types of temporal stability they can detect. In the section we illustrate based on simulated data whether various statistical methods can detect relative stability, mean-level absolute stability or individual-level absolute stability. We simulated data for two repeated measurements of a particular construct and in four scenarios varied the presence or absence of relative stability, mean-level absolute stability and individual absolute stability. The upper row of Figure 1 reports the estimated temporal stability in terms of Pearson correlation coefficient, Cohen's *d* for paired data, consistency ICC, absolute agreement ICC, and the reliable change index (assuming a test reliability of 0.9). The middle row visualizes the individual and mean scores on two repeated measurements. The bottom row shows the individual and mean growth curves.

In the *first* scenario, all individual scores remain constant across time. As expected, all methods suggest temporal stability because there is no intra-individual change and no interindividual differences in this intraindividual change. The *second* scenario also does not involve interindividual differences in change, because the intraindividual change of each individual is similarly positive. Consequently, both the mean-level (Cohen's d) and individual-level (RCI) absolute stability methods indicate that there is significant change across time, while the relative stability methods (Pearson correlation and consistency ICC) suggest perfect relative stability because the ranking of individual scores does not change. As opposed to the consistency ICC, the absolute agreement ICC is sensitive to deviations from absolute stability and therefore does not suggest temporal stability.

Figure 1: Simulated data on two time points, varying in the presence (+) or absence (-) of relative stability, mean-level absolute stability and individual absolute stability. The upper row shows estimated stability statistics. The middle row shows the individual- and mean scores at each time point. The bottom row shows the individual- and mean growth curves.



*ICC = intraclass correlation; * Temporal stability concluded based on this statistic*

In the *third* scenario, the simulated scores on both time points are completely unrelated yet similar on average. As expected, Cohen's d suggests mean-level absolute stability. Pearson correlation and both ICCs indicate no temporal stability because the ranking of individuals completely changes across time. Interestingly, the RCI suggests poor individual-level absolute stability because significant change across time is concluded for more than half (51%) of the individuals. Lastly, in the *fourth* scenario the change across time depends on the score at the first timepoint. The highest baseline scores increase across time; the lowest baseline scores decrease across time; average baseline scores remain stable across time. Cohen's *d* indicates mean-level absolute stability. The RCI suggests no poor individual-level absolute stability because more than half of the individuals (52%) show significant change across time. The ICC's and Pearson correlation coefficient all indicate acceptable temporal stability, yet the ICC's estimates are slightly smaller than the Pearson correlation coefficient, likely because ICC's assumption of equal variances is violated (McGraw & Wong, 1999).

Based on this simulated example we can infer several relations between the various types of temporal stability. First, the presence of perfect individual-level absolute stability implies both relative stability and mean-level absolute stability (column 1). Second, the presence of perfect mean-level absolute stability does not necessarily imply relative stability or individual-level absolute stability (column 2). Third, the presence of perfect relative stability does not necessarily imply mean-level absolute stability (column 3) or individual-level absolute stability (column 4).

Synthesis

Several conclusions can be drawn regarding the reviewed studies on the temporal stability of Type D, NA and SI. First, our simulated data example illustrates why researchers should first explicitly define the type of temporal stability they want to assess and subsequently select one or more models that can adequately detect such stability. Table 2 summarizes for each reviewed method the types of temporal stability that can be detected. If researchers want to know whether individuals do not change in their personality across time, then only using a relative stability method is not sufficient and may better be complemented by an absolute stability method.

Table 2: Characteristics of methods used to assess temporal stability in the reviewed studies

Statistical focus	Statistical method	Handles measurement error in item scores	Detects relative stability	Detects mean absolute stability	Detects individual absolute stability	Detects ipsative stability	Detects genetic stability
Association	Pearson correlation	-	+	-	-	-	-
	ICC (consistency)	-	+	-	-	-	-
	ICC (agreement)	-	+	-	+	-	-
Mean difference	Paired t-test	-	-	+	-	-	-
	Cohen's d	-	-	+	-	-	-
	RM ANOVA	-	-	+	-	-	-
	Reliable change index	+	-	-	+	-	-
Classification	% caseness	-	-	-	-	+	-
	Chi-square test	-	-	-	-	+	-
	Logistic regression	-	-	-	-	+	-
	Cohen's Kappa	-	-	-	-	+	-
Genetic	ACE model	+	-	-	-	-	+

ACE = statistical model used to assess genetic and environmental influences in twin studies; ICC = intraclass correlation; LGCM = latent growth curve model; LMI = Longitudinal measurement invariance; RM ANOVA = repeated measures analysis of variance.

When answering the question whether NA and SI are stable personality traits, we should therefore focus on studies that have reported separate analyses of relative and absolute stability, or on studies that have used analysis that are sensitive to both types of stability, such as the absolute agreement ICC. Three of the reviewed studies assessed both relative and absolute stability on the same sample (Dannemann et al., 2010; Pedersen et al., 2009; Romppel et al., 2012). Mean-level absolute stability was concluded for NA in all studies and for SI in two of the three studies. Regarding the relative stability of these personality traits, the three studies generally showed less than adequate test-retest correlations (*NA*: 0.61, 0.61, 0.85; *SI*: 0.59, 0.60, 0.63). These results suggest that although on average participants did not change in their NA and SI scores over time, the relative ranking of participants showed less temporal stability. Still, the Pearson correlations used to assess this relative

ranking likely suffered from attenuation bias because these correlations were not estimated while taking into account the measurement error in the NA and SI item scores. In sum, based on these studies the temporal stability of both NA and SI appears suboptimal.

However, a limitation of the reviewed studies is that the commonly used methods do not consider the properties of the Type D measurement instrument when assessing temporal stability. The DS14 item scores are not equivalent, generally not normally distributed, and more informative in the higher than in the lower NA and SI score ranges (Emons, Meijer & Denollet, 2007). Such characteristics may contribute to a violation of certain model assumptions (e.g., linearity and homoscedasticity) discussed in this chapter. Another major limitation is that none appropriately tested for longitudinal measurement invariance, though two studies partly investigated this assumption (Romppel et al., 2012; Condén et al., 2014; see Study 2). Longitudinal measurement invariance is often overlooked in longitudinal research and when this assumption is violated then changes on the observed scores do not necessarily reflect changes in the latent constructs of interest. We therefore argue that the starting point of any temporal stability assessment should be a test for longitudinal measurement invariance. After establishing measurement invariance, relative and absolute stability should ideally be investigated using approaches that adjust for measurement error, such as latent variable modeling. Study 2 illustrates three such longitudinal latent variables models that can be used to investigate the temporal stability of constructs measured with ordinal items that often show skewed score distributions.

STUDY 2: LONGITUDINAL LATENT VARIABLE MODELS

INTRODUCTION

A major disadvantage of the temporal stability methods discussed in Study 1 is that they were applied to observed scores and therefore assess change in aggregated item scores (e.g., sum or mean scores) and not in latent construct scores. These observed score methods implicitly assumes that these aggregate scores do not contain measurement error and thus are perfectly reliable measures of the underlying construct at all measurement occasions. However, questionnaire scores generally are imperfect measures of the underlying latent construct. Modern test theory assumes that each questionnaire has measurement properties (e.g., loadings/intercepts/thresholds/residuals in factor models, or discrimination/difficulty parameters in item response models) that relate observed item scores to latent construct(s). Scores on questionnaire items are therefore not exclusively caused by the variation in the latent construct, but also by other factors unique to each particular item.

An assumption underlying most absolute stability methods is that these measurement properties do not change over time, a requirement called longitudinal measurement invariance (Pentz & Chou, 1994; Liu et al., 2017). The fact that factor loadings and other measurement properties *can* change over time, implies that absolute changes in item scores (and therefore also in aggregated scores) are not necessarily the result of changes on the construct level, but may also result from changes in measurement properties. When assessing the temporal stability of constructs, researchers should first test for longitudinal measurement invariance to disentangle changes in the measurement properties from changes on construct level.

As an example, consider a sample of participants completing a seven-item negative affectivity questionnaire in both summer and winter. Further suppose that researchers concluded no absolute stability for negative affectivity, because a paired t-test indicated significantly lower negative affectivity sum scores in winter than in summer. Lastly, suppose

that people did not change in their true negative affectivity scores over time. How is it possible that the paired t-test suggested significant change while there was no true change in negative affectivity over time? One reason could be that the intercept of the negative affectivity item 'I often take a gloomy view of things', was lower in winter than in summer due to people having a lower mood in winter, a phenomenon called the winter blues (Rosenthal, 2012). Such temporary changes in mood involve a different psychological process than scoring high on the personality trait negative affectivity. Using latent variable models to distinguish the latent construct of interest (e.g., negative affectivity) and the individual items measuring it, allows for detecting changes in the item scores due to other influences while the latent negative affectivity scores remain constant. Differences in item intercepts also violate longitudinal measurement invariance (Oort, 2015), further complicating the comparisons of scores over time. Violating this assumption at intercept level would imply that participants scored higher on that particular item in winter than in summer, regardless of their latent negative affectivity score.

Researchers typically test for longitudinal measurement invariance using a series of increasingly restricted structural equation models (SEM; Bollen, 2005). According to these models, each psychological construct is a latent (unobserved) variable and one or more observed item scores reflect the scores on this latent variable. However, each item generally is an imperfect representation of the construct of interest. The variance in an item score not explained by the latent variable is called measurement error or unique variance. By distinguishing the item variance explained by the latent construct from the variance explained by measurement error, latent variable models allow for estimating the association between latent constructs themselves (rather than aggregated observed scores), resulting in estimates that are unaffected by measurement error.

When testing longitudinal measurement invariance using SEM, researchers typically fit a series of nested models to the data. First, a *configural invariance* model is tested to determine whether the factor structure is similar across time. In each step, an additional type of measurement model parameters is constrained to be equal across time (Millsap & Cham, 2012). In the second step, the *weak invariance* model constrains the factor loadings to be equal at each measurement occasion. Next, a *strong invariance* model adds the

constraint that either the intercepts (for continuous data) or the thresholds (for ordered categorical data) are equal across time. Lastly, the *strict invariance* model constrains the residual variances for each item to be invariant across time. These four models are nested because each additional constraint builds upon the already existing constraints of a previous model. In structural equation modeling, such nested models can be compared using for instance a chi-square difference test or likelihood ratio test. Such tests indicate whether the additionally imposed constraints cause a significantly worsening in model fit. If so, then longitudinal measurement invariance is violated for the newly constrained parameter type. Follow-up tests for partial measurement invariance can be performed to investigate whether measurement invariance is still plausible for a subset of the parameters that were not invariant across time.

Regarding the temporal stability of the Type D personality traits, Table 1 in Study 1 shows that longitudinal measurement invariance of NA and SI has been investigated in two studies (Romppel et al., 2012; Conden et al., 2014). However, these studies merely showed that the NA and SI factor loadings were invariant across time and did not test for longitudinal invariance of the intercepts/thresholds and residuals. These tests are essential when assessing temporal stability of psychological constructs, because if researchers want to interpret absolute changes over time as resulting from changes at the latent construct level, then at least strong invariance has to be established for continuous scores and strict invariance for ordinal scores (Liu et al., 2019). Given that the two studies included in our review only investigated a weak invariance model, it is not clear whether the observed changes (or stability) in the NA and SI scores are attributable to changes in the NA and SI constructs, or to changes in the unstudied measurement properties (i.e., intercepts, thresholds, residual variances). Moreover, these studies failed to consider the ordered categorical measurement level of the NA and SI items. Treating ordered categorical data as continuous and normally distributed might cause biased parameter estimates in the structural equation model when there are fewer than five answer categories or when the item scores are not normally distributed (Rhemtulla, Brosseau-Liard, & Savalei, 2012). To solve these problems, the second part of this chapter illustrates a test for the longitudinal measurement invariance of NA and SI that can adequately handle the skewed ordinal nature of these item scores.

In this second study, we assess the temporal stability of the Type D personality traits NA and SI, and depression and anxiety using various latent variable models. These models do not only allow us to determine the relative and absolute stability of the Type D traits, but also assess their relation to the temporal stability of the related psychological states depression and anxiety. Before estimating the longitudinal latent variable models necessary to answer this research question, we will test the assumption of longitudinal measurement invariance separately for each of these four constructs. The next section introduces each of those three latent variable models.

Longitudinal latent variable models

Estimation with ordinal items

When testing the longitudinal measurement invariance of a measurement instrument (and when fitting latent variable models in general), researchers should first evaluate whether the item scores conform to an ordinal or continuous measurement level. Psychological questionnaires often involve Likert-type data with item options ranging between two and nine response categories. Whether these item scores can be considered ordinal or approximately continuous depends not only on the number of response categories, but also on whether the response are approximately normally distributed. Simulation studies have indicated that normally distributed Likert scale data with five or more response categories can be analyzed as continuous variables (Dolan, 1994). Item scores should be analyzed as ordinal variables if they are not (approximately) normally distributed, or if they result from Likert scales with fewer than five response categories. Ignoring the ordinal nature of item scores and treating them as continuous in subsequent analyses results in biased factor loadings and standard errors, especially when the number of response categories is low or the item score distribution is skewed (Rhemtulla, Brosseau-Liard, & Savalei, 2012).

The structural equation models used to test longitudinal measurement invariance on ordinal item scores involve different parameter types than models based continuous items. For continuous item scores, the strong invariance model evaluates the longitudinal invariance of the item *intercepts* (expected item score when the score on the latent construct is zero), while for ordinal item scores, it tests the longitudinal invariance of the item *threshold* parameters. For an ordinal item with X response categories, there are X-1 estimated

threshold parameters that connect the observed ordinal response pattern to an assumed underlying continuous item score. The strong invariance model tests whether the threshold parameters of each ordinal item are invariant across time. Regardless of whether the data are continuous or ordinal, the weak invariance model tests whether the factor loadings are invariant across time and the strict invariance model tests whether the residual variances are invariant across time.

In practice, models based on continuous or ordinal item scores often differ in the method used to estimate the parameters of the structural equation models. For continuous item scores, the model parameters are typically estimated using a full information method such as maximum likelihood (ML) estimation, while the parameters of ordinal item score models are often estimated using limited information methods such as weighted least squares (WLS), diagonally weighted least squares (DWLS) or unweighted least squares (ULS). See Liu and colleagues (2017) for an excellent review of the differences between longitudinal measurement invariance tests based on ordinal or continuous item scores.

Latent growth curves models

After testing longitudinal measurement invariance, we will use latent growth curve (LGC) models to determine the absolute stability of the NA, SI, depression and anxiety, and to investigate how changes over time in one of these constructs relate to changes in the other constructs. LGC models are a special kind of latent variable model where a longitudinal growth curve is estimated for each individual (Hertzog, Lindenberger, Ghisletta & von Oertzen, 2006). These models use latent variables to express the individual differences in the growth curve parameters (i.e., the intercepts and slopes). Different types of growth curve models have been proposed in the literature. A first distinction is between first- and second-order LGC models. First-order models estimate the latent growth curves directly from the observed data (typically repeated measurements of sum scores). Second-order models estimate the latent growth curves based on other latent variables, where each latent variable has a measurement model that connects it to the observed item scores. A second distinction is between univariate and multivariate LGC models. Univariate models concern the longitudinal change in a single latent construct, whereas multivariate models assess change in two or more constructs simultaneously.

LGC models are comparable to longitudinal multilevel models (Curran, Obeidat & Losardo, 2010). Indeed, the least complex LGC model, the univariate single-order model, is mathematically identical to a multilevel model that allows the intercept and slope (effect of time) parameters to vary across individuals (random regression coefficients), instead of assuming they are similar for each individual (fixed regression coefficients). These random regression parameters can be seen as latent variables because they are unobservable, vary across individuals and are estimated from the observed data. We did not use standard multilevel models to assess the temporal stability of psychological constructs because these do not explicitly model measurement error in the item scores.

Both issues are handled by multivariate second-order latent growth models. Being multivariate, these models allow for studying correlated change (Allemand & Martin, 2017) by testing the association between the growth parameters of multiple constructs. In the context of the present study, it would for instance be interesting to assess how individual change over time (the slope parameter) in the latent constructs depression or anxiety relates to individual change in NA. Earlier longitudinal research suggests that NA fluctuates together with the severity of depressive symptoms, indicating that the NA construct may not be temporally stable due to its sensitivity to changes in mood. However, this study (Marchesi et al., 2014) ignored the presence of measurement error in the item scores and did not to test the longitudinal measurement invariance assumption. Consequently, it could be possible that changes in NA over time were not caused by changes in the NA construct, but rather by changes in the measurement model (e.g., factor loadings).

Although the two Type D personality traits are the primary focus of our study, in this study we compare their temporal stability to that of the related constructs depression and anxiety. We first used four univariate second-order latent growth curve models to determine the absolute temporal stability of NA, SI, depression and anxiety. The test whether the average latent slope parameter differs from zero indicates whether a construct shows absolute stability across all participants. The test whether the variance of the latent slope differs from zero indicates whether there are individual differences between individual in the change in the construct over time. Subsequently, we used six multivariate second-order latent growth models to investigate how the change over time in each latent construct relates to change in

the other constructs. Figure 2 visualizes the multivariate growth model of NA and depression. In this model, the correlation between the latent slopes of two constructs indicates to what extent individuals show similar change over time on both constructs.

Figure 2: Multivariate second-order latent growth curve model for negative affectivity and depression. Ordinal item scores are modeled using threshold parameters (τ), mapping for each item the observed ordinal response pattern to an assumed continuous normally distributed latent variable. Change in the latent variables scores across the four time points is modeled using higher-order intercept (i) and slope (s) latent variables. The residuals of the same item at different time points are allowed to correlate. Dotted lines represent fixed parameters and solid lines represent freely estimated parameters.

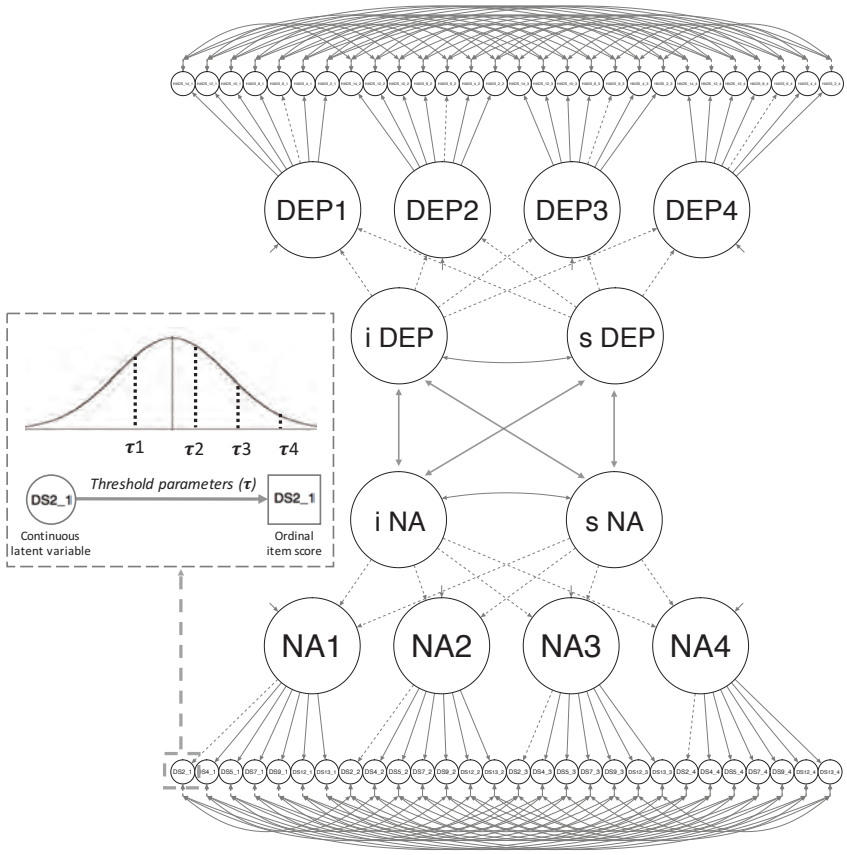
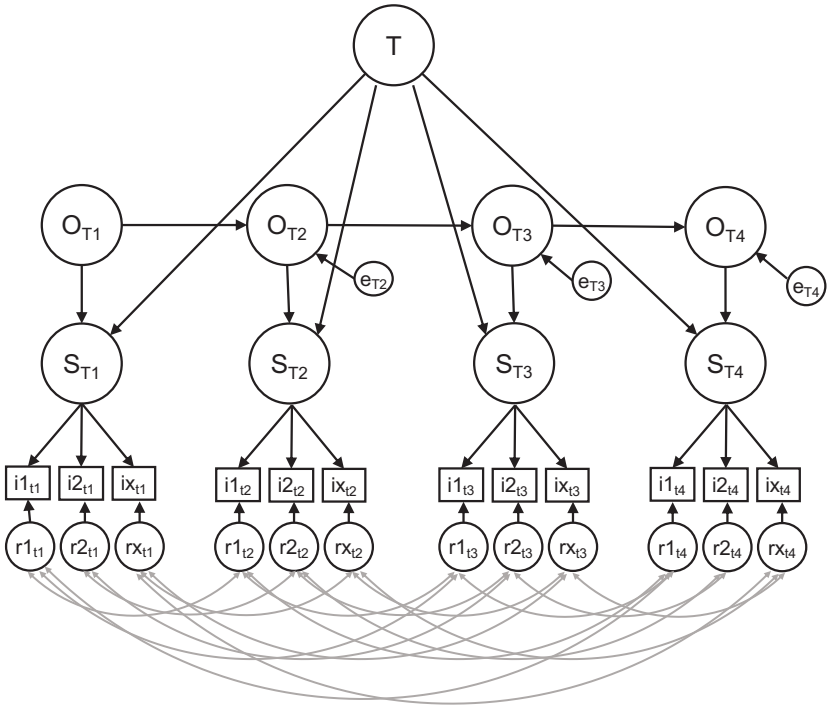


Figure 3: Second-order latent trait-state-occasion model fitted on four repeated measurements of a latent construct. T = Trait; O = Occasion; S = State; T1-T4 = repeated measurements; i = observed item score; e = prediction error variance; r = measurement error variance; The curved arrows indicate that the measurement errors of the same item are allowed to correlate over time.



Latent trait-state-occasion models

Latent trait-state-occasion (LTSO) models are latent variable models used to estimate what proportion of the variance in longitudinal scores can be seen as a stable trait and what proportion as a changeable state. LTSO models are related to various other latent variable models with a similar purpose (e.g., the trait-state-error model or the state-trait model with autoregression), but simulation studies found LTSO models to outperform these other models in decomposing the trait-state variance when the constructs are highly correlated across time (Cole, Martin & Steiger, 2005). LTSO models assume that a latent state is at each repeated measurement occasion explained by two sources of variance: (1) a shared latent

trait variable similar across all measurements; (2) a time-specific latent occasion variable at each measurement. Like LGC models, there exist single-order and second-order LTSO models. We will use a second-order LTSO model to the repeated measurements of NA, SI, depression and anxiety, to determine for each of these constructs the proportion of variance that can be considered trait or state, while dealing with the skewed ordinal nature of the item scores. Figure 3 visualizes an example of such a second-order LTSO model fitted on four repeated measurements of a latent construct.

METHOD

Participants

Data were used from a study conducted using the PROFILES registry (van de Poll-Franse et al., 2011). This population-based longitudinal cohort study assessed patient reported outcomes of colorectal cancer survivors. Eligible participants included all colorectal cancer patients (Stage I to IV) admitted to hospitals in the southern part of the Netherlands between 2000 and 2009. Full details on the inclusion/exclusion criteria and data collection can be found online (https://www.dataarchive.profilesregistry.nl/study_units/view/22). The data collection was approved by the ethics committee of the Maxima Medical Centre in Veldhoven, the Netherlands (approval number 0822). An informed consent statement was signed by all participants. In the present study, we included all participants who completed the psychological questionnaires on at least one time point. Measurements were performed yearly starting in 2010 and ending in 2013. At baseline, the 2625 colorectal cancer survivors were on average 69.4 years old (SD = 9.5, range = 29 to 86). A larger percentage of survivors identified as male (55.1%) than as female (44.9%). On average, they participated 5.2 years since diagnosis (SD = 2.8, range = 1 to 11 years). Part of the sample was lost to follow-up, with 75.8%, 55.5% and 45.3% of the participants responding at the second, third and fourth measurement occasion, respectively. Earlier research on this dataset has indicated that dropouts were significantly more likely to be female, have older age, a lower education and socio-economic status, and show more depressive symptoms than full responders (Ramsey et al., 2019).

Measures

Type D personality

The DS14 questionnaire (Denollet, 2005) was used to measure NA and SI, the two traits underlying Type D personality. Each trait was measured with seven items on a five-point Likert scale, ranging from "false" (0) to "true" (4). The DS14 has been validated in several populations, including the general population (Denollet, 2005) and a breast cancer population (Batselé et al., 2017). Item scores should be considered as having an ordered categorical measurement level, as they are often slightly positively skewed. In the current sample, the estimated McDonald's Omega (total) based on the polychoric correlation matrices of the NA and SI item scores suggested adequate reliability estimates at each of the four measurement occasions (NA = [0.92, 0.93, 0.93, 0.93]; SI = [0.91, 0.91, 0.91, 0.90]).

Symptoms of Depression and Anxiety

The Hospital Anxiety and Depression questionnaire (HADS; Zigmond & Snaith, 1983) was used to measure symptoms of anxiety and depression. Each construct was measured with seven items on a four-point Likert scale, ranging from 0 to 3. These items should be considered ordered categorical scores as they are generally positively skewed and have only four response categories. The HADS questionnaire has been validated in several populations, including in the general population (Spinhoven et al., 1997) and several cancer populations (Bjelland, Dahl, Haug, & Neckelmann, 2002). In the current sample, the estimated McDonald's Omega (total) based on the polychoric correlation matrices of the depression and anxiety item scores suggested adequate reliability estimates at each of the four measurement occasions (Depression = [0.89, 0.88, 0.88, 0.89]; Anxiety = [0.89, 0.88, 0.88, 0.89]).

Statistical analyses

We used to the R-package lavaan for all latent variable models (Version 0.6-4; Rosseel, 2012). The questionnaires used to measure these constructs involve Likert scale items with either 3 or 4 response categories and typically result in positively skewed item scores (see ***Appendix S***). Therefore, in each latent variable model these ordinal item scores will be modeled using threshold parameters and the models will be estimated using DWLS estimation. P-values smaller than .05 were considered statistically significant.

Longitudinal measurement invariance

The R-scripts used to test for longitudinal measurement invariance were based on the scripts reported by Liu and colleagues (2017). Testing for measurement invariance involves fitting a series of nested models, starting with a baseline model. In the present study, this baseline model is a correlated four-factor model, where each factors represent the latent construct at one of the repeated measurements. In each of the three subsequent models a new type of structural equation model parameter is constrained to be invariant across time. This invariance constraint applies to the factor loadings in the weak invariance model, to both the factor loadings and thresholds in the strong invariance model, and to the factor loadings, thresholds and residual variances in the strict invariance model. The chi-square difference (likelihood ratio) test was used to determine whether the more constrained model more poorly fitted the data than the lesser constrained model. Given that the study's large sample size will make even the smallest deviations from measurement invariance statistically significant, we also evaluated the change in the fit indices RMSEA, SRMR and CFI. When the decrease in model fit of the more constrained model was less than 0.015 (RMSEA), 0.030 (SRMR) or 0.002 (CFI), the newly induced longitudinal invariance constraints were considered to show an acceptable fit to the data (Chen, 2007; Meade, Johnson & Braddy, 2008). Fit of the baseline models was evaluated based on the RMSEA (<.07), SRMR (<.10) and CFI (>.95). Missing data on the repeated measurements were handled using available case analysis (pairwise deletion).

Relative stability

After investigating longitudinal measurement invariance, relative stability was investigated by inspecting the correlations between the repeated measures of each latent variable. Another model was fitted to assess relative stability from a different perspective by regressing the latent variable scores at the T2, T3, and T4 on the latent variables score at the preceding time point. The standardized regression coefficients of these autoregressive effects indicate the extent to which the latent variables scores at a certain time point can be predicted from the score on a preceding time point (Borghuis et al., 2017).

Latent growth models

We fitted a univariate growth model for each latent construct and six multivariate growth models for the six pairs of two constructs. The first longitudinal measurement invariance model was used as a baseline model. Assuming conditional independence, the correlations between each construct's repeated latent measurements were fixed to zero and latent intercept and linear slope variables were added to model individual differences in change over time. The factor loadings of these latent growth parameters on the four repeated measurements of each latent construct were fixed to 1 for the latent intercept and to 1, 2, 3, and 4 for the latent slope. In the multivariate models, the correlations between the latent intercepts and slopes of both constructs were freely estimated. Identifying the multivariate growth model required several parameter constraints. First, for each construct the variances of the latent variables were constrained to be equal across the four time points (homogeneity of variances). Second, the correlation between the two latent constructs was constrained to be equal at each time point. Given that second-order multivariate latent growth models require estimation of many parameters, a robustness test was performed by determining the correlation between the latent slopes of two constructs in multivariate *first*-order growth models based on the observed scores of both constructs (i.e. the sum of the item scores at each time point).

Latent Trait-State-Occasion models

A separate LTSO model was fitted for each of the four constructs. As previously, the first longitudinal measurement invariance model was used as a baseline model. For each construct, the correlations between the repeated latent measurements were fixed to zero. To identify the LTSO model, for each latent construct, the latent variable correlations and the state variance at each measurement occasion was fixed to zero (Cole, Martin & Steiger, 2005). Next, for each latent state the factor loadings on both the shared latent trait and the time-specific latent occasion were fixed to one. Autoregressive effects between the four time-specific latent occasion variables were freely estimated (Gana & Broc, 2019). The decomposition into trait and state variance was calculated by dividing respectively the latent state or latent occasion variance at baseline by the sum of these two variances.

Transparency and openness

PROFILES registry data is freely available according to the FAIR (Findable, Accessible, Interoperable, Reusable) data principles for non-commercial (international) scientific research, subject only to privacy and confidentiality restrictions. Data is made available through Questacy (DDI 3.x XML) and can be accessed at (www.profilesregistry.nl). The R-scripts for all analyses reported in this article can be found on the Open Science Framework (https://osf.io/e7ajr/?view_only=41e3de13bb2e4a2eaa4168f0e124fdcc). This study's design and analyses were not preregistered.

RESULTS

Longitudinal measurement invariance

Of all 2625 participants, data from 2597 (99.4%) participants were available for the NA models, 2604 (99.7%) participants for the SI models, 2603 (99.7%) for the Depression models and 2602 (99.6%) for the Anxiety models. Table 3 presents the fit statistics for the models used to test the longitudinal measurement invariance of the DS14 (negative affectivity & social inhibition). For comparison we also included results of the HADS (depression & anxiety). For all constructs, the chi-square statistic of the baseline model was statistically significant, suggesting a strict mismatch between the observed and model implied covariance matrices. However, as the chi-square test is sensitive to large sample sizes, model fit was also evaluated using the RMSEA, SRMR and CFI. Based on these fit indices all baseline models showed adequate fit to the data.

The next step in the longitudinal measurement invariance procedure is to introduce the constraint that the factor loadings of each latent construct are invariant across time (*weak invariance model*). For all constructs, the chi-square difference test was not significant, indicating that this constraint did not lead to deterioration in model fit. This result was corroborated by the small changes in RMSEA, SRMR and CFI compared to the baseline model.

The third step is to constrain the thresholds of each item to be invariant across time (*strong invariance model*). Based on the chi-square difference test, only the depression thresholds appeared to be invariant, yet this test is very sensitive with large sample sizes. The effect size of this invariance violation can be evaluated by computing for each item the change in the estimated response probabilities when freely estimating the thresholds across time, compared to constraining them to be equal. These changes in response probability never exceeded 0.05, suggesting that different item responses than those observed are expected only for a small percentage of participants. Based on guidelines by Liu and colleagues (2017), a change in response probability smaller than 0.05 should be no reason for concern. ***Appendix T*** shows that none of these estimated probabilities exceeded 0.05, indicating that the effect sizes of the invariance violations were very small. Moreover, for all constructs the change in RMSEA, SRMR and CFI relative to the weak invariance model was very small (i.e., < .005), suggesting that these thresholds can be considered invariant across time.

The last step is to constrain the residual variance of the items to be invariant across time (*strict invariance model*). Based on the chi-square difference test this constraint adequately fitted the data of the depression and anxiety models, but did not fit the data for the SI and NA models. However, again the changes in RMSEA, SRMR and CFI are small compared to the strong invariance model, suggesting that the residual variances can also be considered invariant across time.

These findings indicate that it is safe to assume that the properties of the instruments used to measure NA, SI, depression and anxiety are invariant across time. The evidence in favor of longitudinal measurement invariance renders it unlikely that longitudinal changes in the observed scores resulted from changes in the properties of the measurement instrument. Consequently, any reported change (or absence of change) can be interpreted as change in the latent construct.

Table 3: Fit statistics for models testing the longitudinal measurement invariance of the DS14 (negative affectivity & social inhibition) and HADS (depression & anxiety)

Model	df	Δdf	χ^2	$\Delta\chi^2$	RMSEA (95%CI)	ΔRMSEA	SRMR	ΔSRMR	CFI	ΔCFI
Negative affectivity										
1	320	-	998.5	-	.035 [.033 .037]	-	.034	-	.998	-
2	338	18	1016.4	22.5	.033 [.031 .035]	.002	.034	<.001	.998	<.001
3	380	42	1064.0	74.3*	.031 [.029 .033]	.002	.034	<.001	.998	<.001
4	401	21	1186.2	49.1*	.028 [.026 .029]	.003	.035	.001	.998	<.001
Social inhibition										
1	320	-	3338.8	-	.068 [.066 .070]	-	.052	-	.992	-
2	338	18	3352.6	18.7	.065 [.063 .067]	.003	.052	<.001	.992	<.001
3	380	42	3412.1	97.3*	.061 [.059 .063]	.004	.052	<.001	.992	<.001
4	401	21	3783.2	120.6*	.053 [.051 .054]	.008	.054	.002	.991	.001
Depression										
1	299	-	382.0	-	.018 [.016 .021]	-	.028	-	.999	-
2	317	18	400.3	22.2	.018 [.015 .020]	<.001	.028	<.001	.999	<.001
3	359	42	433.9	41.9	.016 [.014 .018]	.002	.029	.001	.999	<.001
4	380	21	497.6	28.4	.014 [.012 .017]	.002	.030	.001	.999	<.001
Anxiety										
1	299	-	759.4	-	.030 [.028 .033]	-	.037	-	.997	-
2	317	18	767.5	11.8	.029 [.027 .031]	.001	.037	<.001	.997	<.001
3	359	42	820.9	65.7*	.027 [.025 .029]	.002	.037	<.001	.997	<.001
4	380	21	899.6	31.8	.023 [.021 .025]	.004	.039	.002	.996	.001

Model 1 (Configural invariance): Baseline model

Model 2 (Weak invariance): Invariant factor loadings

Model 3 (Strong invariance): Invariant factor loadings & thresholds

Model 4 (Strict invariance): Invariant factor loadings, thresholds & residuals

** p < .05 on the scaled chi-squared difference test (Satorra, 2000). Note that these scaled differences are larger than the raw chi-square differences.*

Table 4: Across four measurements of negative affectivity, social inhibition, depression and anxiety, the relative stability, autoregression coefficients, and latent variable mean differences relative to the first time point, including 95% confidence intervals.

Latent construct	Negative affectivity	Social inhibition	Depression	Anxiety
Latent variable correlations				
$r_{(T1, T2)}$	0.780 (0.748, 0.812)	0.803 (0.778, 0.828)	0.821 (0.785, 0.857)	0.825 (0.795, 0.855)
$r_{(T1, T3)}$	0.736 (0.696, 0.776)	0.819 (0.792, 0.847)	0.805 (0.760, 0.851)	0.820 (0.783, 0.857)
$r_{(T1, T4)}$	0.733 (0.678, 0.779)	0.823 (0.795, 0.850)	0.767 (0.712, 0.822)	0.791 (0.749, 0.834)
Autoregression coefficients				
$\beta_{(T1, T2)}$	0.845 (0.822, 0.867)	0.894 (0.875, 0.912)	0.873 (0.849, 0.897)	0.885 (0.860, 0.910)
$\beta_{(T2, T3)}$	0.879 (0.857, 0.900)	0.929 (0.915, 0.944)	0.927 (0.907, 0.948)	0.932 (0.911, 0.952)
$\beta_{(T3, T4)}$	0.893 (0.868, 0.919)	0.928 (0.911, 0.946)	0.904 (0.877, 0.931)	0.915 (0.891, 0.939)

The first rows of Table 4 report for NA, SI, depression and anxiety, the correlations between the baseline estimate of each construct and the estimates at the three later time points. As these estimates concern the correlation between the latent constructs, they are, opposed the correlations between total questionnaire scores, uncontaminated by the presence of measurement error in the item scores. The correlations suggest that the relative stability of these constructs was moderate to high (Shrout, 1998). Interestingly, as the time interval between the repeated measurements increased from one to two and three years, the relative stability of NA, depression and anxiety decreased, while it slightly increased for SI. Lastly, Table 4 also reports the autoregression coefficients, indicating high rank order stability in the latent variable scores at each time point when regressed on scores at a preceding time point.

Latent growth curve models

When assessing absolute temporal stability using latent growth curve models, the mean of the latent slope indicates whether there is absolute stability averaged across all participants, whereas the variance of the latent slope indicates whether this absolute stability is identical for all individuals. Concluding perfect absolute temporal stability would require that both the mean and variance of the estimated latent slope are equal to zero.

Table 5 shows the fit indices for each of the four univariate growth models, as well as results of the Wald tests indicating whether the mean and variance of the latent intercepts and slopes differed significantly from zero. Though each of the four univariate growth models showed misfit based on the significant chi-square test (which is sensitive to misfit in large sample sizes), the RMSEA, SRMR and CFI all suggested good model fit. Figure 4 shows the estimated individual growth curves for NA, SI, depression and anxiety. The red and blue curves represent decreasing and increasing individual trends respectively. The black line indicates the mean estimated latent growth curve across all participants. These plots suggest that on average there were no large changes in these latent constructs over time. Indeed, Table 5 shows that only the average latent slope of depression differed significantly from zero, with a slight increase in depression over time. Based on these findings, the two Type D personality traits NA and SI showed absolute stability when change over time was averaged across all participants. Mean-level stability was also found for anxiety, but not for depression.

The estimated variance of the latent intercept and slope indicate individual deviations from the average intercept and slope. For all four constructs, the variance of the latent intercept was significantly larger than zero, showing that there were significant individual differences in the baseline scores of the four constructs. For NA, depression and anxiety, the estimated variance of the latent slope was small, yet significantly larger than zero, suggesting that for these constructs the estimated individual growth curves deviated from the mean latent slope. Although participants did on average not change in NA and anxiety over time, the significant estimated slope variance indicated that a considerable number of individuals deviated from this pattern showing either positive or negative change over time.
Table 6 reports the estimated correlations between the latent slopes of the four constructs according to both first- and second-order multivariate latent growth models. First-order growth models do not handle measurement error in the item scores, but served as a robustness test. In general, both models resulted in similar estimates of the correlation between the latent slopes. The weakest slope correlation was found for NA and SI ($r = 0.30$, $p = .142$) and the strongest for depression and anxiety ($r = 0.71$, $p < .001$). Interestingly, the NA slopes correlated substantially with the slopes of both depression ($r = 0.49$, $p = .007$) and anxiety ($r = 0.51$, $p = .001$). Similar estimates were found for the slope correlation of SI with

depression (r = .57, p = .221) and anxiety (r = .56, p = .243), though these estimates failed to reach significance due to the very large standard errors. All growth models involving SI resulted in very broad confidence intervals for the correlation between the latent slopes. Although we are not entirely sure how to explain this, it may have been caused by the low variability in the latent SI slopes, as shown by the result of the univariate growth models reported in Table 5.

Figure 4: Estimated individual growth curves for negative affectivity, social inhibition, depression and anxiety. Red and blue curves represent decreasing and increasing individual trends respectively. The black line indicates the mean latent slope across all participants.

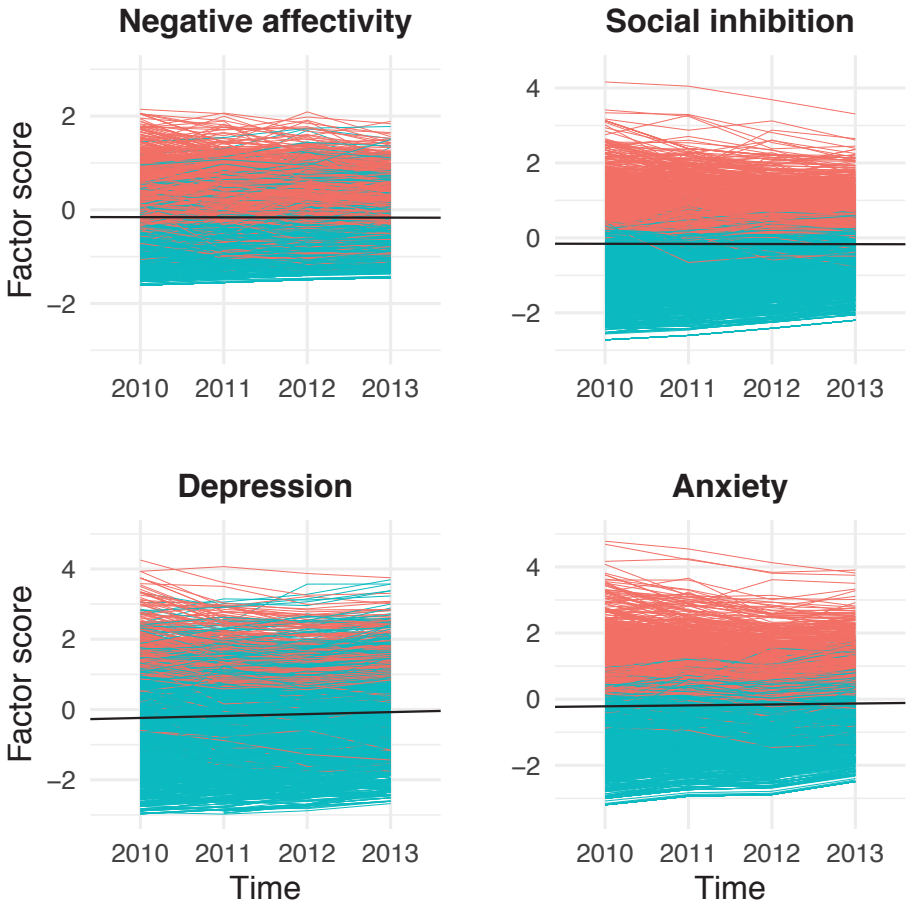


Table 5: Fit indices and individual change in negative affectivity, social inhibition, depression and anxiety in terms of the mean and variance of the latent intercept and slope.

	Negative affectivity	Social inhibition	Depression	Anxiety
Model fit	N=2597	N=2604	N=2602	N=2601
Free parameters	211	211	183	183
χ^2	1262.3*	4016.8*	565.5*	1020.0*
RMSEA (95%CI)	.035 [.033, .037]	.069 [.067, .071]	.018 [.016, .021]	.030 [.028, .032]
SRMR	.034	.052	.029	.037
CFI	.987	.962	.995	.987
Latent growth parameters				
Mean Intercept	-0.253*	-0.291*	-0.527*	-0.401*
Variance Intercept	0.455*	1.111*	1.626*	1.798*
Mean Slope	-0.004	0.027	0.059*	0.035
Variance Slope	0.011*	0.009	0.030*	0.029*

** p < .05*

Table 6: Estimated correlation (95% confidence interval) between the latent slopes of negative affectivity, social inhibition, depression and anxiety, according to both first- and second-order multivariate latent growth models.

Latent slope correlation	First-order growth model	Second-order growth model
NA & SI	0.38 (-0.01, 0.78)	0.30 (-0.10, 0.70)
NA & Depression	0.55 (0.31, 0.79)*	0.49 (0.14, 0.85)*
NA & Anxiety	0.69 (0.45, 0.92)*	0.51 (0.22, 0.80)*
SI & Depression	0.56 (0.20, 0.92)*	0.57 (-0.34, 1.00)
SI & Anxiety	0.51 (0.14, 0.89)*	0.56 (-0.38, 1.00)
Depression & Anxiety	0.66 (0.50, 0.82)*	0.71 (0.36, 1.00)*

** p < .05*

Table 7: Fit indices and trait/state variance proportions (95%CI) for the second-order latent trait-state-occasion models of negative affectivity, social inhibition, depression and anxiety.

	Negative affectivity	Social inhibition	Depression	Anxiety
Model fit				
N	2597	2604	2602	2601
Free parameters	186	186	158	158
χ^2	1059.1*	3533.5*	565.5*	868.4*
RMSEA (95%CI)	.028 [.026, .030]	.059 [.058, .061]	.018 [.016, .021]	.024 [.022, .026]
SRMR	.036	.061	.029	.039
CFI	.991	.967	.995	.991
Proportion explained variance				
Latent trait	.74 [.68, .80]	.83 [.79, .87]	.76 [.69, .83]	.78 [.72, .83]
Latent occasion	.26 [.20, .32]	.17 [.13, .21]	.24 [.17, .31]	.22 [.17, .27]

** $p < .05$*

In sum, the latent growth curves models showed that averaged across all participants, the two Type D personality traits showed absolute stability, as was the case for anxiety. The mean estimated slope of depression differed significantly from zero, yet indicated only a slight increase in depression over time. The variability estimates of the latent slopes suggested that for NA, depression and anxiety the absolute stability did not apply equally to every individual. Lastly, the multivariate growth models revealed that change in these constructs over time was correlated. These results suggest that although NA and SI are personality traits, especially NA appears to covary with changes in depression and anxiety. It seems that part of the NA construct is a stable personality trait, while another part behaves more state-like and is susceptible to internal and external influences. However, the models presented so far do not speak to the extent to which the constructs are traits or states. This last part of this article will investigate this using a Latent trait-state-occasion model.

Latent State-trait-occasion model

Table 7 presents for the LTSO models of NA, SI, depression and anxiety the fit indices and variance estimates of the latent state and trait variables. Similar to the LGC models, each of the four LTSO models showed misfit based on the significant chi-square test (which is very sensitive to misfit under large sample sizes), while the RMSEA, SRMR and CFI all suggested good model fit. For each of the four constructs, after partialling out the measurement error variance, the estimated variance proportions corresponded more to a stable trait than to a changeable state. SI turned out to be most trait-like (83%), followed by anxiety (78%), depression (76%), an NA (74%). However, because the confidence intervals of these percentages showed overlap, the differences between these constructs are likely not statistically significant.

GENERAL DISCUSSION

In this chapter, we reviewed methods commonly used to assess the temporal stability of psychological constructs and focused on Type D as an example. Furthermore, we illustrated how to test the assumption of longitudinal measurement invariance and how to assess temporal stability using various latent variable models that handle skewed and ordinal nature of item scores and measurement error. Based on simulated data we illustrated what types of temporal stability can be detected by several commonly used statistical models. We recommend researchers to explicitly report the type of temporal stability they are interested in and then select a statistical model that can detect such temporal stability. If the researcher is not interested in a specific type of temporal stability, then we recommend them to use multiple stability methods to comprehensively assess individual differences in the change on a construct across time.

In Study 1, our review of temporal stability methods used in the literature on Type D personality covered 24 studies that jointly reported 75 tests for the temporal stability of either Type D personality or its underlying personality traits NA or SI. The review concluded that the temporal stability of both NA and SI was less than optimal based on studies

investigating both the relative and absolute stability. The stability methods encountered in the Type D literature failed to account for measurement error when estimating the relative and absolute stability, thereby risking attenuated stability estimates. Furthermore, the reviewed studies did not test the assumption of longitudinal measurement invariance. When this assumption is violated (or not investigated) researchers cannot exclude the possibility that any observed changes over time were merely caused by changes in the measurement instrument, rather than by change in the psychological construct.

In Study 2, we showed how to handle these issues using various kinds of latent variable models. We assessed both the relative and absolute stability of the personality traits NA and SI and the psychological states depression and anxiety over a period of four yearly measurements. First, we illustrated how latent variable models can take into account the often skewed and ordinal nature of the item scores measuring these constructs. Next, we showed that the assumption of longitudinal measurement invariance was met for all constructs of interest in the current sample of colorectal cancer survivors. Because this assumption was met, any observed change (or stability) in questionnaire scores could be interpreted as being caused by the construct, rather than by the measurement instrument.

Based on the latent variable models, we concluded moderate to good relative stability for NA, SI, depression and anxiety, based on guidelines by Shrout (1998). The four-year relative stability estimates were lowest for NA and highest for SI. This finding is in line with the relative stability estimates discussed in our review, with NA showing a slightly lower median relative stability than SI. Our estimates were often higher than those seen in our review. This may be explained by the fact that the reviewed studies did not adjust for measurement error, thereby risking an underestimation of the true relative stability.

