Latent logistic interaction modeling:

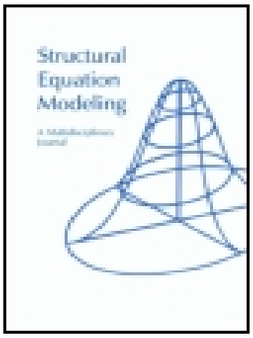
Lodder, Paul; Emons, Wilco H.M.; Denollet, Johan; Wicherts, Jelte M.

Published in:
Structural Equation Modeling

DOI:
10.1080/10705511.2020.1838905

Publication date:
2021

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



Routledge
Taylor & Francis Group

Structural Equation Modeling: A Multidisciplinary Journal

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hsem20

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

Paul Lodder , Wilco H.M. Emons , Johan Denollet & Jelte M. Wicherts

To cite this article: Paul Lodder , Wilco H.M. Emons , Johan Denollet & Jelte M. Wicherts (2020): Latent logistic interaction modeling: A simulation and empirical illustration of Type D personality, Structural Equation Modeling: A Multidisciplinary Journal, DOI: 10.1080/10705511.2020.1838905

To link to this article: https://doi.org/10.1080/10705511.2020.1838905

STRUCTURAL EQUATION MODELING: A MULTIDISCIPLINARY JOURNAL
https://doi.org/10.1080/10705511.2020.1838905

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

Paul Lodder[1], Wilco H.M. Emons[1], Johan Denollet[1†], and Jelte M. Wicherts[1]

[1]Tilburg University

ABSTRACT
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. In this article, we investigated this issue using both a Monte Carlo simulation and an empirical illustration of the effect of Type D personality on cardiac events. 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. Although for most methods bias increased when item scores were skewed and ordinal, SEM produced relatively unbiased interaction effect estimates when items were modeled as ordered categorical.

KEYWORDS
Logistic regression; Factor score regression; Structural equation modeling; Interaction effect; Type D personality

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 et al., 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 (Lomax & Schumacker, 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 (Devlieger & Rosseel, 2017; Lu et al., 2011). 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 the former estimates those parameters in two separate steps.

Although all three approaches are commonly used by applied researchers, it is still unclear how they 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.

CONTACT Paul Lodder p.lodder@uvt.nl Department of Methodology and Statistics, Tilburg School of Social and Behavioral Sciences (TSB), Tilburg University, Tilburg LE 5000, The Netherlands.

[†]This paper is dedicated to the memory of Professor Johan Denollet

Author note
The research of JMW is supported by a Consolidator Grant (IMPROVE) from the European Research Council (ERC; grant no. 726361).

† Prof. Dr. Johan Denollet passed away on October 26, 2019

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., Lodder, 2020a; 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 et al., 2010). These findings were corroborated by a meta-analysis (O'Dell et al., 2011). Another meta-analysis indicated that having a Type D personality imposes an increased mortality risk in people with coronary artery disease (Grande et al., 2012).

Most of the studies included in these meta-analyses 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 et al., 2002).

To answer this criticism, Denollet et al. (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 ξ. The logistic model then equals:

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

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, most importantly, the interaction 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 in ξ_2 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 (x) to the latent construct based on a vector of factor loadings ($\mathbf{\Lambda}_x$) and a vector denoting measurement error ($\boldsymbol{\delta}$):

$$x = \mathbf{\Lambda}_x \xi + \boldsymbol{\delta} \tag{2}$$

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, 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 et al. (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* (Cham et al., 2012; Marsh et al., 2004).

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 et al., 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 (L. K. Muthén & Muthén, 1998–2011). Moreover, in a previous simulation study (Lodder et al., 2019) 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 et al., 2019). To overcome these problems, researchers may use a two-step approach, called factor score regression (Lu et al., 2011).