The univariate LGC models indicated absolute stability for NA, SI, and anxiety. Absolute stability could not be concluded for depression, yet the significant increase in depression over time was small. Earlier research on the current dataset has shown that dropouts were more likely to have high depressive symptoms than full responders (Ramsey et al., 2019). This suggests that the depressive symptoms at later measurement occasions may be overestimated. Indeed, our findings indicate that of all four psychological constructs, only

depression shows a significantly positive latent slope, suggesting that on average these participants increased in their depression during the four-year follow-up. However, this estimate should be interpreted with care, as it is possible that without attrition this latent slope estimate for depression would have been closer to zero, similar to the slopes of anxiety, NA and SI.

SI was the only construct with both absolute stability and no significant individual differences in this stability over time. Although NA and anxiety were on average stable over time, these constructs showed significant individual deviations from this absolute stability on group level. The multivariate LGC models revealed that these significant individual differences in change over time correlated between the constructs. In line with our expectations, individual changes in NA moderately correlated with changes in anxiety and depression. This suggests that NA is not entirely a stable personality trait, but may also in part be susceptible to changes in an individual's life, such as an increase or decrease in depression or anxiety. These findings resonate with earlier research by Ossola and colleagues (2015) who used a repeated measures ANOVA to show that the observed DS14 (NA) sum scores covaried over time with the HADS-D (depression) sum scores. That study also used an exploratory factor analysis on the DS14 and HADS items, revealing that the NA and depression items loaded on the same factor, while the SI and anxiety items all loaded on separate factors.

The LGC models suggested that the NA construct may in part reflect the episodic or transient distress also reflected in psychological states such as anxiety and depression.
The LTSO models highlighted the extent to which each of those constructs can be considered a stable trait or a changeable state. All constructs were more trait than state, with SI being most trait like (83%), followed by anxiety (78%), depression (76%) and NA (74%). The finding that SI corresponds more to a stable trait than NA is in line with our findings regarding the absolute and relative stability of these constructs. Interestingly, together with NA, depression and anxiety also turned out to be less trait-like than SI. According to Baltes (1987) both stable and unstable processes underlie most psychological constructs. Personality traits are known to become less stable as the time between the measurements increases (Roberts & DelVecchio, 2000). Other studies have indicated that both trait and

state processes underlie constructs such as depression (Hartlage, Arduino, & Alloy, 1998) and anxiety (Kantor, Endler, Heslegrave, & Kocovski). Our finding is partly in line with an earlier study showing that anxiety and the personality traits behavioral inhibition and neuroticism are more trait-like than depression, which was found to be more episodic in nature (Prenoveau et al., 2011).

One explanation for this unexpected finding is methodological. In the LTSO model, the stable latent trait variable captures the individual differences in the construct that remain unchanged across time, whereas the time-specific latent occasion variables capture the individual differences in the construct that are unique at each time point. In theory, when none of the participants in a dataset change over time, then the LTSO model would indicate that the construct is a completely stable trait. This could also happen in case of floor effects, when there is so little anxiety or depression in the population that most participants show very low scores on these measurements. This implies that the conclusions resulting from LTSO models should always be interpreted in light of the characteristics of the dataset. If the participants in the current dataset would have shown more changes in depression and anxiety over time, then the LTSO models would likely have estimated a smaller proportion of trait variance. It would therefore be interesting for future research to apply the LTSO models to a dataset where individuals change in their depression or anxiety over time.

A limitation of the present study is that the Type D personality traits, depression, and anxiety were measured only once during four consecutive years. Although personality changes happen across longer timespans, they are often quite small and tend to peak in stability after the age of 50 (Roberts & Nickel, 2017). Therefore, it would be interesting to replicate our findings using data stretching over lengthier time periods (e.g., decades). The long interval between consecutive measurement in the current study limits conclusions regarding the temporal stability of state-like constructs, as these tend to vary over much shorter time frames (e.g., hours, days). Therefore, future research could also investigate whether our results hold when these psychological constructs are more frequently measured than once a year. In recent years, there has been a surge in researchers focusing on time-intensive longitudinal data (e.g., daily depression measurements) and latent variable models can certainly be used in that context (e.g. Vogelsmeier, Vermunt, van

Roekel, & de Roover, 2019). A limitation of the models in the current manuscript is that individual change across time was assumed to be linear due to the relatively few measurements. Future research involving more frequent measurements would enable the estimation of non-linear individual latent growth curves.

A limitation of the univariate LGC model (and all other reviewed methods testing for the absolute stability) is that a non-significant average latent linear slope does not necessarily imply that there is absolute mean-level stability, because the non-significant finding may also be caused by insufficient statistical power. The present study involved over 2000 participants and was sufficiently powered to detect small changes in the constructs over time (see ***Appendix U***). We recommend researchers to evaluate the statistical power when testing for absolute stability. Alternatively, researchers can use Bayesian statistics to directly estimate the evidence in favor of the null hypothesis of no difference between the measurements (Kruschke, 2014). Researchers wishing to stay within a frequentist statistics framework are advised to use equivalence tests (Lakens, 2017) to test the plausibility of absolute stability.

In this chapter, we provided a comprehensive review of the methods traditionally used to assess the temporal stability of psychological constructs. We noted how most of the reviewed methods do not handle the measurement error in the questionnaire item scores. At least in the literature on Type D personality, we observed that the crucial assumption of longitudinal measurement invariance is typically not tested. The strength of the current study is that we illustrated how these issues can be handled using several longitudinal latent variable models. As we focused on commonly used latent variables, other latent variables model such as continuous time models (e.g., Haehner, Kritzler, Fassbender & Numann, 2021) fell beyond the scope of the current study. Nevertheless, our work illustrates the general benefit of latent variable modeling when assessing the temporal stability of psychological constructs.

PART IV

Summary and General Discussion

SUMMARY

The overarching aim of this dissertation was to investigate how to operationalize and model Type D personality in statistical analyses. In ***Chapter 1*** we defined Type D personality as a label referring to individuals with high scores on both the personality traits NA and SI. We argued against conceptualizing Type D personality itself as a latent construct. Instead, we contend that the statistical analysis should focus on the underlying personality traits NA and SI and whether they synergistically influence an outcome. Such a synergistic Type D effect occurs when the interaction between NA and SI is significant in such a way that across the entire observed NA and SI score range, the conditional effect of each trait increases with higher scores on the other trait. Various methods have been proposed to estimate Type D effect, yet it remained unclear to what extent those methods perform adequately in detecting a synergistic Type D effect. According to Smith (2011), at least one commonly used method could not distinguish between several causal mechanisms relating NA and SI to an outcome measure. It therefore remains to be seen whether the Type D effects published in the literature have captured the same underlying causal mechanism. Given that Type D effects are typically estimated with methods that do not explicitly account for measurement errors and skewness in the item scores measuring NA and SI, it is unclear whether earlier published Type D effects can be replicated when using methods that can adequately model such characteristics.

The specific aims of this dissertation were threefold. The aim of **Part I** was to investigate whether several observed score methods commonly used to estimate Type D effects could adequately detect a synergistic Type D effect. After concluding that some methods do not adequately test synergistic Type D effects, the aim of **Part II** was to investigate the extent to which this problem has affected the conclusions drawn from the published Type D literature. Lastly, the aim of **Part III** was to investigate whether the synergistic Type D effect can better be estimated using latent variable models than observed score models, in light of the measurement errors and skewness in the ordinal item scores measuring NA and SI.

Part I: Type D personality effects

In **Part I** we used various computer simulations to study the performance of the 2-group, 4-group, and continuous interaction methods in detecting various causal mechanisms relating NA and SI to an outcome measure. In line with Smith (2011), in ***Chapter 2*** we showed that the 2-group method cannot distinguish between several causal mechanisms relating NA and SI to a simulated outcome. The 2-group method not only indicates significant Type D effects when NA and SI synergistically influence an outcome, but also when only NA *or* SI is causally related to an outcome. We found that the 4-group method is also unable to distinguish between these causal mechanisms, due to the positive correlation between NA and SI. Only if this correlation is equal to zero can the 4-group method distinguish a synergistic Type D effect from the causal influence of NA or SI only.

Several authors have proposed a continuous interaction model to estimate Type D effects without risking the spurious findings caused by dichotomizing NA and SI in personality groups (Ferguson et al., 2009; Smith, 2011). This continuous interaction model includes the sum scores of NA and SI to the regression model together with their interaction, to study the synergistic Type D effect. Both ***Chapters 2 and 3*** showed that this continuous interaction model adequately detects synergistic Type D effects when modeled correctly. To correctly estimate the presence of an interaction between two variables, researchers need to include the two first-order effects in the interaction model and investigate the confounding influence of quadratic effects of the two variables involved in the interaction. When not doing so, spurious interaction effect can arise when only NA or SI is linearly or quadratically related to the outcome. Therefore, similar to the 2-group and 4-group methods, a misspecified continuous method cannot distinguish synergy as a causal mechanism from a causal effect of NA or SI only.

When reviewing the Type D literature in ***Chapter 4,*** we identified various published studies that estimated the interaction between NA and SI without modeling the first-order effects. Re-analysis of these studies is necessary to find out whether these interaction effects remain significant after including the NA and SI main effects in the model. Although most reviewed studies included both the first-order NA and SI effects when estimating the interaction, none have investigated possible quadratic effects for NA and SI that might create a spurious

interaction. To the best of our knowledge, currently two published empirical studies using the continuous method to estimate Type D effects have also investigated the quadratic effects for NA and SI (***Chapters 6 and 7***). This suggests that earlier studies showing a significant interaction effect between NA and SI using the continuous method should be re-analyzed to find out whether these interactions are not confounded by unmodeled quadratic NA or SI effects.

Part II: Reconsidering the Type D literature

In **Part II** we investigated the consequences of using the 2-group, 4-group and continuous interaction methods for substantive conclusions drawn in the Type D literature about the prediction of various medical and psychosocial outcomes.

In ***Chapter 4***, we reviewed Type D studies that have used both the 2-group and continuous interaction methods, and studied whether these methods yielded consistent conclusions regarding the presence of a Type D effect. We observed that half of these published Type D effects based on the 2-group method revealed effects of NA or SI only according to the continuous method. In fact, for 38% of these effects NA appeared to be sufficient in explaining individual differences in the outcome. This for instance applied to outcome measures characterized as negative emotional states, such as PTSD symptoms, perceived stress, or emotional wellbeing. For these outcomes, neither SI nor the interaction between NA and SI (and therefore Type D personality) did not add to the explanatory power of NA. In our systematic review, the true proportion of spurious synergistic Type D effects may be even higher, because all reviewed studies using a continuous interaction model did not include quadratic effects for NA and SI.

In ***Chapter 6*** we reanalyzed an earlier study (Nefs et al., 2012) that found an association between Type D personality and anxiety and depression. We showed that the Type D effect, operationalized as the latent interaction between NA and SI, was no longer statistically significant after including the quadratic NA and SI effects in the model. These quadratic effects better explained variation in depression and anxiety than the interaction between NA and SI. A quadratic effect for both NA and SI implies that both personality traits are still important in explaining variation in the outcome measure. However, the non-significant

interaction effect suggests that high scores on *both* traits are not required for being at increased risk on depression and anxiety. Rather, having high scores on either one of the separate personality traits is sufficient because the risk each trait poses gets larger with higher trait scores, independent of scores on the other trait. This finding supports the plausibility of quadratic effects masquerading as interaction effects. This further highlights that it is not only important to reanalyze earlier published Type D studies that have used the 2-group and 4-group method, but also those using the continuous interaction method, because most of those studies failed to investigate the quadratic NA and SI effects.

In ***Chapter 5*** we presented our first attempt at such a large-scale reanalysis project. We conducted an individual patient meta-analysis, reanalyzing 19 published prospective cohort studies investigating Type D personality as a risk factor for adverse events in cardiovascular disease (CVD) patients. We first concluded that not Type D, but only NA is a risk factor for both all-cause and cardiac mortality. We further found a synergistic Type D effect on the risk on adverse events in coronary artery disease (CAD) patients, but not in heart failure (HF) patients. This latter finding is in line with an earlier meta-analysis based on the 2-group method (Grande, Romppel & Barth, 2012). The next section provides a more fine-grained analysis of the synergistic Type D effect on adverse events in CAD patients.

Part III: A latent variable model of Type D personality

Modeling the Type D effect as an interaction between the total scores of the items measuring NA and SI assumes that each of the DS14 item scores is an equally and perfectly reliable measurement of either the NA or SI construct. This assumption is not very realistic, because psychological measurement instruments show imperfect reliability, as is the case with the DS14 (Denollet, 2005). Furthermore, the estimated factor loadings in our exploratory factor analysis in ***Chapter 1*** suggest differences between items in the extent to which they are related to the latent NA or SI constructs. This shows that it is important to use statistical models that can take into account such differences. Structural equation models do not assume perfectly reliable measurement instruments because they can separate the variation caused by the latent NA and SI constructs from the variation due to measurement error. They also do not need to assume that each item is an equally good measurement of NA or SI. Therefore, in **Part III** we argued for modeling Type D personality

effects as a continuous interaction between the latent NA and SI variables in a structural equation model.

In ***Chapter 6*** we used this latent variable interaction model to investigate the association between Type D personality and depression and anxiety in a sample of 3314 adults with Type 1 or 2 diabetes. In contrast with a previous analysis of this data (Nefs et al., 2014), our reanalysis did not reveal an interaction between NA and SI, but a quadratic association for both NA and SI with depression and anxiety. This quadratic effect for instance implies that the association between NA and depression increases at higher NA scores.

In ***Chapter 7*** we used a similar latent variable interaction model, but now to predict the occurrence of major adverse cardiac events (MACE) and myocardial infarctions (MI) or cardiac mortality in a sample of 541 coronary artery disease patients. In this reanalysis we corroborated the findings of the original study (Denollet et al. 2013b) by showing a significant interaction effect between the latent variables NA and SI on these endpoints. We did not find evidence that this synergistic Type D effect was confounded by quadratic effects of NA and SI.

Modeling interactions between latent variables

There exist various methods to model interactions between latent variables. Because modeling such effects is not straightforward, ***Chapters 6 and 7*** used a combination of simulation studies and empirical applications to investigate the performance of several latent interaction models in the context of structural equation modeling. In ***Chapter 6*** we compared several common methods used to model latent interaction effects on a continuous latent outcome variable, while in ***Chapter 7*** we focused on an observed binary outcome variable. We generated many datasets that varied across many design factors and we analyzed each of those simulated datasets with various method to estimate a Type D effect using the continuous interaction method.

The results of ***Chapter 6*** revealed that latent variable models outperform models that estimate the interaction effect based on a product of sum/total scores or factor scores. Of all latent variable models studied in ***Chapter 6***, the LMS method generally produced the least

biased estimates. However, it still overestimated the interaction effect and had an inflated false-positive rate when the latent NA and SI variables were skewed, especially when item scores were ordinal but less so when they were continuous. This aligns with earlier research indicating that LMS shows biased parameter estimates when skewness is introduced at the latent variable level (Kelava & Nagengast, 2012; Kelava, Nagengast, & Brandt, 2014; Cham, West, Ma, & Aiken, 2012).

A limitation of ***Chapter 6*** is that the LMS method was used in a linear structural equation model that assumes the ordinal items to be continuous. Research has indicated that treating ordinal items with skewed score distributions as continuous in statistical analysis using maximum likelihood estimation can result in biased parameter estimates (Dolan, 1994; Rhemtulla, Brosseau-Liard & Savalei, 2012). This suggests that researchers risk underestimated effects and therefore a lower statistical power when they do not deal with the skewness in the observed ordinal item scores when testing interaction effects. We therefore recommend a categorical structural equation model, that models the skewed ordinal items by relating the ordinal scores to an underlying continuous latent variables using polychoric threshold parameters.

In ***Chapter 6*** we did not manage to model the simulated DS14 items using a categorical SEM because this dramatically increased the computation time of our simulation. Fortunately, the model studied in ***Chapter 7*** was computationally less intensive because the outcome variable was not a latent variable with nine indicators, but a single observed binary variable. This allowed us to compare the performance of the LMS method when treating the ordinal DS14 items as continuous versus modeling them at their ordered categorical measurement level. The results indicated that both LMS methods produced underestimated latent interaction effects when the latent NA and SI traits were skewed. However, when these latent variables were normally distributed and the skewness was *only* present in the item scores, using LMS in a categorical SEM produced relatively unbiased estimates, while linear SEM still resulted in underestimated interaction effects, because these skewed ordinal items were inadequately treated as continuous.

Assessing the temporal stability of latent variables

NA, SI, and other personality traits are generally considered stable within individuals across longer timespans. Researchers have used various methods to assess such temporal stability. However, the observed score methods used in the Type D literature did not consider measurement error and skewness in the observed item scores, potentially leading to underestimated estimates of the change in NA or SI across time. To handle this issue, in ***Chapter 8*** we assessed the temporal stability of NA and SI across a four-year follow-up, using various latent variable models to investigate whether these constructs can better be seen as stable traits or changeable states. We first showed the importance of testing the often-ignored assumption of longitudinal measurement invariance. If this assumption does not hold, then changes in the item scores across time do not necessarily reflect changes in the latent construct, but rather changes in the properties of the measurement instrument (e.g., factor loadings). After establishing longitudinal measurement invariance, we showed that although on average NA and SI were stable across the four-year follow-up, there were significant individual differences in the change on these traits over time and individuals changed in their ranking across the years. Moreover, the proportion of trait-variance estimated by latent trait-state-occasion model suggested that NA behaved less like a stable trait (74%) than SI (83%). The estimated proportion of trait variance in the Type D traits was comparable to depression (76%) or anxiety (78%). Indeed, we discovered that the significant individual differences in the change on NA covaried with similar changes in depression and anxiety. This suggests that at least part of the variance in the personality trait NA (and less so in SI) appears to be susceptible to changes in related psychological states such as depression and anxiety.

GENERAL DISCUSSION

The current interdisciplinary dissertation is a collaboration between researchers experienced in psychometrics on the one hand and medical psychology on the other hand. Indeed, our findings have implications spanning across several fields of research. In the current chapter we discuss the key findings of this dissertation, together with its main implications, potential limitations, and future considerations.

Modeling synergy

A key finding of this dissertation is that the 2-group, 4-group and misspecified continuous methods cannot adequately distinguish a causal effect of only NA or SI from a synergy between NA and SI (***Chapters 2 and 3***). Echoing earlier recommendations (e.g., Ferguson et al., 2009; Smith, 2011) we found that such synergy can best be modeled using a continuous interaction method. Another key finding of this dissertation is that the synergistic Type D effect can best be estimated in a structural equation model as an interaction between the latent NA and SI scores, rather than using observed score methods that model the interaction between the NA and SI total scores (***Chapters 6 and 7***). In the published literature, Type D effects have typically been estimated with methods that do not explicitly account for measurement errors and skewness in the item scores measuring NA and SI. Our simulations in ***Chapters 6 and 7*** indicate that such regression analyses on total NA and SI scores may have produced underestimated interaction effects or even false negative conclusions. We therefore recommend Type D researchers to use structural equation modeling (SEM) to account for the measurement errors in the DS14 item scores. Given that these scores are ordinal and often positively skewed, we do not recommend using the default linear SEM, as this assumes that the DS14 items are continuous and normally distributed, resulting in biased estimates of for instance the interaction effect between NA and SI. Instead, we recommend a categorical structural equation model because adequately handles the skewness and limited response options often encountered in ordinal items. To the best of our knowledge, Mplus is currently the only software package that can estimate a latent interaction effect using the LMS method in a *categorical* structural equation model (CATSEM). Our simulation results may help determining the sample size required to test

latent variable interaction effects with sufficient power (***Chapter 6***: Table 5; ***Chapter 7***: Appendix Q).

Reconsidering the Type D literature

Another key finding of this dissertation is that a major part of the published Type D effects based on the 2-group method may actually represent causal effects of NA or SI only (***Chapter 4***). First, this highlights the importance of reanalyzing all published studies in the Type D literature that have based their findings only on the *2-group method*. Secondly, we argue that studies using the *4-group method* should also be reanalyzed, because our simulation in ***Chapter 2*** suggests that the 4-group method tends to result in Type D effects when in reality only NA or SI is causally related to the outcome. Lastly, we also recommend reanalysis of all studies using a *misspecified continuous method* by either not including the first-order NA and SI effects (e.g., ***Appendix D***) or by not assessing whether quadratic NA or SI effects better fitted the data than an interaction effect between NA and SI.

To the best of our knowledge, this implies that all published Type D effects apart from those in the current dissertation should be reanalyzed to find out whether they indeed involve a synergistic effect between NA and SI, or are better explained by (non)linear effects of NA or SI only. The presence of such nonlinear NA or SI effects are certainty plausible, given that distressed personality traits such as NA and SI are claimed to be especially distressing in their higher score ranges. It is therefore conceivable that the lower score range on these traits is less predictive of adverse health related outcomes than the higher score range, implying a non-linear effect.

Indeed, in ***Chapter 6*** we reanalyzed an earlier published study (Nefs et al., 2012) that found an association between Type D personality and anxiety and depression. Our reanalysis based on a latent variable interaction model did not replicate this synergistic Type D effect, as we discovered that found quadratic effects for NA and SI better explained the data. This highlights that it is not only important to reanalyze earlier published Type D studies that have used the 2-group and 4-group method, but also those using the continuous interaction method, because most of those studies did not investigate the quadratic NA and SI effects.

Methodological implications

Dichotomization

In ***Chapters 2 and 3*** we found that the 2-group and 4-group methods cannot distinguish between various causal mechanisms relating NA and SI to an outcome. Both methods start out by dichotomizing both the continuous NA and SI scores. The cost of dichotomization has been extensively described in the methodological literature (e.g., Royston, Altman & Sauerbrei, 2006). Dichotomizing continuous variables does not only result in less statistical power by reducing many individual differences to only two possible scores (Cohen, 1983), but it may also cause spurious findings, especially when dichotomizing two correlated continuous variables (Maxwell & Delaney, 1993; MacCallum et al., 2002). Indeed, in both ***Chapters 2 and 3*** we illustrated that because of the positive correlation between NA and SI, the 4-group method cannot distinguish a synergistic Type D as a causal mechanism from an effect of NA or SI only.

Despite these warnings in the literature, researchers have not refrained from dichotomizing their variables. In our systematic review in ***Chapter 4*** we for instance noticed that the 2-group method is still often used. Furthermore, in other research areas similar methods are used, for instance when modeling defensive hostility (high scores on both defensiveness and hostility; Helmers et al., 1995; Larson & Langer, 1997), anxious depression (Ionescu, Niciu, Henter & Zarate, 2013), mixed states in bipolar disorder (high scores on both mania and depression; Goldberg et al., 1998), or androgynous gender schemas (high masculinity and femininity gender scores; Bem, 1981).

Arguments in favor of dichotomization are that subsequent statistical analyses often result in odds ratio effect sizes that are more easily interpretable than effect sizes resulting from analyses on continuous measures, such as the correlation coefficient or the standardized mean difference (MacCallum et al., 2002). Another often heard argument is that a dichotomized variable sometimes better represent the phenomenon being studied than a continuous measure because the cutoff value used when dichotomizing has clinical significance or corresponds to a diagnostic criterium (Royston, Altman & Sauerbrei, 2006). For instance, when studying obesity as a predictor of COVID-19 hospitalization, researchers

may dichotomize BMI scores using a cutoff of 30. This obesity predictor then supposedly represents the effect of being obese versus not being obese.

Although the desire to use dichotomized predictors is understandable, we still argue against dichotomization in these situations. Dichotomizing a predictor variable is not necessary because logistic regression analyses can produce odds ratios for the effect of both dichotomous as well as continuous predictors. Compared to a dichotomous predictor, the interpretation of the odds ratio effect for a continuous predictor is slightly more complicated because the effect represents the relative change in odds for a change of one on the scale of the continuous predictor variable. Nevertheless, to facilitate interpretation of this effect, researchers can use the regression model estimates to calculate the expected outcome probabilities given various BMI scores. This would allow researchers to compute the predicted probability of being hospitalized with COVID-19 given a specific BMI score. Such predictions can also be included in risk score calculations or decision trees because these calculations do not necessarily require odds ratios for dichotomous predictors but can equally well handle odds ratios for continuous predictors. Predictions based on continuous variables may therefore better facilitate personalized medicine because the actual rather than dichotomized scores can be used to predict the expected outcome. To help clinicians apply the estimates of prediction models to the situation of an individual patient, we encourage researchers to develop risk score calculators based on their statistical models. Lastly, future research should systematically review all studies using bivariate dichotomization, investigate the correlation between two underlying continuous measures and whenever necessary reanalyze the combined effect estimates using a continuous interaction method.

Quadratic effects in interaction models

Interaction models are not unique to research on Type D personality but commonly used in the medical and social sciences. When testing interaction effects, the methodological literature is clear on the importance of investigating the confounding influence of unmodeled quadratic effects, especially when the two interacting variables are correlated (Belzak & Bauer, 2019; MacCallum & Mar, 1995). In ***Chapter 4*** we found that none of the reviewed Type D studies investigated the possible confounding influence of quadratic NA or

SI effects. The extent of this problem in other fields of research remains unclear. Future research could review the enormous literature investigating interaction effects to identify the studies that have tested an interaction effect between two correlated variables without checking the confounding influence of quadratic effects. In line with our findings, reanalysis of those studies may be necessary when the correlation between the two interaction variables deviates from zero.

A possible issue with adding quadratic effects to an interaction model is that the quadratic terms are correlated with the interaction term because the same variable is used when computing both terms, resulting in larger standard errors for the parameter estimates and therefore in less power to detect the interaction effect (Belzak & Bauer, 2019). Indeed, Table 2 in ***Chapter 5*** shows that the confidence intervals of the interaction effects are narrower when quadratic effects are not included in the model than when they are included. Furthermore, the higher the correlation between the quadratic and interaction terms, the higher the risk on biased parameter estimates due to multicollinearity (MacCallum & Mar, 1995).

Arguably a model including quadratic effects is less parsimonious and less readily interpretable, but one could argue that if they appear robust, they should be explained and taken into account in predicting outcome measures nonetheless. Lubinsky & Humphreys (1990) recommended estimating separate models including either the quadratic effects or the interaction effect and to compare the fit of these models using fit indices or model comparison tests. In addition, MacCallum & Mar (1995) argued that when modeling interactions and quadratic effects it is especially important to use latent variable models because the measurement error in quadratic and interaction terms is considerably higher than the measurement error in individual item scores. Latent variable models can estimate the association between latent variables that are free of measurement error and therefore are not affected by differences in reliability.

Table 1 in ***Chapter 6*** shows an example of how we compared the fit of latent variable models including either interaction or quadratic effects. A limitation of using latent variable models is that they require a larger sample size than regular regression models. If sample size is too

small, then in line with others (Belzak & Bauer, 2019) we recommend testing the models in separate regular regression analysis and compare the R-squared of both models. MacCallum & Mar (1995) further showed that the risk on selecting the wrong model increased when the correlation between the variables involved in the interaction exceeds 0.7. They argue that researchers should in such situations consider past empirical findings and the theoretical plausibility of each model when interpreting their fit to the data. This issue is less relevant to research on Type D personality as the correlation between NA and SI is approximately 0.5. If researchers want to directly analyze the quadratic- and interaction terms in a single model then we recommend using a Bayesian regularization approach in a latent variable model (Brandt, Cambria & Kelava, 2018).

Skewed ordinal item scores

In the medical and social sciences, researchers commonly test interaction effects between latent constructs by modeling the total questionnaire scores in a linear regression analysis. These total scores are often based on questionnaires showing skewed ordinal score distributions (Reise & Waller, 2009). In general, ignoring skewness or measurement error in questionnaire item scores often results in biased associations between latent variables, especially when testing interaction effects (Lodder, 2022). A regression analysis on total scores does not take into account the presence of measurement error and skewness in the ordinal item scores and hence risks underestimated associations. Although a linear SEM can separate measurement error variance from true score variance, by treating positively skewed ordinal item scores as if they were continuous and normally distributed, the interaction effects between latent constructs are underestimated (***Chapter 7***; Lodder, 2022). ***Chapters 6 and 7*** showed that the severity of this issue increases with lower scale reliability and more skewness in the questionnaire item scores. We therefore recommend researchers to use a categorical structural equation model to estimate associations between latent variables measured with ordinal items showing a non-normal score distribution.

The discussion above suggests that many earlier published interaction effects based on either linear SEM or linear regression on observed scores are likely underestimated. In poorly powered studies this underestimation may even have resulted in non-significant interaction effects. Such an increase in Type II errors could help explain why researchers

have not been able to replicate various findings in the medical and psychological sciences (Ioannidis, 2005). Indeed, a many analysts project (Hoogeveen et al., 2022) asked 120 teams consisting of scientists to test the same latent variable interaction effect. A commentary article by Lodder (2022) showed that teams using a linear regression on average estimated smaller interaction effects than teams using a categorical SEM.

In future replications of studies involving interactions between latent constructs measured with skewed ordinal item scores, we recommend researchers to consider using a categorical SEM in order to prevent underestimated associations due to skewness or measurement error (Loken & Gelman, 2017). We recommend conducting large scale reanalysis projects using individual participant data (***Chapter 5***) to efficiently reanalyze multiple studies testing a similar hypothesis, especially because the sample size of individual studies may not be large enough to achieve sufficient power to test the hypothesis with a categorical SEM.

Implications for Type D personality research

The nature of NA and SI

In ***Chapter 8*** we found that longitudinal changes in NA tended to covary with similar changes in depression and anxiety. Covariation of NA with those episodic mood status could imply that the reported relations between NA and various outcome measures (for instance those reviewed in **Chapter 4**) may partly be confounded by individual differences in anxiety or depression. Nevertheless, Denollet & Pedersen (2008) found that Type D personality was still a risk factor of major adverse cardiac events (MACE) after controlling for depressive symptoms and various clinical risk factors. However, their analysis was based on the 2-group method, which cannot distinguish causal effects of only NA or SI from a synergistic Type D effect (***Chapter 2***). Consequently, the incremental validity of Type D personality remains unclear because the findings could equally well reflect that not Type D but only SI predicts MACE after controlling for depression. Therefore, future research could shed light on the underlying causal mechanism by reanalyzing this data using a continuous interaction method.

Both Type D personality traits correlate strongly with both anxiety and depression (Svansdottir et al., 2012). Additionally, NA is strongly associated with neuroticism and SI with introversion (De Fruyt & Denollet, 2002; Denollet, 2005). The conceptual overlap between these constructs raises the question of whether synergistic Type D effects are confounded by interaction effects between constructs strongly related to NA and SI. This incremental validity of Type D personality could for instance be studied by testing whether a synergistic Type D effect can still be found after including the interaction between neuroticism and introversion in the model.

Type D and adverse cardiac events

Based on another reanalysis reported in ***Chapter 5***, we can conclude that CAD patients with a tendency to experience negative emotions (i.e., NA) are more at risk of adverse events when they are also inhibited in self-expression during social interactions (i.e., SI). Given that both NA and SI are measured with seven items that each measures a specific aspect of these personality traits, it would be interesting to investigate whether the synergy between NA and SI is also found at the individual item level. An assumption of a unidimensional reflective latent variable model is that all items are exchangeable measures of the latent construct, except for their precision and difficulty (Rigdon et al., 2011). This may cause researchers to pay less attention in analyses to item scores than to constructs. However, in the context of the latent variable depression, it is often worthwhile to focus on individual item scores, such as the item measuring suicidal ideation. It can be misleading to consider this item and one measuring sleep disturbance as an exchangeable measure of depression (Fried & Nesse, 2015). Therefore, item-level analyses can provide a more fine-grained analysis of whether the associations found at the latent variable level equally apply to all items measuring that latent construct, or whether specific items are driving the latent variable effect.

Item-level interaction between NA and SI

In the context of Type D personality, a more concrete interpretation of the findings of our individual patient data meta-analysis (***Chapter 5*****)** requires investigating how NA and SI interact at the individual item level. Table 2 shows the results of exploratory multilevel logistic regression analyses on 3211 cardiovascular disease patients (***Chapter 5*** data), testing whether interaction effects between individual NA and SI items predict the occurrence of

adverse events. The bold-faced cells indicate the item-level interaction effects that were statistically significant and not confounded by quadratic effects of these two NA and SI items. Note that these are item-level interaction effects, so they are not adjusted for the influence of other item DS14 items and their interactions because these models were not identified. Almost all estimated odds ratios point in the direction of a Type D effect because they are larger than one. However, there are considerable differences between these estimates and only a few items appeared to explain how the NA and SI constructs interact in cardiovascular disease patients for increasing the risk on adverse events.

Two SI items interacted with several NA items. One of those assesses the extent to which respondents *do not make contact easily* (DS1), and the other the extent to which respondents *tend to keep others at a distance* (DS11). The NA item that interacted most with SI items assesses the extent to which respondents *worries about unimportant things* (DS2). Table 1 also reports the results of the interaction effect between individual items and the total score of the other trait. These estimates further support the importance of the DS1, DS2 and DS11 items in explaining the interaction between NA and SI.

Implications of the item-level interactions

This exploratory analysis could indicate that cardiovascular disease patients with a specific personality profile are at increased risk on experiencing adverse events. Those patients tend to indicate they are closed persons that worry a lot about unimportant things, do not easily make contact and keep others at distance. Such individuals may experience mental stress due to excessive worrying and may even develop clinical depression. Both chronic stress (Rosengren et al., 2004) and depression (Gan et al., 2014) are considered important psychosocial risk factors for CAD incidence and progression, alongside many other factors (Piepoli et al., 2016), such as low socio-economic status (Albert, Glynn, Buring & Ridker, 2006), lack of social support (Barth, Schneider & von Känel, 2010), anger and hostility (Chida & Steptoe, 2009).When patients with chronic stress or depression keep others at a distance, then they may not seek help when experiencing physical or mental symptoms. Due to the lack of social support these patients are also at risk of not being surrounded by friends or family who can motivate them to seek help. Together, these factors likely increase the risk on missing early warning signs of future adverse cardiac events.

Future research could also confirm our exploratory item-level analysis, to investigate whether the specific tendencies to keep others at a distance and to take a gloomy view of things are the key drivers in explaining how NA and SI synergistically affect various medical and psychosocial outcomes. Another way to study how NA and SI interact on the item level is by applying a moderated network model to the DS14 item scores and one or more outcome variables. Recent advanced within network modeling have made it possible to test for interaction effects between items on an outcome variable (Haslback, Borsboom & Waldorp, 2019). Such moderated network models can be used to determine the interaction between individual NA and SI items on other items in the network (e.g., items measuring other constructs such as depression or anxiety). Lastly, we recommend researchers studying Type D personality to continue doing large-scale re-analysis projects to find out which published Type D effects on various medical and psychosocial outcomes are still significant when using the continuous interaction model.

Table 1: For 64 multilevel logistic regression analyses, the estimated odds ratios for the interaction effects between individual NA and SI items on the occurrence of adverse events in cardiovascular disease patients. Bold faced cells represent statistically significant interaction effects that were not confounded by quadratic effects of the corresponding NA and SI items.

	SOCIAL INHIBITION							
NEGATIVE AFFECTIVITY	Does not make contact easily (DS1)	Does not often talk to strangers (DS3)	Inhibited in social interactions (DS6)	Difficulties starting a conversation (DS8)	Closed kind of person (DS10)	Keeps others at a distance (DS11)	Does not find things to talk about (DS14)	Sum score
Worries about unimportant things (DS2)	**1.11***	**1.12***	1.08	1.06	**1.10***	**1.14***	**1.17***	**1.17***
Often feels unhappy (DS4)	1.09*	1.04	1.00	1.04	1.08	1.06	1.06	1.11
Is easily irritated (DS5)	1.01	1.02	1.05	1.04	1.02	1.08	1.06	1.06
Takes gloomy view of things (DS7)	**1.18***	1.02	1.07	1.09	1.07	1.10*	1.07	**1.17***
Is often in a bad mood (DS9)	1.04	0.95	1.00	0.99	0.98	0.97	0.97	0.97
Often worries about something (DS12)	1.04	1.05	1.04	1.02	1.07	1.07	1.09	1.09
Is often down in the dumps (DS13)	1.06	1.01	1.04	1.01	1.05	**1.12***	1.02	1.04
Sum score	**1.19***	1.05	1.01	1.04	**1.10***	**1.14***	1.10	**1.19***

* $p < .05$.

Strengths and limitations

Strengths

A key strength of this dissertation is its collaborative nature, not only by being an interdisciplinary project, but also by including a large-scale international collaboration in the form of an individual patient data meta-analysis. Another strength is that it combined empirical analyses with simulation studies to show how methodological differences between studies have likely contributed to the inconsistent findings found in the Type D literature. Furthermore, we aimed investigated the predictive value of Type D personality on various medical and psychological outcomes using state-of-the-art statistical methods and we have used computer simulations to support the adequacy of our chosen methods. Another strength of this dissertation is that we designed our simulations in such a way that the characteristics of the simulated resembled those of the often-muddy empirical datasets encountered in medical and psychological research. Our simulated datasets for instance involve items with skewed ordinal score distributions and scales with suboptimal reliability. As a result, the characteristics of our simulated are more like those encountered in empirical data, strengthening the ecological validity of our research.

Limitations

Dimensional conceptualization

Throughout this dissertation we assumed that NA and SI are dimensional constructs and that the Type D effect is a synergy between these two latent dimensions. These assumptions motivated us to generate the data in our simulations according to a continuous latent interaction model. One could argue that under such a data generating model, it is perhaps not surprising that our simulation studies in ***Chapters 2, 3, 6 and 7*** showed that the continuous interaction method produced the least biased synergistic Type D effects. In **Chapter 1** we argued for why a dimensional conceptualization is more appropriate than a typological or mixture conceptualization. Nevertheless, we concede that *if* Type D personality is a typological or mixture construct, then generating data from such a model may result in lower performance of the continuous interaction method to detect true Type D effects than our simulations have suggested based on the continuous interaction model as the data generating mechanism. Nevertheless, our simulations still show that *if* Type D can

best be modeled as a synergy between NA and SI, then the 2-group, 4-group and misspecified continuous methods should not be used.

Future simulation studies could study whether the continuous interaction method can still adequately detect synergistic Type D effects when other data generating models are hypothesized, such as a mixture model involving several latent personality classes. In one previous study researchers have simulated NA and SI scores from a mixture model (Hillen, 2017), with individuals simulated in the Type D class having higher population mean NA and SI scores than individuals in the non-Type D class. Because this study focused on whether latent variable mixture models could adequately detect the two simulated classes and their within-class differences in NA and SI, future research could further expand this simulation study by investigating whether methods commonly used in Type D research can adequately detect causal relations between the personality classes and simulated outcome variables.

Simulation design

An important limitation relates to the design choices in our simulation studies. We analyzed our simulated datasets using either a logistic or linear regression model. It remains unclear whether our findings generalize to other statistical models frequently used in Type D research. Future research could extend our simulations by focusing on other statistical models, such as the cox regression model or mixed effects models.

Another limitation involves our decision to generate the simulated item scores either according to a normal distribution, or according to a positively skewed normal distribution, because clinical measurement instruments often result in such score distributions (Reise & Waller, 2009). However, it remains unclear how the latent interaction models studied in **Chapters 6 and 7** perform when the item scores are either negatively skewed, or when they are not skewed but show either high or low kurtosis. Researchers could therefore study whether latent interaction methods can also accommodate non-normal item scores beyond those that are positively skewed or normally distributed.

In our simulations we also assumed the same skewness and kurtosis values for all items within a particular scale. In empirical data items frequently differ in their distributions. For

instance, scales measuring depression often include items assessing the symptoms suicidal ideation and tiredness. These items often show different score distributions, with tiredness being more normally distributed and suicidal ideation being highly positively skewed. Our simulations showed that a regression analysis of total scores produces attenuated interaction effects when the items are all positively skewed. However, it remains unclear whether this bias equally applies to other patterns of non-normally distributed item scores, such as when the distributions of only some items are positively skewed, or when all distributions are negatively skewed, or when some distributions are negatively- and some positively skewed, or when the distributions are not skewed but kurtotic. Future simulations can test whether categorical SEM adequately handles more diverse patterns of non-normally distributed item scores within a single scale.

Nonparametric interaction model

In this dissertation we considered the association between Type D personality and other constructs to be a synergistic effect between NA and SI. We concluded such synergy when there was a significant interaction effect between NA and SI, in such a way that the conditional effect of each trait increased across the entire score range of the other trait. A limitation of this definition may be that this synergistic effect does not occur across the *entire* score range of NA and SI, but only in the higher score range of both traits. A problem resulting from such a definition would be that this particular type of synergy is not very well captured using a standard continuous interaction model because such interaction effects occur across the entire score range and cannot be absent at lower scores and start at higher scores. Future research could use more advanced statistical approaches to model this kind of synergy. One option would be to estimate interaction effects using a non-parametric regression analysis. This method does not impose a specific functional from on the effects of the continuous predictors (Cadarso-Suárez, Roca-Pardiñas, & Figueiras, 2006), allowing for smoother associations than the linear or quadratic functions commonly used in parametric regression analyses.

General conclusion

This interdisciplinary dissertation brought together experts from the fields of medical psychology and psychometrics. We performed a psychometric evaluation of the construct Type D personality and studied the way it can best be modeled when testing its role as a risk factor in medical and psychological research. Assuming that the causal mechanism underlying a Type D personality effect is a synergy between NA and SI, we argued that it can best be modeled as a continuous interaction effect. Our simulations showed that this method is better able to detect synergistic Type D effects than personality group methods. This implies that earlier published findings based only on these methods should be reanalyzed with the continuous interaction method. In a first attempt at this endeavor, we were able to replicate the cornerstone of Type D research in a large individual patient-data meta-analysis, namely the risk it poses to adverse events in CAD patients.

We were not the first to argue in favor of using the continuous interaction method to study synergistic effects in Type D research (e.g., Smith, 2011). Although the limitations of modeling sum scores, dichotomizing continuous variables, and the pitfalls of interaction modeling are well known in methodological circles, many applied researchers still appear to be unaware of them. This highlights the importance of building bridges between fields of research that still too often operate in isolation. This interdisciplinary dissertation shows how a collaboration between experts from different fields can result in a synergistic effect, similar to how the effects of NA and SI are considered to be more than the sum of their parts. By combining the perspectives, knowledge, and skills of a diverse set of experts, novel insights can emerge that would have been difficult to achieve otherwise (Weiss, Anderson & Lasker, 2002). Collaboration is one of the pillars of science. We hope this dissertation will inspire many future collaborations between medical psychologists and psychometricians, perhaps even resulting in the emergence of a new field called medical psychometrics.

"Difference is the beginning of synergy" – Stephen Covey (1989)

APPENDIX

Appendix A: Multilevel exploratory factor analysis (chapter 1)

Box 1: Mplus script for multilevel exploratory factor analysis

```
INPUT INSTRUCTIONS
  TITLE: DMEFA
  DATA: FILE IS datm.csv;
  VARIABLE:
  USEVARIABLES
  x1 x2 x3 x4 x5 x6 x7 x8 x9 x10 x11 x12 x13 x14 x15;
  NAMES
  x1 x2 x3 x4 x5 x6 x7 x8 x9 x10 x11 x12 x13 x14 x15;
  CATEGORICAL
  x1 x2 x3 x4 x5 x6 x7 x8 x9 x10 x11 x12 x13 x14 ;
  CLUSTER IS x15;
  ANALYSIS: TYPE IS TWOLEVEL EFA 1 4 1 4;
  PLOT: Type = plot2;
```

Appendix B: Sensitivity analyses (chapter 2)

Figure A1: For each Type D operationalization and for varying levels of the NA and SI main and less than additive interaction effects on a dichotomous outcome, the mean estimated standardized regression coefficient (upper), percentage relative bias (middle) and percentage significant results (lower) of the Type D effect.

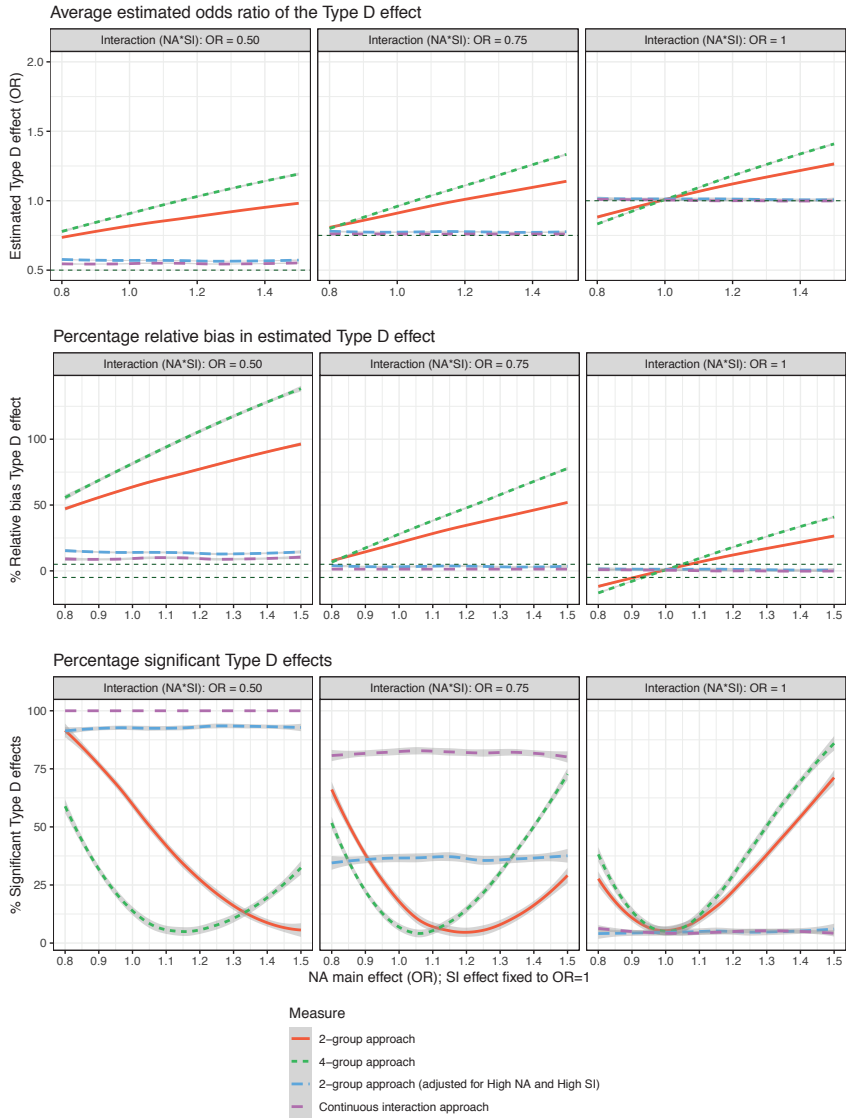


Figure A2: For each Type D operationalization and for varying levels of the NA and SI main and interaction effects on a continuous outcome, the mean estimated standardized regression coefficient (upper), percentage relative bias (middle) and percentage significant results (lower) of the Type D effect.

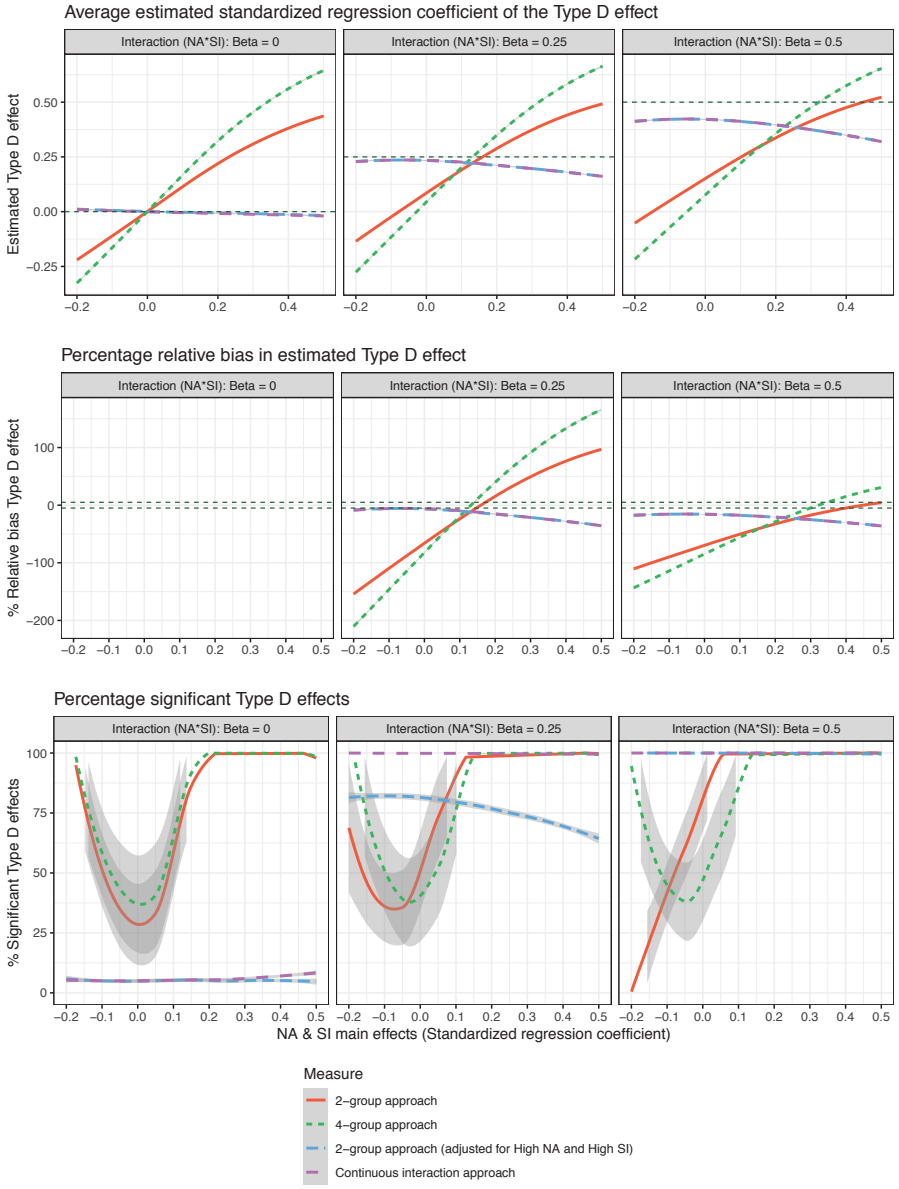
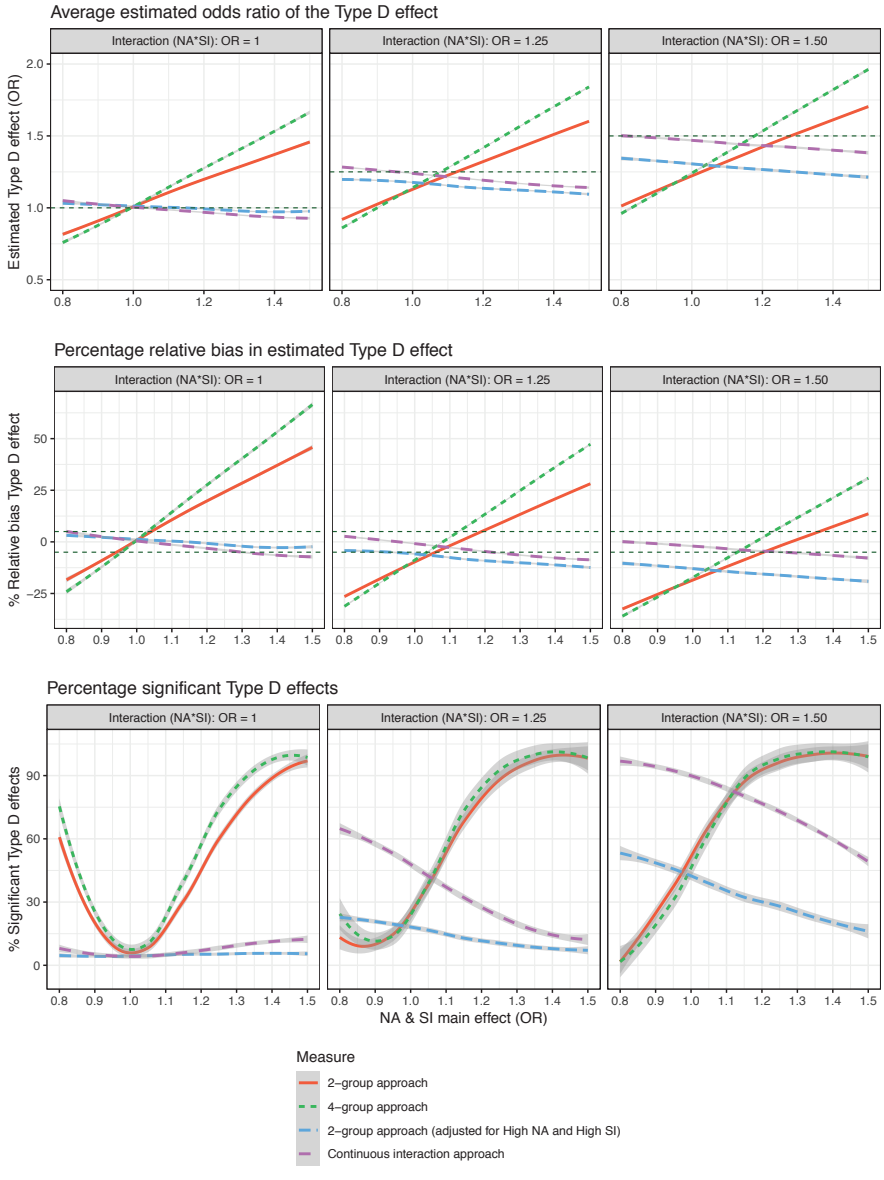


Figure A3: For each Type D operationalization, for varying levels of the NA and SI main and interaction effects, the mean estimated standardized regression coefficient (upper), percentage relative bias (middle) and percentage significant results (lower) of the Type D effect, given positively skewed NA and SI item scores.



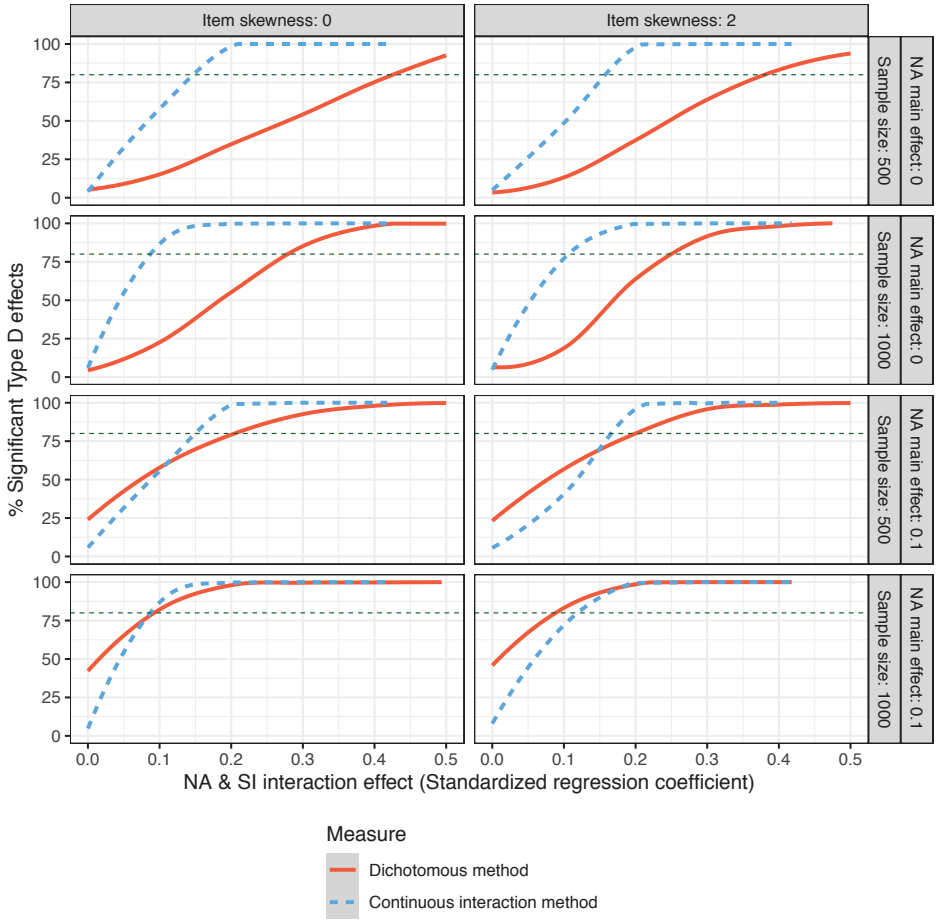
Appendix C: Power analysis continuous method and dichotomous method (*chapter 4*)

The purpose of this small simulation study was to determine the statistical power to detect a Type D effect according to the dichotomous method and the continuous interaction method. 24000 datasets were generated, with 500 datasets in each of the 48 simulation conditions, varying over item skewness (0 or 2), sample size (500, 1000), the standardized regression coefficient of the underlying NA main effect (0 or 0.1), and the standardized regression coefficient of the interaction between NA and SI (0, 0.1, 0.2, 0.3, 0.4, 0.5). In each dataset, the DS14 item scores were generated from a two-factor model with factor loadings ranging between 0.70 and 0.80 for the 7 items loading on each of the NA and SI factors. The factors had a mean of zero and a correlation of 0.5. The generated continuous item scores were converted to either normally distributed or skewed ordinal item scores by using different sets of threshold parameters. In each dataset a continuous outcome variable was generated based on a linear regression model with the regression coefficient of SI fixed to zero, the NA main effect varied (0 or 0.1) and a synergistic Type D effect (positive NA*SI interaction) varying in size (0, 0.1, 0.2, 0.3, 0.4, 0.5) across the simulation conditions. The residual error term was assumed to be normally distributed with a mean of zero and a variance of 2. For each of the 48 simulation conditions the 500 generated datasets were analyzed using a linear regression analysis. The percentage of datasets in which a significant Type D effect was found was determined for each of the three methods and visualized in Figure C1.

The results in the upper two rows concern scenarios in which only a positive interaction between NA and SI was causally related to the outcome. The findings indicate that if a true interaction was present, then the continuous method showed more power to detect these effects than the dichotomous method. Hence, given the same sample size, the dichotomous method showed a higher Type II error rate compared to the continuous method. On the other hand, if a true interaction effect was absent (i.e., zero on the x-axis), the continuous method produced 5% false positive interaction effects, exactly corresponding to the nominal significance level. The dichotomous method also showed acceptable false positive rates, but only when no other causal NA or SI effects were underlying the data. The lower two rows of Figure A1 show that when NA has a main effect, the dichotomous method may cause researchers to falsely conclude the presence of a Type D effect when no such effect (i.e., an interaction between NA and SI) is underlying the data. These findings were robust against

various effect sizes and sample sizes and were not affected by the presence of skewness in the DS14 item scores.

Figure C1: The percentage of significant Type D effects according to dichotomous method (red solid line) and continuous method (blue dashed line), varying over sample size, item skewness, the underlying NA main effect and NA*SI interaction effect. In all conditions, the size of the SI main effect was fixed to zero.



Appendix D: Important excluded studies (*chapter 4*)

Table D1 shows a selection of 14 published studies that were excluded from our review. The authors of these studies took seriously the recommendation to analyze Type D personality continuously by investigating the interaction between NA and SI (except for one study where the sum of NA and SI rather than their product was investigated). However, in these excluded studies, these products (or sums) were used in subsequent analyses without adjusting for the continuous NA and SI main effects. As a result, these analyses may suffer from a problem similar to that of the dichotomous approach: they cannot distinguish between four kinds of underlying effects: [1] NA main effect; [2] SI main effect; [3] Additive Type D effect, or [4] Synergistic Type D effect.

This problem may be circumvented by mean-centering the NA and SI variables before multiplying them (***Chapter 3***). However, the studies in Table D1 did not report that the NA and SI scores were mean-centered before their multiplication. One could argue that the scores were mean-centered and the authors forgot to report it, but there are in fact several reasons to assume that raw scores rather than centered scores were used. First, several studies report the correlations between NA, SI and their product score (Condén, Ekselius & Aslund, 2013; Wiencierz & Williams, 2017; Williams, Bruce & Knapton, 2018). If NA and SI would have been mean-centered before multiplying them, these raw scores would barely correlate with their multiplication. However, the substantial correlations in these studies suggest that raw scores were multiplied instead of the mean-centered scores. Second, several studies (Zuccarella-Hackl et al., 2016; Cho & Kang, 2017; Smith et al., 2018) show that the scale of the product variable was larger than expected. If NA and SI would have been mean-centered before multiplication, then the scale of their product would include both negative as well as positive values, with a mean close to zero. Based on the reported mean scores of two studies (Cho & Kang, 2017; Smith et al., 2018) and the plotted product scores of one study (Zuccarella-Hackl et al., 2016), it appears that the product scores only contain positive values with a mean high above zero. Together these results suggest that raw scores rather than mean-centered scores have been used when multiplying NA and SI. Using only the multiplied raw score in subsequent analyses, without adjusting the effect for

the NA and SI main effects, results in biased estimates of the Type D effect (see ***Chapter 3*** for empirical support for this bias).

As the studies in Table D1 either did not investigate the NA and SI main effects, it remains unclear in what way the two Type D personality traits are related to the dependent measures. However, in some of these studies the bivariate correlation between both traits and the dependent measure were reported. For instance, Williams, Bruce and Knapton (2018) showed that the product of NA and SI was positively correlated with alcohol dependence. However, the bivariate correlations also indicate a significant positive correlation for SI with alcohol dependence, while the correlation with NA is very weak and not statistically significant. This pattern excludes the possibility of an additive Type D effect, or an NA only effect. Yet it remains unclear whether the underlying effect is caused by SI only or by an interaction between NA and SI above and beyond the SI effect. Because the studies listed in Table D1 did not meet our inclusion criteria due to not using the correct continuous method, we have excluded them from our systematic review.

Table D1: Studies in which Type D personality (operationalized as NA+SI or NA*SI) was associated with an outcome while the analyses were not adjusted for NA and SI main effects.

Study	Outcome	Type D personality	Analysis	Statistic	*p*-value	NA & SI main effects	Conclusion
Whitehead et al. (2007)	Cortisol awakening response	NA*SI Product	Linear regression	$\Delta R^2 = .079$	0.008	Not significant	Unclear
Gilmour & Williams (2011)	Wellness maintenance	NA*SI Product	Correlation	$r(198) = .298$	< .01	Not investigated	Unclear
Williams et al. (2011)	Illness perception	NA*SI Product	Correlation	$r(190) = .52$	< .01	Not investigated	Unclear
Damen et al. (2013)	Depression	NA*SI Product	Linear regression	$\Delta R^2 = .46$	< .001	Not investigated	Unclear
Conden et al. (2013)	Sleep hours during weekend	NA+SI sum	Linear regression	OR = 1.683	< .001	Only NA correlates with outcome	NA effect (synergistic Type D effect unclear)
O'leary et al. (2013)	Blood pressure	NA*SI Product	Mixed ANCOVA	F(1,66) = 4.58	0.036	Not investigated	Unclear
Booth & Williams (2015)	Eating behavior	NA*SI Product	Correlation	$r(185) = -.313$	< .001	Not investigated	Unclear
Williams, Abbott & Kerr (2015)	Health behavior	NA*SI Product	Correlation	$r(215) = -.460$	< .001	Not investigated	Unclear
Zuccarella-Hackl et al. (2016)	Cell proliferation rate	NA*SI Product	Linear regression	$\Delta R^2 = .13$	0.022	Both correlate with outcome	Additive Type D effect (synergistic Type D effect unclear)
Wiencierz & Williams (2017)	Self-efficacy	NA*SI Product	Correlation	$r(187) = -.41$	< .001	Not investigated	Unclear
Cho & Kang (2017)	PTSD symptoms	NA*SI Product	Linear regression	$t(178) = 9.43$	< .001	Not investigated	Unclear
Williams et al. (2018)	Alcohol dependence	NA*SI Product	Correlation	$r(136) = .198$	< .05	Only SI correlates with outcome	SI effect (synergistic Type D effect unclear)
Dehghani et al. (2018)	Life satisfaction	NA*SI Product	Correlation	$r(259) = -.50$	< .001	Not investigated	Unclear
Smith et al. (2018)	Physical symptoms	NA*SI Product	Linear regression	$t(99) = 3.18$	0.002	Not investigated	Unclear

Appendix E: Reported Type D effects for all included studies (*chapter 4*)

Table E1: For all studies included in the systematic review, the estimated Type D effect and p-value based on the dichotomous method and continuous method

Study*	N	Dependent measure	dichotomous method		Continuous method						Conclusion
			Type D	p	NA	p	SI	p	NA*SI	p	
Coyne (2011)	706	All-cause mortality	HR=0.78	0.294	HR=0.87	0.156	HR=0.97	0.659	HR=0.9	0.144	No effect
Grande (2011)	1040	All-cause mortality	HR=0.99	0.950	HR=1.01	0.710	HR=1.02	0.590	HR=1.01	0.880	No effect
Howard (2011)	89	Heart rate reactivity	Eta^2=0.05	0.030	NR	NS	NR	NS	R2=0.04	0.036	Synergistic Type D effect
Howard (2011)	89	Cardiac output reactivity	Eta^2=0.07	0.015	NR	NS	NR	NS	R2=0.07	0.008	Synergistic Type D effect
Howard (2011)	89	Total peripheral resistance reactivity	Eta^2=0.05	0.048	NR	NS	NR	NS	R2=0.06	0.020	Synergistic Type D effect
Mommersteeg (2011)	453	Waist circumference	t=0.95	0.331	B=-0.17	0.813	B=-0.17	0.807	B=0.23	0.631	No effect
Mommersteeg (2011)	453	BMI	t=0.45	0.504	B=-0.01	0.967	B=0.13	0.545	B=0.22	0.137	No effect
Mommersteeg (2011)	453	Trigyceride	t=0.01	0.911	B=7.6	0.839	B=-51.6	0.212	B=-25.3	0.278	No effect
Mommersteeg (2011)	453	High density lipoprotein levels	t=4.42	0.036	B=0.45	0.577	B=0.73	0.341	B=-0.34	0.537	Only dichotomous effect
Mommersteeg (2011)	453	Systolic blood pressure	t=0.59	0.445	B=-0.88	0.321	B=0.01	0.992	B=0.19	0.766	No effect
Mommersteeg (2011)	453	Diastolic blood pressure	t=0.09	0.771	B=-0.19	0.732	B=0.01	0.986	B=0.1	0.795	No effect
Mommersteeg (2011)	453	Glucose levels	t=5.14	0.024	B=0	1.000	B=0.02	0.441	B=0.03	0.142	Only dichotomous effect
Mommersteeg (2011)	453	Metabolic syndrome incidence	RR=1.18	0.508	RR=1.28	0.103	RR=0.98	0.891	RR=0.89	0.246	No effect
Rademaker (2011)	410	PTSD symptoms	B=-0.05	0.290	B=-0.16	<.05	B=0.01	NS	B=0.02	NS	Only NA effect
Williams (2011)	192	Medication adherence	t=-6.04	<.01	B=-0.36	<.01	B=-0.17	NS	B=-0.21	<.05	Synergistic Type D effect
Howard (2012)	134	Anxiety	t=3.73	<.001	t=9.85	<.001	t=2.56	0.012	t=-0.78	0.750	Additive Type D effect
Howard (2012)	134	Depression	t=4.95	<.001	t=8.49	<.001	t=1.01	0.316	t=1.71	0.089	Only NA effect
Howard (2012)	134	Perceived stress	t=4.81	<.001	t=10.74	<.001	t=0.80	0.427	t=1.68	0.095	Only NA effect
Howard (2012)	134	Resting cardiac output	t=2.28	0.024	t=1.37	0.175	t=0.20	0.845	t=2.31	0.023	Synergistic Type D Effect

Study*	N	Dependent measure	dichotomous method		Continuous method						Conclusion
			Type D	p	NA	p	SI	p	NA*SI	p	
Howard (2012)	134	Total peripheral resistance	t=2.3	0.027	t=0.82	0.412	t=3.01	0.003	t=0.96	0.342	Only SI effect
Molloy (2012)	165	Medication adherence	B=-0.73	<.01	B=-0.04	0.050	B=-0.02	0.520	B=0	0.700	Only NA effect
Williams (2012)	131	Disability	t=6.94	<.01	B=0.43	<.01	B=0.1	NS	B=0.02	NS	Only NA effect
Williams (2012)	131	Quality of life	t=-7.25	<.01	B=-0.62	<.001	B=0.06	NS	B=0.00	NS	Only NA effect
Compare (2013)	111	Stress cardiomyopathy with vs. without emotional triggers	OR=4.08	0.006	OR=1.05	0.210	OR=1.53	0.040	OR=1.11	0.040	Synergistic Type D effect
Compare (2013)	111	Stress cardiomyopathy with emotional triggers vs. acute MI	OR=6.48	<.001	OR=1	0.380	OR=1.92	0.010	OR=1.86	0.020	Synergistic Type D effect
Compare (2013)	111	Stress cardiomyopathy without emotional triggers vs. acute MI	OR=1.15	0.767	OR=1.22	0.210	OR=0.99	0.320	OR=1.18	0.330	No effect
Conden (2013)	5012	Musculoskeletal symptoms in boys	OR=2.39	<.001	OR=2.08	<.001	OR=0.89	0.231	OR=0.95	0.414	Only NA effect
Conden (2013)	5012	Musculoskeletal symptoms in girls	OR=2.38	<.001	OR=2.19	<.001	OR=0.80	0.001	OR=1.08	0.093	Additive Type D effect
Conden (2013)	5012	Psychosomatic symptoms in boys	OR=5.39	<.001	OR=3.44	<.001	OR=0.98	0.887	OR=0.91	0.157	Only NA effect
Conden (2013)	5012	Psychosomatic symptoms in boys	OR=5.54	<.001	OR=4.40	<.001	OR=0.89	0.068	OR=1.05	0.487	Only NA effect
Denollet (2013b)	541	MACE	OR=1.74	0.016	OR=0.99	0.940	OR=1.16	0.200	OR=1.36	0.003	Synergistic Type D effect
Denollet (2013b)	541	MI	OR=2.35	0.007	OR=1.11	0.560	OR=1.02	0.910	OR=1.48	0.007	Synergistic Type D effect
Husson (2013)	2578	Disease information	B=-0.09	<.01	NR	NS	B=-0.14	<.01	NR	NS	Only SI effect
Husson (2013)	2609	Medical test information	B=-0.12	<.01	B=-0.06	<.01	B=-0.15	<.01	NR	NS	Additive Type D effect
Husson (2013)	470	Treatment information	B=-0.08	<.01	NR	NS	B=-0.12	<.01	NR	NS	Only SI effect
Husson (2013)	2539	Other services information	B=-0.02	NS	NR	NS	B=-0.06	<.01	NR	NS	Only SI effect
Husson (2013)	2682	Satisfaction with information	OR=0.54	<.0001	OR=0.78	<.01	OR=0.85	<.01	NR	NS	Additive Type D effect
Husson (2013)	2533	Helpfulness of information	OR=0.58	<.0001	OR=0.83	<.01	OR=0.82	<.01	NR	NS	Additive Type D effect
Kupper (2013b)	99	Heart rate	F=4.31	0.040	F=5.53	0.020	F=4.52	0.040	F=4.16	0.040	Synergistic Type D effect
Kupper (2013b)	99	Diastolic blood pressure	F=1.77	0.190	F=0	0.960	F=1.34	0.250	F=1.95	0.170	No effect
Kupper (2013b)	99	Systolic blood pressure	F=2.04	0.160	F=1.75	0.190	F=2.01	0.160	F=4.11	0.045	Synergistic Type D effect
Kupper (2013c)	101	Heart period	F=0.45	0.600	NR	NR	NR	NR	F=0.08	0.890	No effect
Kupper (2013c)	101	Respiratory sinus arrhythmia	F=0.45	0.580	NR	NR	NR	NR	F=3.1	0.060	No effect

Study*	N	Dependent measure	dichotomous method		Continuous method						Conclusion
			Type D	p	NA	p	SI	p	NA*SI	p	
Kupper (2013c)	101	Pre-ejection period	F=0.54	0.560	NR	NR	NR	NR	F=3.1	0.051	No effect
Kupper (2013c)	101	Systolic blood pressure	F=2.68	0.070	NR	NR	NR	NR	F=0.26	0.770	No effect
Kupper (2013c)	101	Diastolic blood pressure	F=4.64	0.015	NR	NR	NR	NR	F=0.25	0.740	Only dichotomous effect
Larson (2013)	304	CHD risk	t=-0.42	0.680	B=-0.12	0.254	B=0.09	0.384	B=0.01	0.750	No effect
Smeijers (2013)	1279	Non cardiac chest pain	OR=1.71	0.140	OR=2.31	<.001	OR=0.67	0.07	OR=1.12	0.34	Only NA effect
VandeVen (2013)	188	Medication adherence	B=-0.17	<.01	B=-0.19	0.019	B=-0.08	0.322	B=-0.03	0.630	Only NA effect
Compare (2014)	75	Coronary artery plaque presence	OR=1.19	0.049	OR=1.13	0.57	OR=2.14	0.02	OR=1.15	0.32	Only SI effect
Hoogwegt (2014)	64	Autonomic control SDNN	B=-0.25	0.043	NR	NS	NR	NS	B=0.63	0.100	Only dichotomous effect
Hoogwegt (2014)	64	Autonomic control SDANN	B=-0.33	0.010	NR	NS	NR	NS	B=0.61	0.110	Only dichotomous effect
Hoogwegt (2014)	64	Autonomic control HRV	B=-0.21	0.090	NR	NS	NR	NS	B=0.66	0.080	No effect
Kelly-Hughes (2014)	67	Background stress	t=-2.32	<.05	B=0.28	<.05	B=0.34	<.01	B=-0.69	NS	Additive Type D effect
Kelly-Hughes (2014)	67	Pretest Alerness	NR	NS	B=-0.38	<.01	B=-0.08	NS	B=0.45	NS	Only NA effect
Kelly-Hughes (2014)	67	Posttest Alerness	NR	NS	B=0.05	NS	B=-0.24	NS	B=0.22	NS	No effect
Kelly-Hughes (2014)	67	Pretest Connectedness	t=2.71	<.01	B=-0.47	<.001	B=-0.12	NS	B=-0.14	NS	Only NA effect
Kelly-Hughes (2014)	67	Posttest Connectedness	NR	NS	B=-0.1	NS	B=-0.23	NS	B=0.38	NS	No effect
Kelly-Hughes (2014)	67	Pretest Calmness	NR	NS	B=-0.02	NS	B=0.06	NS	B=0.39	NS	No effect
Kelly-Hughes (2014)	67	Posttest Calmness	NR	NS	B=-0.01	NS	B=-0.04	NS	B=-0.74	NS	No effect
Kelly-Hughes (2014)	67	Pretest Anxiousness	NR	NS	B=0.11	NS	B=-0.04	NS	B=-0.18	NS	No effect
Kelly-Hughes (2014)	67	Posttest Anxiousness	NR	NS	B=-0.17	NS	B=0.37	<.01	B=0.19	NS	Only SI effect
Kelly-Hughes (2014)	67	Pretest Relaxedness	NR	NS	B=-0.05	NS	B=0.37	<.01	B=-0.93	<.05	Synergistic Type D effect
Kelly-Hughes (2014)	67	Posttest Relaxedness	t=2.58	<.05	B=-0.22	NS	B=-0.08	NS	B=-0.86	<.05	Synergistic Type D effect
Kelly-Hughes (2014)	67	Pretest Stressfulness	NR	NS	B=0.2	NS	B=-0.22	NS	B=0.13	NS	No effect
Kelly-Hughes (2014)	67	Posttest Stressfulness	NR	NS	B=-0.16	NS	B=0.36	<.01	B=0.32	NS	Only SI effect
Kelly-Hughes (2014)	67	Pretest Happiness	NR	NS	B=-0.21	NS	B=0.08	NS	B=-0.61	NS	No effect

Study*	N	Dependent measure	dichotomous method		Continuous method						Conclusion
			Type D	p	NA	p	SI	p	NA*SI	p	
Kelly-Hughes (2014)	67	Posttest Happiness	NR	NS	B=0.03	NS	B=0.06	NS	B=-0.86	NS	No effect
Kelly-Hughes (2014)	67	Pretest State anxiety	NR	NS	B=0.4	<.01	B=-0.01	NS	B=-0.34	NS	Only NA effect
Kelly-Hughes (2014)	67	Posttest State anxiety	t=-2.03	<.05	B=0.17	NS	B=0.22	NS	B=-0.1	NS	Only dichotomous effect
Kelly-Hughes (2014)	67	Pretest Arousal	NR	NS	B=-0.17	NS	B=-0.04	NS	B=0.1	NS	No effect
Kelly-Hughes (2014)	67	Posttest Arousal	NR	NS	B=0.06	NS	B=-0.17	NS	B=-0.32	NS	No effect
Kelly-Hughes (2014)	67	Pretest Stressfulness	NR	NS	B=0.04	NS	B=-0.01	NS	B=0.16	NS	No effect
Kelly-Hughes (2014)	67	Posttest Stressfulness	NR	NS	B=0.04	NS	B=0.14	NS	B=0.89	<.05	Synergistic Type D effect
Kelly-Hughes (2014)	67	Perceived workload	NR	NS	NR	NS	NR	NS	NR	NS	No effect
Kelly-Hughes (2014)	67	Heart rate baseline	NR	NS	NR	NS	NR	NS	NR	NS	No effect
Kelly-Hughes (2014)	67	Heart rate demo	NR	NS	NR	NS	NR	NS	NR	NS	No effect
Kelly-Hughes (2014)	67	Heart rate task	NR	NS	NR	NS	NR	NS	NR	NS	No effect
Kelly-Hughes (2014)	67	Heart rate recovery	NR	NS	NR	NS	NR	NS	NR	NS	No effect
Kelly-Hughes (2014)	67	Pretest Systolic blood pressure	NR	NS	B=-0.03	NS	B=-0.05	NS	B=-1.02	<.05	Synergistic Type D effect
Kelly-Hughes (2014)	67	Pretest Diastolic blood pressure	NR	NS	B=-0.04	NS	B=0.03	NS	B=-0.89	NS	No effect
Kelly-Hughes (2014)	67	Demo Systolic blood pressure	NR	NS	B=-0.03	NS	B=0.08	NS	B=-1.08	<.05	Synergistic Type D effect
Kelly-Hughes (2014)	67	Demo Diastolic blood pressure	NR	NS	B=0.11	NS	B=0	NS	B=-1.57	<.01	Synergistic Type D effect
Kelly-Hughes (2014)	67	Task Systolic blood pressure	NR	NS	B=-0.03	NS	B=-0.04	NS	B=-1.02	NS	No effect
Kelly-Hughes (2014)	67	Task Diastolic blood pressure	NR	NS	B=0.08	NS	B=-0.13	NS	B=-1.3	<.01	Synergistic Type D effect
Kelly-Hughes (2014)	67	Posttest Systolic blood pressure	NR	NS	B=-0.14	NS	B=-0.02	NS	B=-0.89	NS	No effect
Kelly-Hughes (2014)	67	Posttest Diastolic blood pressure	NR	NS	B=-0.01	NS	B=-0.1	NS	B=-1.33	<.01	Synergistic Type D effect
Marchesi (2014)	250	Depressive disorder	OR=1.91	0.110	OR=1.11	0.290	OR=1.1	0.300	OR=0.99	0.680	No effect
Meyer (2014)	465	All-cause mortality	OR=0.88	0.861	HR=0.99	0.874	HR=0.99	0.812	HR=1	0.743	No effect
Meyer (2014)	465	MACE	OR=0.91	0.841	HR=1.07	0.074	HR=0.92	0.027	HR=1	0.287	Only SI effect
Stevenson (2014)	177	Quality of life	t=5.91	<.001	B=-0.45	<.01	B=-0.23	<.01	B=-0.12	NS	Additive Type D effect
Stevenson (2014)	177	Total symptoms	t=4.09	<.001	B=0.26	<.01	B=0.099	NS	B=-0.18	NS	Only NA effect

Study*	N	Dependent measure	dichotomous method		Continuous method						Conclusion
			Type D	p	NA	p	SI	p	NA*SI	p	
Wu (2014)	84	Medication adherence MMAS-4	B=-0.239	0.042	B=0.29	0.026	B=0.008	0.949	B=0.03	0.911	Only NA effect
Wu (2014)	84	Medication adherence MEMS	B=0.247	0.043	B=-0.25	0.038	B=0.00	0.980	B=0.04	0.900	Only NA effect
Dulfer (2015)	1190	All-cause mortality	HR=1.2	0.211	HR=0.78	0.027	HR=0.9	0.592	HR=1.18	0.229	Only NA effect
Ginting (2016)	386	Addictive behaviors	B=0.17	0.005	NR	NS	NR	NS	NR	NS	Only dichotomous effect
Ginting (2016)	386	Healthy food consumption	B=-0.19	0.001	B=-0.13	0.030	B=-0.12	0.028	NR	NS	Additive Type D effect
Ginting (2016)	386	Unhealthy food consumption	B=0.18	0.001	NR	NS	B=0.12	0.028	NR	NS	Only SI effect
Ginting (2016)	386	Exercise	B=-0.16	0.007	NR	NS	NR	NS	NR	NS	Only dichotomous effect
Ginting (2016)	386	Weight control	B=-0.19	0.001	NR	NS	B=-0.11	0.050	B=-0.09	0.075	Only SI effect
Ginting (2016)	386	Medication adherence	B=-0.04	NS	NR	NS	NR	NS	NR	NS	No effect
Ginting (2016)	386	Social support family	B=-0.28	<.001	B=-0.23	0.007	B=-0.17	<.001	NR	NS	Additive Type D effect
Ginting (2016)	386	Social support friends	B=-0.21	<.001	NR	NS	B=-0.26	<.001	NR	NS	Only SI effect
Ginting (2016)	386	Social support others	B=-0.24	<.001	B=-0.17	0.043	B=-0.13	0.004	NR	NS	Additive Type D effect
Horwood (2016)	389	Health behavior	R^2=0.02	0.004	B=-0.32	<.0001	B=-0.03	0.630	B=0.1	0.012	Synergistic Type D effect (wrong direction)
Horwood (2016)	389	Social support	R^2=0.20	<.001	B=-0.34	<.0001	B=-0.28	<.0001	B=-0.04	0.322	Additive Type D effect
Horwood (2016)	389	Physical symptoms	R^2=0.06	<.001	B=0.33	<.0001	B=-0.03	0.162	B=-0.05	0.183	Only NA effect
Horwood (2016)	389	Psychological symptoms	R^2=0.22	<.001	B=0.69	<.0001	B=0.02	0.814	B=-0.03	0.515	Only NA effect
Van Middendorp (2016)	315	Mental health composite	t=11.56	<.0001	B=-0.55	<.001	NR	0.760	NR	0.310	Only NA effect
Van Middendorp (2016)	315	Emotional well-being	t=15.43	<.0001	B=-0.66	<.001	NR	>.27	NR	>.14	Only NA effect
Van Middendorp (2016)	315	Emotional role limitations	t=7.87	<.0001	NR	<.001	NR	>.27	NR	>.14	Only NA effect
Van Middendorp (2016)	315	Social functioning	t=3.10	0.002	B=-0.17	<.001	NR	>.27	NR	>.14	Only NA effect
Van Middendorp (2016)	315	Energy	t=4.69	<.0001	NR	<.001	NR	>.27	NR	>.14	Only NA effect
Van Middendorp (2016)	315	Physical health composite	t=3.33	<.001	B=-0.21	<.001	NR	0.720	NR	0.920	Only NA effect
Van Middendorp (2016)	315	Physical functioning	t=0.02	0.866	NR	0.880	NR	>.23	NR	>.07	No effect
Van Middendorp (2016)	315	Physical role limitations	t=1.88	0.061	NR	<.004	B=0.13	0.020	NR	>.07	Additive Type D effect

Study*	N	Dependent measure	dichotomous method		Continuous method						Conclusion
			Type D	p	NA	p	SI	p	NA*SI	p	
Van Middendorp (2016)	315	Pain	t=1.89	0.059	B=-0.15	<.004	NR	>.23	NR	>.07	Only NA effect
Van Middendorp (2016)	315	General health perceptions	t=6.153	<.0001	B=-0.25	<.004	NR	>.23	NR	>.07	Only NA effect
Wang (2016)	109	Lipid plaque	OR=4.87	0.025	OR=3.43	0.018	OR=1.8	0.250	OR=0.66	0.540	Only NA effect
Wang (2016)	109	Thin cap fibroatheroma	OR=3.84	0.011	OR=2.2	0.026	OR=0.9	0.710	OR=1.17	0.550	Only NA effect
Wang (2016)	109	Fibrous Cap Thickness	B=-1.43	0.001	B=-0.05	0.030	B=0.02	0.480	B=0.01	0.690	Only NA effect
Conden (2017)	946	Recurrent MI	HR=1.16	0.501	HR=1	1.000	HR=1.01	0.522	HR=1	1.000	No effect
Conden (2017)	946	All-cause mortality	HR=0.94	0.911	HR=1	1.000	HR=1.02	<.05	HR=1	1.000	Only SI effect
Jeon (2017)	236	Atrial fibrillation recurrence	NR	NS	NR	NS	NR	NS	NR	NS	No effect
Li (2017)	330	HbA1c level	B=0.36	<.001	B=-0.01	NS	B=0.034	<.001	B=0.00	<0.05	Synergistic Type D effect
Bekendam (2018)	547	Need for repeat cardiac testing	HR=1.1	0.672	HR=1.05	0.636	HR=1.09	0.413	HR=0.93	0.515	No effect
Bekendam (2018)	547	Emergency department admission	HR=1.15	0.565	HR=1.06	0.601	HR=1.05	0.655	HR=0.89	0.296	No effect
Borkoles (2018)	482	Problem Focused Coping	F=7.38	0.020	B=-0.14	<.01	B=-0.13	<.01	NR	NS	Additive Type D effect
Borkoles (2018)	482	Emotion Focused Coping	F=6.97	0.009	B=0.29	<.01	B=0.12	<.05	B=0.39	<.01	Synergistic Type D effect
Borkoles (2018)	482	Avoidance Coping	F=13.46	<.001	B=0.30	<.01	NR	NS	B=-0.33	<.05	Synergistic Type D effect
Borkoles (2018)	482	Problem Focused Coping Effectiveness	F=16.4	<.001	B=-0.14	<.01	B=-0.13	<.05	B=0	NS	Additive Type D effect
Borkoles (2018)	482	Emotion Focused Coping Effectiveness	F=0	NS	B=0	NS	B=0	NS	B=0	NS	No effect
Borkoles (2018)	482	Avoidance Coping Effectiveness	F=4	0.046	B=0	NS	B=0	NS	B=0	NS	No effect
Lin (2018)	152	Lipid plaque	OR=4.56	0.010	OR=3.92	0.002	OR=1.81	0.360	OR=0.49	0.120	Only NA effect
Lin (2018)	152	Thin cap fibroatheroma	OR=3.15	0.003	OR=1.77	0.030	OR=1.19	0.450	OR=1.38	0.130	Only NA effect
Lin (2018)	152	Rupture	OR=2.52	0.020	OR=1.66	0.060	OR=1.73	0.049	OR=0.95	0.830	Only SI effect
Lin (2018)	152	Fibrous Cap Thickness	B=-0.03	<.001	B=-0.01	0.020	B=-0.01	0.130	B=0.00	0.300	Only NA effect
Schoormans (2018)	1883	Leukocyte telomere length	B=-47.37	0.010	B=-14.3	0.210	B=5.95	0.630	B=-20.2	0.040	Synergistic Type D effect
Talaei-Khoei (2018)	102	Pain interference	B=5.9	<.001	B=0.34	0.002	B=0.26	0.028	B=0.01	0.582	Additive Type D effect
Wang (2018)	173	In-stent restenosis	OR=4.92	0.017	OR=6.93	0.016	OR=0.16	0.140	OR=1.13	0.800	Only NA effect
Allen (2019)	160	Total symptoms	t=-5.41	<.001	B=1.15	<.001	B=-0.07	0.550	B=0.05	0.004	Synergistic Type D effect

Study*	N	Dependent measure	dichotomous method		Continuous method						Conclusion
			Type D	p	NA	p	SI	p	NA*SI	p	
Allen (2019)	160	Cardiac/sympathetic symptoms	t=-3.13	0.002	B=0.17	<.001	B=-0.02	0.590	B=0.01	0.016	Synergistic Type D effect
Allen (2019)	160	Muscular symptoms	t=-4.22	<.001	B=0.17	<.001	B=0.02	0.580	B=0.01	0.026	Synergistic Type D effect
Allen (2019)	160	Metabolic symptoms	t=-5.76	<.001	B=0.35	<.001	B=-0.04	0.220	B=0.01	0.041	Synergistic Type D effect
Allen (2019)	160	Gastrointestinal symptoms	t=-2.78	0.006	B=0.13	<.001	B=-0.03	0.220	B=0.01	0.057	Only NA effect
Allen (2019)	160	Vasovagal symptoms	t=-3.35	0.001	B=0.14	<.001	B=0.00	0.930	B=0.01	0.004	Synergistic Type D effect
Allen (2019)	160	Cold	t=-1.98	0.049	B=0.08	<.001	B=-0.01	0.740	B=0.00	0.410	Only NA effect
Allen (2019)	160	Headache	t=-4.77	<.001	B=0.10	<.001	B=0.01	0.430	B=0.01	0.004	Synergistic Type D effect
Bouwens (2019)	294	Health related quality of life	B=-0.04	0.001	B=-0.03	0.004	B=-0.01	0.43	B=0.00	0.91	Only NA effect
de Vroege (2019)	212	Physical symptoms remission	OR=0.58	NS	OR=0.90	NS	OR=1.01	NS	OR=0.98	NS	No effect
de Vroege (2019)	212	Anxiety remission	OR=0.33	<.05	OR=0.92	NS	OR=1.00	NS	OR=0.99	NS	Only dichotomous effect
de Vroege (2019)	212	Depression remission	OR=0.36	NS	OR=0.99	NS	OR=0.95	NS	OR=1.00	NS	No effect
de Vroege (2019)	212	Physcial symptoms response	OR=0.44	NS	OR=0.94	NS	OR=0.98	NS	OR=1.00	NS	No effect
de Vroege (2019)	212	Anxiety response	OR=0.54	NS	OR=0.91	NS	OR=1.00	NS	OR=0.99	NS	No effect
de Vroege (2019)	212	Depression response	OR=0.58	NS	OR=0.98	NS	OR=0.96	NS	OR=1.00	NS	No effect
Lee (2019)	191	In-stent neoatherosclerosis	OR=2.99	0.007	OR=1.85	0.002	OR=0.98	0.940	OR=0.71	0.110	Only NA effect
Lee (2019)	191	Thin cap fibroatheroma	OR=2.81	0.010	OR=1.77	0.006	OR=0.94	0.750	OR=0.72	0.120	Only NA effect
Matsuishi (2019)	142	Delirium/coma days	OR=2.8	0.009	OR=1.09	0.002	OR=1.05	0.044	OR=0.9	NS	Additive Type D effect

N = Sample size; NA = Negative affectivity; NI = Not investigated; NR = Not reported; NS = Not significant; SI = Social inhibition

Appendix F: Sensitivity analysis (*chapter 4*)

Table F1: The association between the statistical significance of the dichotomous method effects and various continuous methods effects. Cohen's kappa is reported for all outcomes combined and separately for the outcome types (1) mortality, (2) cardiometabolic, and (3) psychosocial.

Continuous method effect	Overall N=158	Mortality* N=5	Cardiometabolic N=59	Psychosocial N=94
NA effect	0.55	-	0.43	0.58
SI effect	0.19	-	0.24	0.14
Additive Type D effect	0.14	-	0.08	0.13
Synergistic Type D effect	0.08	-	0.11	0.07
Any continuous effect	0.59	-	0.59	0.59

** Cohen's kappa could not be estimated due to sparse cell counts in the cross table*

Appendix G: Estimating the Type D effect with the 2-group & 4-group method (*chapter 5*)

In line with earlier simulation studies (***Chapter 2 and 3***) we expect that the oversensitive 2-group and 4-group methods will suggest a Type D effect in situations where the continuous method points to an effect of NA or SI only. The 2-group method first computes total scores for Type D's subcomponents NA and SI by summing the scores of the items assessing each construct. Both personality traits are typically measured using the DS14 questionnaire, that measures each trait with seven ordinal items on a 0-4 Likert scale (Denollet, 2005). Subsequently the 2-group method dichotomizes the NA and SI sum scores into a high vs. low score using a predetermined cut-off score of 10. These dichotomized NA and SI variables are recoded into a single dichotomous variable by assigning a value of 1 to people who score high on both constructs and otherwise assigning a value of 0. The 4-group method uses the dichotomized NA and SI scores to classify people in four rather than two different groups: (1) High NA and SI scores (Type D personality); (2) Only high NA scores (NA+SI-); (3) Only high SI scores (NA-SI+); (4) Low NA and SI scores (NA-SI-). These four groups are recoded into three dummy variables indicating whether someone is classified in group 1, 2, or 3. The fourth group then serves as a reference category.

Table G1 shows for each endpoint, the estimated odds ratios [95% Bayesian credible interval] of demographic predictors, the Type D effects according to 2-group and 4-group methods. The 95%CI of bold cells does not include an odds ratio of one. The estimates according to the 2-group and 4-group method were partly in line with those of the continuous method. Whenever the continuous method indicated a Type D effect (i.e., on MACE and any adverse event), the 2-group and 4-group method would also show an effect. However, the 2-group method also suggested a Type D effect on all-cause mortality, cardiac mortality and PCI, and the 4-group method indicated that, compared to the reference group (NA-SI-), the Type D group had a higher odds on every endpoints except CABG. For the outcomes all-cause mortality, cardiac mortality and PCI, the continuous method did not identify Type D personality, but only NA as a risk factor. These findings are in line with earlier research indicating that the 2-group and 4-group methods are not only sensitive to detect Type D effects, but also to main or quadratic effects of NA or SI only (***Chapters 2 & 3***).