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 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 based on the item 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 et al. (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 & Rabe-Hesketh, 2004), and (4) *Bias correcting* method (Croon, 2002). Devlieger et al. (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.

$$A_\xi = var(\xi)\mathbf{\Lambda}_x\Sigma_x^{-1} \tag{3}$$

Equation (3) shows how this *Regression* approach estimates the factor scores of a latent variable (A_ξ) based on the variance of the latent factor $var(\xi)$, the vector with factor loadings ($\mathbf{\Lambda}_x$) and the inverse of the model implied covariance matrix of the indicator variables $\left(\Sigma_x^{-1}\right)$. 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; B. 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, 2010). 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

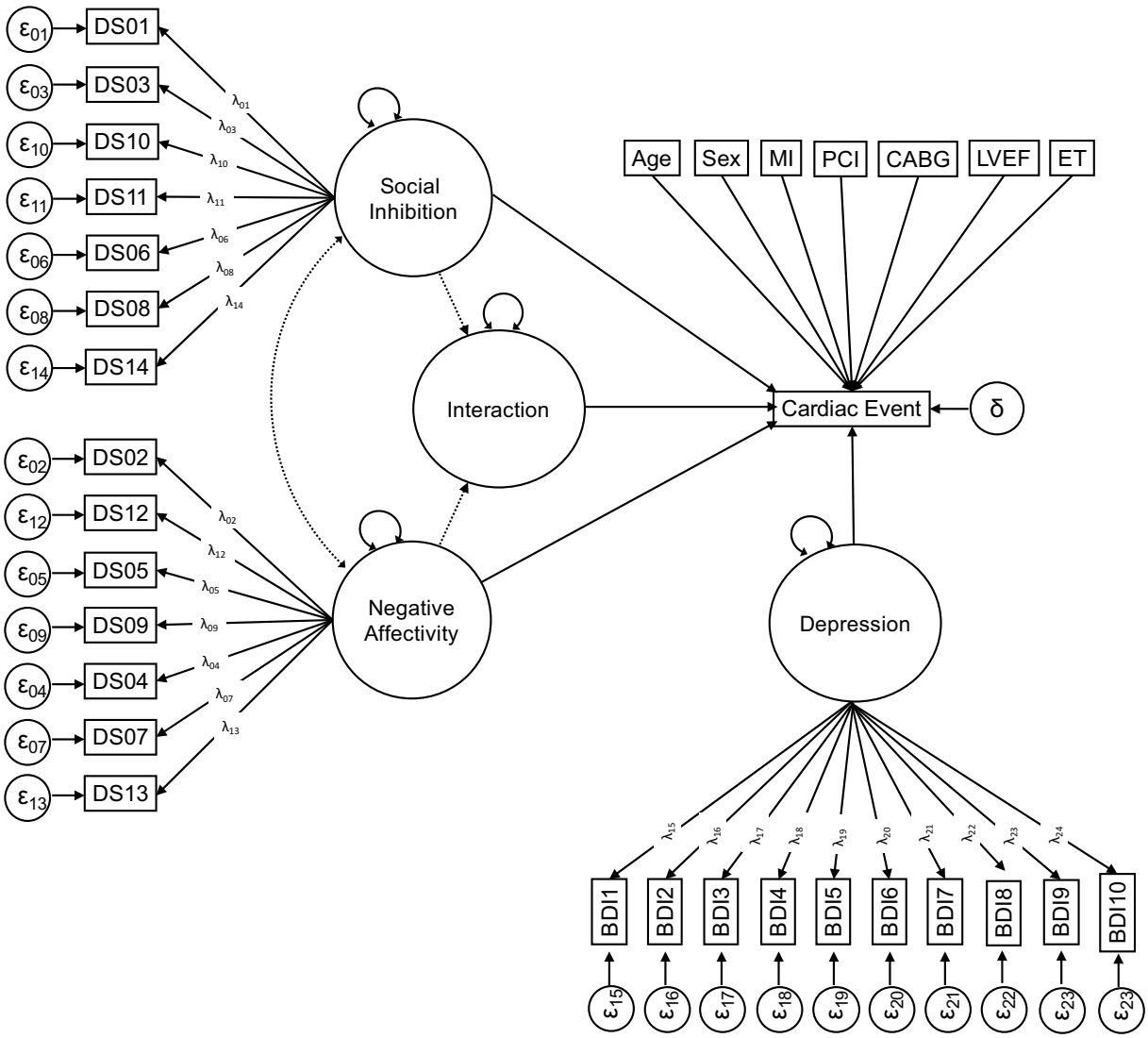


Figure 1. Structural equation model of the relation between Type D personality (modeled as the interaction between NA and SI) and cardiac events, while controlling for the covariates depression, age, sex, MI at baseline, Percutaneous Coronary Intervention (PCI) at baseline, Coronary Artery Bypass Grafting (CABG) at baseline, Left Ventricular Ejection Fracture (LVEF), and Poor Exercise Tolerance (ET).

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).

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.

Depression. Depressive symptoms were measured using the ten-item version of the Beck Depression Inventory (BDI10; Denollet, Martens et al., 2010), 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, Martens et al., 2010). 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 (Lodder, 2020b; MacCallum & Mar, 1995), 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; L. K. Muthén & Muthén, 1998–2011) 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 et al., 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 et al., 2019). Lastly, Mplus provided the pooled results for the latent interaction models (Asparouhov & Muthén, 2010). As only a few observations were missing, the test statistics and fit indices were similar