Table G1: For each endpoint, the estimated odds ratios [95% Bayesian credible interval] of demographic predictors, the Type D effects according to 2-group and 4-group methods. The 95%CI of bold cells does not include an odds ratio of one

Outcome	All-cause mortality	Cardiac mortality	Myocardial infarction	CABG	PCI	MACE	Any adverse event
Sample size	N = 10255	N = 5774	N = 5877	N = 2440	N = 2448	N = 3923	N = 3935
Predictor	OR [95%CI}	OR [95%CI}	OR [95%CI}	OR [95%CI}	OR [95%CI}	OR [95%CI}	OR [95%CI}
2-group method [m1]							
Type D	**1.437 [1.016, 2.098]**	**2.001 [1.23, 3.44]**	1.369 [0.97, 1.985]	1.705 [0.789, 3.178]	**1.614 [1.023, 2.389]**	**1.858 [1.445, 2.404]**	**1.876 [1.445, 2.524]**
4-group method [m2]							
Type D vs. NA- SI-	**1.309 [1.049, 1.684]**	**1.912 [1.271, 3.04]**	**1.433 [1.029, 2.049]**	1.536 [0.751, 3.165]	**1.587 [1.056, 2.457]**	**1.923 [1.453, 2.639]**	**1.883 [1.451, 2.488]**
Type D vs. NA- SI+	**1.497 [1.147, 2.058]**	**2.326 [1.323, 4.673]**	1.362 [0.891, 2.132]	3.067 [0.892, 14.286]	**2.07 [1.222, 3.69]**	**2.222 [1.548, 3.3]**	**2.203 [1.608, 3.049]**
Type D vs. NA+ SI-	1.175 [0.9, 1.592]	2.02 [0.989, 5.848]	1.269 [0.801, 2.347]	2.146 [0.682, 9.174]	1.529 [0.921, 2.688]	**1.6 [1.172, 2.208]**	**1.577 [1.117, 2.222]**

CABG = coronary artery bypass grafting; MACE = major adverse cardiac event; OR = Odds ratio; *PCI = percutaneous coronary intervention*
m1: Model = Age + Men + Type D (2-group)
m2: Model = Age + Men + Type D (4-group)

Appendix H: Quality assessment for studies included in the individual patient data meta-analysis (*Chapter 5*)

Table H1 (part 1): For each included study, a risk of bias assessment due to study participation, attrition, prognostic factor measurement and outcome measurement. Bias assessment inspired on the QUIPS tool [30]

Quality assessment criteria	Denollet et al. (1996)	Denollet et al. (2000)	Denollet et al. (2006)	Martens et al. (2010)	Pelle et al. (2010)	Schmidt et al. (2011)	Coyne et al. (2011)	Grande et al. (2011)	Denollet et al. (2013a)	Denollet et al. (2013b)	Meyer et al. (2014)	Sumin et al. (2015)	Dulfer et al. (2015)	Herrmann-Lingen et al. (2016)	Gostoli et al. (2016)	Pushkarev et al. (2017)	Conden et al. (2017)	Lin et al. (2019)	Lv et al. (2020)
Adequate participation in the study by eligible persons	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Description of the source population or population of interest	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Description of the baseline study sample	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+
Adequate description of the sampling frame and response rate	-	-	-	-	+	-	-	-	-	-	-	-	+	+	-	-	+	-	-
Adequate description of the period and place of recruitment	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Adequate description of inclusion and exclusion criteria	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	-	+	+	+
Risk of bias due to study participation	**L**	**L**	**L**	**L**	**L**	**L**	**L**	**M**	**L**	**M**	**L**	**L**	**L**	**L**	**L**	**M**	**L**	**L**	**L**
Less than 10% dropout across time for study participants	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Description of attempts to collect information on who dropped out	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	-	+	-	+
Reasons for loss to follow-up are provided	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+
Adequate description of participants lost to follow-up	+	+	+	+	+	-	+	+	+	+	+	-	-	+	+	-	+	-	+
No important differences between completers and dropouts	+	+	+	+	+	?	+	-	+	+	+	?	?	+	+	?	+	?	+
Risk of bias due to attrition	**L**	**L**	**L**	**L**	**L**	**M**	**L**	**M**	**L**	**L**	**L**	**M**	**M**	**L**	**L**	**M**	**L**	**M**	**L**

Table H1 (part 2): For each included study, a risk of bias assessment due to study participation, attrition, prognostic factor measurement and outcome measurement. Bias assessment inspired on the QUIPS tool [30]

Quality assessment criteria	Denollet et al. (1996)	Denollet et al. (2000)	Denollet et al. (2006)	Martens et al. (2010)	Pelle et al. (2010)	Schmidt et al. (2011)	Coyne et al. (2011)	Grande et al. (2011)	Denollet et al. (2013a)	Denollet et al. (2013b)	Meyer et al. (2014)	Sumin et al. (2015)	Dulfer et al. (2015)	Herrmann-Lingen et al. (2016)	Gostoli et al. (2016)	Pushkarev et al. (2017)	Conden et al. (2017)	Lin et al. (2019)	Lv et al. (2020)
A clear definition or description of the PF is provided	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	+	+	+
Method of PF measurement is adequately valid and reliable	?	+	+	+	+	+	+	+	+	+	+	+	+	?	?	?	+	+	+
Same method and setting of measurement of PF for all participants	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
More than 95% of the study sample has complete data for the PF	+	+	+	+	+	+	-	+	-	+	+	-	+	+	+	+	-	-	+
Complete data or adequate imputation methods for missing PF data	+	+	+	-	-	+	-	+	-	+	-	-	+	+	+	+	-	-	+
Risk of bias due to prognostic factor measurement	**M**	**L**	**L**	**L**	**L**	**L**	**M**	**L**	**M**	**L**	**L**	**M**	**L**	**L**	**L**	**L**	**M**	**M**	**L**
A clear definition of the outcome is provided	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
outcome measurement used is adequately valid and reliable	+/-	+/-	+/-	+	+/-	+	+	+	+/-	+/-	+/-	+/-	+	+/-	+/-	+/-	+	+	+/-
Same method and setting outcome measurement for all participants	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Risk of bias due to outcome measurement	**M**	**M**	**M**	**L**	**M**	**L**	**L**	**L**	**M**	**M**	**M**	**M**	**L**	**M**	**M**	**M**	**L**	**L**	**M**

+ = criterium met
+/- = criterium partially met
- = criterium not met
? = insufficient information
L = Low risk of bias
M = Medium risk of bias
H = High risk of bias

Table H2: For each included study, a summary of the motivation for a possible increased risk on bias as assessed in Table S3.

Study	Evaluation
Denollet et al. (1996)	Unclear sampling frame: How many respondents were eligible? No relibaility analysis. Cardiac mortality can be difficult to determine. Deaths in first 5 years were excluded.
Denollet et al. (2000)	Unclear sampling frame: How many respondents were eligible? Cause specific mortality can be difficult to determine.
Denollet et al. (2006)	Unclear sampling frame: How many respondents were eligible? Cause specific mortality can be difficult to determine.
Martens et al. (2010)	Unclear sampling frame: How many respondents were eligible? Estimate based on small subset = 73% of approached participants agreed to participate.
Pelle et al. (2010)	Nonparticipants having higher rates of hypercholesterolemia, kidney disease, statin and nitrate prescription, and a lower prescription rate for angio- tensin-converting enzyme inhibitors. Cause specific mortality can be difficult to determine.
Schmidt et al. (2011)	Unclear sampling frame: How many respondents were eligible? Reason for lost to follow-up not described, nor patient characteristics, though only 8% of respondents dropped out.
Coyne et al. (2011)	Unclear sampling frame: How many respondents were eligible? Participants excluded because missing DS14 or CESD scores had higher antidepressant prescriptions.
Grande et al. (2011)	Target population is cardiac patients, but no list of eligible diagnoses reported. Unclear sampling frame: How many respondents were eligible? Participants with unknown vital status were younger, more likely to be female, live without a partner, and report current smoking than patients with known vital status.
Denollet et al. (2013a)	Unclear sampling frame: How many respondents were eligible? Cause specific mortality can be difficult to determine.
Denollet et al. (2013b)	Unclear sampling frame: How many respondents were eligible? No sample baseline characteristics. Cause specific mortality can be difficult to determine.
Meyer et al. (2014)	Unclear sampling frame: How many respondents were eligible? Cause specific mortality can be difficult to determine.

Study	Evaluation
Sumin et al. (2015)	Unclear sampling frame: How many respondents were eligible? Patients with incomplete data excluded and characteristics not described and compared with included participants. Cause specific mortality can be difficult to determine.
Dulfer et al. (2015)	Characteristics of excluded patients due to missing mortality data not described and compared with included participants.
Herrmann-Lingen et al. (2016)	Type D personality not defined and reliability of NA/SI measurements not reported, but understandable because not primary focus of trial. Cause specific mortality can be difficult to determine.
Gostoli et al. (2016)	Type D personality not defined and reliability of NA/SI measurements not reported. Cause specific mortality can be difficult to determine.
Pushkarev et al. (2017)	Unclear sampling frame and response rate. Inclusion and exclusion criteria not explicitly mentioned. Type D personality not defined and reliability of NA/SI measurements not reported. Cause specific mortality can be difficult to determine.
Conden et al. (2017)	6% of patients excluded due to missing DS14 scores. Compared to those excluded, the included patients were less often undergoing hypertension treatment, had a higher education level, were more physical active during leisure time, and had less often suffered from a prior MI.
Lin et al. (2019)	Unclear sampling frame: How many respondents were eligible? 8% of participants removed by listwise deletion for missing questionnaire scores. 9% removed due to dropout. No description of differences between these and included patients.
Lv et al. (2020)	Unclear sampling frame: How many respondents were eligible? Cause specific mortality can be difficult to determine.

Appendix I: Standard deviation estimates for all random predictors in multilevel model (chapter 5)

Table I1: For each endpoint, the standard deviation [95% Bayesian credible interval] of all random predictor effects according to the continuous method. The 95%CI of bold cells does not include zero.

Outcome	*All-cause mortality*	*Cardiac mortality*	*Myocardial infarction*	*CABG*	*PCI*	*MACE*	*Any adverse event*
Sample size	*N = 10255*	*N = 5774*	*N = 5877*	*N = 2440*	*N = 2448*	*N = 3923*	*N = 3935*
Random effect	*OR [95%CI]*	*OR [95%CI]*	*OR [95%CI]*	*OR [95%CI]*	*OR [95%CI]*	*OR [95%CI]*	*OR [95%CI]*
Intercept	**2.49 [1.77, 3.55]**	**1.21 [0.7, 2.04]**	**2.37 [1.51, 3.82]**	**0.86 [0.06, 2.36]**	**0.38 [0.02, 0.99]**	**0.89 [0.5, 1.5]**	**6.01 [3.89, 9.62]**
Age	**0.34 [0.18, 0.57]**	**0.2 [0.02, 0.53]**	**0.43 [0.11, 0.89]**	**0.48 [0.03, 1.28]**	**0.15 [0.01, 0.46]**	**0.34 [0.11, 0.75]**	**0.27 [0.02, 0.62]**
Men	**0.23 [0.02, 0.56]**	**0.24 [0.01, 0.73]**	**0.35 [0.01, 1.07]**	**0.84 [0.03, 2.58]**	**0.53 [0.05, 1.21]**	**0.31 [0.02, 0.82]**	**0.3 [0.02, 0.77]**
NA	0.09 [0, 0.24]	0.1 [0, 0.29]	0.1 [0, 0.28]	**0.3 [0.01, 0.98]**	**0.17 [0.01, 0.48]**	0.09 [0, 0.24]	0.09 [0, 0.25]
SI	0.1 [0, 0.27]	**0.21 [0.02, 0.48]**	0.08 [0, 0.26]	**0.23 [0.01, 0.77]**	**0.13 [0.01, 0.41]**	0.07 [0, 0.2]	0.07 [0, 0.21]
NA^2	0.04 [0, 0.12]	0.08 [0, 0.26]	0.06 [0, 0.18]	**0.23 [0.01, 0.76]**	0.1 [0, 0.34]	0.04 [0, 0.12]	0.03 [0, 0.1]
SI^2	0.07 [0, 0.19]	0.08 [0, 0.25]	**0.15 [0.01, 0.37]**	**0.24 [0.01, 0.73]**	0.09 [0, 0.31]	0.07 [0, 0.2]	0.06 [0, 0.16]
NA * SI	0.07 [0, 0.18]	0.1 [0, 0.31]	0.08 [0, 0.22]	**0.18 [0.01, 0.6]**	**0.16 [0.01, 0.54]**	**0.11 [0.01, 0.29]**	0.09 [0, 0.23]

CABG = coronary artery bypass grafting; MACE = major adverse cardiac event; OR = Odds ratio; PCI = percutaneous coronary intervention

Appendix J: Simple slope analysis for NA*SI interaction effect in individual patient data meta-analysis 1 (*Chapter 5*)

Table J1: Estimates and 95% confidence intervals for the simple slope effects of SI and NA on any adverse event in cardiovascular disease patients, conditional on scores on the other personality trait. Confidence intervals of bold faced cells do not contain a slope of zero (no effect).

SI simple slopes			NA simple slopes		
NA (Z)	NA (raw)	Estimate [95%CI]	SI (Z)	SI (raw)	Estimate [95%CI]
-1	2.7	-0.12 [-0.24, 0]	-1	3.1	0.06 [-0.07, 0.18]
-0.75	4.3	-0.10 [-0.21, 0.02]	-0.75	4.6	0.1 [-0.01, 0.21]
-0.5	5.9	-0.05 [-0.15, 0.05]	-0.5	6.2	**0.14 [0.04, 0.24]**
-0.25	7.4	-0.02 [-0.11, 0.08]	-0.25	7.7	**0.17 [0.08, 0.26]**
0	9	0.02 [-0.07, 0.1]	0	9.2	**0.2 [0.12, 0.29]**
0.25	10.6	0.05 [-0.04, 0.14]	0.25	10.7	**0.24 [0.15, 0.32]**
0.5	12.2	0.08 [-0.01, 0.17]	0.5	12.3	**0.27 [0.18, 0.36]**
0.75	13.8	**0.12 [0.02, 0.21]**	0.75	13.8	**0.3 [0.21, 0.4]**
1	15.3	**0.15 [0.15, 0.15]**	1	15.3	**0.34 [0.23, 0.44]**
1.25	16.9	**0.18 [0.18, 0.18]**	1.25	16.8	**0.37 [0.37, 0.37]**
1.5	18.5	**0.22 [0.09, 0.34]**	1.5	18.4	**0.41 [0.28, 0.53]**
1.75	20.1	**0.25 [0.11, 0.39]**	1.75	19.9	**0.44 [0.3, 0.58]**
2	21.7	**0.28 [0.28, 0.28]**	2	21.4	**0.47 [0.32, 0.63]**
2.25	23.3	**0.32 [0.15, 0.49]**	2.25	22.9	**0.51 [0.51, 0.51]**
2.5	24.8	**0.35 [0.17, 0.53]**	2.5	24.5	**0.53 [0.35, 0.72]**
2.75	26.4	**0.38 [0.18, 0.58]**	2.75	26	**0.57 [0.57, 0.57]**
3	28	**0.42 [0.42, 0.42]**	3	27.5	**0.61 [0.61, 0.61]**

Appendix K: Two-step individual patient meta-analysis (chapter 5)

Figure K1: Fixed effects meta-analysis of studies on the association between Type D personality (continuous method) and the occurrence of all-cause mortality

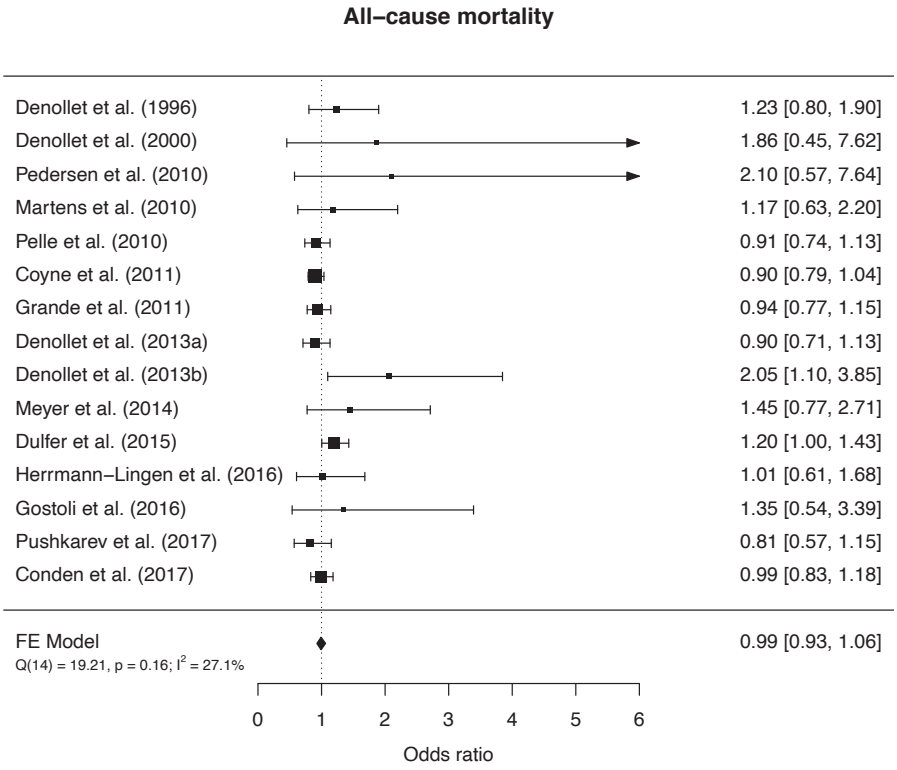


Figure K2: Fixed effects meta-analysis of studies on the association between Type D personality (continuous method) and the occurrence of cardiac mortality

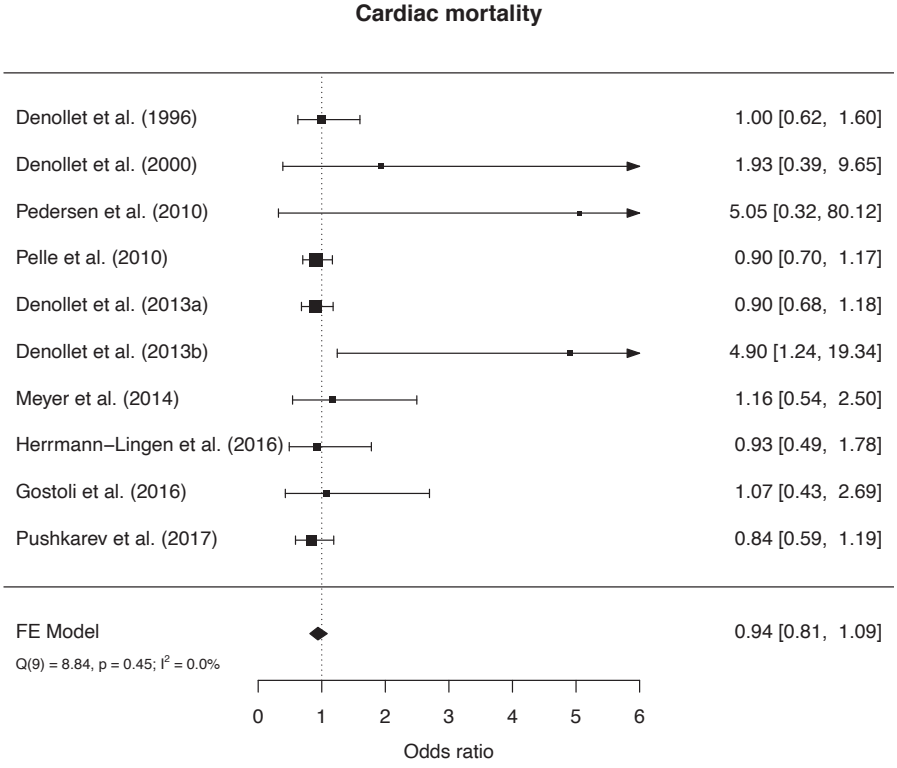


Figure K3: Fixed effects meta-analysis of studies on the association between Type D personality (continuous method) and the occurrence of myocardial infarction

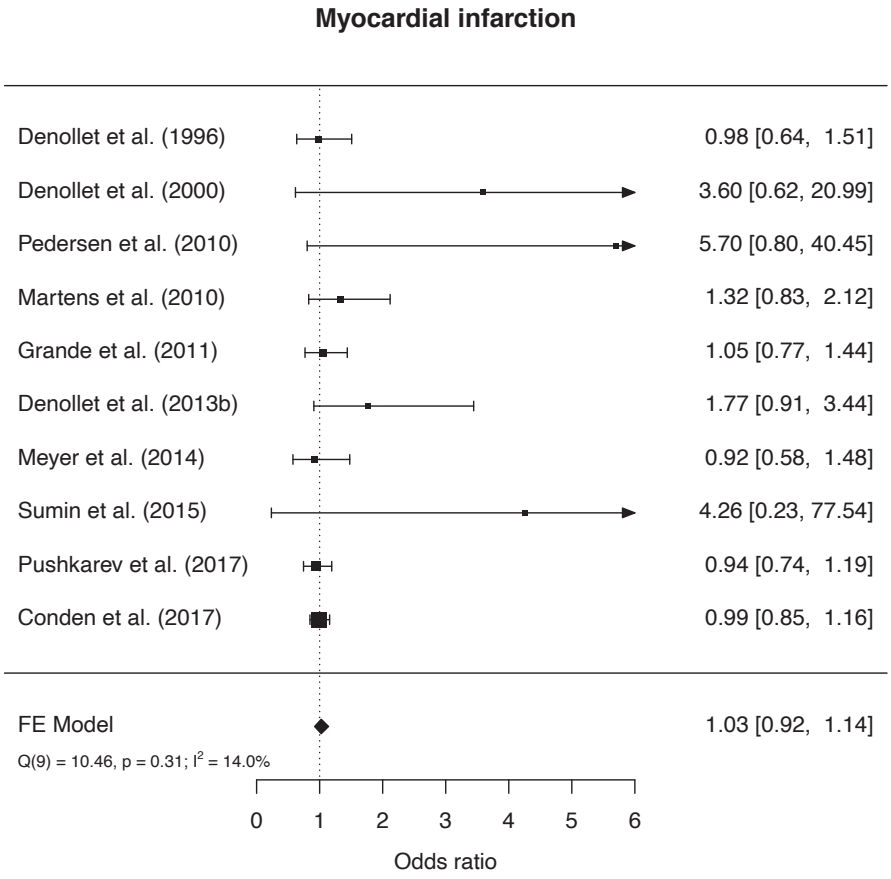


Figure K4: Fixed effects meta-analysis of studies on the association between Type D personality (continuous method) and the occurrence of coronary artery bypass grafting

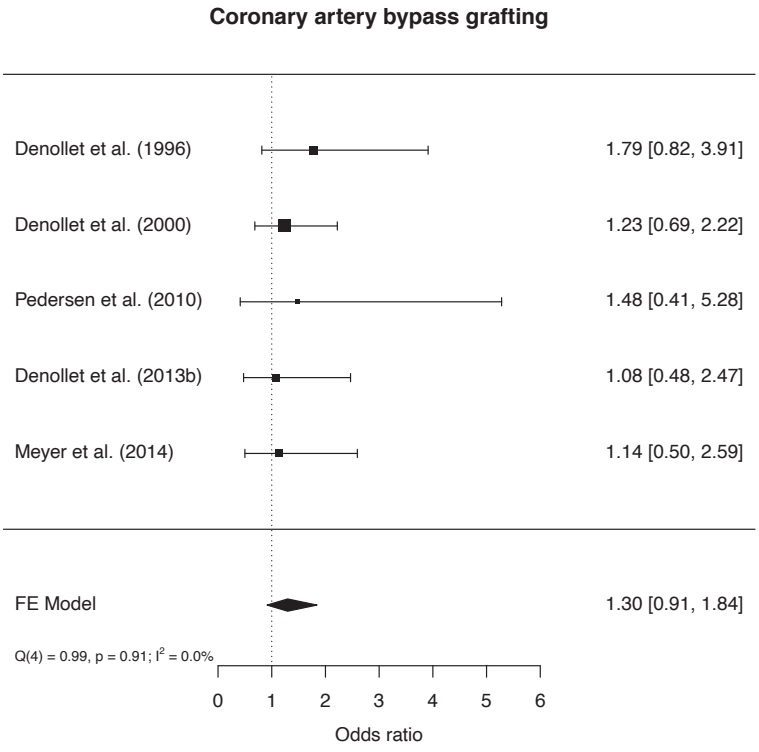


Figure K5: Fixed effects meta-analysis of studies on the association between Type D personality (continuous method) and the occurrence of percutaneous coronary intervention

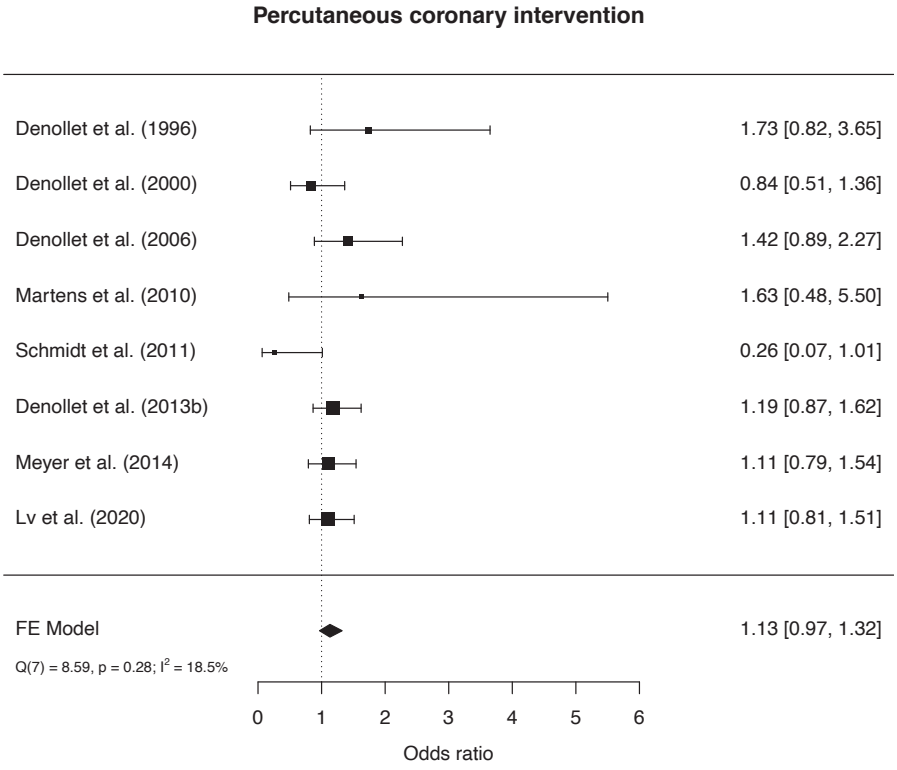


Figure K6: Fixed effects meta-analysis of studies on the association between Type D personality (continuous method) and the occurrence of major adverse cardiac events

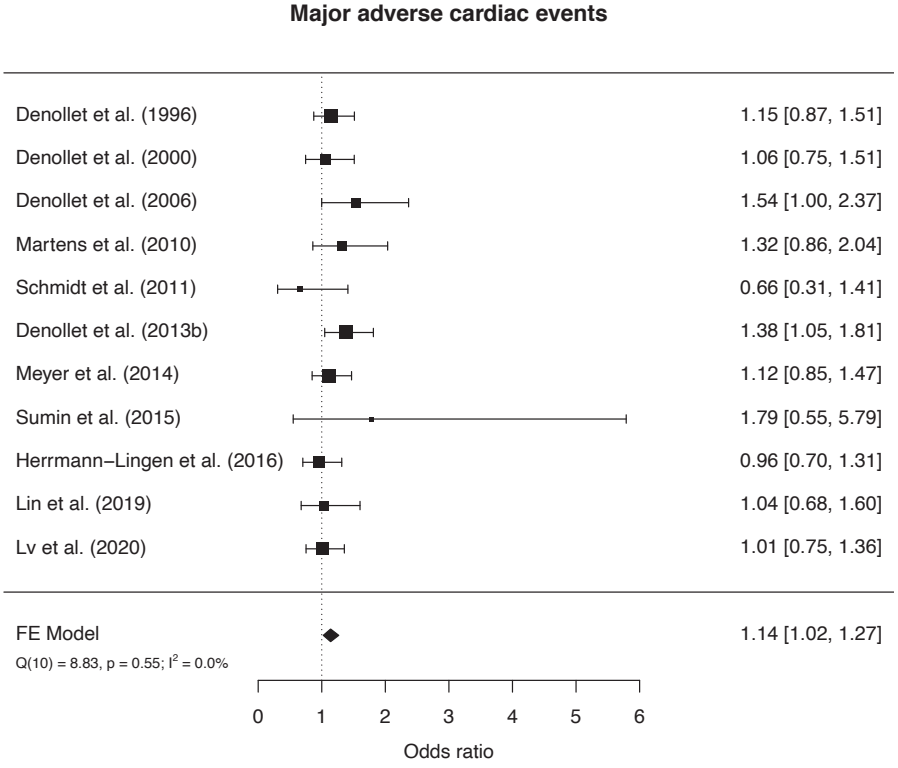
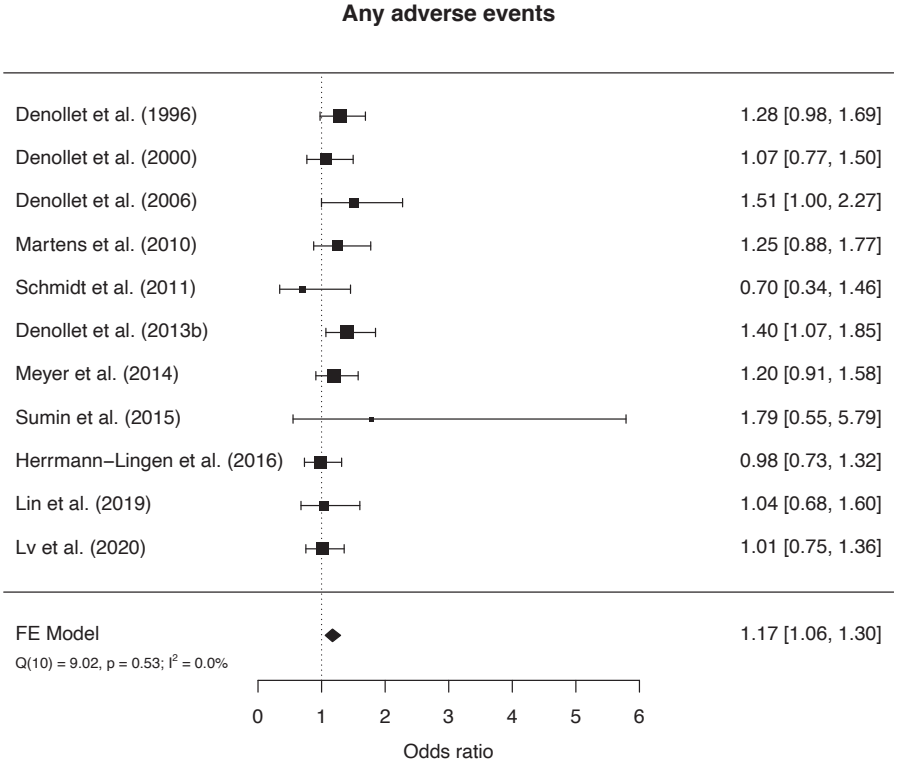


Figure K7: Fixed effects meta-analysis of studies on the association between Type D personality (continuous method) and the occurrence of adverse events



Appendix L: Sensitivity analyses (*Chapter 5*)

Table L1: Overall fixed effects model odds ratio estimates [95%CI] for the interaction effect between NA and SI on each endpoint, when excluding a particular study from the meta-analysis. Bold-faced estimates indicate a different conclusion after leaving out that study.

Excluded study	All-cause mortality	Cardiac mortality	Myocardial infarction	CABG	PCI	MACE	Any adverse event
Denollet et al. (1996)	0.99 [0.92, 1.06]	0.93 [0.80, 1.09]	1.03 [0.92, 1.15]	1.2 [0.81, 1.77]	1.11 [0.95, 1.3]	1.14 [1.01, 1.28]	1.15 [1.03, 1.29]
Denollet et al. (2000)	0.99 [0.92, 1.06]	0.93 [0.81, 1.08]	1.02 [0.92, 1.14]	1.33 [0.86, 2.07]	1.17 [0.99, 1.38]	1.15 [1.03, 1.28]	1.18 [1.06, 1.32]
Denollet et al. (2006)	0.99 [0.92, 1.06]	0.94 [0.81, 1.08]	1.02 [0.92, 1.14]	1.28 [0.89, 1.85]	1.10 [0.93, 1.3]	1.12 [1.00, 1.25]	1.15 [1.04, 1.28]
Martens et al. (2010)	0.99 [0.92, 1.06]	-	1.01 [0.90, 1.13]	-	1.12 [0.96, 1.32]	1.13 [1.01, 1.26]	1.17 [1.05, 1.3]
Pelle et al. (2010)	1.00 [0.93, 1.08]	0.96 [0.80, 1.14]	-	-	-	-	-
Coyne et al. (2011)	1.02 [0.95, 1.11]	-	-	-	-	-	-
Grande et al. (2011)	1.00 [0.93, 1.08]	-	1.02 [0.91, 1.15]	-	-	-	-
Schmidt et al. (2011)	-	-	-	-	1.15 [0.99, 1.35]	1.15 [1.03, 1.28]	1.18 [1.07, 1.31]
Denollet et al. (2013a)	1.00 [0.93, 1.08]	0.96 [0.81, 1.14]	-	-	-	-	-
Denollet et al. (2013b)	0.98 [0.92, 1.05]	0.92 [0.80, 1.07]	1.01 [0.91, 1.13]	1.35 [0.91, 1.99]	1.11 [0.93, 1.33]	**1.10 [0.98, 1.24]**	1.14 [1.02, 1.27]
Meyer et al. (2014)	0.99 [0.92, 1.06]	0.93 [0.80, 1.08]	1.03 [0.92, 1.15]	1.33 [0.90, 1.97]	1.14 [0.95, 1.36]	1.14 [1.02, 1.28]	1.17 [1.04, 1.31]
Sumin et al. (2015)	-	-	1.02 [0.92, 1.14]	-	-	1.14 [1.02, 1.26]	1.17 [1.05, 1.3]
Dulfer et al. (2015)	0.96 [0.89, 1.03]	-	-	-	-	-	-
Herrmann-Lingen et al. (2016)	0.99 [0.93, 1.06]	0.94 [0.81, 1.09]	-	-	-	1.16 [1.04, 1.3]	1.20 [1.08, 1.34]
Gostoli et al. (2016)	0.99 [0.92, 1.06]	0.94 [0.81, 1.08]	-	-	-	-	-
Pushkarev et al. (2017)	1.00 [0.93, 1.07]	0.96 [0.82, 1.13]	1.05 [0.93, 1.18]	-	-	-	-
Conden et al. (2017)	0.99 [0.92, 1.07]	-	1.06 [0.91, 1.23]	-	-	-	-
Lin et al. (2019)	-	-	-	-	-	1.15 [1.03, 1.28]	1.18 [1.06, 1.31]
Lv et al. (2020)	-	-	-	-	1.14 [0.95, 1.37]	1.16 [1.03, 1.30]	1.20 [1.07, 1.34]

CABG = coronary artery bypass grafting; MACE = major adverse cardiac event; PCI = percutaneous coronary intervention

Table L2: For each endpoint, the estimated odds ratios [95% Bayesian credible interval] of demographic predictors, the Type D effects according to the continuous method, according to three different prior distributions for the regression coefficient of the interaction effect between NA and SI. The 95%CI of bold cells does not include an odds ratio of one.

Outcome	All-cause mortality	Cardiac mortality	Myocardial infarction	CABG	PCI	MACE	Any adverse event
Sample size	N = 10647	N = 6166	N = 6269	N = 2832	N = 2840	N = 4315	N = 6013
Prior for interaction effect	OR [95%CI}	OR [95%CI}	OR [95%CI}	OR [95%CI}	OR [95%CI}	OR [95%CI}	OR [95%CI}
$N\ (\mu = 0, \sigma = 1)$	0.998 [0.919, 1.092]	1.000 [0.857, 1.196]	1.071 [0.941, 1.234]	1.183 [0.868, 1.689]	1.115 [0.892, 1.389]	1.141 [0.992, 1.296]	**1.168 [1.042, 1.313]**
$N\ (\mu = 0, \sigma = 2)$	0.996 [0.918, 1.092]	0.996 [0.851, 1.191]	1.069 [0.937, 1.232]	1.171 [0.844, 1.666]	1.112 [0.896, 1.351]	**1.140 [1.001, 1.286]**	**1.167 [1.033, 1.313]**
$N\ (\mu = 0, \sigma = 4)$	0.996 [0.913, 1.089]	1.000 [0.853, 1.194]	1.068 [0.937, 1.239]	1.173 [0.857, 1.639]	1.115 [0.897, 1.371]	1.139 [0.997, 1.297]	**1.170 [1.036, 1.313]**

CABG = coronary artery bypass grafting; MACE = major adverse cardiac event; N = Normal distribution; OR = Odds ratio; PCI = percutaneous coronary intervention

Table L3: For each endpoint and method to estimate the Type D effect, the brier scores, representing the accuracy of predicting the observed endpoint based on the model estimates. The minimum brier score of 0 indicates a perfect model prediction of the observed events, whereas the maximum brier score of 1 indicates the worst possible prediction.

Outcome	All-cause mortality	Cardiac mortality	Myocardial infarction	CABG	PCI	MACE	Any adverse event
Model	N = 10647	N = 6166	N = 6269	N = 2832	N = 2840	N = 4315	N = 6013
Age + Men + TypeD	0.085	0.038	0.050	0.023	0.081	0.118	0.092
Age + Men + TypeD + HighNA + HighSI	0.086	0.038	0.050	0.022	0.081	0.118	0.092
Age + Men + NA + SI + NA*SI	0.085	0.038	0.049	0.022	0.080	0.117	0.092

CABG = coronary artery bypass grafting; MACE = major adverse cardiac event; N = Normal distribution; OR = Odds ratio; PCI = percutaneous coronary intervention

Appendix M: DS14 scale information for Study 1 (*Chapter 6*)

DS14	*Negative Affectivity*							*Social Inhibition*						
	DS2	DS4	DS5	DS7	DS9	DS12	DS13	DS1	DS3	DS6	DS8	DS10	DS11	DS14
DS2	1													
DS4	**0.35**	1												
DS5	**0.43**	**0.52**	1											
DS7	**0.45**	**0.64**	**0.54**	1										
DS9	**0.36**	**0.51**	**0.6**	**0.57**	1									
DS12	**0.52**	**0.6**	**0.54**	**0.68**	**0.51**	1								
DS13	**0.4**	**0.71**	**0.51**	**0.73**	**0.58**	**0.72**	1							
DS1	0.09	0.23	0.16	0.27	0.24	0.21	0.24	1						
DS3	0.05	0.15	0.1	0.18	0.18	0.15	0.14	**0.63**	1					
DS6	0.32	0.46	0.39	0.5	0.42	0.44	0.46	**0.54**	**0.4**	1				
DS8	0.22	0.31	0.27	0.39	0.36	0.32	0.33	**0.64**	**0.51**	**0.64**	1			
DS10	0.12	0.3	0.26	0.33	0.35	0.3	0.31	**0.57**	**0.45**	**0.52**	**0.55**	1		
DS11	0.19	0.34	0.33	0.4	0.4	0.38	0.38	**0.51**	**0.42**	**0.53**	**0.53**	**0.66**	1	
DS14	0.2	0.37	0.29	0.4	0.39	0.36	0.4	**0.59**	**0.47**	**0.63**	**0.7**	**0.56**	**0.56**	1
Skewness	0.11	0.79	0.18	0.63	0.86	0.25	0.93	0.67	0.22	0.48	0.62	0.45	0.37	0.72
Kurtosis	-0.97	-0.37	-0.92	-0.67	-0.08	-1.07	-0.1	-0.32	-0.76	-0.83	-0.69	-1.02	-0.92	-0.47

Appendix N: PHQ-9 & GAD-7 scale information for Study 1 (*Chapter 6*)

	PHQ9_1	PHQ9_2	PHQ9_3	PHQ9_4	PHQ9_5	PHQ9_6	PHQ9_7	PHQ9_8	PHQ9_9
PHQ9_1	1								
PHQ9_2	0.72	1							
PHQ9_3	0.42	0.44	1						
PHQ9_4	0.54	0.52	0.55	1					
PHQ9_5	0.44	0.45	0.4	0.48	1				
PHQ9_6	0.53	0.64	0.36	0.43	0.45	1			
PHQ9_7	0.49	0.51	0.38	0.49	0.39	0.46	1		
PHQ9_8	0.39	0.43	0.33	0.37	0.34	0.4	0.48	1	
PHQ9_9	0.38	0.48	0.22	0.27	0.29	0.43	0.3	0.26	1
Skewness	1.66	2.06	0.93	0.73	1.94	2.19	1.77	3.1	4.99
Kurtosis	2.63	4.28	-0.22	-0.5	3.2	4.4	2.66	10.19	29.49

	GAD7_1	GAD7_2	GAD7_3	GAD7_4	GAD7_5	GAD7_6	GAD7_7
GAD7_1	1						
GAD7_2	0.71	1					
GAD7_3	0.68	0.77	1				
GAD7_4	0.67	0.66	0.67	1			
GAD7_5	0.49	0.48	0.47	0.58	1		
GAD7_6	0.51	0.5	0.52	0.57	0.48	1	
GAD7_7	0.52	0.54	0.53	0.46	0.37	0.42	1
Skewness	1.76	2.06	1.44	1.5	2.26	1.36	2.62
Kurtosis	3.27	4.34	2.02	1.99	5.41	1.81	7.42

Appendix O: Additional simulation results (*Chapter 6*)

Table O1: Mean standardized regression coefficient [including 95% confidence interval] of the interaction effect between NA and SI on depression for all **continuous** simulation scenarios.

N	Skewness	True size interaction	*Sum score regression*	Latent variable models		
				LMS MLR	*Single PI MLR*	*Matched PI MLR*
250	0	0	0 [-.08 / .08]	0 [-.08 / .08]	0 [-.1 / .1]	0 [-.1 / .1]
		.1035	.12 [.04 / .2]	.12 [.04 / .2]	.11 [.01 / .21]	.14 [.04 / .24]
		.207	.22 [.14 / .3]	.23 [.15 / .31]	.21 [.09 / .33]	.27 [.17 / .37]
		.414	.41 [.33 / .49]	.42 [.36 / .48]	.39 [.25 / .53]	.49 [.39 / .59]
	2	0	.02 [-.06 / .1]	.02 [-.06 / .1]	0 [-.22 / .22]	0 [-.1 / .1]
		.1035	.15 [.05 / .25]	.14 [.06 / .22]	.2 [-.43 / .83]	.16 [.04 / .28]
		.207	.26 [.16 / .36]	.25 [.17 / .33]	.34 [-.19 / .87]	.29 [.17 / .41]
		.414	.44 [.32 / .56]	.44 [.36 / .52]	.58 [-.48 / 1.64]	.49 [.37 / .61]
	3	0	.02 [-.08 / .12]	.01 [-.07 / .09]	0 [-.41 / .41]	0 [-.12 / .12]
		.1035	.16 [.04 / .28]	.13 [.05 / .21]	.29 [-.81 / 1.39]	.18 [.04 / .32]
		.207	.29 [.15 / .43]	.26 [.18 / .34]	.47 [-.71 / 1.65]	.33 [.17 / .49]
		.414	.48 [.3 / .66]	.44 [.36 / .52]	.68 [-.81 / 2.17]	.54 [.36 / .72]
500	0	0	0 [-.06 / .06]	0 [-.06 / .06]	0 [-.06 / .06]	0 [-.06 / .06]
		.1035	.12 [.06 / .18]	.12 [.06 / .18]	.11 [.05 / .17]	.14 [.08 / .2]
		.207	.22 [.16 / .28]	.23 [.19 / .27]	.21 [.13 / .29]	.27 [.21 / .33]
		.414	.41 [.35 / .47]	.43 [.39 / .47]	.39 [.29 / .49]	.49 [.43 / .55]
	2	0	.02 [-.04 / .08]	.02 [-.04 / .08]	0 [-.12 / .12]	0 [-.08 / .08]
		.1035	.15 [.07 / .23]	.14 [.08 / .2]	.19 [-.05 / .43]	.16 [.08 / .24]
		.207	.27 [.19 / .35]	.25 [.19 / .31]	.37 [-.24 / .98]	.3 [.2 / .4]
		.414	.45 [.35 / .55]	.44 [.38 / .5]	.58 [-.11 / 1.27]	.5 [.4 / .6]
	3	0	.02 [-.04 / .08]	.01 [-.03 / .05]	0 [-.43 / .43]	0 [-.08 / .08]
		.1035	.17 [.09 / .25]	.14 [.08 / .2]	.32 [-.74 / 1.38]	.18 [.08 / .28]
		.207	.3 [.2 / .4]	.25 [.19 / .31]	.6 [-.97 / 2.17]	.33 [.21 / .45]
		.414	.49 [.35 / .63]	.44 [.38 / .5]	.8 [-.75 / 2.35]	.55 [.41 / .69]
3000	0	0	0 [-.02 / .02]	0 [-.02 / .02]	0 [-.02 / .02]	0 [-.02 / .02]
		.1035	.11 [.09 / .13]	.12 [.1 / .14]	.11 [.09 / .13]	.14 [.12 / .16]
		.207	.22 [.2 / .24]	.24 [.22 / .26]	.22 [.18 / .26]	.27 [.25 / .29]
		.414	.41 [.39 / .43]	.43 [.41 / .45]	.39 [.35 / .43]	.49 [.45 / .53]
	2	0	.02 [0 / .04]	.01 [-.01 / .03]	0 [-.04 / .04]	0 [-.04 / .04]
		.1035	.15 [.13 / .17]	.14 [.12 / .16]	.18 [.1 / .26]	.17 [.13 / .21]
		.207	.27 [.23 / .31]	.26 [.24 / .28]	.34 [.2 / .48]	.31 [.27 / .35]
		.414	.45 [.41 / .49]	.44 [.42 / .46]	.54 [.34 / .74]	.51 [.47 / .55]
	3	0	.01 [-.01 / .03]	.01 [-.01 / .03]	0 [-.08 / .08]	0 [-.04 / .04]
		.1035	.17 [.13 / .21]	.14 [.12 / .16]	.35 [-.22 / .92]	.19 [.15 / .23]
		.207	.31 [.23 / .39]	.25 [.23 / .27]	.67 [-.6 / 1.94]	.35 [.27 / .43]
		.414	.5 [.42 / .58]	.45 [.43 / .47]	.93 [-.21 / 2.07]	.56 [.48 / .64]

Table O2: Mean standardized regression coefficient [including 95% confidence interval] of the interaction effect between NA and SI on depression for all **ordinal** simulation scenarios.

N	Skewness	True size Interaction	*Sum score regression*	Latent variable models				
				LMS MLR	*Single PI MLR*	*Single PI DWLS*	*Matched PI MLR*	*Matched PI DWLS*
250	0	0	0 [-.08 / .08]	0 [-.08 / .08]	0 [-.08 / .08]	0 [-.08 / .08]	0 [-.1 / .1]	0 [-.1 / .1]
		.1035	.11 [.03 / .19]	.12 [.04 / .2]	.08 [0 / .16]	.08 [0 / .16]	.14 [.04 / .24]	.13 [.03 / .23]
		.207	.21 [.13 / .29]	.23 [.15 / .31]	.15 [.07 / .23]	.15 [.07 / .23]	.26 [.16 / .36]	.25 [.15 / .35]
		.414	.38 [.3 / .46]	.41 [.33 / .49]	.27 [.17 / .37]	.27 [.17 / .37]	.47 [.37 / .57]	.46 [.36 / .56]
	2	0	.03 [-.07 / .13]	.04 [-.06 / .14]	.01 [-.07 / .09]	.01 [-.07 / .09]	.01 [-.11 / .13]	.03 [-.09 / .15]
		.1035	.14 [.04 / .24]	.15 [.07 / .23]	.07 [-.01 / .15]	.08 [0 / .16]	.15 [.03 / .27]	.16 [.04 / .28]
		.207	.22 [.14 / .3]	.24 [.14 / .34]	.12 [.04 / .2]	.12 [.04 / .2]	.24 [.14 / .34]	.25 [.13 / .37]
		.414	.33 [.25 / .41]	.37 [.27 / .47]	.18 [.1 / .26]	.19 [.09 / .29]	.37 [.27 / .47]	.38 [.26 / .5]
	3	0	.03 [-.07 / .13]	.04 [-.06 / .14]	0 [-.08 / .08]	.01 [-.07 / .09]	.01 [-.13 / .15]	.03 [-.09 / .15]
		.1035	.14 [.04 / .24]	.15 [.07 / .23]	.06 [-.02 / .14]	.07 [-.01 / .15]	.15 [.03 / .27]	.16 [.04 / .28]
		.207	.22 [.14 / .3]	.23 [.13 / .33]	.11 [.03 / .19]	.11 [.03 / .19]	.24 [.12 / .36]	.25 [.13 / .37]
		.414	.32 [.22 / .42]	.36 [.24 / .48]	.16 [.08 / .24]	.17 [.07 / .27]	.36 [.24 / .48]	.37 [.25 / .49]
500	0	0	0 [-.06 / .06]	0 [-.06 / .06]	0 [-.06 / .06]	0 [-.06 / .06]	0 [-.08 / .08]	0 [-.08 / .08]
		.1035	.11 [.05 / .17]	.12 [.06 / .18]	.08 [.02 / .14]	.07 [.01 / .13]	.14 [.06 / .22]	.13 [.07 / .19]
		.207	.21 [.15 / .27]	.23 [.17 / .29]	.15 [.09 / .21]	.15 [.09 / .21]	.26 [.18 / .34]	.25 [.17 / .33]
		.414	.38 [.32 / .44]	.41 [.35 / .47]	.27 [.19 / .35]	.26 [.18 / .34]	.47 [.39 / .55]	.46 [.38 / .54]
	2	0	.03 [-.03 / .09]	.04 [-.02 / .1]	.01 [-.05 / .07]	.01 [-.05 / .07]	.01 [-.09 / .11]	.03 [-.05 / .11]
		.1035	.14 [.08 / .2]	.15 [.09 / .21]	.07 [.01 / .13]	.08 [.02 / .14]	.15 [.07 / .23]	.16 [.08 / .24]
		.207	.22 [.16 / .28]	.24 [.18 / .3]	.12 [.06 / .18]	.12 [.06 / .18]	.24 [.16 / .32]	.25 [.17 / .33]
		.414	.33 [.27 / .39]	.37 [.31 / .43]	.18 [.12 / .24]	.19 [.13 / .25]	.37 [.29 / .45]	.38 [.3 / .46]
	3	0	.04 [-.04 / .12]	.04 [-.02 / .1]	.01 [-.05 / .07]	.01 [-.05 / .07]	.02 [-.08 / .12]	.03 [-.07 / .13]
		.1035	.14 [.08 / .2]	.15 [.09 / .21]	.07 [.01 / .13]	.07 [.01 / .13]	.15 [.07 / .23]	.16 [.08 / .24]
		.207	.22 [.16 / .28]	.23 [.17 / .29]	.1 [.04 / .16]	.11 [.05 / .17]	.24 [.16 / .32]	.24 [.16 / .32]
		.414	.31 [.25 / .37]	.35 [.27 / .43]	.16 [.1 / .22]	.17 [.11 / .23]	.35 [.27 / .43]	.37 [.29 / .45]
3000	0	0	0 [-.02 / .02]	0 [-.02 / .02]	0 [-.02 / .02]	0 [-.02 / .02]	0 [-.02 / .02]	0 [-.02 / .02]
		.1035	.11 [.09 / .13]	.12 [.1 / .14]	.08 [.06 / .1]	.08 [.06 / .1]	.14 [.1 / .18]	.13 [.09 / .17]
		.207	.21 [.19 / .23]	.23 [.21 / .25]	.15 [.13 / .17]	.15 [.13 / .17]	.26 [.24 / .28]	.25 [.21 / .29]
		.414	.38 [.36 / .4]	.41 [.39 / .43]	.27 [.23 / .31]	.27 [.23 / .31]	.46 [.42 / .5]	.46 [.44 / .48]
	2	0	.03 [.01 / .05]	.04 [.02 / .06]	.01 [-.01 / .03]	.01 [-.01 / .03]	.01 [-.03 / .05]	.03 [-.01 / .07]
		.1035	.14 [.12 / .16]	.15 [.13 / .17]	.07 [.05 / .09]	.08 [.06 / .1]	.14 [.1 / .18]	.16 [.12 / .2]
		.207	.22 [.2 / .24]	.24 [.22 / .26]	.12 [.1 / .14]	.12 [.1 / .14]	.24 [.22 / .26]	.25 [.23 / .27]
		.414	.33 [.31 / .35]	.37 [.35 / .39]	.18 [.16 / .2]	.19 [.17 / .21]	.37 [.35 / .39]	.38 [.36 / .4]
	3	0	.04 [.02 / .06]	.04 [.02 / .06]	.01 [-.01 / .03]	.01 [-.01 / .03]	.01 [-.03 / .05]	.03 [-.01 / .07]
		.1035	.14 [.12 / .16]	.15 [.13 / .17]	.06 [.04 / .08]	.07 [.05 / .09]	.15 [.11 / .19]	.15 [.11 / .19]
		.207	.22 [.2 / .24]	.23 [.21 / .25]	.1 [.08 / .12]	.11 [.09 / .13]	.23 [.19 / .27]	.24 [.2 / .28]
		.414	.31 [.29 / .33]	.35 [.31 / .39]	.16 [.12 / .2]	.17 [.15 / .19]	.35 [.31 / .39]	.37 [.33 / .41]

Table O3: For each simulation condition, the squared distance between the estimated value of the interaction effect and the true value of the interaction, averaged across 500 replications (mean squared error).

N	Skew ness	True size Inter-action	Continuous item scores				Ordinal item scores					
			Sum score regression	*LMS MLR*	*Single PI MLR*	*Matched PI MLR*	*Sum score regression*	*LMS MLR*	*Single PI MLR*	*Single PI WLS*	*Matched PI MLR*	*Matched PI WLS*
250	0	0	0.002	0.002	0.002	0.003	0.002	0.002	0.002	0.001	0.003	0.003
		.1035	0.002	0.002	0.002	0.003	0.002	0.002	0.002	0.002	0.004	0.004
		.207	0.002	0.002	0.003	0.006	0.002	0.002	0.005	0.005	0.006	0.004
		.414	0.002	0.001	0.006	0.007	0.003	0.001	0.022	0.024	0.006	0.005
	1	0	0.002	0.002	0.012	0.003	0.003	0.004	0.001	0.002	0.004	0.005
		.1035	0.005	0.003	0.11	0.007	0.004	0.004	0.003	0.002	0.005	0.006
		.207	0.006	0.004	0.092	0.011	0.002	0.003	0.01	0.009	0.004	0.005
		.414	0.004	0.002	0.319	0.011	0.01	0.004	0.058	0.053	0.005	0.004
	2	0	0.003	0.002	0.044	0.003	0.004	0.004	0.001	0.002	0.004	0.005
		.1035	0.007	0.002	0.348	0.01	0.004	0.004	0.003	0.002	0.005	0.006
		.207	0.012	0.004	0.428	0.02	0.002	0.003	0.012	0.011	0.005	0.005
		.414	0.012	0.002	0.652	0.023	0.011	0.006	0.066	0.061	0.007	0.006
500	0	0	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
		.1035	0.001	0.001	0.001	0.002	0.001	0.001	0.001	0.002	0.002	0.002
		.207	0.001	0.001	0.002	0.005	0.001	0.001	0.004	0.004	0.004	0.003
		.414	0.001	0.001	0.003	0.006	0.002	0.001	0.022	0.024	0.004	0.003
	1	0	0.001	0.001	0.004	0.001	0.002	0.003	0.001	0.001	0.002	0.003
		.1035	0.004	0.002	0.022	0.005	0.003	0.003	0.002	0.002	0.003	0.004
		.207	0.006	0.003	0.119	0.011	0.001	0.002	0.009	0.008	0.003	0.003
		.414	0.004	0.002	0.149	0.01	0.008	0.003	0.056	0.051	0.003	0.002
	2	0	0.001	0.001	0.048	0.001	0.003	0.003	0.001	0.001	0.002	0.003
		.1035	0.006	0.002	0.336	0.009	0.003	0.003	0.002	0.002	0.003	0.004
		.207	0.011	0.003	0.799	0.019	0.001	0.001	0.011	0.01	0.003	0.003
		.414	0.011	0.002	0.777	0.023	0.011	0.005	0.066	0.06	0.005	0.004
3000	0	0	0	0	0	0	0	0	0	0	0	0
		.1035	0	0	0	0.001	0	0	0.001	0.001	0.001	0.001
		.207	0	0.001	0	0.004	0	0.001	0.003	0.003	0.003	0.003
		.414	0	0	0.001	0.005	0.001	0	0.02	0.022	0.003	0.002
	1	0	0	0	0	0	0.001	0.002	0	0	0	0.001
		.1035	0.003	0.001	0.007	0.004	0.002	0.002	0.001	0.001	0.002	0.003
		.207	0.005	0.002	0.021	0.01	0	0.001	0.009	0.007	0.001	0.002
		.414	0.002	0.001	0.026	0.009	0.008	0.002	0.056	0.051	0.002	0.001
	2	0	0	0	0.002	0	0.001	0.002	0	0	0.001	0.001
		.1035	0.005	0.001	0.146	0.008	0.002	0.002	0.002	0.001	0.002	0.003
		.207	0.013	0.002	0.632	0.023	0	0.001	0.011	0.009	0.001	0.002
		.414	0.009	0.001	0.609	0.023	0.01	0.004	0.066	0.06	0.004	0.002

Appendix P: DS14 scale information for Study 1 (*Chapter 7*)

DS14	*Negative Affectivity*							*Social Inhibition*						
item	DS2	DS4	DS5	DS7	DS9	DS12	DS13	DS1	DS3	DS6	DS8	DS10	DS11	DS14
DS2	1													
DS4	0.35	1												
DS5	0.49	0.39	1											
DS7	0.4	0.64	0.45	1										
DS9	0.41	0.49	0.54	0.53	1									
DS12	0.52	0.53	0.46	0.57	0.39	1								
DS13	0.42	0.71	0.44	0.7	0.52	0.59	1							
DS1	0.04	0.19	0.08	0.19	0.1	0.15	0.18	1						
DS3	-0.06	0.09	-0.05	0.1	0.03	0.04	0.08	0.61	1					
DS6	0.25	0.41	0.34	0.44	0.36	0.33	0.38	0.5	0.25	1				
DS8	0.16	0.29	0.15	0.33	0.26	0.23	0.29	0.55	0.43	0.58	1			
DS10	0.14	0.23	0.16	0.29	0.22	0.21	0.21	0.56	0.44	0.52	0.59	1		
DS11	0.12	0.2	0.18	0.23	0.16	0.19	0.2	0.5	0.38	0.45	0.48	0.51	1	
DS14	0.09	0.31	0.14	0.33	0.25	0.23	0.28	0.49	0.38	0.51	0.66	0.5	0.41	1
Skewness	0.02	1.08	0.13	0.93	0.94	0.05	1.1	0.54	0.14	0.58	0.55	0.36	0.16	0.66
Kurtosis	-1.13	0.25	-1.1	-0.34	0.11	-1.22	0.14	-0.63	-0.99	-0.69	-0.8	-1.12	-0.79	-0.45

Appendix Q: Additional simulation results (*Chapter 7*)

Figure Q1: Comparison of four methods to estimate interaction effects between NA and SI on a binary outcome in scenarios with **continuous items**, varying over the true interaction size (x axis), reliability of the NA and SI scales, amount of **latent skewness** (columns) and sample size (rows). Each data point shows the mean bias (including 95% confidence interval) in the standardized regression coefficient of the estimated **NA main effect**.

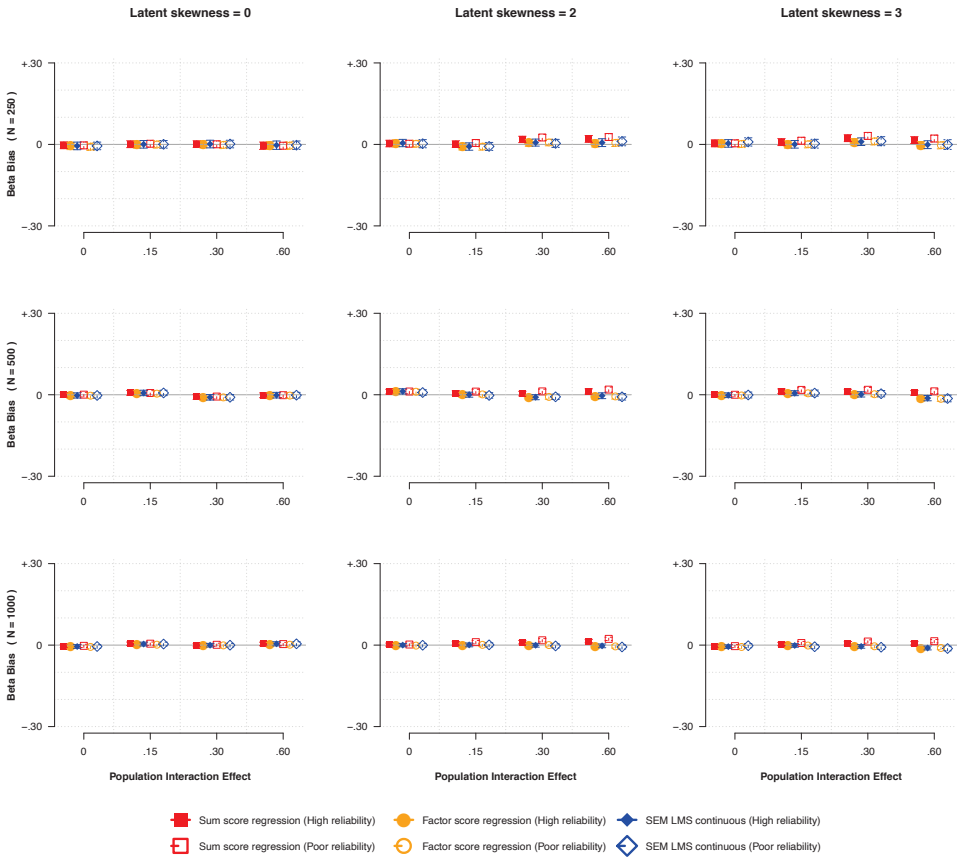


Figure Q2: Comparison of four methods to estimate interaction effects between NA and SI on a binary outcome in scenarios with **ordinal items**, varying over the true interaction size (x axis), reliability of the NA and SI scales, amount of **latent skewness** (columns) and sample size (rows). Each data point shows the mean bias (including 95% confidence interval) in the standardized regression coefficient of the estimated **NA main effect**.

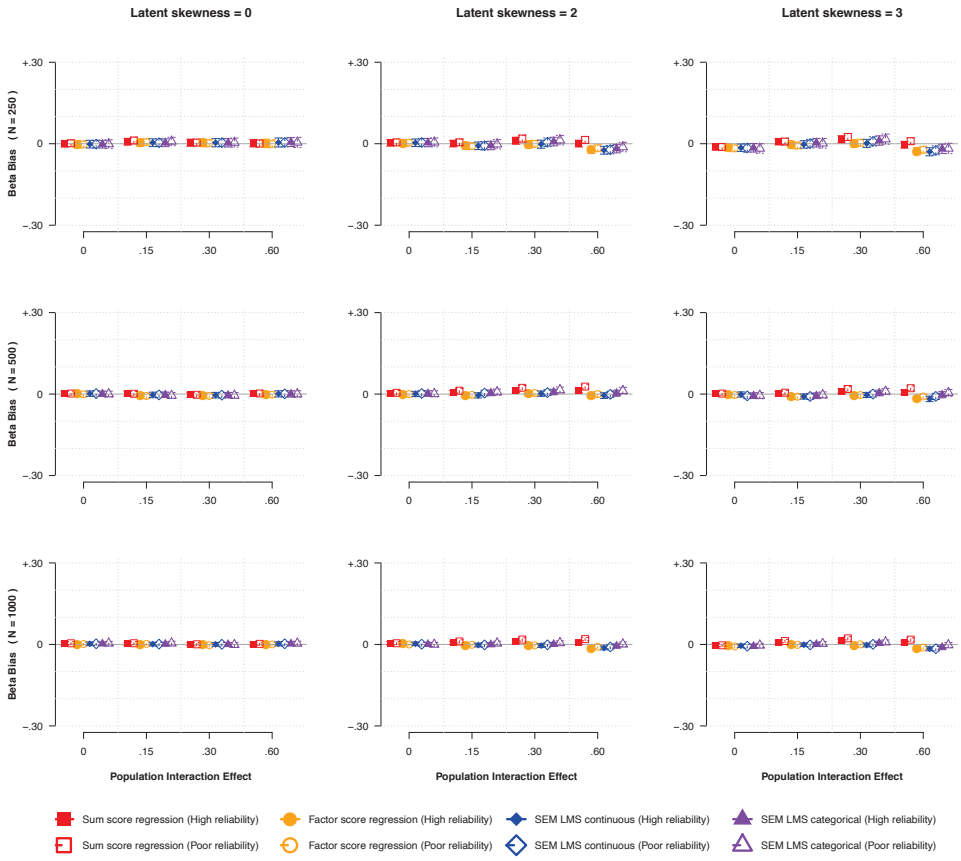


Figure Q3: Comparison of four methods to estimate interaction effects between NA and SI on a binary outcome in scenarios with **ordinal items**, varying over the true interaction size (x axis), reliability of the NA and SI scales, amount of **item skewness** (columns) and sample size (rows). Each data point shows the mean bias (including 95% confidence interval) in the standardized regression coefficient of the estimated **NA main effect**.

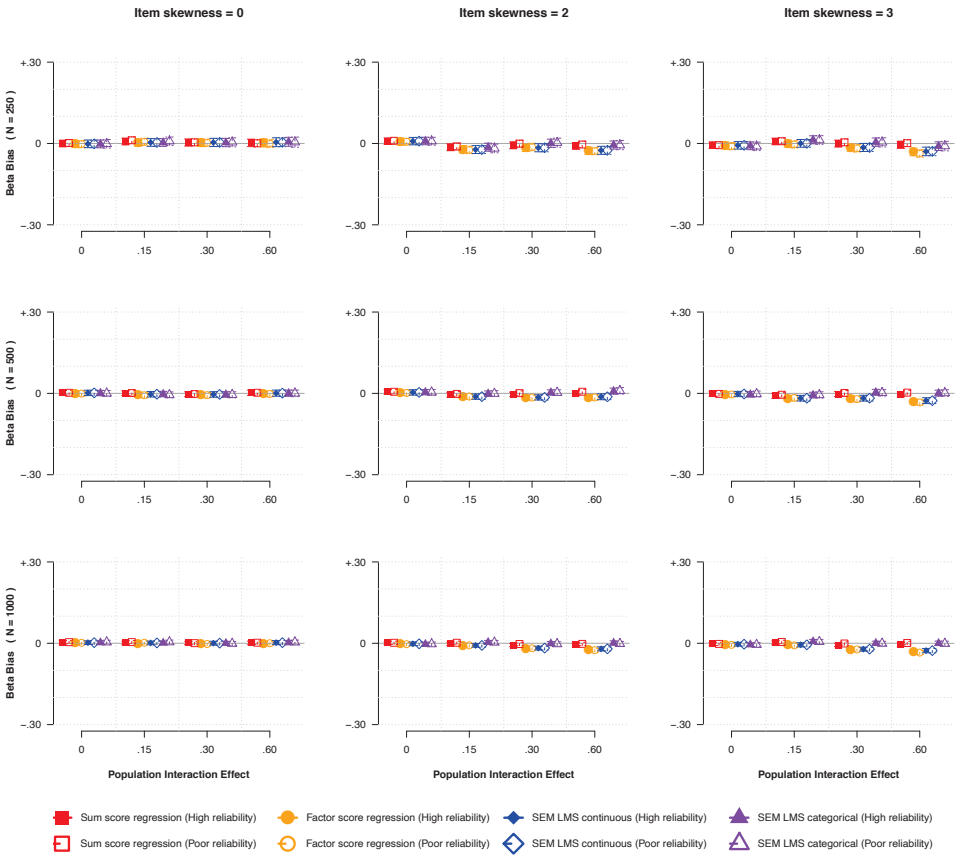


Figure Q4: Comparison of four methods to estimate interaction effects between NA and SI on a binary outcome in scenarios with **continuous items**, varying over the true interaction size (x axis), reliability of the NA and SI scales, amount of **latent skewness** (columns) and sample size (rows). Each data point shows the mean bias (including 95% confidence interval) in the standardized regression coefficient of the estimated **SI main effect**.

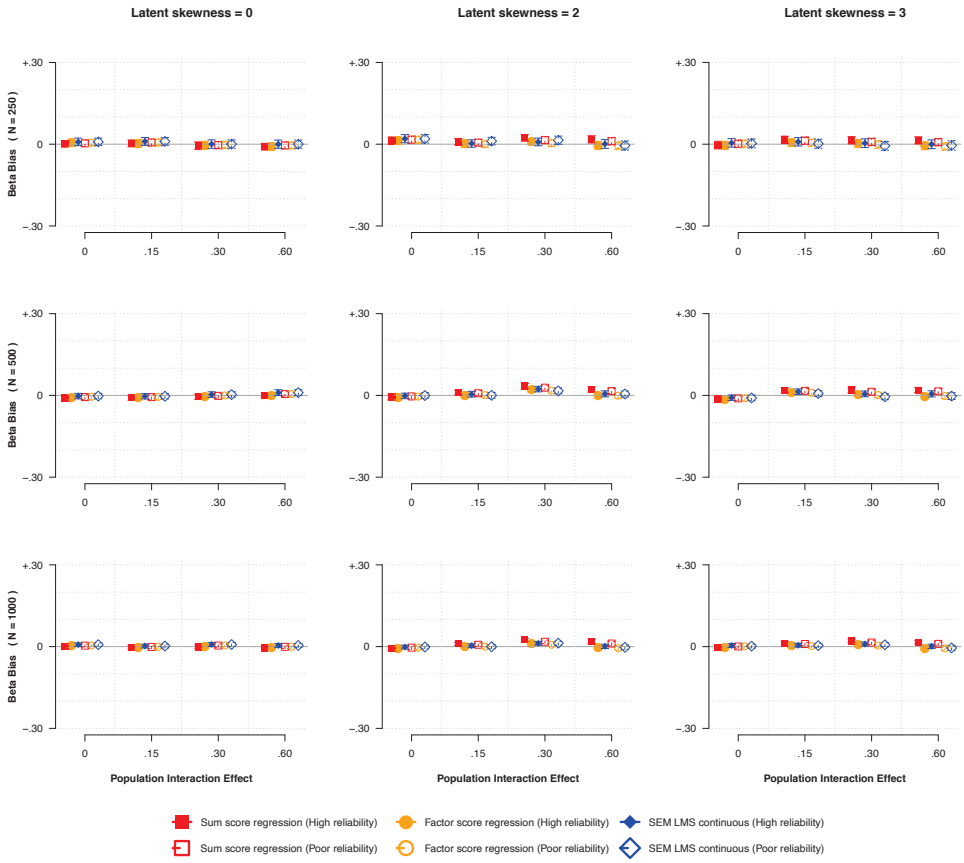


Figure Q5: Comparison of four methods to estimate interaction effects between NA and SI on a binary outcome in scenarios with **ordinal items**, varying over the true interaction size (x axis), reliability of the NA and SI scales, amount of **latent skewness** (columns) and sample size (rows). Each data point shows the mean bias (including 95% confidence interval) in the standardized regression coefficient of the estimated **SI main effect**.

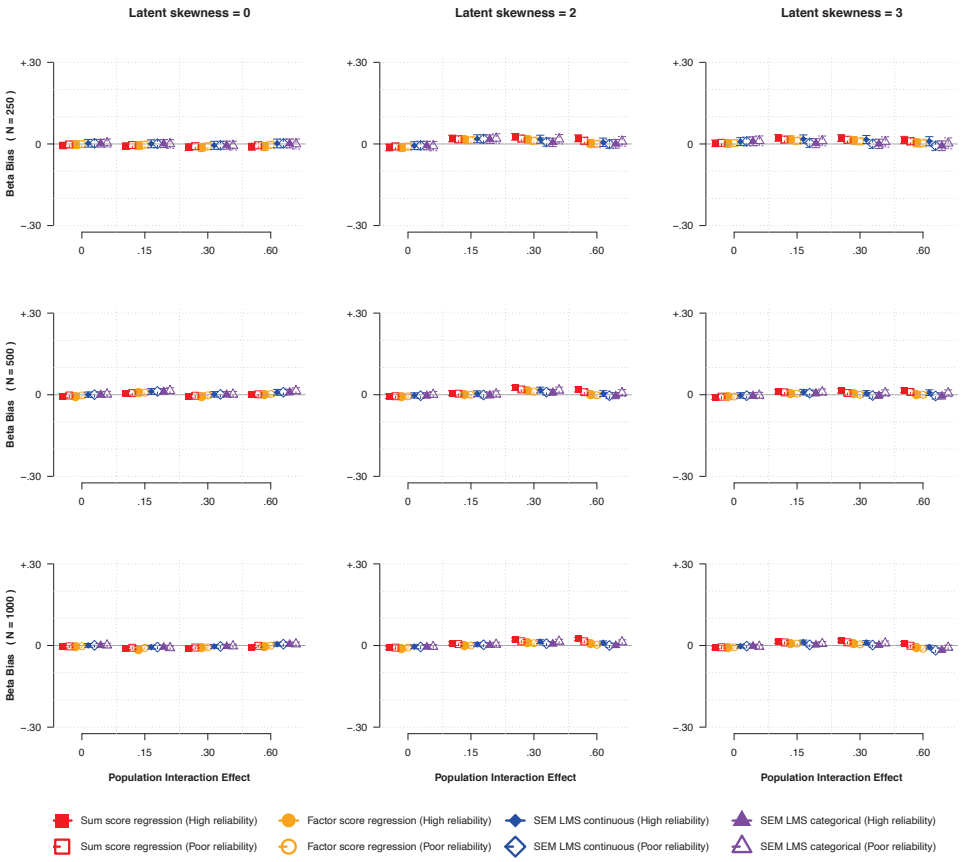


Figure Q6: Comparison of four methods to estimate interaction effects between NA and SI on a binary outcome in scenarios with **ordinal items**, varying over the true interaction size (x axis), reliability of the NA and SI scales, amount of **item skewness** (columns) and sample size (rows). Each data point shows the mean bias (including 95% confidence interval) in the standardized regression coefficient of the estimated **SI main effect**.

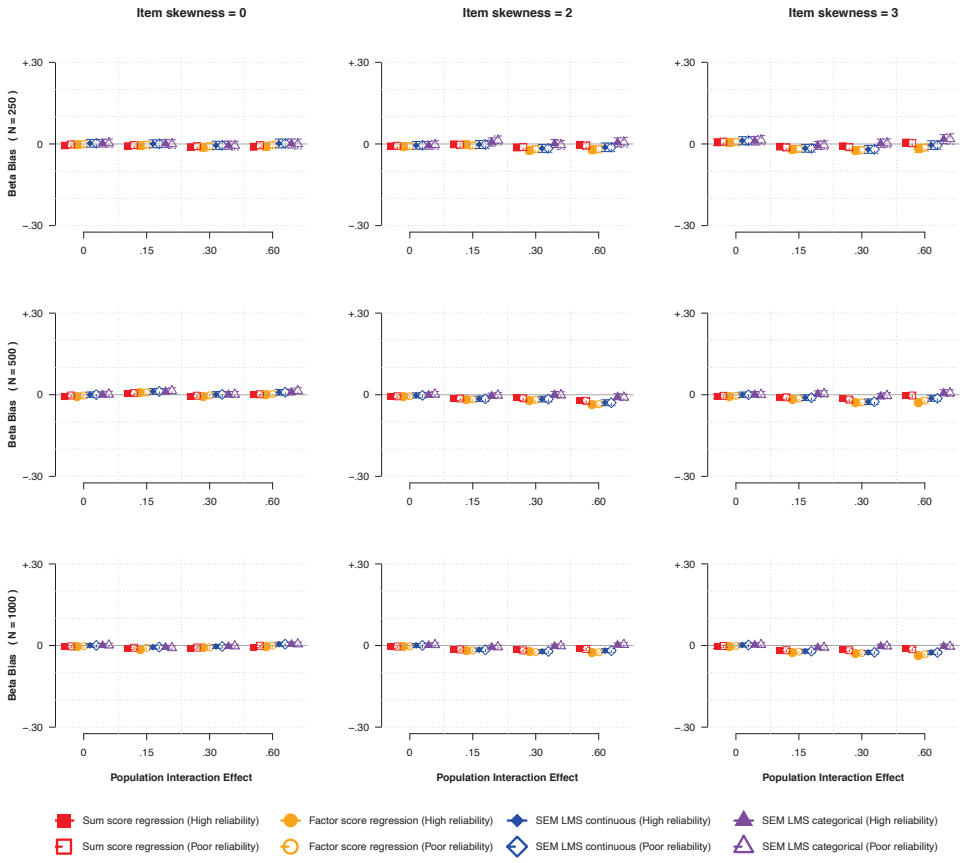


Table Q1: Mean squared error (x1000) for all simulation scenarios involving continuous item scores and skewness generated at the latent variable level.

N	N/p ratio	Skewness	Interaction	Sum scores		Factor scores		LMS (continuous)	
				a = .87	a = .60	a = .87	a = .60	a = .87	a = .60
250	18	0	0	17	17	16	16	21	21
			0.15	18	19	17	17	22	22
			0.31	19	20	18	19	27	27
			0.62	31	44	30	42	42	42
		2	0	18	17	17	16	22	21
			0.15	17	17	17	16	24	24
			0.31	20	21	20	21	28	28
			0.62	43	62	41	59	52	57
		3	0	18	18	18	16	23	19
			0.15	20	18	20	17	28	29
			0.31	26	26	26	25	39	36
			0.62	50	70	48	66	58	63
500	36	0	0	8	8	7	7	9	9
			0.15	8	9	7	8	9	9
			0.31	9	11	10	11	13	13
			0.62	20	34	20	33	23	23
		2	0	8	8	8	7	10	9
			0.15	8	9	8	9	11	11
			0.31	11	14	11	13	14	16
			0.62	26	46	24	43	27	28
		3	0	7	7	7	7	9	7
			0.15	9	9	9	9	13	14
			0.31	11	13	12	13	15	14
			0.62	31	57	29	53	28	33
1000	72	0	0	4	4	4	4	5	5
			0.15	4	5	4	5	5	5
			0.31	6	8	6	8	6	6
			0.62	14	30	13	28	10	10
		2	0	4	4	4	4	5	4
			0.15	4	4	4	4	5	5
			0.31	6	9	6	9	6	6
			0.62	18	40	16	36	11	12
		3	0	3	3	3	3	4	3
			0.15	4	4	4	4	5	5
			0.31	7	10	7	10	8	9
			0.62	29	59	26	54	16	20

Table Q2: Mean squared error (x1000) for all simulation scenarios involving ordinal item scores and skewness generated at the latent variable level.

N	N/p ratio	Skewness	Interaction	Sum scores		Factor scores		LMS (continuous)		LMS (categorical)	
				a = .87	a = .60	a = .87	a = .60	a = .87	a = .60	a = .87	a = .60
250	18	0	0	19	19	18	16	24	24	28	28
			0.15	16	17	15	15	20	20	26	26
			0.31	19	23	18	22	25	25	36	36
			0.62	31	48	30	48	43	43	69	69
		2	0	18	18	17	16	24	23	26	26
			0.15	18	18	18	17	27	25	34	34
			0.31	20	23	20	23	28	27	39	39
			0.62	46	66	45	64	46	48	66	66
		3	0	17	18	16	17	24	19	24	24
			0.15	15	16	15	15	24	23	34	34
			0.31	21	24	21	23	29	27	38	38
			0.62	56	76	56	75	49	57	69	69
500	36	0	0	10	9	9	8	12	12	14	14
			0.15	9	9	8	8	11	11	14	14
			0.31	10	14	10	14	12	12	16	16
			0.62	20	38	19	36	21	21	31	31
		2	0	8	8	8	7	11	10	11	11
			0.15	8	9	8	8	11	10	13	13
			0.31	12	16	12	15	13	12	16	16
			0.62	40	63	39	61	24	30	37	37
		3	0	8	7	8	7	12	9	11	11
			0.15	8	8	8	8	12	10	13	13
			0.31	13	16	12	15	14	11	14	14
			0.62	58	82	56	81	31	48	52	52
1000	72	0	0	4	4	3	3	4	4	5	5
			0.15	4	5	4	5	5	5	6	6
			0.31	6	9	5	9	6	6	8	8
			0.62	16	36	15	35	9	9	14	14
		2	0	4	4	4	3	5	5	5	5
			0.15	4	4	4	4	6	5	6	6
			0.31	7	12	7	11	6	6	7	7
			0.62	37	61	34	58	14	21	25	25
		3	0	3	3	3	3	5	4	4	4
			0.15	4	5	4	4	6	4	5	5
			0.31	9	13	9	13	7	7	8	8
			0.62	53	78	52	77	23	39	40	40

Table Q3: Mean squared error (x1000) for all simulation scenarios involving ordinal item scores and skewness generated at the item score level.

N	N/p ratio	Skewness	Interaction	Sum scores		Factor scores		LMS (continuous)		LMS (categorical)	
				a = .87	a = .60	a = .87	a = .60	a = .87	a = .60	a = .87	a = .60
250	18	0	0	19	19	18	16	24	28	23	28
			0.15	16	17	15	15	20	26	20	26
			0.31	19	23	18	22	25	40	26	36
			0.62	31	48	30	48	43	65	47	69
		2	0	19	20	18	18	25	32	24	30
			0.15	20	21	20	19	27	31	26	31
			0.31	21	25	22	25	28	41	28	39
			0.62	43	57	42	54	45	77	51	71
		3	0	18	18	18	17	25	31	25	31
			0.15	19	21	18	19	24	33	28	35
			0.31	25	27	25	28	33	48	32	42
			0.62	55	69	55	70	69	97	60	95
500	36	0	0	10	9	9	8	12	13	12	14
			0.15	9	9	8	8	11	13	11	14
			0.31	10	14	10	14	12	17	12	16
			0.62	20	38	19	36	21	36	23	31
		2	0	8	8	7	7	10	13	9	12
			0.15	9	10	8	9	11	14	11	13
			0.31	14	17	13	16	14	16	14	17
			0.62	33	51	31	48	23	38	26	34
		3	0	8	8	8	7	10	13	12	14
			0.15	9	10	9	9	11	15	11	14
			0.31	16	19	15	18	15	21	14	18
			0.62	38	55	37	53	29	41	26	35
1000	72	0	0	4	4	3	3	4	5	4	5
			0.15	4	5	4	5	5	6	5	6
			0.31	6	9	5	9	6	7	6	8
			0.62	16	36	15	35	9	13	10	14
		2	0	4	4	4	4	5	6	5	6
			0.15	5	6	5	5	6	6	5	6
			0.31	8	12	8	12	6	9	6	7
			0.62	28	47	25	42	11	15	11	15
		3	0	4	4	4	4	5	7	5	7
			0.15	6	7	6	7	6	7	6	7
			0.31	11	15	11	15	8	10	8	9
			0.62	37	55	33	49	15	18	12	15

Table Q4: The percentage of statistically significant (p < .05) interaction effects for all simulation scenarios with continuous item scores and skewness at the latent variable level. Bold faced cells indicate a power > .80.

N	N/p ratio	Skewness	Interaction	Sum scores		Factor scores		LMS (continuous)	
				a = .87	a = .60	a = .87	a = .60	a = .87	a = .60
250	18	0	0	5.4	5	5	5	5.4	5.4
			0.15	18	13.8	16.4	13.2	16.6	16.6
			0.31	58.4	51	55.2	47.2	57.2	57.2
			0.62	**96.2**	**91.8**	**95.6**	**90.4**	**95.4**	**95.4**
		2	0	4.6	4	3.2	2.2	3.4	3.4
			0.15	15	13.2	12	11.4	14.8	13.2
			0.31	44.4	38.2	39.8	33	45.8	40.4
			0.62	**90**	79.6	**89**	79	**89.2**	78.6
		3	0	4.6	4.8	3	2.8	3.8	3
			0.15	15.2	13.6	11.6	10.4	14.4	12.2
			0.31	41.8	33	36.8	29.6	42.4	34.4
			0.62	**80.4**	68.4	76.8	66.6	79.2	67.6
500	36	0	0	4.6	4.6	4	3.2	3.8	3.8
			0.15	35	29	32.8	28.4	32.8	32.8
			0.31	**88.6**	**80.8**	**89.6**	**82**	**89.4**	**89.4**
			0.62	**99.8**	**99.8**	**99.8**	**99.6**	**99.8**	**99.8**
		2	0	6	6	5.2	5	5.4	6.2
			0.15	31.4	27.4	32	28.4	33.4	33
			0.31	79.8	71.6	79.2	70.4	**82**	**80.2**
			0.62	**99.8**	**99**	**99.6**	**99.2**	**99.8**	**99.6**
		3	0	4	4.2	2.6	3.2	3.6	4.42
			0.15	31.8	28	31	28.6	34.2	35.6
			0.31	78.4	68.8	75.8	68.8	**81.16**	77.96
			0.62	**99.4**	**96.2**	**99.4**	**96.4**	**99**	**96.19**
1000	72	0	0	6	5.8	5.4	4.8	5.4	5.4
			0.15	63.4	54.2	63.6	57.6	63.6	63.6
			0.31	**99.6**	**97.2**	**99.6**	**99.2**	**99.6**	**99.6**
			0.62	**100**	**100**	**100**	**100**	**100**	**100**
		2	0	6.2	5.8	6.2	5.8	6.6	5.8
			0.15	67	59.6	65.8	60.4	68.2	69.2
			0.31	**99.2**	**96.6**	**98.8**	**97.4**	**98.6**	**98.6**
			0.62	**100**	**100**	**100**	**100**	**100**	**100**
		3	0	5.2	3.8	4.4	3.4	5.65	4.08
			0.15	65.8	58.8	64.4	58.6	68.69	71.05
			0.31	**98.2**	**94.2**	**97.6**	**95.4**	**97.97**	**98.36**
			0.62	**100**	**100**	**100**	**100**	**100**	**100**

Table Q5: The percentage of statistically significant ($p < .05$) interaction effects for all simulation scenarios with ordinal item scores and skewness generated at the latent variable level. Bold faced cells indicate a power > .80.

N	N/p ratio	Skewness	Interaction	Sum scores		Factor scores		LMS (continuous)		LMS (categorical)	
				a = .87	a = .60	a = .87	a = .60	a = .87	a = .60	a = .87	a = .60
250	18	0	0	5.4	5.8	4.2	3.4	4.6	4.6	4.4	3.8
			0.15	18	14	15.6	12.2	14.6	14.6	15.2	10.8
			0.31	49.8	40.4	47.2	38.2	48.4	48.4	47	35.6
			0.62	**97**	**89.8**	**98.2**	**90**	**97.6**	**97.6**	**96.4**	**84.6**
		2	0	5.6	5.2	5.2	4.4	5.8	5.2	5.6	2.8
			0.15	19.4	17	17	14	17.6	17	17.4	12.4
			0.31	46.4	36.2	44	32.4	46.8	42.4	40	28.4
			0.62	**90**	**82.2**	**88.2**	**79.6**	**89.4**	**86.2**	**84.8**	66.8
		3	0	4.6	4	3.2	2.8	3.4	2.6	3.4	2.8
			0.15	16.2	15.6	12.6	10.4	15.2	19	19.2	13.2
			0.31	44.2	35.6	39.4	28.8	42.6	43	41.8	28.2
			0.62	**85.8**	73.8	**82.6**	71.4	**86.6**	**81.2**	77.8	60
500	36	0	0	6.4	6.6	6	7.4	6.4	6.4	7	6
			0.15	36.8	27.8	37.4	28.6	36.8	36.8	35.2	26.4
			0.31	**83.4**	72.8	**85.6**	73.4	**85.8**	**85.8**	**84.2**	70.4
			0.62	**100**	**99.6**	**100**	**100**	**100**	**100**	**100**	**99.8**
		2	0	4.8	4.4	4	3.8	3.8	3.6	3.6	4
			0.15	33.4	30	33.6	28.8	34	36.8	39.2	31
			0.31	77	66.8	75.6	66.8	77.8	77.8	**77.6**	65.2
			0.62	**99.6**	**98.2**	**99.6**	**98.6**	**99.8**	**99.2**	**99.4**	**98**
		3	0	6	4.2	5.4	3.8	5.2	4.2	4.4	4.2
			0.15	36.6	31.4	35.2	31.8	37.6	40	40.8	36
			0.31	75.2	65	75.4	67.4	79.2	**82.2**	**82.4**	74.2
			0.62	**99.2**	**96.4**	**98.8**	**96.4**	**99.4**	**99**	**98.6**	**94.8**
1000	72	0	0	4.4	5	4.6	4	4.2	4.2	4.4	4
			0.15	56.6	47.6	59.4	49.6	59.4	59.4	58.4	48.2
			0.31	**99.2**	**96.2**	**99.6**	**98.2**	**99.6**	**99.6**	**99.6**	**97.8**
			0.62	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
		2	0	4.6	4.6	5	4.2	5.2	6	6	4.6
			0.15	65.4	56.4	65.2	59.2	65.2	66.8	67.6	66
			0.31	**97.6**	**93**	**97.8**	**94.4**	**97.8**	**98.2**	**98.4**	**97**
			0.62	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
		3	0	4	4.2	3.8	3.8	4	5.4	4.2	4.2
			0.15	58.6	52.2	57.2	52.8	59.2	70.8	73.4	68.8
			0.31	**96**	**92.4**	**96.4**	**93.8**	**97.2**	**98**	**98.4**	**97.2**
			0.62	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**

Table Q6: The percentage of statistically significant (p < .05) interaction effects for ordinal item scores and skewness generated at the item score level. Bold faced cells indicate a power > .80.

N	N/p ratio	Skewne	Interaction	Sum scores		Factor scores		LMS (continuous)		LMS (categorical)	
				a =	a = .60	a = .87	a = .60	a = .87	a = .60	a = .87	a = .60
250	18	0	0	5.4	5.8	4.2	3.4	4.6	3.8	4.4	3.8
			0.15	18	14	15.6	12.2	14.6	10.0	15.2	10.8
			0.31	49.8	40.4	47.2	38.2	48.4	38.6	47	35.6
			0.62	**97**	**89.8**	**98.2**	**90**	**97.6**	**86.4**	**96.4**	**84.6**
		2	0	6.6	7.4	5.6	4.8	5	3.2	3.6	2.6
			0.15	17.4	13	15.4	11.8	16.6	9.2	15.8	11.2
			0.31	48	38.8	43.6	37.4	42.8	34.8	47.4	35.6
			0.62	**92.6**	**86**	**91.8**	**85.2**	**90.2**	**80.0**	**91.2**	**82.2**
		3	0	6	5.8	4.2	4.6	4.8	3.2	3.8	2.6
			0.15	13.6	12.6	10.2	10.4	10	6.2	15	10.4
			0.31	39.2	32.8	35.4	30.4	34.2	23.2	45.6	33.2
			0.62	**87.4**	78.4	**84**	75.4	**80.8**	63.6	**92**	77.8
500	36	0	0	6.4	6.6	6	7.4	6.4	4.4	7	6
			0.15	36.8	27.8	37.4	28.6	36.8	23.8	35.2	26.4
			0.31	**83.4**	72.8	**85.6**	73.4	**85.8**	73.2	**84.2**	70.4
			0.62	**100**	**99.6**	**100**	**100**	**100**	**99.8**	**100**	**99.8**
		2	0	4	3.4	4	2.6	3.8	4	2.8	3.2
			0.15	26.4	22.8	25	23	24.8	21.4	30.4	23.2
			0.31	71.8	66.2	73.4	67.8	72.6	63.8	77.4	70.8
			0.62	**100**	**99.2**	**99.6**	**99.2**	**99.6**	**98.4**	**100**	**100**
		3	0	4.2	4	4.4	3.4	4.8	3.4	6.2	3.6
			0.15	25.6	21.6	25.2	21.2	24.4	20.4	29.8	22.6
			0.31	67	60.8	67.2	62.2	66.8	55.8	**80**	67.8
			0.62	**99.6**	**98.4**	**99.4**	**98.6**	**99.4**	**94.8**	**100**	**99.2**
1000	72	0	0	4.4	5	4.6	4	4.2	3.2	4.4	4
			0.15	56.6	47.6	59.4	49.6	59.4	49.6	58.4	48.2
			0.31	**99.2**	**96.2**	**99.6**	**98.2**	**99.6**	**97.8**	**99.6**	**97.8**
			0.62	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
		2	0	4.8	5.2	3.8	4.8	4	3.4	3	3.6
			0.15	49.6	43.4	50.2	47.4	51.4	42.4	56.8	49.6
			0.31	**96.8**	**92.8**	**96.8**	**94.6**	**96.8**	**94.2**	**98.2**	**96.4**
			0.62	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
		3	0	4.8	4.4	4.6	4.4	4.6	4.8	4.4	3.8
			0.15	42.6	37	41	37	41.2	37.2	54.8	45.4
			0.31	**93.4**	**89.2**	**93**	**90**	**93**	**88**	**98**	**94**
			0.62	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**

Appendix R: Mathematical details behind two temporal stability methods (*chapter 8*)

Cohen's d (for paired data)

Cohen's d can be calculated by standardizing the raw difference in mean scores between the two measurements ($M_2 - M_1$) by a pooled standard deviation for paired data (formula 1; Borenstein, Hedges, Higgins & Rothkins, 2011, page 29), where SD_{T1} and SD_{T2} are the standard deviations of the scores at each measurement and where $r_{(T1,T2)}$ denotes the correlation between the two measurements. Because the numerator is a single difference between average scores, this method only assesses mean-level absolute stability and does therefore not necessarily detect individual-level absolute stability.

$$d = \frac{M_2 - M_1}{\left(\frac{\sqrt{\left(SD_{T_1}^2 + SD_{T_2}^2 - 2 * r_{(T1,T2)} * SD_{T_1} * SD_{T_2}\right)}}{\sqrt{2 * (1 - r_{(T1,T2)})}}\right)} \quad (1)$$

Reliable change index (RCI)

The RCI can be computed using Formula 2 (Christensen, 1986), where D_i denotes the difference between scores on two measurements for individual *i* ($D_i = X_{(T2)i} - X_{(T1)i}$) and S_E the standard error of measurement, which can be calculated using formula 3, where $S_{(T1)}$ indicates the standard deviation of scores at T1 and $r_{(T1,T2)}$ the test-rest reliability.

$$RCI = \frac{D_i}{\sqrt{(2(S_E{}^2))}} \quad (2)$$

$$S_E = S_{(T1)} * \sqrt{\left(1 - r_{(T1,T2)}\right)} \quad (3)$$

The last step involves an interpretation of the RCI's computed for each individual. An RCI larger than 1.96 suggests significant increase, an RCI smaller than -1.96 suggests significant decrease, and an RCI between -1.96 and 1.96 indicates no significant change at a significance level of .05.

Appendix S: Item characteristics (*Chapter 8)*

Table S1: Item baseline characteristics for the DS14 (negative affectivity & social inhibition) and HADS (depression and anxiety) questionnaires.

Construct	Item	n	mean	sd	median	min	max	skew	kurtosis
NA	DS_2	2542	1.51	1.34	2	0	4	0.37	-1.01
	DS_4	2555	0.84	1.13	0	0	4	1.14	0.34
	DS_5	2553	1.12	1.19	1	0	4	0.68	-0.59
	DS_7	2555	0.92	1.17	0	0	4	1.04	0.06
	DS_9	2562	0.65	0.95	0	0	4	1.39	1.28
	DS_12	2562	1.37	1.31	1	0	4	0.5	-0.92
	DS_13	2560	0.76	1.07	0	0	4	1.31	0.86
SI	DS_1	2579	0.99	1.09	1	0	4	0.72	-0.33
	DS_3	2554	1.67	1.29	2	0	4	0.26	-0.86
	DS_6	2563	0.97	1.15	1	0	4	0.92	-0.15
	DS_8	2560	1	1.18	1	0	4	0.9	-0.24
	DS_10	2562	1.14	1.26	1	0	4	0.73	-0.64
	DS_11	2561	1.15	1.18	1	0	4	0.66	-0.54
	DS_14	2564	0.93	1.13	0	0	4	0.96	-0.01
DEP	HADS_2	2572	0.63	0.75	0	0	3	1.02	0.52
	HADS_4	2575	0.43	0.65	0	0	3	1.41	1.59
	HADS_6	2572	0.39	0.7	0	0	3	1.82	2.74
	HADS_8	2572	1.05	0.82	1	0	3	0.63	0.06
	HADS_10	2573	0.56	0.74	0	0	3	1.3	1.37
	HADS_12	2575	0.63	0.79	0	0	3	1.11	0.6
	HADS_14	2582	0.72	0.9	0	0	3	0.96	-0.22
ANX	HADS_1	2571	0.8	0.73	1	0	3	0.82	0.83
	HADS_3	2572	0.71	0.85	0	0	3	0.98	0.08
	HADS_5	2556	0.73	0.8	1	0	3	0.93	0.31
	HADS_7	2574	0.62	0.78	0	0	3	1	0.09
	HADS_9	2570	0.45	0.64	0	0	3	1.37	1.7
	HADS_11	2567	0.9	0.98	1	0	3	0.78	-0.52
	HADS_13	2578	0.43	0.61	0	0	3	1.41	2.26

Appendix T: Additional results longitudinal measurement invariance (*chapter 8*)

Table T1: Difference in item response probabilities between the models with and without the constraint that item threshold parameters are invariant over time.