across the imputed datasets. Our analysis scripts can be found on the open science framework: 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, Table 1 shows the results of the sum score and factor score methods to model interaction effects for the MACE and MI endpoints, while Table 2 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 −2LL 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, $\chi^2(1) = 6.489$, $p = .011$. 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 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]). 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,

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

	Sum score logistic regression			Factor score logistic regression		
Method	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
MACE						
−2 log likelihood	509.13	507.24	500.75	512.19	508.29	502.22
Difference test ‡	χ^2 (2) = 2.58	χ^2 (2) = 1.88	χ^2 (1) = 6.49*	χ^2 (2) = 1.99	χ^2 (2) = 3.89	χ^2 (1) = 6.07*
Type D Personality						
NA	0.84 (0.63, 1.12)	0.81 (0.60, 1.09)	0.79 (0.58, 1.08)	0.77 (0.49, 1.20)	0.70 (0.43, 1.12)	0.68 (0.41, 1.11)
SI	1.17 (0.93, 1.48)	1.12 (0.88, 1.43)	1.14 (0.88, 1.48)	1.18 (0.90, 1.53)	1.13 (0.85, 1.50)	1.16 (0.85, 1.58)
NA^2	-	1.08 (0.89, 1.31)	0.95 (0.75, 1.19)	-	1.15 (0.93, 1.41)	0.97 (0.74, 1.26)
SI^2	-	1.10 (0.91, 1.33)	0.97 (0.78, 1.22)	-	1.11 (0.91, 1.36)	0.92 (0.69, 1.21)
NA x SI	-	-	1.41 (1.06, 1.87)*	-	-	1.48 (1.03, 2.14)*
MI						
−2 log likelihood	271.86	268.97	257.85	272.87	269.22	259.80
Difference test ‡	χ^2 (2) = 0.11	χ^2 (2) = 2.88	χ^2 (1) = 11.12*	χ^2 (2) = 0.42	χ^2 (2) = 3.65	χ^2 (1) = 9.42*
Type D Personality						
NA	0.94 (0.76, 1.16)	0.84 (0.61, 1.17)	0.88 (0.74, 1.06)	0.82 (0.42, 1.58)	0.70 (0.34, 1.39)	0.75 (0.35, 1.62)
SI	1.04 (0.87, 1.24)	1.09 (0.90, 1.33)	1.04 (0.81, 1.33)	1.03 (0.69, 1.55)	1.10 (0.70, 1.73)	1.03 (0.57, 1.86)
NA^2	-	1.24 (1.12, 1.38)*	0.91 (0.75, 1.10)	-	1.31 (0.96, 1.77)	0.90 (0.56, 1.43)
SI^2	-	0.87 (0.64, 1.19)	0.53 (0.31,.90)*	-	0.86 (0.60, 1.23)	0.48 (0.24,.94)*
NA x SI	-	-	2.47 (1.27, 4.81)*	-	-	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.