Time	Response	Item						
Negative affectivity		DS_2	DS_4	DS_5	DS_7	DS_9	DS_12	DS_13
T1	False	0.00	-0.01	0.00	0.00	-0.01	0.01	0.00
	Rather false	0.00	0.01	0.00	0.00	0.00	-0.01	0.00
	Neutral	-0.01	0.00	0.00	0.00	0.00	0.00	0.00
	Rather true	0.01	0.00	0.01	0.02	0.00	0.01	0.01
	True	-0.01	0.00	0.00	-0.01	0.00	-0.01	-0.01
T2	False	-0.01	0.00	0.00	0.01	0.00	0.00	0.00
	Rather false	0.00	0.00	0.00	0.00	0.01	0.00	0.01
	Neutral	0.01	0.01	0.01	0.01	0.00	0.01	0.00
	Rather true	-0.01	0.00	-0.01	-0.01	0.00	0.00	0.00
	True	0.01	-0.01	0.00	0.00	-0.01	0.00	-0.01
T3	False	0.01	0.01	0.00	0.00	0.01	-0.01	0.00
	Rather false	0.00	0.00	0.01	0.00	0.00	0.01	0.01
	Neutral	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Rather true	0.01	0.00	0.01	-0.01	0.00	-0.01	-0.01
	True	-0.01	-0.01	-0.01	0.01	0.00	0.01	0.01
T4	False	0.00	0.01	-0.01	0.00	0.01	-0.01	0.01
	Rather false	0.00	0.00	0.01	0.01	-0.01	0.01	-0.01
	Neutral	0.01	0.01	0.01	0.01	0.00	0.01	0.00
	Rather true	-0.01	-0.01	-0.01	-0.02	-0.01	0.00	-0.01
	True	0.00	0.00	-0.01	0.01	0.00	-0.01	0.01
Social inhibition		DS_1	DS_3	DS_6	DS_8	DS_10	DS_11	DS_14
T1	False	-0.01	-0.01	0.01	0.01	0.00	0.01	0.00
	Rather false	0.01	0.01	-0.01	-0.01	0.00	-0.01	0.00
	Neutral	-0.01	-0.01	0.00	0.00	0.00	0.00	0.00
	Rather true	0.00	0.01	0.01	0.01	0.00	0.01	0.01
	True	0.00	-0.01	-0.01	-0.01	0.01	-0.01	0.00
T2	False	0.00	-0.01	0.00	-0.01	0.00	0.00	0.00
	Rather false	0.00	0.01	0.00	0.01	0.00	0.00	0.00
	Neutral	0.01	0.01	0.01	0.00	0.00	0.01	0.00
	Rather true	0.00	0.00	0.00	-0.01	0.01	-0.01	-0.02
	True	-0.01	0.00	-0.01	0.00	-0.01	0.01	0.01
T3	False	0.01	0.00	0.00	0.00	-0.01	0.00	0.00
	Rather false	-0.01	0.00	0.00	0.00	0.01	0.01	0.00
	Neutral	0.00	0.01	0.00	0.00	0.00	0.00	0.00
	Rather true	0.00	-0.01	-0.01	0.00	0.01	0.00	0.01
	True	-0.01	0.01	0.01	0.00	-0.01	0.00	-0.01

Social inhibition		DS_1	DS_3	DS_6	DS_8	DS_10	DS_11	DS_14
T4	False	0.01	0.01	-0.01	0.00	-0.01	-0.02	0.00
	Rather false	-0.01	-0.02	0.01	0.00	0.01	0.02	0.00
	Neutral	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	Rather true	-0.01	-0.01	-0.03	-0.01	-0.01	0.00	-0.01
	True	0.00	0.00	0.01	0.00	0.00	-0.01	-0.01

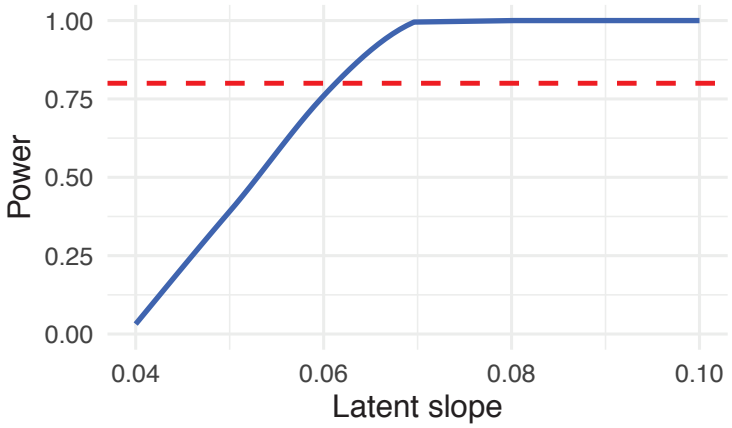
Depression		HADS_2	HADS_4	HADS_6	HADS_8	HADS_10	HADS_12	HADS_14
T1	Not at all	0.00	0.00	0.01	0.00	-0.01	0.00	0.01
	Sometimes	0.00	0.00	-0.01	0.00	0.00	0.00	-0.01
	A lot of the time	0.00	0.01	0.00	0.00	0.01	0.00	0.00
	Most of the time	0.00	0.00	0.00	-0.01	0.00	0.00	0.00
T2	Not at all	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Sometimes	0.00	-0.01	0.00	0.00	-0.01	0.00	-0.01
	A lot of the time	0.00	0.00	0.00	0.00	0.01	-0.01	0.01
	Most of the time	0.01	0.00	0.00	0.00	0.00	0.00	-0.01
T3	Not at all	-0.01	0.01	0.01	-0.01	0.01	0.01	0.00
	Sometimes	0.02	0.01	0.01	0.02	0.02	0.00	0.02
	A lot of the time	0.01	-0.01	-0.01	0.00	-0.01	0.01	0.00
	Most of the time	-0.03	-0.01	-0.01	0.00	-0.02	-0.01	-0.01
T4	Not at all	0.00	0.01	0.01	0.00	0.01	-0.01	-0.01
	Sometimes	0.03	0.04	0.03	0.01	0.04	0.03	0.04
	A lot of the time	0.01	-0.02	-0.02	-0.01	-0.01	-0.01	0.01
	Most of the time	-0.04	-0.02	-0.02	0.00	-0.04	-0.01	-0.04

Anxiety		HADS_1	HADS_3	HADS_5	HADS_7	HADS_9	HADS_11	HADS_13
T1	Not at all	0.01	0.01	-0.01	-0.01	0.00	-0.01	0.00
	Sometimes	-0.01	0.00	0.01	0.00	-0.01	0.01	-0.01
	A lot of the time	0.00	-0.01	0.00	0.01	0.00	0.00	0.00
	Most of the time	0.00	0.00	-0.01	0.00	0.00	0.00	0.00
T2	Not at all	-0.01	0.00	0.01	0.00	-0.02	0.00	0.01
	Sometimes	0.02	0.02	0.00	0.01	0.02	0.01	0.01
	A lot of the time	0.00	0.00	0.00	-0.01	-0.01	0.00	-0.01
	Most of the time	-0.01	-0.01	-0.01	0.00	0.00	-0.01	-0.01
T3	Not at all	-0.02	0.00	0.00	0.02	0.02	-0.01	0.01
	Sometimes	0.04	0.02	0.02	0.01	0.00	0.02	0.03
	A lot of the time	0.01	0.02	-0.01	-0.02	-0.01	0.01	-0.01
	Most of the time	-0.04	-0.04	0.00	-0.01	-0.01	-0.02	-0.03
T4	Not at all	-0.01	0.00	-0.01	0.01	0.02	0.01	0.01
	Sometimes	0.04	0.02	0.03	0.02	0.01	0.01	0.04
	A lot of the time	0.00	0.01	-0.01	0.01	-0.02	0.00	0.00
	Most of the time	-0.03	-0.04	-0.01	-0.03	-0.01	-0.01	-0.05

Appendix U: Power analysis for the univariate latent growth curve model (*Chapter 8*)

We conducted a power analysis for the univariate LGC model to test the null hypothesis that the average latent slope parameter is equal to zero. We used the R-package *simsem* to conduct the power analysis for our LGC model fitted in study 2. We assumed a significance level of 0.05 and investigated the minimal size of the latent slope that could be detected with sufficient power given the current sample size of 2600 participants. The figure below indicates that this number of participants is enough to detect a latent slope of 0.06 or larger with a power of 0.80. A latent linear slope of 0.06 implies that on average participants change 0.18 on the scale of the latent variable over a four-year period. The estimated variance of the latent NA factor varies across repeated measurements between 0.17 and 0.22, corresponding to standard deviations of 0.40 and 0.47. The most conservative estimate is that our study was sufficiently powered to detect a latent slope corresponding to a change in NA of 0.18 / 0.40 = 0.45 standard deviations.

Figure U1*:* Power to detect various values of a mean latent slope effect given N=2600.



NEDERLANDSE SAMENVATTING (Dutch summary)

Introductie

Mensen met een Type D ("Distressed") persoonlijkheid tonen een hoge score op de twee persoonlijkheidstrekken negatieve affectiviteit (NA) en sociale inhibitie (SI). NA drukt de neiging uit tot het ervaren van negatieve gevoelens, emoties en gedachten. Mensen die hoog scoren op SI zijn teruggetrokken en hebben moeite met het uiten van hun gevoelens in sociale situaties. In de medische en psychologische literatuur is Type D gerelateerd aan een groot aantal uitkomstmaten. Zo wordt Type D gezien als een risicofactor voor het optreden van cardiale gebeurtenissen bij mensen met hart- en vaatziekten (Grande et al., 2012).

Er bestaat veel discussie over hoe Type D persoonlijkheid het best kan worden gemodelleerd in statistische analyses (Ferguson et al., 2009; Smith, 2011; Coyne & de Voogd, 2012). De persoonlijkheidstrekken NA en SI kunnen als continue dimensies worden beschouwd, waarop mensen in kleine gradaties van elkaar kunnen verschillen. Echter, het woord "type" suggereert dat het construct Type D persoonlijkheid niet continue is maar categorisch, waarbij ieder persoon wordt ingedeeld in een bepaald persoonlijkheids*type* zoals Type D. In dit proefschrift beargumenteren we dat het Type D construct vooral een pragmatisch middel is om een beschrijving geven van bepaald scorepatroon op de persoonlijkheidstrekken NA en SI. Gegeven dat het construct Type D instrumenteel van aard is, stellen we dat statistische analyses zich primair zouden moeten focussen op NA en SI en de manier waarop de combinatie van hoge scores op deze persoonlijkheidstrekken uitkomstmaten beïnvloedt.

NA en SI worden door de DS14 vragenlijst elk gemeten met zeven vragen op een antwoordschaal van 0 tot 4 (Denollet, 2005). Onderzoekers berekenen vervolgens apart voor NA en SI een totaalscore op basis van de zeven vragenlijstscores, elk met een bereik van 0 tot 28 waarbij een hogere score meer NA of meer SI voorstelt. Deze totaalscores correleren rond de 0.4 en kunnen op verschillende manieren in statistische analyses worden gebruikt. De uitdaging is om de scores op zo'n manier te analyseren dat onderzocht kan worden of deelnemers met hoge scores op beide persoonlijkheidstrekken in het bijzonder risico lopen op negatieve gezondheidsuitkomsten. Type D theorie stelt namelijk dat het hebben van

hoge scores op *beide* persoonlijkheidstrekken een extra risico oplevert, oftewel er vindt synergie plaats tussen NA en SI. Er is sprake van een synergistisch Type D effect wanneer de invloed van beide persoonlijkheidstrekken op een uitkomstmaat sterker wordt met hogere scores van de andere persoonlijkheidstrek over het gehele scorebereik.

Als eerste poging om dit synergistische Type D effect te ontdekken maakt de meest gebruikte *2-groeps methode* eerst een nieuwe variabele aan en geeft alle deelnemers het label "Type D" wanneer ze een score van 10 of hoger hebben op zowel NA en SI. Alle andere deelnemers krijgen het label "Niet Type D". Deze dichotome variabele met twee mogelijke waarden wordt vervolgens in de statistische analyse gerelateerd aan andere variabelen. Deze methode is in eerder onderzoek reeds bekritiseerd omdat deze ook een Type D effect kan aangeven wanneer niet zowel NA en SI, maar slechts één van deze persoonlijkheidstrekken een causale invloed heeft op een uitkomstmaat (Smith, 2011).

Om dit probleem op te lossen is er een variant op de 2-groeps methode ontwikkeld, namelijk de *4-groeps methode*, waarbij wederom een "Type D" groep wordt gebruikt, maar waarbij de "Niet Type D" groep verder wordt opgesplitst in drie groepen: (1) Alleen op NA een score van 10 of hoger, maar op SI een score lager dan 10. (2) Alleen op SI een score van 10 of hoger, maar op NA een score lager dan 10. (3) Op zowel NA en SI een score lager dan 10. De resulterende categorische variabelen met vier mogelijke waarden wordt vervolgens gebruikt in statistische analyses. Een mogelijk nadeel van de 4-groeps methode is echter dat alle individuele verschillen op de persoonlijkheidstrekken worden gereduceerd tot vier mogelijke types, wat ervoor kan zorgen dat deze methode minder gevoelig is om Type D effect te ontdekken in statistische analyses.

De derde *continue interactie methode* verdeelt mensen niet in groepen, maar onderzoekt de totaalscores van NA en SI direct in de statistische analyse, samen met hun interactie effect. Onderzoekers hebben namelijk beargumenteerd dat als Type D persoonlijkheid inhoudt dat mensen hoge scores hebben op zowel NA en SI, dat dit statistisch het best kan worden gemodelleerd met een interactie effect tussen NA en SI (Ferguson et al., 2009; Smith, 2011). Daarnaast houdt deze methode in de statistische analyse ook rekening met alle individuele verschillen tussen mensen op de persoonlijkheidstrekken, waardoor deze methode

sensitiever zou zijn dan de 2-groeps en 4-groeps methodes. Ondanks eerdere kritieken worden de 2-groeps en 4-groeps methode echter nog in veel studies toegepast. Daarom is er onderzoek nodig dat concreet aantoont wat de gevolgen zijn van het toepassen van deze methodes op het onterecht concluderen van Type D effecten.

Een nadeel van alle drie de beschreven methodes is dat ze aannemen dat de berekende NA en SI totaalscores een perfecte meting zijn van de constructen NA en SI. Deze constructen kunnen net als de meeste psychologische constructen niet direct worden geobserveerd en worden daarom latente variabelen genoemd. In de psychologie worden vaak vragenlijsten gebruikt waarin met een aantal vragen getracht wordt individuele verschillen op een latente variabele zoals NA of SI te meten. De vragen in dit soort vragenlijsten kunnen echter verschillen in de mate waarin ze een goede meting zijn van het construct. Tijdens het invullen van de vragenlijst kunnen er ook allerlei factoren anders dan het construct dat gemeten wordt invloed hebben op het antwoord dat iemand geeft. Deze factoren worden meetfout genoemd. Wanneer onderzoekers een totaalscore van vragenlijst antwoorden gebruiken in analyses, dan nemen ze aan dat er geen meetfout is en de vragenlijst dus perfect betrouwbaar is. Deze aanname gaat echter in de praktijk niet op voor de meeste psychologische meetinstrumenten. Daarom is een statistische analyse nodig die individuele verschillen in vragenlijstscores verklaard door de latente variabelen NA en SI kunnen scheiden van de verschillen die voortkomen uit meetfout. Dat kan met latente variabele modellen zoals structurele vergelijkingsmodellen (SEM).

Doel van dit proefschrift

Het doel van dit proefschrift is te onderzoeken welke methodes het best gebruikt kunnen worden om Type D persoonlijkheid te bestuderen in statistische analyses. In **Deel I** hebben we met behulp van een groot aantal computer gegenereerde datasets de mate bestudeerd waarin de 2-groeps, 4-groeps en continue interactie methode in staat zijn om verschillende causale invloeden van NA en SI op een uitkomstmaat te ontdekken. We onderzochten met name of deze methodes correct aangeven dat er geen synergistisch Type D effect is, wanneer dat daadwerkelijk afwezig was in de gesimuleerde data. Omdat bleek dat vergeleken met de continue interactie methode, de 2-groeps en 4-groeps methode niet goed de onderliggende causale mechanismes konden blootleggen, hebben we in **Deel II**

onderzocht wat de impact van deze discrepantie is op de conclusies die getrokken zijn in de Type D literatuur. Ook presenteren we de bevindingen van een internationale samenwerking waarbij we 18 datasets van eerder gepubliceerde artikelen hebben samengevoegd en opnieuw hebben geanalyseerd met de continue interactie methode.

Ten slotte hebben we in **Deel III** onderzocht of latente variabele modellen gebruikt kunnen worden om bij het statistisch analyseren van Type D persoonlijkheid rekening te houden met de meetfout en de vaak scheef verdeelde item scores van de DS14 vragenlijst. Omdat het toetsen van interactie effecten relatief nieuw is binnen de latente variabele theorie, is er nog weinig bekend over hoe methodes die latente interactie effecten schatten presteren bij het gebruik van vragenlijsten zoals de DS14. De methodes om interacties tussen latente variabelen te schatten zijn namelijk ontwikkeld voor vragenlijstscores scores die een groot aantal verschillende waarden kunnen aannemen en normaal verdeeld zijn. De DS14 vragen kennen echter maar 5 mogelijke antwoordopties en deze scores volgen vaak geen normaalverdeling. Met onze computersimulaties hebben we onderzocht welke methode binnen de latente variabele modellen goed in staat is om de interactie effecten tussen latente variabelen zoals NA en SI te ontdekken. Ten slotte hebben we in het laatste hoofdstuk van deel III bestudeerd hoe latente variabele modellen kunnen worden gebruikt om te onderzoeken in welke mate de persoonlijkheidstrekken NA en SI stabiel zijn over tijd. Over een periode van vier jaarlijkse metingen hebben we individuele verandering in deze persoonlijkheidstrekken door middel van latente variabele modellen vergeleken met gelijktijdige metingen van depressie en angst symptomen.

Deel I: Type D persoonlijkheid effecten

In **Deel 1** hebben we bestudeerd in welke mate de meest gebruikt methodes om een synergistisch Type D effect te schatten in staat zijn om dit effect wel of niet te vinden, gegeven verschillende onderliggende causale mechanismes. In **Hoofdstuk 2** hebben we met name gefocust op de beperkingen van de 2-groeps en 4-groeps methode. We hebben de gesimuleerde datasets gevarieerd in de aanwezigheid van een causale invloed van NA *of* SI op een uitkomstmaat. Allereerst vonden we in lijn met eerdere studies over gedichotomiseerde continue variabelen (Cohen, 1983; Royston, Altman & Sauerbrei, 2006), dat de 2-groeps en 4-groeps methode door hun informatie reductie minder gevoelig waren

voor het ontdekken van de onderliggende effecten dan de continue interactiemethode. Verder bleek dat wanneer alleen één van deze persoonlijkheidstrekken een causale invloed heeft op de uitkomstmaat, dat in die situatie de 2-groeps methode, waarbij mensen met Type D vergeleken worden met mensen zonder Type D, ook vaak een significant effect opleverde. Deze methode kan dus geen onderscheid maken tussen een situatie waarin alleen NA of SI, of zowel NA en SI causaal gerelateerd zijn aan de uitkomstmaat. Deze bevinding sluit aan bij eerdere kritiek op deze methode door Smith (2011). De 4-groeps methode was eerder slechts bekritiseerd voor een lagere gevoeligheid om ware effecten te ontdekken en er werd aangenomen dat deze methode wel onderscheid kon maken tussen verschillende causale mechanismes. In **Hoofdstuk 2** toonden we echter aan dat *ook* de 4-groeps methode een significant Type D effect kan tonen wanneer alleen NA of SI causaal gerelateerd is aan de uitkomstmaat. Dit probleem trad echter wel in mindere mate op dan bij de 2-groepsmethode. Het bleek dat door de positieve correlatie tussen NA en SI er regelmatig een effect voor de combinatie van beide persoonlijkheidstrekken werd geobserveerd wanneer in werkelijkheid alleen NA of SI causale invloed had.

Deze resultaten laten zien dat de 2-groeps en 4-groeps methodes beiden niet gebruikt kunnen worden om te ontdekken of de combinatie hoge scores op beide persoonlijkheidstrekken invloed heeft op een uitkomstmaat. In eerdere studies werd beargumenteerd dat Type D onderzoekers de continue interactiemethode zouden moeten gebruiken. Bij het testen van interactie-effecten wordt vaak een vermenigvuldiging van de scores van de twee variabelen gemodelleerd als interactieterm. Het is in de analyse belangrijk om niet alleen de interactieterm maar ook de eerste-orde effecten van de onderliggende variabelen mee te analyseren (Aiken & West, 1991). Daarnaast toont eerder onderzoek aan dat wanneer de interactie wordt onderzocht tussen twee continue variabelen die gecorreleerd zijn (zoals NA en SI), het belangrijk is om de invloed van kwadratische effecten van deze variabelen te onderzoeken (Maxwell & Delaney, 1983; Belzak & Bauer, 2019). Omdat in Type D onderzoek niet standaard aan beide voorwaarden is voldaan hebben we in **Hoofdstuk 3** met gesimuleerde data aangetoond wat de gevolgen zijn van het onjuist toepassen van de continue interactiemethode.

Onze gesimuleerde datasets varieerden over de aanwezigheid van een lineair NA effect, een kwadratisch NA effect, of een interactie-effect tussen NA en SI. Het bleek dat wanneer alleen NA (en dus niet SI) causaal gerelateerd was aan de uitkomstmaat, dat het alleen modelleren van een interactie-effect zonder hoofdeffecten resulteerde in fout positieve interactie-effecten. Ook bleek dat wanneer alleen NA kwadratisch gerelateerd was aan de uitkomstmaat, dat het niet modelleren van een kwadratisch effect dan ook resulteerde in foutpositieve interactie-effecten. Dit probleem trad niet op wanneer de gesimuleerde correlatie tussen NA en SI gelijk was aan 0 en het trad vaker op naarmate de correlatie tussen NA en SI positiever of negatiever werd. Deze problemen traden niet op wanneer de continue interactiemethode juist werd toegepast door de hoofdeffecten, kwadratische effecten en interactie tussen NA en SI te modelleren.

Conclusie deel 1

In deel 1 van dit proefschrift hebben we aan de hand van computersimulaties aangetoond dat de 2-groeps en 4-groeps methodes foutpositieve Type D effecten kunnen opleveren wanneer alleen NA of SI causaal gerelateerd was aan een uitkomstmaat. We adviseren dat deze methodes niet meer gebruikt worden in statistische analyses. Ook bleek dat de continue interactiemethode kan zorgen voor foutpositieve Type D effecten wanneer de eerste-orde effecten van NA en SI niet worden gemodelleerd, of wanneer niet wordt uitgezocht of er een mogelijke invloed is van kwadratische NA en SI effecten.

Deel II: Het herzien van de Type D persoonlijkheid literatuur

Gegeven dat van de bestudeerde methodes de 2-groeps methode het minst in staat is om onderscheid te maken tussen verschillende causale mechanismes, hebben we in **Hoofdstuk 4** onderzocht wat de discrepantie is in de conclusies die in de Type D literatuur zijn getrokken op basis van de 2-groeps methode, vergeleken met de conclusies in diezelfde studies op basis van de continue interactiemethode. We hebben een systematische literatuurstudie uitgevoerd naar alle empirische Type D studies waarin onderzoekers zowel de 2-groeps als de continue interactiemethode hebben gebruikt. We hebben 44 gepubliceerde studies bestudeerd. In deze studies werd op basis van de 2-groeps methodes 89 keer een Type D effect op een uitkomstmaat geconcludeerd. Wanneer deze effecten echter met de continue interactiemethode werden geanalyseerd dan bleek slechts 41.6%

een Type D effect te tonen. Verder bleek 47.2% van de significante 2-groeps effecten slechts een effect van NA *of* SI te zijn volgens de continue methode. Dit toont aan dat mogelijk de helft van de gepubliceerde studies waarin een Type D effect wordt geconcludeerd op basis van de 2-groeps methode, waarschijnlijk slechts effecten van NA *of* SI zijn. Dit suggereert dat in veel studies niet de combinatie van NA en SI voorspellend is en daardoor het construct Type D niets toevoegt boven de ene persoonlijkheidstrek die wel invloed had. Het blijft echter onduidelijk voor welke studies in de literatuur dit geldt, waardoor het belangrijk is om studies in de Type D literatuur opnieuw te analyseren met de continue interactiemethode.

Een eerste aanzet hiertoe is gedaan in **Hoofdstuk 5**, waarin we een meta-analyse hebben uitgevoerd over individuele patiënt data voortkomend uit 18 prospectieve studies naar het verband tussen Type D persoonlijkheid en het optreden van negatieve gezondheids-gebeurtenissen bij patiënten met hart en vaatziekten. Uit een eerdere meta-analyse bleek dat patiënten met Type D een grotere risico lopen op dergelijke gebeurtenissen dan patiënten zonder Type D persoonlijkheid (Grande et al, 2012). Een nadeel van deze meta-analyse is echter dat de geïncludeerde studies uitsluitend gebruik hebben gemaakt van de 2-groeps methode, welke het onderliggende causale mechanisme dus niet kan blootleggen. In onze meta-analyse hebben we data van 11151 patiënten met hart- en vaatziekten geanalyseerd. Het bleek dat de significante interactie tussen NA en SI voorspellend bleek voor het optreden van negatieve gezondheidsgebeurtenissen gedurende een gemiddelde volgtijd van vier jaar. Dit effect werd niet verklaard door een kwadratische invloed van NA of SI op deze gebeurtenissen. In een vervolganalyse werd het effect wel gevonden bij mensen met kransslagaderziekte, maar niet bij mensen met hartfalen. Wat betreft mortaliteit bleek niet de interactie tussen NA en SI, maar alleen NA het risico te verhogen op zowel cardiale mortaliteit als mortaliteit door andere oorzaken.

Conclusie deel 2

In deel 2 van dit proefschrift hebben we aangetoond dat er een grote discrepantie is in de gepubliceerde Type D literatuur tussen de conclusies getrokken op basis van de 2-groeps methode en de continue interactiemethode. Mogelijk de helft van alle gepubliceerde Type D effecten op basis van de 2-groepsmethode blijkt een effect van NA *of* SI te zijn volgens de continue interactiemethode. Dit toont het belang aan van het opnieuw analyseren van de

Type D literatuur met de correct gespecificeerde continue interactiemethode. We hebben laten zien dat met deze methode nog steeds een verband werd gevonden tussen Type D persoonlijkheid en negatieve gezondheidsgebeurtenissen bij mensen met kransslagaderziekte. De link tussen Type D en mortaliteit die op basis van de 2-groeps methode werd gevonden bleek slechts een effect van NA te zijn.

Deel III: Een latente variabele model van Type D persoonlijkheid

Een nadeel van de 2-groeps, 4-groeps en continue interactiemethodes zoals ze tot nu toe in dit proefschrift zijn gebruikt is dat ze aannemen dat de berekende NA en SI totaalscores een perfecte meting zijn van de constructen NA en SI. Deze aanname is vaak niet realistisch aangezien vragenlijsten in de psychologie vaak meetfout bevatten. Methodes die dit probleem negeren ontdekken vaak zwakkere effecten dan er in werkelijkheid bestaan tussen de constructen (Spearman, 1904). Daarnaast is vaak geen rekening gehouden met de scheve verdelingen van de item scores die NA en SI meten. Zowel de scheve scoreverdelingen als de meetfout kan worden aangepakt met latente variabele modellen zoals structurele vergelijkingsmodellen (SEM). Het doel van **Deel III** van dit proefschrift is daarom om uit te zoeken in welke mate Type D goed geanalyseerd kan worden latente variabele modellen.

In **Hoofdstuk 6** hebben we onderzocht of de continue interactiemethode in SEM tot accuratere schattingen van een Type D effect leidt dan de klassieke continue interactiemethode gebaseerd op een lineaire regressie van totaalscores. Daarnaast hebben we binnen de techniek SEM een aantal methodes vergeleken die gebruikt kunnen worden om een interactie-effect tussen latente variabelen te schatten. Het was namelijk nog onduidelijk of deze methodes goede schattingen opleveren wanneer de vragenlijstscores zoals bij de DS14 niet normaal verdeeld zijn en slechts 5 antwoordopties kennen. We hebben een groot aantal datasets gesimuleerd waarbij we varieerden over de steekproefgrootte, de grootte van het Type D effect op een continue uitkomstmaat, de mate waarin de constructen NA en SI normaal verdeeld waren, en of ze met slechts vijf of talloze antwoordopties werden gemeten. Elk van deze datasets werden geanalyseerd met de traditionele continue interactie tussen totaalscores, en met drie SEM methodes: (1) *LMS;* (2) *Single product indicator;* (3) *Matched product indicators.* Zoals verwacht bleek dat de traditionele continue interactiemethode zorgde voor een onderschatting van het Type D

effect ten opzichte van de SEM methodes. Binnen de SEM methodes presteerde de single product indicator methode het slechtst en zorgde de LMS methode voor de meest accurate schattingen van de interactie tussen NA en SI. Wanneer de persoonlijkheidstrekken niet normaal waren verdeeld toonde LMS wel foutpositieve Type D effecten. Dit probleem trad het minst op bij de matched product indicators methode.

Een nadeel van **Hoofdstuk 6** is dat alleen is gefocust op continue latente variabelen als uitkomstmaat. Het is nog onduidelijk of deze resultaten gegeneraliseerd kunnen worden naar een situatie waarin de uitkomstmaat dichotoom is en direct geobserveerd. Een voorbeeld van zo'n uitkomstmaat in Type D onderzoek is het wel of niet optreden van bepaalde hartproblemen. In **Hoofdstuk 7** werd daarom voor dit soort uitkomstmaten de mate vergeleken waarin vier continue interactiemethodes accurate schattingen van de interactie tussen NA en SI opleverden: (1) de continue interactie tussen totaalscores; (2) continue interactie tussen factor scores; (3) LMS methode in een continue SEM; (4) LMS methode in een categorische SEM. Wederom werden datasets gesimuleerd onder verschillende omstandigheden en werd elk van deze datasets geanalyseerd met de te vergelijken methodes. Zoals verwacht bleek dat de continue interactie tussen totaalscores zorgde voor een onderschatting van het Type D effect. Ook de interactie tussen geschatte factor scores toonde deze onderschatting. Beide SEM LMS methodes zorgden voor de meest accurate schattingen van de interactie tussen NA en SI wanneer de vragenlijstscores normaal verdeeld waren. Bij niet normaal verdeelde scores toonde categorische SEM de best schattingen.

Naast het gebruik van latente variabele modellen in onderzoek waarin interactie-effecten worden geschat, kunnen deze modellen ook worden toegepast in andere situaties. In Type D onderzoek is vaak bestudeerd of de onderliggende persoonlijkheidstrekken NA en SI stabiel blijven wanneer deelnemers meerdere malen worden gemeten over tijd. In **Hoofdstuk 8** hebben we een overzicht gegeven van de methodes die hier doorgaans voor zijn gebruikt in het Type D veld. Een nadeel van deze methodes is dat ze geen rekening houden met meetfout in de vragenlijstscores. Dit kan zorgen voor een onderschatting van de sterkte van verandering over tijd en dus mogelijk onterecht tot de conclusie dat er geen verandering is in persoonlijkheid over tijd. In **Hoofdstuk 8** hebben we daarom ook latente variabelen

modellen gebruikt om te onderzoeken of NA en SI stabiel zijn over een periode van vier jaar. Ook onderzochten we in hoeverre deze stabiliteit of verandering ook optrad bij de gerelateerde constructen depressie en angst. NA, SI, depressie en angst zijn gedurende deze periode viermaal gemeten bij 2625 mensen die darmkanker overleefd hebben. Allereerst hebben we vastgesteld dat de psychometrische kenmerken van de meetinstrumenten stabiel bleven over tijd. Hierdoor konden we aannemen dat eventuele verandering over tijd daadwerkelijk veroorzaakt werden door de latente constructen en niet door veranderingen in de manier waarop deelnemers de vragenlijst interpreteren. Vervolgens hebben we aangetoond dat de vier psychologische constructen gemiddeld genomen stabiel bleven gedurende de vier jaar. Op individueel niveau waren er significante verschillen tussen deelnemers in of ze verandering toonden over tijd. Verder bleek dat veranderingen in NA samenhingen met veranderingen in depressie en angst. Dit suggereert dat NA niet een volledig stabiele persoonlijkheidstrek is, maar enigszins samenhangt met veranderingen in depressie en angst. Ten slotte hebben we ook laten zien dat van alle vier de constructen, NA de minste stabiliteit toont (74%), gevolgd door angst (76%), depressie (78%) en SI (83%).

Conclusie deel 3

In **Deel 3** hebben we laten zien hoe latente variabele modellen kunnen worden toegepast in Type D onderzoek, niet alleen om een synergistisch Type D effect door middel van de latente interactie tussen NA en SI op een accurate manier te schatten, maar ook om te onderzoeken of NA en SI stabiel zijn over tijd. Op basis van onze simulaties raden we Type D onderzoekers aan om latente variabelen modellen te gebruiken mits een steekproef van voldoende grote beschikbaar is.

Algemene conclusie

Het doel van dit proefschrift was te onderzoeken welke methodes het best gebruikt kunnen worden om Type D persoonlijkheid te bestuderen in statistische analyses. Op basis van meerdere computersimulaties hebben we laten zien dat Type D effecten het best geanalyseerd kunnen worden met een continue interactiemethode. Ook hebben we aangetoond dat deze methode het best kan worden toegepast in een categorische SEM, om zo rekening te houden met zowel meetfout en de niet normaal verdeelde vragenlijstscores. Verder bleek dat de tot nu toe gepubliceerde Type D effecten bijna uitsluitend zijn

gebaseerd op de 2-groeps en 4-groeps methode, of een onjuist gespecificeerd continue interactiemethode. Deze methodes bleken slecht in staat te zijn om de onderliggende causale mechanismes bloot te leggen. Dit geeft het belang aan van het opnieuw analyseren van de Type D literatuur met een juist gespecificeerde continue interactiemethode. In een grote internationale samenwerking hebben we aangetoond dat op basis van deze methode Type D effect nog steeds een significante risicofactor is op negatieve gezondheidsuitkomsten bij patiënten met hart- en vaatziekten. We hopen dat dit eerste grootschalige heranalyse project andere onderzoekers zal inspireren soortgelijke projecten uit te voeren.

Onze bevindingen hebben ook implicaties voor onderzoeken die niet gaan over Type D persoonlijkheid. De literatuur kent voorbeelden van studies die net als Type D onderzoek een synergistisch effect hebben onderzocht met een 2-groeps, 4-groeps, of incorrect gespecifieerde continue interactiemethode. Toekomstige studies zouden kunnen uitzoeken in welke mate de onderliggende causale mechanismes in andere onderzoeksvelden ook niet adequaat zijn vastgesteld. Vervolgens zouden deze studies opnieuw kunnen worden geanalyseerd met een correct gespecificeerde continue interactiemethode.

REFERENCES

Agresti A. (2018). An introduction to categorical data analysis. New York: Wiley.

Aiken, L. S., West, S. G., & Reno, R. R. (1991). *Multiple regression: Testing and interpreting interactions.* Sage.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723. https://doi.org/10. 1109/TAC.1974.1100705

Albert, M. A., Glynn, R. J., Buring, J., & Ridker, P. M. (2006). Impact of traditional and novel risk factors on the relationship between socioeconomic status and incident cardiovascular events. *Circulation, 114(24),* 2619-2626.

Albus, C., Beutel, M. E., Deter, H. C., Fritzsche, K., Hellmich, M., Jordan, J., ... & Herrmann-Lingen, C. (2011). A stepwise psychotherapy intervention for reducing risk in coronary artery disease (SPIRR-CAD)—rationale and design of a multicenter, randomized trial in depressed patients with CAD. *Journal of psychosomatic research*, *71*(4), 215-222.

Alçelik, A., Yildirim, O., Canan, F., Eroglu, M., Aktas, G., & Savli, H. (2012). A preliminary psychometric evaluation of the type D personality construct in Turkish hemodialysis patients. *Journal of Mood Disorders, 2(1),* 1.

Allen, S. F., Wetherell, M. A., & Smith, M. A. (2019). A one-year prospective investigation of Type D personality and self-reported physical health. *Psychology & health, 34(7),* 773-795.

Aluja, A., Malas, O., Lucas, I., Worner, F., & Bascompte, R. (2019). Assessment of the Type D personality distress in coronary heart disease patients and healthy subjects in Spain. *Personality and Individual Differences, 142,* 301-309.

American Heart Association (2008). Heart disease and stroke statistics – 2008 update. A *American journal of psychology, 15(1),* 72-101.

Asendorpf, J. B., & Denissen, J. J. (2006). Predictive validity of personality types versus personality dimensions from early childhood to adulthood: Implications for the distinction between core and surface traits. *Merrill-Palmer Quarterly (2006),* 486-513.

Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J., Fiedler, K., ... & Perugini, M. (2013). Recommendations for increasing replicability in psychology. *European journal of personality, 27(2),* 108-119.

Asparouhov, T., & Muthén, B. (2010a). *Multiple imputation with Mplus*. MPlus Web Notes.

Asparouhov, T., & Muthén, B. (2010b). Simple second order chi-square correction. Mplus Technical Appendix.

Babakus, E., Ferguson, C. E., Jr., & Jo€reskog, K. G. (1987). The sensitivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distributional assumptions. Journal of Marketing Research, 37, 222–228. doi:10.2307/3151512

Bagherian, R., & Ehsan, H. B. (2011). Psychometric properties of the Persian version of type D personality scale (DS14). *Iranian journal of psychiatry and behavioral sciences, 5(2),* 12.

Baltes, P. B. (1987). Theoretical propositions of life-span developmental psychology: On the dynamics between growth and decline. *Developmental psychology, 23(5),* 611.

Barefoot, J. C., Peterson, B. L., Harrell Jr, F. E., Hlatky, M. A., Pryor, D. B., Haney, T. L., ... & Williams Jr, R. B. (1989). Type A behavior and survival: a follow-up study of 1,467

patients with coronary artery disease. *The American journal of cardiology, 64(8),* 427-432.

Barth, J., Schneider, S., & Von Känel, R. (2010). Lack of social support in the etiology and the prognosis of coronary heart disease: a systematic review and meta-analysis. *Psychosomatic medicine, 72(3),* 229-238.

Bartko, J. J. (1976). On various intraclass correlation reliability coefficients. Psychological bulletin, 83(5), 762.

Bartlett, M. S. (1937). The statistical conception of mental factors. *British Journal of Psychology. General Section, 28*, 97–104. :doi: 10.1111/j.2044- 8295.1937.tb00863.x

Batselé, E., Denollet, J., Lussier, A., Loas, G., Vanden Eynde, S., Van de Borne, P., & Fantini-Hauwel, C. (2017). Type D personality: application of DS14 French version in general and clinical populations. *Journal of health psychology, 22(8),* 1075-1083.

Bauman, A. E., Sallis, J. F., Dzewaltowski, D. A., & Owen, N. (2002). Toward a better understanding of the influences on physical activity: the role of determinants, correlates, causal variables, mediators, moderators, and confounders. *American journal of preventive medicine, 23(2),* 5-14.

Bekendam, M. T., Kop, W. J., Barzilay, S., Widdershoven, J. W., Aarnoudse, W., Denollet, J., & Mommersteeg, P. M. (2018). The predictive value of positive affect and Type D personality for adverse cardiovascular clinical outcomes in patients with non-obstructive coronary artery disease. *Journal of psychosomatic research, 104,* 108-114.

Belzak, W. C., & Bauer, D. J. (2019). Interaction effects may actually be nonlinear effects in disguise: A review of the problem and potential solutions. *Addictive Behaviors, 94*, 99–108. https://doi.org/10.1016/j.addbeh.2018.09.018.

Bem, S. L. (1981). Gender schema theory: A cognitive account of sex typing. Psychological Review, 88(4), 354–364.

Beutel, M. E., Wiltink, J., Till, Y., Wild, P. S., Münzel, T., Ojeda, F. M., ... & Michal, M. (2012). Type D personality as a cardiovascular risk marker in the general population: results from the Gutenberg health study. *Psychotherapy and psychosomatics, 81(2),* 108-117.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397–479). Reading, MA: Addison- Wesley.

Bjelland, I., Dahl, A. A., Haug, T. T., & Neckelmann, D. (2002). The validity of the Hospital Anxiety and Depression Scale: an updated literature review. *Journal of psychosomatic research, 52(2),* 69-77.

Blount, R. L., Davis, N., Powers, S. W., & Roberts, M. C. (1991). The influence of environmental factors and coping style on children's coping and distress. *Clinical Psychology Review, 11(1),* 93-116.

Bollen, K. A. (1989). *Structural equations with latent variables*. John Wiley & Sons.

Bollen, K. A. (2005). Structural equation models. *Encyclopedia of biostatistics, 7.*

Bollen, K. A., & Paxton, P. (1998). Two stage least squares estimation of interaction effects. In R. E. Schumacker & G. A. Marcoulides (Eds.), *Interaction and nonlinear effects in structural equation modeling* (pp. 125–152). Psychology Press.

Bollen, K. A., & Stine, R. A. (1993). Bootstrapping goodness-of-fit measures in structural equation models. Sociological Methods & Research, 21(2), 205–229. doi: 10.1177/0049124192021002004

Booth, L., & Williams, L. (2015). Type D personality and dietary intake: The mediating effects of coping style. *Journal of Health Psychology, 20(6),* 921-927.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. John Wiley & Sons.

Borkoles, E., Kaiseler, M., Evans, A., Ski, C. F., Thompson, D. R., & Polman, R. C. (2018). Type D personality, stress, coping and performance on a novel sport task. *PloS one, 13(4),* e0196692.

Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2003). The theoretical status of latent variables. Psychological Review, 110(2), 203. doi:10.1037/0033-295X. 110.2.203

Bouwens, E., van Lier, F., Rouwet, E. V., Verhagen, H. J., Stolker, R. J., & Hoeks, S. E. (2019). Type D personality and health-related quality of life in vascular surgery patients. *International Journal of Behavioral Medicine, 26*(4), 343–351. https://doi.org/10.1007/s12529-018-09762-3.

Brandt, H., Cambria, J., & Kelava, A. (2018). An adaptive Bayesian lasso approach with spike-and-slab priors to identify multiple linear and nonlinear effects in structural equation models*. Structural Equation Modeling: A Multidisciplinary Journal, 25(6),* 946-960.

Briley, D., & Tucker-Drob, E. (2014). Genetic and environmental continuity in personality development: a meta-analysis. *Psychological bulletin, 140 (5),* 1303-1331.

Brosschot, J. F., & Thayer, J. F. (1998). Anger inhibition, cardiovascular recovery, and vagal function: A model of the link between hostility and cardiovascular disease. *Annals of Behavioral Medicine, 20(4),* 326-332.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. British Journal of Mathematical and Statistical Psychology, 37(1), 62–83. doi:10.1111/j.2044-8317.1984.tb00789.x

Bundgaard, J. S., Østergaard, L., Gislason, G., Thune, J. J., Nielsen, J. C., Haarbo, J., ... & Mogensen, U. M. (2019). Association between Type D personality and outcomes in patients with non-ischemic heart failure. *Quality of Life Research*, *28*(11), 2901-2908.

Bunevicius, A., Brozaitiene, J., Staniute, M., Gelziniene, V., Duoneliene, I., Pop, V. J., ... & Denollet, J. (2014). Decreased physical effort, fatigue, and mental distress in patients with coronary artery disease: importance of personality-related differences. *International journal of behavioral medicine*, *21*(2), 240-247.

Bunevicius, A., Staniute, M., Brozaitiene, J., Stropute, D., Bunevicius, R., & Denollet, J. (2013). Type D (distressed) personality and its assessment with the DS14 in Lithuanian patients with coronary artery disease. *Journal of health psychology, 18(9),* 1242-1251.

Burke, D. L., Ensor, J., & Riley, R. D. (2017). Meta-analysis using individual participant data: one-stage and two-stage approaches, and why they may differ. *Statistics in medicine, 36(5),* 855-875.

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of statistical software*, *80*(1), 1-28.

Busemeyer, J. R., & Jones, L. E. (1983). Analysis of multiplicative combi- nation rules when the causal variables are measured with error. *Psychological Bulletin*, *93*, 549–562. https://doi.org/10.1037/0033- 2909.93.3.549

Cadarso-Suárez, C., Roca-Pardiñas, J., & Figueiras, A. (2006). Effect measures in non-parametric regression with interactions between continuous exposures. *Statistics in medicine, 25(4),* 603-621.

Canty, A., & Ripley, B. D. (2019). *boot: Bootstrap R (S-Plus) functions* (R package version 1.3-24). https://cran.r-project.org/web/packages/ boot/citation.html

Caspi, A., Roberts, B. W., & Shiner, R. L. (2005). Personality development: Stability and change. *Annual Review of Psychology, 56,* 453-484.

Cham, H., West, S. G., Ma, Y., & Aiken, L. S. (2012). Estimating latent variable interactions with nonnormal observed data: A comparison of four approaches. *Multivariate Behavioral Research*, *47*, 840–876. https://doi.org/10.1080/00273171.2012.732901

Chapman, B. P., Roberts, B., & Duberstein, P. (2011). Personality and longevity: knowns, unknowns, and implications for public health and personalized medicine. *Journal of aging research*, *2011*.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14(3),* 464-504. http://dx.doi.org/10.1080/10705510701301834

Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008). An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. Sociological methods & research, 36(4), 462–494.

Chida, Y., & Steptoe, A. (2009). The association of anger and hostility with future coronary heart disease: a meta-analytic review of prospective evidence. *Journal of the American college of cardiology, 53(11),* 936-946.

Cho, G. J., & Kang, J. (2017). Type D personality and post-traumatic stress disorder symptoms among intensive care unit nurses: The mediating effect of resilience*. PloS one, 12(4),* e0175067.

Christensen, L. (1986). A method of assessing change in a single subject: An alteration of the RC index. *Behav Ther*, *17*, 305-308.

Cohen, J. (1983). The cost of dichotomization. *Applied psychological measurement, 7(3),* 249-253.

Cole, D. A., & Preacher, K. J. (2014). Manifest variable path analysis: Potentially serious and misleading consequences due to uncorrected measurement error. Psychological Methods, 19(2), 300. doi:10.1037/a0033805

Cole, D. A., Martin, N. C., & Steiger, J. H. (2005). Empirical and conceptual problems with longitudinal trait-state models: introducing a trait-state-occasion model. *Psychological methods, 10(1),* 3.

Collins, L. M., & Lanza, S. T. (2009). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences* (Vol. 718). John Wiley & Sons.

Compare, A., Bigi, R., Orrego, P. S., Proietti, R., Grossi, E., & Steptoe, A. (2013). Type D personality is associated with the development of stress cardiomyopathy following emotional triggers. *Annals of Behavioral Medicine, 45(3),* 299-307.

Compare, A., Mommersteeg, P. M., Faletra, F., Grossi, E., Pasotti, E., Moccetti, T., & Auricchio, A. (2014). Personality traits, cardiac risk factors, and their association with presence and severity of coronary artery plaque in people with no history of cardiovascular disease. *Journal of Cardiovascular Medicine, 15(5)*, 423-430.

Condén, E., Ekselius, L., & Åslund, C. (2013). Type D personality is associated with sleep problems in adolescents. Results from a population-based cohort study of Swedish adolescents. *Journal of psychosomatic research, 74(4),* 290-295.

Condén, E., Leppert, J., Ekselius, L., & Åslund, C. (2013). Type D personality is a risk factor for psychosomatic symptoms and musculoskeletal pain among adolescents: a cross-

sectional study of a large population-based cohort of Swedish adolescents. *BMC pediatrics, 13(1)*, 11.

Condén, E., Rosenblad, A., Ekselius, L., & Åslund, C. (2014). Prevalence of Type D personality and factorial and temporal stability of the DS 14 after myocardial infarction in a Swedish population. *Scandinavian Journal of Psychology, 55(6),* 601-610.

Condén, E., Rosenblad, A., Wagner, P., Leppert, J., Ekselius, L., & Åslund, C. (2017). Is type D personality an independent risk factor for recurrent myocardial infarction or all-cause mortality in post-acute myocardial infarction patients?. *European journal of preventive cardiology*, *24*(5), 522-533.

Conraads, V. M., Denollet, J., De Clerck, L. S., Stevens, W. J., Bridts, C., & Vrints, C. J. (2006). Type D personality is associated with increased levels of tumour necrosis factor (TNF)-α and TNF-α receptors in chronic heart failure. *International journal of cardiology, 113(1),* 34-38.

Coyne, J. C., & de Voogd, J. N. (2012). Are we witnessing the decline effect in the type D personality literature? What can be learned? *Journal of Psychosomatic Research, 73*(6), 401–407. https://doi.org/10.1016/j.jpsychores.2012.09.016.

Coyne, J. C., Jaarsma, T., Luttik, M. L., van Sonderen, E., van Veldhuisen, D. J., & Sanderman, R. (2011). Lack of prog- nostic value of Type D personality for mortality in a large sample of heart failure patients. *Psychosomatic Medicine, 73(7),* 557–562. doi:10.1097/PSY.0b013e318227ac75

Croon, M. (2002). Using predicted latent scores in general latent structure models. In G. Marcoulides & I. Moustaki (Eds.), *Latent variable and latent structure modeling* (pp. 195–223).Mahwah, NJ: Erlbaum

Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve frequently asked questions about growth curve modeling. *Journal of cognition and development, 11(2),* 121-136.

Damen, N. L., Versteeg, H., van Helmondt, S. J., de Jaegere, P. P., van Geuns, R. J. M., Meine, M. M., ... & Pedersen, S. S. (2014). The distressed (Type D) personality mediates the relationship between remembered parenting and psychological distress in cardiac patients. *Psychology & health, 29(3),* 318-333.

Dannemann, S., Matschke, K., Einsle, F., Smucker, M. R., Zimmermann, K., Joraschky, P., ... & Köllner, V. (2010). Is type-D a stable construct? An examination of type-D personality in patients before and after cardiac surgery. *Journal of psychosomatic research, 69(2),* 101-109.

De Boeck, P., Wilson, M., & Acton, G. S. (2005). A conceptual and psychometric framework for distinguishing categories and dimensions. *Psychological Review, 112(1),* 129.

De Fruyt, F., & Denollet, J. (2002). Type D personality: A five-factor model perspective. Psychology and Health, 17(5), 671–683. doi:10.1080/08870440290025858

De Fruyt, F., Bartels, M., Van Leeuwen, K. G., De Clercq, B., Decuyper, M., & Mervielde, I. (2006). Five types of personality continuity in childhood and adolescence. *Journal of Personality and Social Psychology, 91(3),* 538.

De Vroege, L., de Heer, E., van der Thiel, E., van den Broek, K., Van Eck Van Der Sluijs, J., & van der Feltz-Cornelis, C. (2019). Type D personality, concomitant depressive and anxiety disorders, and treatment outcome in Somatic Symptom and Related Disorders: an observational longitudinal cohort study. *Frontiers in psychiatry, 10,* 417.

Dehghani, F. (2018). Type D personality and life satisfaction: The mediating role of social support. *Personality and Individual Differences, 134*, 75–80. https://doi.org/10.1016/j. paid.2018.06.005.

Denollet, J. (1993). Biobehavioral research on coronary heart disease: where is the person?. *Journal of behavioral medicine, 16(2),* 115-141.

Denollet, J. (1998). Personality and coronary heart disease: the type-D scale-16 (DS16). *Annals of Behavioral Medicine, 20(3),* 209-215.

Denollet, J. (2000). Type D personality: A potential risk factor refined. *Journal of psychosomatic research*, *49*(4), 255-266.

Denollet, J. (2005). DS14: standard assessment of negative affectivity, social inhibition, and Type D personality. *Psychosomatic medicine, 67(1),* 89-97. http://dx.doi.org/10.1097/01.psy.0000149256.81953.49

Denollet, J., Holmes, R. V., Vrints, C. J., & Conraads, V. M. (2007). Unfavorable outcome of heart transplantation in recipients with type D personality. *The Journal of heart and lung transplantation*, *26*(2), 152-158.

Denollet, J., Martens, E. J., Smith, O. R., & Burg, M. M. (2010). Efficient assessment of depressive symptoms and their prognostic value in myocardial infarction patients. *Journal of Affective Disorders*, *120*, 105–111. https://doi.org/http://dx.doi.org/10.1016/j. jad.2009.04.013

Denollet, J., & Pedersen, S. S. (2008). Prognostic value of Type D personality compared with depressive symptoms. *Archives of Internal Medicine*, *168*(4), 431-432.

Denollet, J., Pedersen, S. S., Ong, A. T., Erdman, R. A., Serruys, P. W., & van Domburg, R. T. (2006a). Social inhibition modulates the effect of negative emotions on cardiac prognosis following percutaneous coronary intervention in the drug-eluting stent era. *European Heart Journal, 27*(2), 171–177. https://doi.org/10.1093/eurheartj/ehi616.

Denollet, J., Pedersen, S. S., Vrints, C. J., & Conraads, V. M. (2006). Usefulness of type D personality in predicting five-year cardiac events above and beyond concurrent symptoms of stress in patients with coronary heart disease. *The American journal of cardiology, 97(7),* 970-973.

Denollet, J., Pedersen, S. S., Vrints, C. J., & Conraads, V. M. (2013b). Predictive value of social inhibition and negative affectivity for cardi- ovascular events and mortality in patients with coronary artery disease: The type D personality construct. *Psychosomatic Medicine*, *75*, 873–881. https://doi.org/10.1016/S0140-6736(96)90007-0

Denollet, J., Rombouts, H., Gillebert, T. C., Brutsaert, D. L., Sys, S. U., & Stroobant, N. (1996). Personality as independent predictor of long-term mortality in patients with coronary heart disease. *The Lancet, 347*(8999), 417–421. https://doi.org/10.1016/ S0140-6736(96)90007-0.

Denollet, J., Schiffer, A. A., Kwaijtaal, M., Hooijkaas, H., Hendriks, E. H., Widdershoven, J. W., & Kupper, N. (2009). Usefulness of Type D personality and kidney dysfunction as predictors of interpatient variability in inflammatory activation in chronic heart failure. *The American Journal of Cardiology*, *103*(3), 399-404.

Denollet, J., Schiffer, A. A., & Spek, V. (2010). A general propensity to psychological distress affects cardiovascular outcomes: Evidence from research on the Type D (distressed) personality profile. *Circulation: Cardiovascular Quality and Outcomes*, *3*, 546–557. https://doi.org/10. 1161/CIRCOUTCOMES.109.934406

Denollet, J., Sys, S. U., & Brutsaert, D. L. (1995). Personality and mortality after myo- cardial infarction. *Psychosomatic Medicine, 57*(6), 582–591. https://doi.org/10.1097/ 00006842-199511000-00011.

Denollet, J., Rombouts, H., Gillebert, T. C., Brutsaert, D. L., Sys, S. U., & Stroobant, N. (1996). Personality as independent predictor of long-term mortality in patients with coronary heart disease. *The Lancet*, *347*(8999), 417-421.

Denollet, J., Tekle, F. B., Pedersen, S. S., van der Voort, P. H., Alings, M., & van den Broek, K. C. (2013a). Prognostic importance of distressed (Type D) personality and shocks in patients with an implantable cardioverter defibrillator. *International journal of cardiology*, *167*(6), 2705-2709.

Denollet, J., Vaes, J., & Brutsaert, D. L. (2000). Inadequate response to treatment in coronary heart disease: adverse effects of type D personality and younger age on 5-year prognosis and quality of life. *circulation*, *102*(6), 630-635.

Denollet, J., van Felius, R. A., Lodder, P., Mommersteeg, P. M., Goovaerts, I., Possemiers, N., Vanhees, L., Beckers, P., Pattyn, N., & Van Craenenbroeck, E. M. (2018). Predictive value of Type D personality for impaired endothelial function in patients with coronary artery disease. *International Journal of Cardiology*, *259*, 205–210. https://doi.org/10.1016/j.ijcard.2018.02.064

Devlieger, I., & Rosseel, Y. (2017). Factor score path analysis. *Methodology*, *13*, 31–38. https://doi.org/10.1027/1614-2241/a000130 Devlieger, I., Talloen, W., & Rosseel, Y. (2019). New developments in factor score regression: Fit indices and a model comparison test. *Educational and Psychological Measurement*, *79*, 1017–1037. https:// doi.org/10.1177/0013164419844552

Devlieger, I., Mayer, A., & Rosseel, Y. (2016). Hypothesis testing using factor score regression: A comparison of four methods. *Educational and Psychological Measurement*, *76*, 741–770. https://doi.org/10.1177/ 0013164415607618

D'haenens, E., Van Damme, J., & Onghena, P. (2010). Multilevel exploratory factor analysis: Illustrating its surplus value in educational effectiveness research. School Effectiveness and School Improvement, 21(2), 209-235.

Dohrenwend, B. S., & Dohrenwend, B. P. (1974). *Stressful life events: Their nature and effects.* John Wiley & Sons.

Dolan, C. V. (1994). Factor analysis of variables with 2, 3, 5 and 7 response categories: A comparison of categorical variable estimators using simulated data. British Journal of Mathematical and Statistical Psychology, 47(2), 309–326. doi:10.1111/j.2044-8317.1994.tb01039.x

Dolan, C. V. (1994). Factor analysis of variables with 2, 3, 5 and 7 response categories: A comparison of categorical variable estimators using simulated data. *British Journal of Mathematical and Statistical Psychology*, *47*, 309–326. https://doi.org/10.1111/j.2044-8317.1994. tb01039.x

Dolan, C. V., & van der Maas, H. L. (1998). Fitting multivariate normal finite mixtures subject to structural equation modeling. Psychometrika, 63(3), 227–253. doi:10.1007/ BF02294853

Du, J., Zhang, D., Yin, Y., Zhang, X., Li, J., Liu, D., ... & Chen, W. (2016). The personality and psychological stress predict major adverse cardio- vascular events in patients with coronary heart disease after percuta- neous coronary intervention for five years. *Medicine, 95*. https://doi. org/10.1097/MD.0000000000003364

Dulfer, K., Hazemeijer, B. A. F., Van Dijk, M. R., Van Geuns, R. J. M., Daemen, J., Van Domburg, R. T., & Utens, E. M. W. J. (2015). Prognostic value of type D personality for 10-year mortality and subjective health status in patients treated with percutaneous coronary intervention. *Journal of psychosomatic research*, *79*(3), 214-221.

Embretson, S. E. (1996). Item response theory models and spurious interaction effects in factorial ANOVA designs. *Applied Psychological Measurement*, *20*, 201–212. https://doi.org/10.1177/ 014662169602000302

Emons, W. H., Meijer, R. R., & Denollet, J. (2007). Negative affectivity and social inhibition in cardiovascular disease: evaluating type-D personality and its assessment using item response theory. *Journal of psychosomatic research*, *63*(1), 27-39.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. *Psychological methods*, *12*(2), 121.

Epskamp, S., Waldorp, L. J., Mõttus, R., & Borsboom, D. (2018). The Gaussian graphical model in cross-sectional and time-series data. *Multivariate Behavioral Research, 53(4),* 453-480.

Faber, J. (1988). Consistent estimation of correlations between observed interval variables with skewed distributions. *Quality and Quantity*, *22*(4), 381-392.

Ferreira-González, I., Permanyer-Miralda, G., Domingo-Salvany, A., Busse, J. W., Heels-Ansdell, D., Montori, V. M., ... & Guyatt, G. H. (2007). Problems with use of composite end points in cardiovascular trials: systematic review of randomised controlled trials. *Bmj*, *334*(7597), 786.

Ferguson, E., Williams, L., O'Connor, R. C., Howard, S., Hughes, B. M., Johnston, D. W., ... O'Carroll, R. E. (2009). A taxometric analysis of type-D personality. *Psychosomatic Medicine, 71*(9), 981–986. https://doi.org/10.1097/PSY.0b013e3181bd888b.

Fiedorowicz, J. G. (2020). Paul Lodder, 2020 EAPM Elsevier young investigator award recipient, on how to assess a type D personality effect. *Journal of Psychosomatic Research, 136.* 110197.

Field, A. (2010). *Discovering statistics using IBM SPSS statistics* (3rd ed.). Sage. Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*, 466–491. https://doi.org/10.1037/1082- 989X.9.4.466

Figueredo, A. J., de Baca, T. C., & Black, C. (2014). No matter where you go, there you are: The genetic foundations of temporal stability. *Journal of Methods and Measurement in the Social Sciences, 5(2),* 76-106.

Fitzsimons, GJ. (2008). Death to dichotomizing. *Journal of consumer research, 35,* 5-8

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory fac- tor analysis with ordinal data. Psychological Methods, 9(4), 466. doi:10.1037/1082-989X.9.4.466

Flora, D. B., LaBrish, C., & Chalmers, R. P. (2012). Old and new ideas for data screening and assumption testing for exploratory and confirmatory factor analysis. Frontiers in Psychology, 3, 55. http://dx.doi.org/10.3389/ fpsyg.2012.00055

Forero, C. G., & Maydeu-Olivares, A. (2009). Estimation of IRT graded response models: Limited versus full informa- tion methods. Psychological Methods, 14(3), 275. doi: 10.1037/a0015825

Forman, E. M., Herbert, J. D., Moitra, E., Yeomans, P. D., & Geller, P. A. (2007). A randomized controlled effectiveness trial of acceptance and commitment therapy and cognitive therapy for anxiety and depression. *Behavior modification*, *31*(6), 772-799.

Frasure-Smith, N., & Lesperance, F. (2008). Depression and anxiety as predictors of 2-year cardiac events in patients with stable coronary artery disease. Archives of General Psychiatry, 65(1), 62–71. doi:10.1001/archgenpsychiatry.2007.4

Friedman, H. S., & Booth-Kewley, S. (1987). The" disease-prone personality": A meta-analytic view of the construct. *American psychologist, 42(6*), 539.

Friedman, M., & Rosenman, R. H. (1959). Association of specific overt behavior pattern with blood and cardiovascular findings: blood cholesterol level, blood clotting time, incidence of arcus senilis, and clinical coronary artery disease. *Journal of the American medical association, 169(12),* 1286-1296.

Gan, Y., Gong, Y., Tong, X., Sun, H., Cong, Y., Dong, X., ... & Lu, Z. (2014). Depression and the risk of coronary heart disease: a meta-analysis of prospective cohort studies. *BMC psychiatry, 14(1),* 1-11.

Gana, K., & Broc, G. (2019). *Structural equation modeling with lavaan.* John Wiley & Sons.

Gilman, S. E., Sucha, E., Kingsbury, M., Horton, N. J., Murphy, J. M., & Colman, I. (2017). Depression and mortality in a longitudinal study: 1952–2011. *Cmaj*, *189*(42), E1304-E1310.

Gilmour, J., & Williams, L. (2012). Type D personality is associated with maladaptive health-related behaviours. *Journal of health psychology, 17(4),* 471-478.

Ginting, H., van de Ven, M., Becker, E. S., & Näring, G. (2016). Type D personality is associated with health behaviors and perceived social support in individuals with coronary heart disease. *Journal of health psychology, 21(5),* 727-737.

Goldberg, J. F., Garno, J. L., Leon, A. C., Kocsis, J. H., & Portera, L. (1998). Association of recurrent suicidal ideation with nonremission from acute mixed mania. American Journal of psychiatry, 155(12), 1753-1755.

Gostoli, S., Bonomo, M., Roncuzzi, R., Biffi, M., Boriani, G., & Rafanelli, C. (2016). Psychological correlates, allostatic overload and clinical course in patients with implantable cardioverter defibrillator (ICD). *International journal of cardiology*, *220*, 360-364.

Grande, G., Jordan, J., Kümmel, M., Struwe, C., Schubmann, R., Schulze, F., ... & Herrmann-Lingen, C. (2004). Evaluation of the German Type D Scale (DS14) and prevalence of the Type D personality pattern in cardiological and psychosomatic patients and healthy subjects. Psychotherapie, Psychosomatik, Medizinische Psychologie, 54(11), 413-422.

Grande, G., Romppel, M., & Barth, J. (2012). Association between type D personality and prognosis in patients with cardiovascular diseases: A systematic review and meta-analysis. *Annals of Behavioral Medicine, 43*(3), 299–310. https://doi.org/10.1007/s12160-011-9339-0.

Grande, G., Romppel, M., Glaesmer, H., Petrowski, K., & Hermann-Lingen, C. (2010). The type-D scale (DS14)-norms and prevalence of type-D personality in a population-based representative sample in Germany. *Personality and Individual Differences, 48*(8), 935–939. https://doi.org/10.1016/j.paid.2010.02.026.

Grande, G., Romppel, M., Vesper, J. M., Schubmann, R., Glaesmer, H., & Herrmann-Lingen, C. (2011). Type D personality and all-cause mortality in cardiac patients-data from a German cohort study. Psychosomatic Medicine, 73(7), 548–556. doi:10.1097/PSY.0b013e318227a9bc

Gremigni, P., & Sommaruga, M. (2005). Type D personality, a relevant construct in cardiology. Preliminary study of validation of the Italian questionnaire. *Cognitive and Behavioral Psychotherapy , 11 (1),* 7-18.

Gu, Z., Emons, W. H., & Sijtsma, K. (2018). Review of issues about classical change scores: A multilevel modeling perspective on some enduring beliefs. Psychometrika, 83(3), 674-695.

Hallquist, M., & Wiley, J. (2011). MplusAutomation: Automating Mplus model estimation and interpretation. R package version 0.6. Retrieved from http://cran.r-project.org/web/packages/MplusAutomation/index.html

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). Fundamentals of item response theory (Vol. 2). Thousand Oaks, CA: Sage.

Hartlage, S., Arduino, K., & Alloy, L. B. (1998). Depressive personality characteristics: State dependent concomitants of depressive disorder and traits independent of current depression. *Journal of Abnormal Psychology, 107(2),* 349.

Haslbeck, J. M., Borsboom, D., & Waldorp, L. J. (2019). Moderated network models. *Multivariate behavioral research*, 1-32.

Hayden, J. A., van der Windt, D. A., Cartwright, J. L., Côté, P., & Bombardier, C. (2013). Assessing bias in studies of prognostic factors. *Annals of internal medicine*, *158*(4), 280-286.

Hayduk, L. A., & Littvay, L. (2012). Should researchers use single indica- tors, best indicators, or multiple indicators in structural equation models? *BMC Medical Research Methodology*, *12*, 159. https://doi.org/ 10.1186/1471-2288-12-159

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach.* Guilford publications.

Haynes, S., Feinleib, M., & Kannel, W. B. (1980). The relationship of psychosocial factors to coronary heart disease in the Framingham Study. III. Eight-year incidence of coronary heart disease. American journal of epidemiology, 111(1), 37-58.

Hedges, L. V., & Olkin, I. (1980). Vote-counting methods in research synthesis*. Psychological bulletin, 88(2),* 359.

Helmers, K. F., Krantz, D. S., Merz, C., Klein, J., Kop, W. J., Gottdiener, J. S., & Rozanski, A. (1995). Defensive hostility: Relationship to multiple markers of cardiac ischemia in patients with coronary disease. *Health Psychology*, *14*(3), 202.

Hertzog, C., Lindenberger, U., Ghisletta, P., & von Oertzen, T. (2006). On the power of multivariate latent growth curve models to detect correlated change. *Psychological methods, 11(3),* 244.

Hillen, R. (2017). *Modeling psychological attributes: Merits and drawbacks of taxometrics and latent variable mixture models. S.l.: [s.n.]* (pp. 51–68). https://pure.uvt.nl/ws/portalfiles/portal/17484431/Hillen_Modeling_23_06_2017.pdf

Honaker, J., King, G., & Blackwell, M. (2011). Amelia II: A program for missing data. *Journal of Statistical Software*, *45*, 1–47. https://doi.org/ 10.18637/jss.v045.i07

Hoogeveen, S., Sarafoglou, A., Aczel, B., Aditya, Y., Alayan, A. J., Allen, P. J., ... & Nilsonne, G. (2022). A many-analysts approach to the relation between religiosity and well-being. *Religion, Brain & Behavior*, 1-47.

Hoogwegt, M. T., Pedersen, S. S., Theuns, D. A., & Kupper, N. (2014). Relation between emotional distress and heart rate variability in patients with an implantable cardioverter-defibrillator. *Psychophysiology, 51(2),* 187-196.

Horsten, M., Erigson, M., Perski, A., Wamala, S. P., Schenck-Gustafsson, K., & Orth-Gomér, K. (1999). Psychosocial factors and heart rate variability in healthy women. *Psychosomatic Medicine*, *61*(1), 49-57.

Horwood, S., Anglim, J., & Tooley, G. (2015). Type D personality and the Five-Factor Model: A facet-level analysis. *Personality and Individual Differences, 83*, 50–54. https://doi.org/10.1016/j.paid.2015.03.041.

Horwood, S., Anglim, J., & Tooley, G. (2016). Statistically modelling the relationships between Type D personality and social support, health behaviors and symptom severity in chronic illness groups. *Psychology & health, 31(9),* 1047-1063.

Howard, S., & Hughes, B. M. (2012). Construct, concurrent and discriminant validity of Type D personality in the general population: associations with anxiety, depression, stress and cardiac output. *Psychology & health, 27(2),* 242-258.

Howard, S., Hughes, B. M., & James, J. E. (2011). Type D personality and hemodynamic reactivity to laboratory stress in women*. International Journal of Psychophysiology,* 80(2), 96-102.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal, 6(1), 1–55. doi: 10.1080/10705519909540118

Husson, O., Denollet, J., Oerlemans, S., & Mols, F. (2013). Satisfaction with information provision in cancer patients and the moderating effect of Type D personality. *Psycho-Oncology, 22(9),* 2124-2132.

Imbalzano, E., Vatrano, M., Quartuccio, S., Ceravolo, R., Ciconte, V. A., Rotella, P., ... & Mandraffino, G. (2018). Effect of type D personality on smoking status and their combined impact on outcome after acute myocardial infarction. *Clinical cardiology*, *41*(3), 321-325.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS medicine, 2(8),* e124.

Ioannidis, J. P. (2012). Scientific inbreeding and same-team replication: type D personality as an example. *Journal of psychosomatic research*, *73*(6), 408-410.

Ionescu, D. F., Niciu, M. J., Henter, I. D., & Zarate, C. A. (2013). Defining anxious depression: a review of the literature. *CNS spectrums, 18(5),* 252-260.

Jaccard, J., & Wan, C. K. (1995). Measurement error in the analysis of interaction effects between continuous predictors using multiple regression: Multiple indicator and structural equation approaches. *Psychological bulletin*, *117*(2), 348.

Jaccard, J., Wan, C. K., & Turrisi, R. (1990). The detection and interpretation of interaction effects between continuous variables in multiple regression. *Multivariate Behavioral Research, 25*(4), 467–478. https://doi.org/10.1207/ s15327906mbr2504_4.

Jacobson, N. S., & Truax, P. (1992). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of consulting and clinical psychology, 59 (1),* 12-19.

Jak, S., Oort, F. J., & Dolan, C. V. (2013). A test for cluster bias: Detecting violations of measurement invariance across clusters in multilevel data. *Structural Equation Modeling: A Multidisciplinary Journal, 20(2),* 265-282.

Jeon, S. W., Lim, H. E., Yoon, S., Na, K. S., Ko, Y. H., Joe, S. H., & Kim, Y. H. (2017). Does type D personality impact on the prognosis of patients who underwent catheter ablation for atrial fibrillation? A 1-year follow-up study. *Psychiatry investigation, 14(3),* 281.

Jöreskog, K. G., & Sörbom, D. (1993). LISREL 8: Structural equation modeling with the SIMPLIS command language. *Scientific Software International.*

Jöreskog, K. G., & Yang, F. (1996). Nonlinear structural equation models: The Kenny-Judd model with interaction effects. In R. E. Schumacker & G. A. Marcoulides (Eds.),

Advanced structural equation modeling: Issues and techniques (pp. 57–88). Lawrence Erlbaum.

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2019). *semTools: Useful tools for structural equation modeling* (R package version 0.5-2). https://CRAN.R-project.org/package= semTools

Jorm, A. F., Christensen, H., Henderson, A. S., Jacomb, P. A., Korten, A. E., & Rodgers, B. (2000). Predicting anxiety and depression from personality: Is there a synergistic effect of neuroticism and extraversion? *Journal of Abnormal Psychology, 109*(1), 145. https://doi.org/10.1037//002I-843X.109.1.145.

Kang, S. M., & Waller, N. G. (2005). Moderated multiple regression, spurious interaction effects, and IRT. *Applied Psychological Measurement*, *29*, 87–105. https://doi.org/10.1177/ 0146621604272737

Kantor, L., Endler, N. S., Heslegrave, R. J., & Kocovski, N. L. (2001). Validating self-report measures of state and trait anxiety against a physiological measure. *Current Psychology, 20(3),* 207-215.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors*. Journal of the american statistical association, 90(430),* 773-795.

Kelava, A., & Nagengast, B. (2012). A Bayesian model for the estimation of latent interaction and quadratic effects when latent variables are non-normally distributed. *Multivariate Behavioral Research*, *47*, 717–742. https://doi.org/10.1080/00273171.2012.715560

Kelava, A., Nagengast, B., & Brandt, H. (2014). A Nonlinear structural equation mixture modeling approach for nonnormally distributed latent predictor variables. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*, 468–481. https://doi.org/10.1080/ 10705511.2014.915379

Kelly-Hughes, D. H., Wetherell, M. A., & Smith, M. A. (2014). Type D personality and cardiovascular reactivity to an ecologically valid multitasking stressor. *Psychology & health, 29(10),* 1156-1175.

Kenny, D. A., & Judd, C. M. (1984). Estimating the nonlinear and inter- active effects of latent variables. *Psychological Bulletin*, *96*, 201–210. https://doi.org/10.1037/0033-2909.96.1.201

Kessing, D., Denollet, J., Widdershoven, J., & Kupper, N. (2017). Self-care and health-related quality of life in chronic heart failure: A longitudinal analysis. *European Journal of Cardiovascular Nursing*, *16*(7), 605-613.

Klein, A., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrica*, *65*, 457–474. https://doi.org/10.1007/BF02296338

Kline, R. B. (2010). *Principles and practice of structural equation modeling. 2011*. Guilford Press.

Kobayashi, M., Voors, A. A., Girerd, N., Billotte, M., Anker, S. D., Cleland, J. G., ... & Ferreira, J. P. (2020). Heart failure etiologies and clinical factors precipitating for worsening heart failure: Findings from BIOSTAT-CHF. *European Journal of Internal Medicine*, *71*, 62-69.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. Journal of General Internal Medicine, 16(9), 606–613. doi: 10.1046/j.1525-1497.2001.016009606.x

Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. London: Academic Press.

Kubzansky, L. D., Cole, S. R., Kawachi, I., Vokonas, P., & Sparrow, D. (2006). Shared and unique contributions of anger, anxiety, and depression to coronary heart disease: a prospective study in the normative aging study. Annals of Behavioral Medicine, 31(1), 21–29. doi:10.1207/ s15324796abm3101_5

Kupper, N., & Denollet, J. (2007). Type D personality as a prognostic factor in heart disease: Assessment and mediating mechanisms. *Journal of Personality Assessment, 89*(3), 265–276. https://doi.org/10.1080/00223890701629797.

Kupper, N., & Denollet, J. (2016). Explaining heterogeneity in the pre- dictive value of Type D personality for cardiac events and mortality. *International journal of cardiology, 224*, 119–124. https://doi.org/10.1016/j.ijcard.2016.09.006

Kupper, N., & Denollet, J. (2018). Type D personality as a risk factor in coronary heart disease: a review of current evidence. Current cardiology reports, 20(11), 104.

Kupper, N., Boomsma, D. I., de Geus, E. J., Denollet, J., & Willemsen, G. (2011). Nine-year stability of type D personality: contributions of genes and environment. *Psychosomatic Medicine, 73(1),* 75-82.

Kupper, N., Pedersen, S. S., Höfer, S., Saner, H., Oldridge, N., & Denollet, J. (2013a). Cross-cultural analysis of type D (distressed) personality in 6222 patients with ischemic heart disease: a study from the International HeartQoL Project. *International journal of cardiology, 166(2),* 327-333.

Kupper, N., Denollet, J., Widdershoven, J., & Kop, W. J. (2013b). Type D personality is associated with low cardio- vascular reactivity to acute mental stress in heart failure patients. International Journal of Psychophysiology, 90(1), 44–49. doi:10.1016/j.ijpsycho.2013.01.011

Kupper, N., Pelle, A., & Denollet, J. (2013c). Association of Type D personality with the autonomic and hemodynamic response to the cold pressor test. *Psychophysiology, 50(12),* 1194-1201.

Kupper, N., Lodder, P., Habibovic, M., Spek, V. M., & Denollet, J. Type D personality traits are associated with impaired well-being self-management skills in the general po-pulation: A LifeLines Cohort study. Retrieved from https://osf.io/r8cvj/.

Lagakos, S. W. (1988). Effects of mismodelling and mismeasuring explanatory variables on tests of their association with a response variable. *Statistics in medicine*, *7*(1-2), 257-274.

Lakens, D. (2017). Equivalence tests: a practical primer for t tests, correlations, and meta-analyses. *Social psychological and personality science, 8(4),* 355-362.

Larson, N. C., Barger, S. D., & Sydeman, S. J. (2013). Type D personality is not associated with coronary heart disease risk in a North American sample of retirement-aged adults. *International journal of behavioral medicine, 20(2),* 277-285.

Lee, R., Yu, H., Gao, X., Cao, J., Tao, H., Yu, B., ... & Lin, P. (2019). The negative affectivity dimension of Type D personality is associated with in-stent neoatherosclerosis in coronary patients with percutaneous coronary intervention: An optical coherence tomography study. *Journal of psychosomatic research, 120,* 20-28.

Leszko, M., Elleman, L. G., Bastarache, E. D., Graham, E. K., & Mroczek, D. K. (2016). Future directions in the study of personality in adulthood and older age. *Gerontology, 62(2),* 210-215.

Leu, H. B., Yin, W. H., Tseng, W. K., Wu, Y. W., Lin, T. H., Yeh, H. I., ... & Chen, J. W. (2019). Impact of type D personality on clinical outcomes in Asian patients with stable coronary artery disease. *Journal of the Formosan Medical Association*, *118*(3), 721-729.

Levy, S. M. (1985). *Behavior and Cancer*. San Francisco (Jossey-Bass) 1985.

Li-Gao, R., Boomsma, D., Dolan, C.V., de Geus, E.J.C., Denollet, J., Kupper, N. (2021) Genetic and environmental contributions to stability and change in social inhibition across the adolescent and adult lifespan. *Preprint received through personal communication.*

Li, X., Gao, M., Zhang, S., Xu, H., Zhou, H., Wang, X., ... & Tian, D. (2017). Medication adherence mediates the association between type D personality and high HbA1c level in Chinese patients with type 2 diabetes mellitus: a six-month follow-up study. *Journal of diabetes research,* 2017.

Lim, H. E., Lee, M. S., Ko, Y. H., Park, Y. M., Joe, S. H., Kim, Y. K., ... & Denollet, J. (2011). Assessment of the type D personality construct in the Korean population: a validation study of the Korean DS14. *Journal of Korean medical science, 26(1),* 116-123.

Lin, P., Li, L., Wang, Y., Zhao, Z., Liu, G., Chen, W., ... & Gao, X. (2018). Type D personality, but not Type A behavior pattern, is associated with coronary plaque vulnerability. *Psychology, health & medicine, 23(2),* 216-223.

Lin, T. K., You, K. X., Hsu, C. T., Li, Y. D., Lin, C. L., Weng, C. Y., & Koo, M. (2019). Negative affectivity and social inhibition are associated with increased cardiac readmission in patients with heart failure: A preliminary observation study. *Plos one*, *14*(4), e0215726.

Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*, 1198–1202. https://doi.org/10.1080/01621459.1988. 10478722

Liu, W., Zhang, Y., Yu, C. M., Ji, Q. W., Cai, M., Zhao, Y. X., & Zhou, Y. J. (2015). Current understanding of coronary artery calcification. *Journal of geriatric cardiology: JGC*, *12*(6), 668.

Liu, Y., Millsap, R. E., West, S. G., Tein, J. Y., Tanaka, R., & Grimm, K. J. (2017). Testing measurement invariance in longitudinal data with ordered-categorical measures. *Psychological Methods, 22(3),* 486.

Lodder, P. (2013). To impute or not impute: That's the question. In G. J. Mellenbergh & H. J. Adèr (Eds.), *Advising on research methods: Selected topics 2013*, (pp. 1–7). Johannes van Kessel Publishing.

Lodder, P. (2022). Why researchers should not ignore measurement error and skewness in questionnaire item scores. *Commentary in MARP special issue*. *Religion, Brain & Behavior.*

Lodder, P., Wicherts, J. M., Denollet, J., & Kupper, N. Type D personality as a risk factor for adverse outcome in patients with cardiovascular disease: An individual patient data meta-analysis. PROSPERO preregistration under review. Retrieved from https://osf.io/czmhs/.

Loftus, G. R. (1978). On interpretation of interactions. *Memory & Cognition, 6(3),* 312-319.

Loken, E., & Gelman, A. (2017). Measurement error and the replication crisis. *Science, 355(6325),* 584-585.

Lomax, R. G., & Schumacker, R. E. (2004). *A beginner's guide to structural equation modeling*. Psychology Press.

Loosman, W. L., de Jong, R. W., Haverkamp, G. L., van den Beukel, T. O., Dekker, F. W., Siegert, C. E., & Honig, A. (2018). The stability of type D personality in dialysis patients. *International journal of behavioral medicine, 25(1),* 85-92.

Löwe,B.,Decker,O.,Mu€ller,S.,Br€ahler,E.,Schellberg,D., Herzog, W., & Herzberg, P. Y. (2008). Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. Medical Care, 46(3), 266–274. doi:10.1097/MLR.0b013e318160d093

Lu, I. R., Kwan, E., Thomas, D. R., & Cedzynski, M. (2011). Two new methods for estimating structural equation models: An illustration and a comparison with two established methods. *International Journal of Research in Marketing*, *28*, 258–268. https://doi.org/10.1016/j.ijresmar. 2011.03.006

Lubinski, D., & Humphreys, L. G. (1990). Assessing spurious "moderator effects": Illustrated substantively with the hypothesized ("synergistic") relation between spatial and mathematical ability. Psychological Bulletin, 107(3), 385. doi:10.1037/0033-2909.107.3.385

Lv, H., Tao, H., Wang, Y., Zhao, Z., Liu, G., Li, L., ... & Lin, P. (2020). Impact of type D personality on major adverse cardiac events in patients undergoing percutaneous coronary intervention: The mediating role of cognitive appraisal and coping style. *Journal of Psychosomatic Research*, *136*, 110192.

Maassen, G. H., Bossema, E., & Brand, N. (2009). Reliable change and practice effects: Outcomes of various indices compared. *Journal of clinical and experimental neuropsychology*, *31*(3), 339-352.

MacCallum, R. C., & Mar, C. M. (1995). Distinguishing between mod- erator and quadratic effects in multiple regression. *Psychological Bulletin*, *118*, 405–421. https://doi.org/10.1037/0033-2909.118.3.405

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7*, 19–40. https://doi.org/10.1037/1082-989X.7.1.19

Mahase, E. (2019). Cancer overtakes CVD to become leading cause of death in high income countries. BMJ: British Medical Journal (Online), 366.

Marchesi, C., Ossola, P., Scagnelli, F., Paglia, F., Aprile, S., Monici, A., ... & Ardissino, D. (2014). Type D personality in never depressed patients at their first acute coronary syndrome. *Psychotherapy and Psychosomatics, 83,* 190:1.

Marsh, H. W., Wen, Z., & Hau, K. T. (2004). Structural equation models of latent interactions: Evaluation of alternative estimation strategies and indicator construction. *Psychological Methods*, *9*, 275–300. https://doi. org/10.1037/1082-989X.9.3.275

Martens, E. J., Kupper, N., Pedersen, S. S., Aquarius, A. E., & Denollet, J. (2007). Type-D personality is a stable taxonomy in post-MI patients over an 18-month period. *Journal of psychosomatic research, 63(5),* 545-550.

Martens, E. J., Mols, F., Burg, M. M., & Denollet, J. (2010). Type D personality predicts clinical events after myocardial infarction, above and beyond disease severity and depression. *The Journal of clinical psychiatry*, *71*(6), 0-0.

Martin, A., Rief, W., Klaiberg, A., & Braehler, E. (2006). Validity of the brief patient health questionnaire mood scale (PHQ-9) in the general population. General Hospital Psychiatry, 28(1), 71–77. doi:10.1016/ j.genhosppsych.2005.07.003

Maslowsky, J., Jager, J., & Hemken, D. (2015). Estimating and interpreting latent variable interactions: A tutorial for applying the latent moderated structural equations

method. International Journal of Behavioral Development, 39(1), 87–96. doi:10.1177/0165025414552301

Matsuishi, Y., Shimojo, N., Unoki, T., Sakuramoto, H., Tokunaga, C., Yoshino, Y., ... & Hiramatsu, Y. (2019). Type D personality is a predictor of prolonged acute brain dysfunction (delirium/coma) after cardiovascular surgery. *BMC psychology, 7(1),* 27.

Matthews, K. A., Owens, J. F., Kuller, L. H., Sutton-Tyrrell, K., & Jansen-McWilliams, L. (1998). Are hostility and anxiety associated with carotid atherosclerosis in healthy postmenopausal women?. *Psychosomatic Medicine, 60(5),* 633-638.

Maxwell, S. E., & Delaney, H. D. (1993). Bivariate median splits and spurious statistical significance. *Psychological Bulletin, 113*(1), 181. https://doi.org/10.1037/0033-2909.113.1.181.

May-Ling, J. L., Loxton, D., & McLaughlin, D. (2015). Trauma exposure and the subsequent risk of coronary heart disease among mid-aged women. *Journal of behavioral medicine, 38(1),* 57-65.

McArdle, J. J. (1986). Latent variable growth within behavior genetic models. *Behavior Genetics, 16(1),* 163-200.

McCrae, R. R., & Costa Jr, P. T. (1994). The stability of personality: Observations and evaluations. *Current directions in psychological science, 3(6),* 173-175.

Meehl, P. E. (1995). Bootstraps taxometrics: Solving the classification problem in psychopathology. *American psychologist, 50(4),* 266.
Methods, 19(2), 300. http://dx.doi.org/10.1037/a0033805

Meyer, T., Hussein, S., Lange, H. W., & Herrmann-Lingen, C. (2014). Type D personality is unrelated to major adverse cardiovascular events in patients with coronary artery disease treated by intracoronary stenting. *Annals of Behavioral Medicine, 48(2),* 156-162.

Millsap, R. E., & Cham, H. (2012). Investigating factorial invariance in longitudinal data. In B. Laursen, T. D. Little, & N. A. Card (Eds.), *Handbook of developmental research methods* (pp. 109–127). New York, NY: Guilford Press.

Modica, M., Carabalona, R., Spezzaferri, R., Tavanelli, M., Torri, A., Ripamonti, V., ... & Ferratini, M. (2012). Psychological profiles derived by cluster analysis of Minnesota Multiphasic Personality Inventory and long term clinical outcome after coronary artery by pass grafting. *Monaldi Archives for Chest Disease*, *78*(1).

Molloy, G. J., Randall, G., Wikman, A., Perkins-Porras, L., Messerli-Bürgy, N., & Steptoe, A. (2012). Type D personality, self-efficacy, and medication adherence following an acute coronary syndrome. *Psychosomatic medicine, 74(1),* 100-106.

Mols, F., & Denollet, J. (2010). Type D personality in the general population: a systematic review of health status, mechanisms of disease, and work-related problems. Health and Quality of Life Outcomes, 8(1), 9. doi:10.1186/ 1477-7525-8-9

Mommersteeg, P. M., Herr, R., Bosch, J., Fischer, J. E., & Loerbroks, A. (2011). Type D personality and metabolic syndrome in a 7-year prospective occupational cohort. *Journal of psychosomatic research, 71(5),* 357-363.

Montero, P., Bermúdez, J., & Rueda, B. (2017). Adaptation to Spanish of the DS-14 Scale ("Type D Scale-14") for the measurement of type D personality. *Journal of Psychopathology and Clinical Psychology , 22 (1),* 55-67.

Moser, B. K., & Coombs, L. P. (2004). Odds ratios for a continuous outcome variable without dichotomizing. *Statistics in Medicine, 23(12),* 1843-1860.

Mroczek, D. K., & Spiro III, A. (2003). Modeling intraindividual change in personality traits: Findings from the Normative Aging Study. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 58(3),* P153-P165.

Mroczek, D. K., Graham, E. K., Turiano, N. A., & Aro-Lambo, M. O. (2021). Personality development in adulthood and later life. In O. P. John & R. W. Robins (Eds.), *Handbook of personality: Theory and research* (pp. 336–351). The Guilford Press.

Muchinsky, P. M. (1996). The correction for attenuation. *Educational and psychological measurement, 56(1),* 63-75.

Muthén, B.O. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. Psychometrika, 49(1), 115–132. doi:10.1007/BF02294210

Muthén, B.O., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology, 38*(2), 171–189. https://doi.org/10.1111/j.2044-8317.1985.tb00832.x.

Muthén, B.O., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and experimental research*, *24*(6), 882-891.

Muthén, L. K., & Muthén, B. O. (1998). 2011). Mplus user's guide. 6th ed. Los Angeles, CA: Muthén & Muthén.

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and deter- mine power. Structural Equation Modeling, 9(4), 599–620. doi:10.1207/S15328007SEM0904_8

Neale, M. C., & McArdle, J. J. (2000). Structured latent growth curves for twin data. *Twin Research and Human Genetics, 3(3),* 165-177.

Nefs, G., Bot, M., Browne, J. L., Speight, J., & Pouwer, F. (2012). Diabetes MILES–The Netherlands: Tationale, design and sample characteristics of a national survey examining the psychosocial aspects of living with diabetes in Dutch adults. BMC Public Health, 12(1), 1. http://dx. doi.org/10.1186/1471-2458-12-925

Nefs, G., Pouwer, F., Pop, V., & Denollet, J. (2012). Type D (distressed) personality in primary care patients with type 2 diabetes: Validation and clinical correlates of the DS14 assessment. *Journal of psychosomatic research*, *72*, 251–257. https://doi.org/10.1016/j.jpsychores.2012.01.006

Nefs, G., Speight, J., Pouwer, F., Pop, V., Bot, M., & Denollet, J. (2015). Type D per- sonality, suboptimal health behaviors and emotional distress in adults with diabetes: Results from Diabetes MILES–The Netherlands. *Diabetes Research and Clinical Practice, 108*(1), 94–105. https://doi.org/10.1016/j.diabres.2015.01.015.

Nivard, M. G., Dolan, C. V., Kendler, K. S., Kan, K. J., Willemsen, G., Van Beijsterveldt, C. E. M., ... & Boomsma, D. I. (2015). Stability in symptoms of anxiety and depression as a function of genotype and environment: a longitudinal twin study from ages 3 to 63 years. Psychological medicine, 45(5), 1039-1049.

O'Leary, É. D., Howard, S., Hughes, B. M., & James, J. E. (2013). An experimental test of blunting using sleep-restriction as an acute stressor in Type D and non-Type D women. *International Journal of Psychophysiology, 90(1),* 37-43.

O'Dell, K. R., Masters, K. S., Spielmans, G. I., & Maisto, S. A. (2011). Does type-D personality predict outcomes among patients with cardiovascular disease? A meta-ana- lytic

review. Journal of Psychosomatic Research, 71(4), 199–206. doi:10.1016/j.jpsychores.2011.01.009

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. Psychometrika, 44(4), 443–460. http://dx.doi.org/10.1007/BF02296207

Oort, F. J. (2005). Using structural equation modeling to detect response shifts and true change. Quality of Life Research, 14(3), 587-598.

Ossola, P., De Panfilis, C., Tonna, M., Ardissino, D., & Marchesi, C. (2015). DS 14 is more likely to measure depression rather than a personality disposition in patients with acute coronary syndrome. *Scandinavian Journal of Psychology, 56(6),* 685-692.

Parker, G., Malhi, G., Mitchell, P., Wilhelm, K., Austin, M. P., Crawford, J., & Hadzi-Pavlovic, D. (2005). Progressing a spectrum model for defining non-melancholic depression. *Acta Psychiatrica Scandinavica*, *111*(2), 139-143.

Pedersen, S. S., & Denollet, J. (2003). Type D personality, cardiac events, and impaired quality of life: A review. *European Journal of Cardiovascular Prevention & Rehabilitation, 10*(4), 241–248. https://doi.org/10.1097/00149831-200308000-00005.

Pedersen, S. S., van Domburg, R. T., Theuns, D. A., Jordaens, L., & Erdman, R. A. (2004). Type D personality is associated with increased anxiety and depressive symp- toms in patients with an implantable cardioverter defib- rillator and their partners. Psychosomatic Medicine, 66(5), 714–719. doi:10.1097/01.psy.0000132874.52202.21

Pedersen, S. S., Yagensky, A., Smith, O. R., Yagenska, O., Shpak, V., & Denollet, J. (2009). Preliminary evidence for the cross-cultural utility of the type D personality construct in the Ukraine. *International journal of behavioral medicine, 16(2)*, 108.

Pedersen, S. S., van den Broek, K. C., Erdman, R. A., Jordaens, L., & Theuns, D. A. (2010). Pre-implantation implantable cardioverter defibrillator concerns and Type D personality increase the risk of mortality in patients with an implantable cardioverter defibrillator. *Europace*, *12*(10), 1446-1452.

Pelle, A. J., Erdman, R. A., van Domburg, R. T., Spiering, M., Kazemier, M., & Pedersen, S. S. (2008). Type D patients report poorer health status prior to and after cardiac rehabilitation compared to non-type D patients. *Annals of Behavioral Medicine, 36(2),* 167-175.

Pelle, A. J., Pedersen, S. S., Schiffer, A. A., Szabó, B., Widdershoven, J. W., & Denollet, J. (2010). Psychological distress and mortality in systolic heart failure. Circulation: Heart Failure, 3(2), 261–267. doi:10.1161/CIRCHEARTFAILURE.109.871483

Pentz, M. A., & Chou, C. P. (1994). Measurement invariance in longitudinal clinical research assuming change from development and intervention. *Journal of Consulting and Clinical Psychology, 62(3)*, 450.

Piepoli, M. F., Hoes, A. W., Agewall, S., Albus, C., Brotons, C., Catapano, A. L., ... & Verschuren, W. M. (2016). Guidelines: Editor's choice: 2016 European Guidelines on cardiovascular disease prevention in clinical practice: The Sixth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of 10 societies and by invited experts) Developed with the special contribution of the European Association for Cardiovascular Prevention & Rehabilitation (EACPR). *European heart journal, 37(29),* 2315.

Ping, R. A. (1995). A parsimonious estimating technique for interaction and quadratic latent variables. Journal of Marketing Research, 32(3), 336–347. doi:10.2307/3151985

Potthast, M. J. (1993). Confirmatory factor analysis of ordered categorical variables with large models. British Journal of Mathematical and Statistical Psychology, 46(2), 273–286. doi:10.1111/j.2044-8317.1993.tb01016.x

Prenoveau, J. M., Craske, M. G., Zinbarg, R. E., Mineka, S., Rose, R. D., & Griffith, J. W. (2011). Are anxiety and depression just as stable as personality during late adolescence? Results from a three-year longitudinal latent variable study. *Journal of Abnormal Psychology, 120(4),* 832.

Pushkarev, G. S., Kuznetsov, V. A., Yaroslavskaya, E. I., & Bessonov, I. S. (2017). Prognostic significance of psychosocial risk factors in patients with ischemic heart disease after percutaneous coronary interventions. *Kardiologiya*, *57*(6), 11-15.

R Development Core Team. (2008). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07- 0. http://www.R-project.org

R Development Core Team. (2008). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. 3-900051-07-0. http://www.R-project.org

Rademaker, A. R., van Zuiden, M., Vermetten, E., & Geuze, E. (2011). Type D personality and the development of PTSD symptoms: A prospective study. *Journal of Abnormal Psychology, 120(2),* 299.

Ramsey, I., de Rooij, B. H., Mols, F., Corsini, N., Horevoorts, N. J., Eckert, M., & van de Poll-Franse, L. V. (2019). Cancer survivors who fully participate in the PROFILES registry have better health-related quality of life than those who drop out. *Journal of Cancer Survivorship, 13(6),* 829-839.

Rasch, G. (1960). Probabilistic models for some intelligence and attain- ment tests. Copenhagen, Denmark: Paedagogiske Institut

Raykh, O. I., Sumin, A. N., Kokov, A. N., Indukaeva, E. V., & Artamonova, G. V. (2020). Association of type D personality and level of coronary artery calcification. *Journal of Psychosomatic Research*, *139*, 110265.

Reise, S. P., & Waller, N. G. (2009). Item response theory and clinical measurement. Annual Review of Clinical Psychology, 5(1), 27–48. doi:10.1146/annurev.clinpsy.032408.153553

Reise, S. P., & Waller, N. G. (2009). Item response theory and clinical measurement. *Annual Review of Clinical Psychology*, *5*, 27–48. https://doi.org/10.1146/annurev.clinpsy.032408.153553

Rhemtulla, M., Brosseau-Liard, P. E., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*, 354. https://doi.org/10.1037/ a0029315

Rice, K., Higgins, J. P., & Lumley, T. (2018). A re-evaluation of fixed effect (s) meta-analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *181*(1), 205-227.

Rigdon, E. E., Preacher, K. J., Lee, N., Howell, R. D., Franke, G. R., & Borsboom, D. (2011). Avoiding measurement dogma: a response to Rossiter. *European Journal of Marketing*.

Rijsdijk, F. V., & Sham, P. C. (2002). Analytic approaches to twin data using structural equation models. *Briefings in bioinformatics, 3(2),* 119-133.

Riley, R. D., Lambert, P. C., & Abo-Zaid, G. (2010). Meta-analysis of individual participant data: Rationale, conduct, and reporting. *Bmj, 340*, c221. https://doi.org/10.1136/bmj.c221.

Riordan, A. O., Howard, S., & Gallagher, S. (2019). Social context and sex moderate the association between type D personality and cardiovascular reactivity. *Applied psychophysiology and biofeedback*, *44*(4), 321-330.

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: a quantitative review of longitudinal studies. *Psychological bulletin, 126(1),* 3.

Roberts, B. W., & Mroczek, D. (2008). Personality trait change in adulthood. *Current directions in psychological science, 17(1),* 31-35.

Roberts, B. W., & Wood, D. (2006). Personality Development in the Context of the Neo-Socioanalytic Model of Personality. In D. K. Mroczek & T. D. Little (Eds.), *Handbook of personality development (p. 11–39).* Lawrence Erlbaum Associates Publishers.

Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: a meta-analysis of longitudinal studies. Psychological bulletin, 132(1), 1.

Roberts, B. W., Wood, D., & Caspi, A. (2008). The development of personality traits in adulthood. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (p. 375–398).

Roest, A. M., Martens, E. J., de Jonge, P., & Denollet, J. (2010). Anxiety and risk of incident coronary heart dis- ease: A meta-analysis. Journal of the American College of Cardiology, 56(1), 38–46. doi:10.1016/j.jacc.2010.03.034

Romppel, M., Herrmann-Lingen, C., Vesper, J. M., & Grande, G. (2012). Six year stability of Type-D personality in a German cohort of cardiac patients. *Journal of Psychosomatic Research, 72*, 136–141. https://doi. org/10.1016/j.jpsychores.2011.11.009

Rosengren, A., Hawken, S., Ôunpuu, S., Sliwa, K., Zubaid, M., Almahmeed, W. A., ... & INTERHEART investigators. (2004). Association of psychosocial risk factors with risk of acute myocardial infarction in 11 119 cases and 13 648 controls from 52 countries (the INTERHEART study): case-control study. *The Lancet, 364(9438)*, 953-962. https://doi.org/10.1016/S0140-6736(04)17019-0

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*, 1–36. https://doi.org/10. 18637/jss.v048.i02

Royston, P., Altman, D. G., & Sauerbrei, W. (2006). Dichotomizing con- tinuous predictors in multiple regression: A bad idea. *Statistics in Medicine*, *25*, 127–141. https://doi.org/10.1002/sim.2331

Rutledge, T., Linke, S. E., Krantz, D. S., Johnson, B. D., Bittner, V., Eastwood, J. A., ... & Vido, D. A. (2009). Comorbid depression and anxiety symptoms as predictors of cardiovascular events: results from the NHLBI-sponsored Women's Ischemia Syndrome Evaluation (WISE) study. *Psychosomatic medicine, 71(9), 958.*

Samejima, F. (1997). Graded response model. In W. J. Van der Linden & R. K. Hambleton. Handbook of Modern Item Response Theory (pp. 85–100). New-York: Springer-Verlag.

Satorra, A. (2000). Scaled and adjusted restricted tests in multi-sample analysis of moment structures. In *Innovations in multivariate statistical analysis* (pp. 233-247). Springer, Boston, MA.

Satorra, A., & Bentler, P. M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. American statistical association 1988 proceedings of the business and economics sections (pp. 308–313). Alexandria, VA: American Statistical Association.

Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika*, *75*, 243–248. https:// doi.org/10.1007/s11336-009-9135-y

Schiffer, A., Pedersen, S., Widdershoven, J., Hendriks, E., Winter, J., & Denollet, J. (2005). Type D personality is independently associated with impaired health status, and depressive symptoms in chronic heart failure. European Journal of Cardiovascular Prevention & Rehabilitation, 12(3), 282–283. http://dx.doi.org/10.1097/01.hjr.0000173107. 76109.6c

Schmidt, M. M., Quadros, A. S., Abelin, A. P., Minozzo, E. L., Wottrich, S. H., Kunert, H. Z., ... & Gottschall, C. A. (2011). Psychological characteristics of patients undergoing percutaneous coronary interventions. *Arquivos Brasileiros de Cardiologia*, *97*, 331-337.

Schoormans, D., Verhoeven, J. E., Denollet, J., van de Poll-Franse, L., & Penninx, B. W. J. H. (2018). Leukocyte telomere length and personality: associations with the Big Five and Type D personality traits. *Psychological medicine, 48(6),* 1008-1019.

Schwabe, I., & van den Berg, S. M. (2014). Assessing genotype by environment interaction in case of hetero- geneous measurement error. Behavior Genetics, 44(4), 394–406. http://dx.doi.org/10.1007/s10519-014-9649-7

Schwabe, I., Gu, Z., Tijmstra, J., Hatemi, P., & Pohl, S. (2019). Psychometric modelling of longitudinal genetically-informative twin data. *Frontiers in genetics, 10,* 837.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*, 461–464. https://doi.org/10.1214/aos/1176344136

Schweig, J. (2013). Testing the Assumption of Cross-Level Measurement Invariance in Multilevel Models: Evidence from School and Classroom Environment Surveys. *CRESST Report 829. National Center for Research on Evaluation, Standards, and Student Testing (CRESST).*

Shekelle, R. B., Gale, M., Norusis, M., & Aspirin Myocardial Infarction Study Research Group. (1985). Type A score (Jenkins Activity Survey) and risk of recurrent coronary heart disease in the Aspirin Myocardial Infarction Study. *The American journal of cardiology, 56(4),* 222-225.

Shrout, P. E. (1998). Measurement reliability and agreement in psychiatry. *Statistical methods in medical research, 7(3),* 301-317.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: a key to the file-drawer. *Journal of experimental psychology: General, 143(2),* 534.

Skrondal, A., & Rabe-Hesketh, S. (2004). Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models. Miami: CRC Press.

Smeijers, L., van de Pas, H., Nyklicek, I., Notten, P. J., Pedersen, S. S., & Kop, W. J. (2014). The independent association of anxiety with non-cardiac chest pain. *Psychology & health, 29(3),* 253-263.

Smith, M. A., Riccalton, V. C., Kelly-Hughes, D. H., Craw, O. A., Allen, S. F., O'Connor, D. B., & Wetherell, M. A. (2018). The relationship between type D personality and physical health complaints is mediated by perceived stress and anxiety but not diurnal cortisol secretion. *Stress, 21*(3), 229–236. https://doi.org/10.1080/ 10253890.2018.1435637.

Smith, T. W. (2011). Toward a more systematic, cumulative, and applicable science of personality and health: Lessons from type D personality. *Psychosomatic Medicine, 73*(7), 528–532. https://doi.org/10.1097/PSY.0b013e31822e095e.

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, *15*, 72–101. https://doi. org/10.2307/1412159

Spindler, H., Kruse, C., Zwisler, A. D., & Pedersen, S. S. (2009). Increased anxiety and depression in Danish cardiac patients with a type D personality: cross-validation of the Type D Scale (DS14). *International journal of behavioral medicine, 16(2),* 98-107.

Spinhoven, P. H., Ormel, J., Sloekers, P. P. A., Kempen, G. I. J. M., Speckens, A. E. M., & Van Hemert, A. M. (1997). A validation study of the Hospital Anxiety and Depression Scale (HADS) in different groups of Dutch subjects*. Psychological medicine, 27(2),* 363-370.

Spitzer, R. L., Kroenke, K., Williams, J. B., & Lo€we, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. Archives of Internal Medicine, 166(10), 1092–1097. doi:10.1001/archinte.166.10.1092

Steiger, J. H. (2006a). Things we could have known: Some thoughts on seeing the future and rediscovering the past in data analysis and model selection. Presidential Address, Society of Multivariate Experimental Psychology, Lawrence, Kansas, October 2006.

Steiger, J. H. (2006b). Things we could have known: Some thoughts on seeing the future and avoiding regret in data analysis and model selection. Keynote Address, International Meeting of the Psychometric Society, Montreal, CA, June 2006.

Stevenson, C., & Williams, L. (2014). Type D personality, quality of life and physical symptoms in the general population: A dimensional analysis. *Psychology & health*, *29(3),* 365-373.

Straat, J. H., van der Ark, L. A., & Sijtsma, K. (2012). Multi-method analysis of the internal structure of the Type D Scale-14 (DS14). *Journal of Psychosomatic Research, 72(4),* 258-265.

Sumin, A. N., Raikh, O. I., Gaifulin, R. A., Korok, E. V., Bezdenezhnykh, A. V., Ivanov, S. V., & Barbarash, O. L. (2015). Predisposition to psychological distress in patients after coronary bypass surgery: relation to one year prognosis. *Kardiologiia*, *55*(10), 76-82.

Suls, J. (2014). An Informative Failure to Replicate—the Type D is Distressed: a Comment on Meyer et al. *Annals of Behavioral Medicine, 48(2),* 140-141.

Svansdottir, E., Karlsson, H. D., Gudnason, T., Olason, D. T., Thorgilsson, H., Sigtryggsdottir, U., ... & Denollet, J. (2012). Validity of Type D personality in Iceland: association with disease severity and risk markers in cardiac patients. *Journal of behavioral medicine, 35(2),* 155-166.

Svansdottir, E., van den Broek, K. C., Karlsson, H. D., Olason, D. T., Thorgilsson, H., & Denollet, J. (2013). The distressed (Type D) and Five-Factor Models of personality in young, healthy adults and their association with emotional inhibition and distress. *Personality and individual differences*, *55*(2), 123-128.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*. Allyn & Bacon/Pearson Education.

Talaei-Khoei, M., Mohamadi, A., Fischerauer, S. F., Ring, D., & Vranceanu, A. M. (2018). Type D personality in patients with upper extremity musculoskeletal illness: Internal consistency, structural validity and relationship to pain interference. *General hospital psychiatry, 50,* 38-44.

Team, R. C. (2013). R: A language and environment for statistical computing.

ten Holt, J. C. T., van Duijn, M. A., & Boomsma, A. (2010). Scale construction and evaluation in practice: A review of factor analysis versus item response theory applications. Psychological Test and Assessment Modeling, 52(3), 272–297.

Thoresen, M. (2019). Spurious interaction as a result of categorization. *BMC medical research methodology*, *19*(1), 1-8.

Thurstone, L. (1935). *The vectors of mind*. Chicago, IL: University of Chicago Press.

Traub, R. (1997). Classical Test Theory in Historical Perspective. Educational Measurement: Issues and Practice 16 (4), 8–14. doi:doi:10.1111/j.1745-3992.1997.tb00603.x

Vale, C. D., & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika*, *48*, 465–471. https://doi.org/10.1007/ BF02293687

van Assen, M. A., van Aert, R., & Wicherts, J. M. (2015). Meta-analysis using effect size distributions of only statistically significant studies. *Psychological methods, 20(3),* 293.

van de Poll-Franse, L. V., Horevoorts, N., van Eenbergen, M., Denollet, J., Roukema, J. A., Aaronson, N. K., ... & Mols, F. (2011). The patient reported outcomes following initial treatment and long term evaluation of survivorship registry: scope, rationale and design of an infrastructure for the study of physical and psychosocial outcomes in cancer survivorship cohorts. *European Journal of Cancer, 47(14),* 2188-2194.

van de Ven, M. O., Witteman, C. L., & Tiggelman, D. (2013). Effect of Type D personality on medication adherence in early adolescents with asthma. *Journal of psychosomatic research, 75(6),* 572-576.

van den Berg, S. M., Glas, C. A., & Boomsma, D. I. (2007). Variance decomposition using an IRT measurement model. *Behavior genetics, 37(4),* 604-616.

van Dooren, F. E., Verhey, F. R., Pouwer, F., Schalkwijk, C. G., Sep, S. J., Stehouwer, C. D., ... & Denollet, J. (2016). Association of Type D personality with increased vulnerability to depression: Is there a role for inflammation or endothelial dysfunction?–The Maastricht Study. *Journal of affective disorders*, *189*, 118-125.

van Middendorp, H., Kool, M. B., van Beugen, S., Denollet, J., Lumley, M. A., & Geenen, R. (2016). Prevalence and relevance of Type D personality in fibromyalgia. *Generalhospital psychiatry, 39,* 66-72.

van Montfort, E., Kupper, N., Widdershoven, J., & Denollet, J. (2018). Person-centered analysis of psychological traits to explain heterogeneity in patient-reported outcomes of coronary artery disease–the THORESCI study. *Journal of affective disorders, 236,* 14-22.

van Steenbergen-Weijenburg, K. M., de Vroege, L., Ploeger, R. R., Brals, J. W., Vloedbeld, M. G., Veneman, T. F., ... van der Feltz-Cornelis, C. M. (2010). Validation of the PHQ-9 as a screening instrument for depression in diabetes patients in specialized outpatient clinics. BMC Health Services Research, 10(1), 235. doi:10.1186/ 1472-6963-10-235

van Zwet, E. (2019). A default prior for regression coefficients. *Statistical methods in medical research*, *28*(12), 3799-3807.

Verdinelli, I., & Wasserman, L. (1995). Computing Bayes factors using a generalization of the Savage-Dickey density ratio. *Journal of the American Statistical Association*, *90*(430), 614-618.

Versteeg, H., Spek, V., Pedersen, S. S., & Denollet, J. (2012). Type D personality and health status in cardiovascular disease populations: a meta-analysis of prospective studies. *European journal of preventive cardiology*, *19*(6), 1373-1380.

Vogelsmeier, L. V., Vermunt, J. K., van Roekel, E., & De Roover, K. (2019). Latent Markov factor analysis for exploring measurement model changes in time-intensive longitudinal studies. *Structural Equation Modeling: A Multidisciplinary Journal, 26(4),* 557-575.

Volz, A., Schmid, J. P., Zwahlen, M., Kohls, S., Saner, H., & Barth, J. (2011). Predictors of readmission and health related quality of life in patients with chronic heart failure: a comparison of different psychosocial aspects. *Journal of behavioral medicine*, *34*(1), 13-22.

Waller, N. G. (2016). *Fungible coefficients and Monte Carlo functions*. R-package. https://cran.r-project.org/web/packages/fungible/index. html

Wang, Y., Liu, G., Gao, X., Zhao, Z., Li, L., Chen, W., ... & Lin, P. (2018). Prognostic value of type D personality for in-stent restenosis in coronary artery disease patients treated with drug-eluting stent. *Psychosomatic medicine*, *80*(1), 95-102.

Wang, Y., Zhao, Z., Gao, X., Li, L., Liu, G., Chen, W., ... & Lin, P. (2016). Type D personality and coronary plaque vulnerability in patients with coronary artery disease: an optical coherence tomography study. *Psychosomatic Medicine*, *78*(5), 583-592.

Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. The Journal of Strength & Conditioning Research, 19(1), 231-240.

Weiss, E. S., Anderson, R. M., & Lasker, R. D. (2002). Making the most of collaboration: exploring the relationship between partnership synergy and partnership functioning. *Health Education & Behavior, 29(6),* 683-698.

Whetzel, D. L., McDaniel, M. A., Yost, A. P., & Kim, N. (2010). Linearity of personality performance relationships: A large-scale examination. *International Journal of Selection and Assessment*. https://doi.org/10.1111/j.1468-2389.2010.00514.x.

Whitehead, A. N. (1925). *Science and the modern world*. Macmillan.

Whitehead, D. L., Perkins-Porras, L., Strike, P. C., Magid, K., & Steptoe, A. (2007). Cortisol awakening response is elevated in acute coronary syndrome patients with type-D personality. Journal of Psychosomatic Research, 62(4), 419–425. doi:10.1016/j.jpsychores.2006. 11.005

Wiencierz, S., & Williams, L. (2017). Type D personality and physical inactivity: the mediating effects of low self-efficacy. Journal of health psychology, 22(8), 1025-1034.

Williams, L., Abbott, C., & Kerr, R. (2016). Health behaviour mediates the relationship between Type D personality and subjective health in the general population. *Journal of health psychology, 21(10),* 2148-2155.

Williams, L., Bruce, G., & Knapton, C. (2018). Type D personality is associated with increased desire for alcohol in response to acute stress. *Stress and Health, 34(3),* 411-415.

Williams, L., O'Connor, R. C., Grubb, N. R., & O'Carroll, R. E. (2011a). Type D personality and illness perceptions in myocardial infarction patients. *Journal of Psychosomatic Research, 70(2),* 141-144.

Williams, L., O'Connor, R. C., Grubb, N. R., & O'Carroll, R. E. (2011b). Type D personality predicts poor medication adherence in myocardial infarction patients. *Psychology & health, 26(6),* 703-712.

Williams, L., O'Connor, R. C., Grubb, N. R., & O'Carroll, R. E. (2012). Type D personality and three-month psychosocial outcomes among patients post-myocardial infarction. *Journal of psychosomatic research, 72(6),* 422-426.

Wright, D. B. (2003). Making friends with your data: Improving how statistics are conducted and reported 1. *British Journal of Educational Psychology, 73(1),* 123-136.

Wu, J. R., & Moser, D. K. (2014). Type D personality predicts poor medication adherence in patients with heart failure in the USA. *International journal of behavioral medicine, 21(5),* 833-842.

Wulsin, L. R., & Singal, B. M. (2003). Do depressive symptoms increase the risk for the onset of coronary disease? A systematic quantitative review. Psychosomatic Medicine, 65(2), 201–210. doi:10.1097/01.PSY.0000058371. 50240.E3

York, K. L., & John, O. P. (1992). The four faces of eve: a typological analysis of women's personality at midlife. *Journal of personality and social psychology*, *63*(3), 494.

Yu, D. S. F., Thompson, D. R., Yu, C. M., Pedersen, S. S., & Denollet, J. (2010). Validating the Type D personality construct in Chinese patients with coronary heart disease. *Journal of psychosomatic research, 69(2),* 111-118.

Yuan, K. H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. Sociological Methodology, 30(1), 165–200. doi:10.1111/0081-1750.00078

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta psychiatrica scandinavica, 67(6)*, 361-370.

Zohar, A. H. (2016). Is type-D personality trait (s) or state? An examination of type-D temporal stability in older Israeli adults in the community. *PeerJ, 4*, e1690.

Zuccarella-Hackl, C., von Känel, R., Thomas, L., Kuebler, P., Schmid, J. P., Mattle, H. P., ... & Wirtz, P. H. (2016). Higher macrophage superoxide anion production in coronary artery disease (CAD) patients with Type D personality. *Psychoneuroendocrinology, 68,* 186-193

LIST OF PUBLICATIONS

Publications included in this thesis:

1. **Lodder**, P. (2020). Modeling synergy: How to assess a Type D personality effect? *Journal of Psychosomatic Research.*
2. **Lodder**, P. (2020). A re-evaluation of the Type D personality effect. *Personality and Individual Differences*.
3. **Lodder**, P., Kupper, N., Antens, M., & Wicherts, J.M. (2021). A systematic review comparing two popular methods to assess a Type D personality effect. *General Hospital Psychiatry.*
4. **Lodder,** P, Kupper, N., Antens, ..., …, & Wicherts, J.M. (2022). Type D personality as a risk factor for adverse outcome in patients with cardiovascular disease: an individual patient data meta-analysis *Psychosomatic Medicine.*
5. **Lodder**, P., Denollet, J., Emons, W.H.M., Nefs, G., Pouwer, F., Speight, J., & Wicherts, J.M. (2019). Modeling interactions between latent variables in research on Type D personality: A Monte Carlo simulation and clinical study of depression and anxiety. *Multivariate Behavioral Research*.
6. **Lodder**, P., Denollet, J., Emons, W.H.M, & Wicherts, J.M. (2020). Latent logistic interaction modeling: A simulation and empirical illustration of Type D personality. *Structural Equation Modeling.*
7. **Lodder**, P., Kupper, N., Mols, F., Emons, W.H.M., & Wicherts, J.M. (2022). Assessing the temporal stability of psychological constructs: An illustration of Type D personality, anxiety and depression. *Journal of Research in Personality.*

Other publications:

8. **Lodder**, P. (2022) Why researchers should not ignore skewness and measurement error in questionnaire item scores. *Religion, Brain & Behavior.*
9. Van Helmondt, S.J., **Lodder**, P., van Woezik, R., van der Lee, M.L., & de Vries. J. (2022). CBT-based online self-help training to reduce fear and distress after cancer (CAREST randomized trial): 24 months follow-up using latent growth models and latent class analysis. *Annals of Behavioral Medicine.*
10. Mertens, G., **Lodder**, P., Smeets, T., & Duijndam, S.N.C. (2022). Pandemic panic? Results of a 6-month longitudinal study on fear of COVID-19. *Journal of Affective Disorders.*
11. Mertens, G., **Lodder**, P., Smeets, T., & Duijndam, S.N.C. (2022). Fear of COVID-19 predicts vaccination willingness 14 months later. *Journal of Anxiety Disorders.*
12. Yang, F., **Lodder**, P., Fu, M., & Guo, J. (2022). Six-year Changes in the Comorbidity of Posttraumatic Stress Symptoms and Depressive Symptoms among Chinese Earthquake Survivors: A network analysis. *Journal of Affective Disorders.*
13. Hoogeveen, S., Sarafoglu, A., …. , … , **Lodder**, P., …., …., & Wagenmakers, E.J. (2022). A many-analysts approach to the relation between religiosity and well-being. *Religion, Brain & Behavior.*
14. Firanescu, C.E., Venmans, A., de Vries, J., **Lodder**, P.,..., & Lohle, P. (2022). Predictive factors for sustained pain 12 months after an acute osteoporotic vertebral fracture in patients treated with conservative therapy, sham procedure

or vertebroplasty. Combined results from the VERTOS 2 and VERTOS 4 trials. *CardioVascular and Interventional Radiology.*
15. Hulschbosch, L.P., Boekhorst, M.G.B.M., **Lodder**, P., Potharst, E., Nyklicek, I., Bergink, V., Guid Oei, S., Verhoeven, C.J., & Pop, V.J.M. (2022). High levels of comorbid anxiety and depression symptoms throughout pregnancy decrease the likelihood of physiological birth. *BJOG.*
16. Hulsbosch, L.P., Boekhorst, M.G.B.M., **Lodder**, P., Potharst, E., Nyklicek, I., Bergink, V., Guid Oei, S., Verhoeven, C.J., & Pop, V.J.M. (2022). Antenatal anxiety and depressive symptoms and physiological birth. *BJOG.*
17. Yang, F., Wen, J., Huang, N., Riem, M.M.E., **Lodder**, P., & Guo, J. (2022). Prevalence and Related Factors of Child Posttraumatic Stress Disorder during COVID-19 Pandemic: A Systematic Review and Meta-analysis. *European Psychiatry.*
18. Riem, M.M.E., **Lodder**, P., Guo, J., Vrielink-Verpaalen, M., van IJzerendoorn, M.R., Bakermans-Kranenburg, M.J., de Carli, P. (2021) Predictive models of maternal harsh parenting during COVID-19 in China, Italy, and the Netherlands. *Frontiers in Psychiatry: child and adolescent psychiatry.*
19. Persons, J.E., **Lodder**, P., Coryell, W.H., Nurnberger, J.I., & Fiedorowicz, J.G. (2021). Estimating the Contribution of Symptom Clusters to Risk of Suicidal Ideation or Behavior in Bipolar Disorder. *Psychiatry research.*
20. Guo, J., de Carli, P., **Lodder**, P., Bakermans-Kranenburg, M.J., Riem, M.M.E. (2021) Maternal mental health during the COVID-19 lockdown in China, Italy, and the Netherlands: A cross validation study. *Psychological Medicine*.
21. Visser, E., den Oudsten, B.L, **Lodder**, P., Gosens, T., & de Vries, J. (2021). Psychological risk factors that characterize trajectories of posttraumatic stress disorder after injury: a study using latent class analysis. *European Journal of Psychotraumatology.*
22. Nguyen, L.A., Pouwer, F., **Lodder**, P., Hartman, E.E., Winterdijk, P., Aanstoot, H.-J., & Nefs, G. (2021). Depression and anxiety in adolescents with type 1 diabetes and their parents: a longitudinal study. *Pediatric research.*
23. Van Helmondt, S.J., van der Lee, M.L., Bisseling, E.M, **Lodder**, P., de Vries. J. (2021). Factor analysis of the Dutch version of the Fear of Cancer Recurrence Inventory (FCRI-NL): secondary analysis using three predominantly breast cancer samples. *European Journal of Cancer Care.*
24. Riem, M.M.E., De Carli, P., Guo, J., Bakermans-Kranenburg, M.J., van IJzendoorn, M.H., **Lodder**, P. (2021). Internet searches for terms related to child maltreatment increase during COVID-19. *JMIR Pediatrics and Parenting*.
25. Mertens, G., Duijndam, S.N.C., Smeets, T., & **Lodder**, P. (2021). The latent and item structure of COVID-19 fear: A comparison of four COVID-19 fear questionnaires using SEM and network analyses. *Journal of Anxiety Disorders*.
26. Gu, J., Yang, F., Fu, M., Huang, N., Ahmed, F., Shahid, M., Zhang, B. & **Lodder**, P. (2021). Network Analysis of COVID-19-related PTSD Symptoms in China: the similarities and differences between the general population and PTSD sub-population. *European Journal of Psychotraumatology.*
27. Peters, C.M.L., **Lodder**, P., de Vries, J., Steunenberg, S.L., Veen, E.J., de Groot, H.G.W., & van der Laan, L. (2020). Two-year Outcome of Quality of Life and

Health Status of treatment for the Elderly with Critical Limb Ischemia. *Clinical Interventions in Aging.*
28. Hafkamp, F.J., **Lodder**, P., de Vries, J., Gosens, T., den Oudsten, B.L. (2020). Characterizing patients' expectations in hip and knee osteoarthritis. *Clinical Orthopaedics and Related Research.*
29. Janssen, T.L., **Lodder**, P., de Vries, J., van Hoof - de Lepper, C.C.H.A., Gobardhan, P.D., Ho, G.H., van der Laan, L. (2020). Caregiver strain index and factors influencing strain on informal caregivers when providing care for elderly patients undergoing major abdominal surgery. *BMC Geriatrics.*
30. Van der Duin, P., **Lodder**, P., & Snijders, D. (2020). Dutch doubts and desires: Exploring citizen opinions on future and technology. *Futures.*
31. Smaardijk, V.R., Maas, A., **Lodder**, P., Kop, W.J., van Gennep, B., & Mommersteeg, P. (2020). Sex and gender-stratified risks of psychological factors for ischemic heart disease and prognosis: a meta-analysis. *International Journal of Cardiology.*
32. Janssen, T.L., de Vries, J., **Lodder**, P., Ho, G.H., Gobardhan, P.D., van der Laan, L. (2020). The impact of major abdominal surgery and postoperative delirium on long-term quality of life, cognitive dysfunction and depressive symptoms in elderly surgical patients. *Aging and Mental Health.*
33. Broers, E.R., Widdershoven, J., Denollet, J., **Lodder**, P., Kop, W.J., Wetzels, M., Ayoola, I., Piera-Jimenez, J., Habibovic, M. (2020). Personalized eHealth Program for Lifestyle Change: Results from the 'Do Cardiac Health Advanced New Generated Ecosystem (Do CHANGE 2)'- Randomized Controlled Trial. *Psychosomatic Medicine.*
34. Visser, E., den Oudsten, B.L., Gosens, T., **Lodder**, P. & de Vries, J. (2020). Psychological risk factors characterize the trajectories of quality of life after a physical trauma: a longitudinal study using latent class analysis. *Quality of Life Research.*
35. de Mol, M., Visser, S., Aerts, J., **Lodder**, P., van Walree, N., Belderbos, H., & den Oudsten, B. (2020). Depressive symptoms and performance status are associated with (Health-Related) Quality of Life in patients with advanced stage lung cancer: an observational multi-center cohort study. *BMC Cancer.*
36. Roijers, J.P., van den Houten, M.M., Hopmans, C.J., Vriens, W.H.E., Willigendael, E.M., **Lodder**, P., de Vries, J., Teijink, J.A., & van der Laan, L. (2020). A Comparison of Quality of Life in Patients with Claudication and Critical Limb Ischemia. *Annals of Vascular Surgery.*
37. **Lodder**, P., Ong, H.H, Grasman, R.P.P.P., & Wicherts, J.M. (2019). A Comprehensive Meta-analysis of Money Priming. *Journal of Experimental Psychology: General.*
38. Peerbooms, J.C., **Lodder**, P., den Oudsten, B.L., Faber, F., Schuller, H.M., van der Hoeven, H., & Gosens, T. (2019). Use of platelet rich plasma to treat plantar fasciitis: a multi-centre randomized controlled trial. *American Journal of Sports Medicine.*
39. Smaardijk, V.R., **Lodder**, P., Kop, W.J., Maas, A., & Mommersteeg, P. (2019). Sex and gender-sensitive risks of psychosocial factors for ischemic heart disease incidence and prognosis: A systematic review and meta-analysis. *Journal of the American Heart Association.*

40. Van der Duin, P., & **Lodder**, P. (2019). De toekomst is ook maar een mening. In: Hodemaekers, L. (Ed.) *STAD 2019: Technologische en sociale innovatie*. Publicatiereeks Overheid & Arbeid, 2019, Nummer 50.
41. Van der Duin, P., Snijders, D., & **Lodder**, P. (2019). De Nationale Toekomstmonitor 2018. Hoe denken Nederlanders over technologie en toekomst? *Stichting Toekomstbeeld der Techniek.*
42. Broers, E.R., **Lodder**, P., Spek, V.R.M., Denollet, J., Widdershoven, J.W.M.G., Habibovic, M. (2019). Healthcare utilization in patients with first-time implantable cardioverter defibrillators (Data from the WEBCARE study*). Pacing and Clinical Electrophysiology.*
43. Raaijmakers, C.P.A.M., Lohle, P.N.M., **Lodder**, P., de Vries. J. (2019). Quality of Life and Clinical Outcome after Traumatic Spleen Injury (SPLENIQ study): Protocol for an Observational Retrospective and Prospective Cohort Study. *JMIR Research Protocols.*
44. Van Helmondt, S.J., van Woezik, R., **Lodder**, P., van der Lee, M.L., & de Vries, J. (2019). No effect of CBT-based online self-help training to reduce fear of cancer recurrence: First results of the CAREST multicenter randomized controlled trial. *Psycho-oncology.*
45. Voortman, M., Hendriks, C.M.R., **Lodder**, P., Drent, M., & de Vries, J. (2019). Quality of life of couples living with sarcoidosis. *Respiration.*
46. Firanescu, C.E., de Vries, J., **Lodder**, P., et al. (2019). Percutaneous vertebroplasty is no risk factor for new vertebral fractures, and protects against further height loss (VERTOS IV). *CardioVascular and Interventional Radiology.*
47. Peters, C.M.L., de Vries, J., Lodder, P., & van der Laan, L. (2019). Is a good quality of life and health status possible in elderly patients dying from critical limb ischemia: a prospective clinical study. *Annals of Vascular Surgery.*
48. De Mol, M., Visser, S., Den Oudsten, B.L., **Lodder**, P., Van Walree, N.C., Belderbos, H.N.A., & Aerts, J.G.J.V. (2019). Frequency of low grade adverse events and quality of life during chemotherapy determine patients' judgement about treatment in advanced stage thoracic cancer. *Supportive Care in Cancer.*
49. Steunenberg, S.L., de Vries, J., Raats, J.W., Verbogt, N., **Lodder**, P., et al. (2019) Important differences between quality of life and health status in elderly patients suffering from critical limb ischemia. *Clinical interventions in aging.*
50. Peters, C.M.L., de Vries, J., Veen, E.J., de Groot, H.G., Gwan, H., **Lodder**, P., Steunenberg, S.L., & van der Laan, L. (2019). Is amputation in the elderly with critical limb ischemia acceptable in the Long-term? *Clinical interventions in aging.*
51. Van Elk, M. & **Lodder**, P. (2018). Experimental Manipulations of Personal Control do Not Increase Magical Thinking and Illusory Pattern Perception. *Collabra: Psychology.*
52. Firanescu, C.E., de Vries, J., **Lodder**, P., et al. (2018). A randomized sham controlled trial of vertebroplasty for painful acute osteoporotic vertebral compression fractures (VERTOS IV). *British Medical Journal.*
53. Peters, C.M.L., de Vries, J., **Lodder**, P., Steunenberg, S.L., Veen, E.J., de Groot, H.G., Gwan, H. & van der Laan, L. (2018). Quality of life and not health status improves after major amputation in the elderly critical limb ischemia patient. *European Journal of Vascular and Endovascular Surgery.*

54. Denollet, J., van Felius, R.A., **Lodder**, P., Mommersteeg, P.M., Goovaerts, I. et al. (2018) Predictive value of Type D personality for impaired endothelial function in coronary artery disease – The SAINTEX-CAD Study. *International Journal of Cardiology.*
55. De Mol, M., Visser, S., Aerts, J.G.J.V., **Lodder**, P., De Vries, J., Den Oudsten, B.L. (2018). Satisfactory results of a psychometric analysis and calculation of minimal clinically important differences of the World Health Organization quality of life-BREF questionnaire in an observational cohort study with lung cancer and mesothelioma patients. *BMC Cancer.*
56. van Broekhoven, K. E. M., Karreman, A., Hartman, E. E., **Lodder**, P., Endendijk, J. J., Bergink, V., & Pop, V. (2018). Obsessive-compulsive personality disorder symptoms as a risk factor for postpartum depressive symptoms. *Archives of Women's Mental Health.*
57. Peters, C.M.L., de Vries, J., Steunenberg, S.L., Ho, G.H, **Lodder**, P. & van der Laan, L. (2018). Is there an important role for anxiety and depression in the elderly patient with critical limb ischemia, especially after major amputation? *Annals of Vascal Surgery.*
58. Steunenberg, S.L., de Vries, J., Raats, J.W., Thijsse, W.J., Verbogt, N.,**Lodder**, P. et al. (2018). Quality of life after endovascular, surgical or conservative treatment of elderly patients suffering from critical limb ischemia, short-term results. *Annals of Vascular Surgery.*
59. Creemers, A., Ter Veer, E., de Waal, L., **Lodder**, P., Hooijer, G.K.J., van Grieken, N.C.T., Bijlsma, M.F., Meijer, S.L., van Oijen, M.G.H., van Laarhoven, H.W.M. (2017). Discordance in HER2 Status in Gastro-esophageal Adenocarcinomas: A Systematic Review and Meta-analysis. *Scientific Reports*.
60. de Rooij, B.H., Ezendam, N.P.M., Nicolaije, K.A.H., **Lodder**, P., et al.; M. (2017). Survivorship Care Plans have a negative impact on long-term quality of life and anxiety through more threatening illness perceptions - the ROGY care trial. *Quality of Life research.*
61. ter Veer, E., Mohammad, N.H., **Lodder**, P., Ngai, L., Samaan, M., van Oijen, M. & van Laarhoven, H. (2016). The Efficacy and Safety of S-1 based regimens in the First-line treatment of Advanced Gastric Cancer: A Systematic Review and Meta-analysis. *Gastric Cancer.*
62. **Lodder**, P. (2014). Comparing item imputation methods in questionnaire research. In: G.J. Mellenbergh & H.J. Adèr (Eds.). *Advising on research methods: Selected topics 2013*, Huizen: Johannes van Kessel Publishing.
63. **Lodder**, P., Rotteveel, M. & van Elk, M. (2014). Enactivism and neonatal imitation: conceptual and empirical considerations and clarifications. *Frontiers in Psychology.*

DANKWOORD (acknowledgements)

In 1675 schreef Isaac Newton in een brief aan een collega: 'Als ik verder heb gezien, dan was dit doordat ik op de schouders van reuzen stond'. De inhoud van dit proefschrift valt niet te vergelijken met het werk van Isaac Newton, maar ook mijn bijdrage aan de wetenschap had niet zo mooi kunnen zijn zonder het werk en de steun van anderen. Elke interactie met anderen heeft me gevormd tot wie ik nu ben, maar in dit dankwoord wil ik graag een aantal personen in het bijzonder bedanken omdat ze een onmisbare rol hebben gespeeld in mijn leven, of in de totstandkoming van dit proefschrift, of beide.

Prof. Dr. Denollet, beste *Johan,* wat doet het me verdriet dat u ons hebt moeten verlaten. De wereld is een top wetenschapper en een heel warm mens kwijtgeraakt. Ik heb niet alleen veel van u geleerd over persoonlijkheid en cardiologie, maar heb ook erg fijne gesprekken met u gehad en veel nieuwe muziek ontdekt. Regelmatig leende u me een CD uit uw uitgebreide collectie waarvan u dacht dat ik die wel kon waarderen. En dat klopte zonder uitzondering! Als begeleider stond u altijd volledig voor mij klaar en bleef u interesse in mij tonen, zelfs toen uw eigen gezondheid steeds minder werd. Tot op het laatste moment reisde u vanuit Vlaanderen naar Tilburg om vol passie mee te denken over het onderzoek waar u zelf de grondlegger van bent. Ik zal u nooit vergeten en ben dankbaar voor de jaren waarin ik u als promotor heb mogen meemaken. Ik draag dit proefschrift op aan u.

Prof. Dr. Wicherts*,* beste *Jelte*, toen ik in 2009 begon met de studie Psychologie aan de Universiteit van Amsterdam was jij van mijn eerste docenten. Ik had die opleiding niet gekozen voor de methoden en statistiek vakken, maar mede dankzij jouw colleges ontdekte ik hoe belangrijk het is om met een kritische blik te kijken naar deze aspecten van wetenschappelijk onderzoek. In het laatste jaar van mijn Bachelor inspireerde je me opnieuw, maar ditmaal door me te leren over meta-analyse, wat ertoe heeft geleid dat je twee jaar later de begeleider werd van mijn masterthese over dit onderwerp. Ik ontdekte dat je hele onderzoekslijn aansloot bij mijn interesses en ik was dan ook ontzettend blij dat je me uitnodigde om als je PhD student te beginnen in Tilburg. Ik keek altijd erg uit naar onze meetings. Niet alleen heb ik enorm veel van je geleerd over een divers aantal

methodologische onderwerpen, maar ik heb ook erg veel gelachen om je mooie en grappige verhalen over je vele opmerkelijke ervaringen in de academische wereld. Ik hoop dan ook dat er nog veel nieuwe samenwerkingen gaan volgen.

Dr. Emons, beste *Wilco*, bedankt voor je waardevolle hulp tijdens mijn promotietraject. Je hebt vooral in de begin- en eindfase ervan een belangrijke rol gespeeld. Op het einde heb je in vrij korte tijd het hele proefschrift doorgenomen en tot in de details voorzien van nuttige feedback. Met je expertise in psychometrie en eerdere ervaring met onderzoek naar Type D persoonlijkheid ben je als geen ander geschikt om dit proefschrift te begeleiden. Ik ben blij dat je na een aantal jaar bij het CITO gewerkt te hebben nu weer terug bent in Tilburg en ik kijk uit naar de toekomstige samenwerkingen die zeker nog zullen volgen.

Dr. Kupper, beste *Nina*, met name in de eindfase van dit promotietraject heb jij een belangrijke rol gespeeld. Na het overlijden van Johan was er een nieuwe begeleider nodig met verstand van Type D persoonlijkheid en medische psychologie in het algemeen. Het is niet zomaar wat om in de voetsporen te treden van iemand zoals Johan, maar ik denk dat hij het met me eens zou zijn geweest dat niemand zijn rol beter had kunnen overnemen dan jij. En dat vind ik niet alleen op inhoudelijk vlak maar ook als begeleider, waarbij je net als Johan altijd klaar staat voor je promovendi. Veder bewonder ik hoe je in je eigen onderzoek regelmatig geavanceerde statistiek en psychometrie toepast. Bedankt voor alles en ik kijk uit naar onze verdere samenwerkingen.

Many thanks to all members of the promotion committee: *Dr. Stefanie Duijndam*, *Prof. dr. Brian Hughes*, *Prof. dr. Timothy Smith*, *Dr. Mathilde Verdam* and *Prof. dr. Jeroen Vermunt*. Thank you for taking the time and effort to read my dissertation. I am looking forward to discussing my dissertation with you during the public defence on January 27th 2023.

Over the years I have had the privilege to collaborate with so many stellar scientists. First, I would like to thank those who collaborated with me on this dissertation: *Prof. dr. Albus, Mr. Antens*, *Dr. Bessonov, Dr. Dulfer, Dr. Condén, Dr. Gostoli, Dr. Grande, Dr. Hedberg, Prof. dr. Herrmann-Lingen, Prof. dr. Jaarsma, Dr. Koo, Dr. Lin, Prof. dr. Meyer, Dr. Mols*, *Dr. Nefs*, *Prof. Dr. Pouwer*, *Dr. Pushkarev, Dr. Rafanelli, Dr. Raykh, Dr. Schaan de Quadros, Dr. Schmidt,*

Prof. dr. Speight, *Dr. Sumin, Prof. dr. Utens, Prof. dr. van Veldhuisen,* and *Dr. Wang*. Regarding the studies not included in this dissertation, I would like to thank all collaborators for involving me in their excellent research. Special thanks to *Prof. dr. Jolanda de Vries* for introducing me as a statistician in various medical research projects. I would also like to thank *Dr. Michiel van Elk* for being an excellent mentor when I was still studying at the University of Amsterdam and for helping me write my first academic publication. Finally, thanks to *Prof. dr. Jess Fiedorowicz* for putting your trust in me as the statistical editor of the Journal of Psychosomatic Research.

Many thanks to all (ex-)colleagues at the Department of Methodology and Statistics and the Department of Medical and Clinical Psychology. Thanks for the pleasant work environment, for the nice conversations over lunch or coffee, and for inspiring me with your excellent research. Special thanks to *Marcel* for letting us know when it is time to have lunch ("happie hap"), to *Wijo* for sharing great music (e.g., Laurie Anderson's Big Science), and to *Guy* for being a great colleague and mentor when it comes to teaching. Last but not least, thanks to *Anne-Marie*, *Saskia* and *Kitty* for being the heart of our departments.

A shout-out to the many roomies I have had for the past several years. *Belle, Edoardo, Elise, Eva, Florian, Hannah, Leonie, Pia, Ruifang, Wietske,* en *Xynthia,* thanks for making work even more fun than it already is. Special thanks to my current roomies *Lisa* ("Eindhovuuuh") and *Caspar* ("fellow synth geek"). I would also like to thank *Chris* for the many inspiring conversations and for raving together at the Berghain. Also special thanks to my former roomies *Niek, IJsbrand,* and *Soogeun* for the many laughs and beers. We should meet again someday to finish our track "Dikke pijp", which will definitely crush all hit charts across the globe. *IJsbrand*, I'm honored that you will perform as *DJ IJsbeer* at my promotion party.

MKPartypeople, zonder jullie had promoveren een stuk saaier geweest. *Anouk, Belle, Cynthia, Daniëlle, Dinah, Emma, Eva, Eveline, Frederique, Gabriela, Gaetan, Isabel, Ivon, Ivy, Janniek, Jori, Josine, Kiki, Laura (Likelihood), Lianne, Linh, Lisa, Lotte, Manon, Maria, Marijn, Milou, Myrthe, Nico, Noor, Sandra, Simone, Sophie, Stefanie, Tom IJ, Tom R, en Veerle*, bedankt voor alle borrels (zelfs op dinsdagavond, waardoor ik woensdag vrij moest nemen), voor de drie fantastische schrijfweken en het dansen in dierenpak tijdens de jungleparty.

Robbie en *Tom*, bedankt dat jullie mijn paranimfen willen zijn. *Robbie*, ik heb je leren kennen als een heel gezellige collega met een nuchtere humor die ik erg kan waarderen. Je bent iemand die zijn zaken goed op orde heeft, altijd bereid om met me mee te denken, en als geen ander weet je de weg te vinden door het doolhof van het intranet als ik weer eens een formulier nodig heb. Je bent één van de sportiefste collega's die ik ken en ik hoop ooit nog eens de Warandeloop mee te kunnen lopen. Met de "dynaband instructievideo" heb je laten zien een terechte Fit-at-Work ambassadeur te zijn. Ik heb erg genoten van ons weekendje met Niek naar Chris in Berlijn en ik hoop dat we nog lang collega's zullen blijven.

Tom, ik vind het ontzettend tof dat we op 27 januari allebei ons proefschrift mogen verdedigen en dit samen groots gaan vieren. Toen we samen zochten naar een geschikte fotograaf en locaties voor een borrel, diner en feest, voelde het bijna alsof we zouden gaan trouwen. Ik ben blij met de vriendschap ("bromance") die de afgelopen jaren is ontstaan. We hebben dezelfde smaak wat betreft bier, muziek en kerstbomen ("breed in de heupen"). Ondanks dat we geen collega's meer zijn ben ik ervan overtuigd dat we in de toekomst nog veel lekkere biertjes zullen drinken (waaronder het bier dat we zelf voor de promotie hebben gebrouwen) en nog veel plaatjes zullen draaien als dj *t*-test en dj *p*-value (ik heb de diepe dreunen playlist weer aangevuld, dus let's go!).

Als afleiding van het drukke leven als promovendus bracht ik graag tijd door in de natuur, het liefst in het bijzijn van vrienden. Ik wil daarom *de Pastoor, Panzermeier, Zwerver, Pooly, Mindster* en *Vorticon* bedanken voor de fantastische bergtochten die we hebben gelopen de afgelopen jaren, van Winterberg (zou die banaan nog aan de muur zitten?), tot de Spaanse Pyreneeën (wijn uit de kelk), de Sloveense Alpen (zelfgestookte likeur als ontbijt) en de Tour de Mont Blanc (soupe de carottes and cheese with spocke). Hopelijk gaan er nog veel mooie tochten volgen!

In het bijzonder wil ik ook graag mijn goede vrienden *Bertje, Jody, Mattie*, en *Yoshi* bedanken voor alle gezelligheid en goede gesprekken. *Matts, Tinyee, Bart, Sifra, Bregje,* en *Ruben,* bedankt voor de vele leuke uitjes en weekendjes weg (iets met een dolfijnentafel, duizenden vliegen en urenlang tafeltennissen). *Matts,* ik voel me vereerd dat jij als bet-achter-achter-achter-neef plaatjes wil draaien op mijn promotiefeest.

Lieve *Papa* en *Mama*, dit proefschrift zou zonder jullie nooit hebben bestaan. Bedankt voor jullie continue vertrouwen in mij en dat jullie me de kans hebben gegeven om na mijn opleiding Bedrijfskunde weer opnieuw te beginnen aan een opleiding Psychologie. Ik ben jullie eeuwig dankbaar en ben blij dat ik binnenkort weer bij jullie in de buurt zal wonen. Ook wil ik graag mijn lieve zusje *Ilse* bedanken voor haar onuitputtelijke interesse en positiviteit. We zien elkaar door de lange afstand niet vaak, maar als ik je zie dan geniet ik altijd weer van onze gesprekken over allerlei (vaak diepgaande) onderwerpen.

Er bestaan talloze grappen over moeilijke schoonmoeders, maar ik kan me geen betere schoonmoeder wensen dan *Els*. Bedankt dat je altijd voor mij en mijn gezin klaarstaat. Zonder je hulp met oppassen was dit proefschrift minimaal een hoofdstuk minder lang geweest. Verder wil ik *Oma Stad, Martijn, Henk*, *Wilma*, *Marijn, Nieke, Jesse, Denise, Bregje, Ruben* en *Ton* bedanken dat jullie een geweldige (schoon)familie zijn. Jullie bezoeken voelt iedere keer weer als thuiskomen.

Lieve *Noam* en *Juno*, jullie zijn allebei geboren in de jaren dat papa bezig was met dit proefschrift. Ik ben blij dat ik sinds 2020 het grootste deel van die tijd thuis heb kunnen werken. Na urenlang hard te hebben gewerkt is er niets mooier dan dat jullie me naar beneden roepen (“Paul, eten!”) en me daarna omhelzen alsof ik een week weg ben geweest. Jullie hebben me meer over mezelf geleerd dan een studie Psychologie ooit zou kunnen doen en jullie herinneren me elke dag aan wat er echt belangrijk is in het leven.

Lieve *Japke*, dit proefschrift bevat meer dan 125.000 woorden maar toch heb ik veel moeite om mijn oneindige dankbaarheid en liefde voor jou in woorden uit te drukken. Woorden zijn niet meer dan een zwakke afspiegeling van mijn werkelijke gevoelens voor jou en zullen daarom nooit de volledige lading dekken. Maar onze liefde heeft geen woorden nodig. Zoals Ludwig Wittgenstein zei: “Whereof one cannot speak, thereof one must be silent.”

Paul
November 2022

ABOUT THE AUTHOR

Paul Lodder was born on September 24th 1986 on the Dutch island Goeree-Overflakkee. After graduating pre-university education at the RGO in Middelharnis, he obtained a Bachelor's degree in Business Administration and a Master's in Human Resource Management at Erasmus University Rotterdam. His Master's thesis was a conceptual replication of Ab Dijksterhuis' experiment testing the influence of unconscious thought on decision making. Around that time Paul developed a passion for psychological science, motivating him to further his education. At the University of Amsterdam, Paul obtained a Bachelor's degree and a Research Master's degree in Psychology (both cum laude). For several years, he was the chair of 'de Methodologiewinkel', a group of students who voluntarily give methodological and statistical advice to psychology students at the University of Amsterdam. In 2016, Paul's thesis supervisor Jelte Wicherts offered him a position as PhD candidate at Tilburg University. Paul has received the Elsevier Young Investigator Award from the European Association of Psychosomatic Medicine (EAPM) for research included in this dissertation. Currently, he works as an assistant professor in methodology and statistics at Tilburg University. His research focuses on applied statistics and psychometrics. He is also currently the statistical editor of the Journal of Psychosomatic Research and statistical consultant at the Elizabeth-TweeSteden Hospital.

PAUL LODDER

MEDICAL PSYCHOMETRICS

A psychometric evaluation of Type D personality and its predictive value in medical research

Type D personality–a combination of high negative affectivity and high social inhibition–has been identified as a risk factor for adverse outcome in various patient populations. However, common methods used to establish the predictive value of Type D personality have been criticized and several recent studies were not able to replicate previous findings. To explain these inconsistencies, this interdisciplinary dissertation brings together experts from the fields of medical psychology and psychometrics. It presents a psychometric evaluation of the construct Type D personality and illustrates how it can best be modeled in medical and psychological research. Based on thousands of computer-simulated datasets, as well as empirical data from patients with various types of diseases, this dissertation shows why most published research testing a Type D personality effect should be reanalyzed using modern psychometric and statistical methods. It also presents a first attempt at this endeavor by reanalyzing various earlier published datasets, showing that coronary artery disease patients with Type D personality are at increased risk for adverse outcome.