Model 0: Covariates only (not shown in table); Model 1: Covariates + NA & SI main effects; Model 2: Covariates + NA & SI main + quadratic effects; Model 3: Covariates + NA & SI main + quadratic + interaction effects

* p <.05. 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.

‡ 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 Cardiac Events (**MACE**) and Myocardial Infarction (**MI**), modeling the interaction with SEM and the LMS approach, assuming the ordinal items to be continuous or categorical.

Method	SEM LMS (continuous)			SEM LMS (categorical)		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
MACE						
−2 log likelihood	−13749.54	−13747.61	−13744.86	−12493.70	−12491.78	−12487.36
Difference test ‡	χ^2 (2) = 1.04	χ^2 (2) = 1.94	χ^2 (1) = 2.74	χ^2 (2) = 0.98	χ^2 (2) = 1.92	χ^2 (1) = 4.42*
AIC	27671.08	27671.21	27667.73	25215.40	25215.56	25208.72
BIC	28040.32	28049.03	28049.84	25704.85	25713.60	25711.04
SABIC	27767.32	27769.69	27767.32	25342.97	25345.37	25339.65
Type D Personality						
NA	0.76 (0.47, 1.23)	0.69 (0.42, 1.11)	0.66 (0.39, 1.1)	0.8 (0.51, 1.24)	0.81 (0.53, 1.24)	0.76 (0.47, 1.24)
SI	1.19 (0.92, 1.54)	1.14 (0.86, 1.51)	1.19 (0.85, 1.66)	1.15 (0.88, 1.51)	1.15 (0.89, 1.48)	1.17 (0.84, 1.63)
NA^2	-	1.16 (0.94, 1.43)	0.95 (0.71, 1.28)	-	1.12 (0.94, 1.34)	0.85 (0.62, 1.17)
SI^2	-	1.12 (0.91, 1.37)	0.9 (0.66, 1.21)	-	1.13 (0.94, 1.34)	0.85 (0.62, 1.17)
NA x SI	-	-	1.58 (1.04, 2.4)*	-	-	1.85 (1.11, 3.08)*
MI						
−2 log likelihood	−13629.62	−13630.12	−13623.19	−12372.21	−12370.93	−12367.70
Difference test ‡	χ^2 (2) = 0.21	χ^2 (2) = 0.50	χ^2 (1) = 6.93*	χ^2 (2) = 0.12	χ^2 (2) = 1.28	χ^2 (1) = 3.23
AIC	27431.23	27436.24	27424.38	24972.42	24973.86	24969.40
BIC	27800.47	27814.06	27806.50	25461.87	25471.85	25471.33
SABIC	27527.47	27534.72	27523.98	25100.00	25103.67	25100.33
Type D Personality						
NA	0.81 (0.41, 1.61)	0.7 (0.34, 1.42)	0.75 (0.34, 1.68)	0.89 (0.47, 1.68)	0.9 (0.5, 1.62)	0.92 (0.44, 1.93)
SI	1.03 (0.75, 1.42)	1.1 (0.73, 1.66)	1.03 (0.51, 2.08)	0.99 (0.71, 1.38)	0.99 (0.67, 1.45)	0.9 (0.44, 1.83)
NA^2	-	1.32 (1.01, 1.72)*	0.82 (0.5, 1.34)	-	1.24 (1.01, 1.52)*	0.72 (0.44, 1.18)
SI^2	-	0.85 (0.6, 1.19)	0.36 (0.15, 0.85)	-	0.87 (0.65, 1.17)	0.36 (0.13, 1.02)
NA x SI	-	-	3.71 (1.25, 11.02)*	-	-	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.

Model 0: Covariates only (not shown in table); Model 1: Covariates + NA & SI main effects; Model 2: Covariates + NA & SI main + quadratic effects; Model 3: Covariates + NA & SI main + quadratic + interaction effects

* p <.05. 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.

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

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 −2LL 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 = .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

These results 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

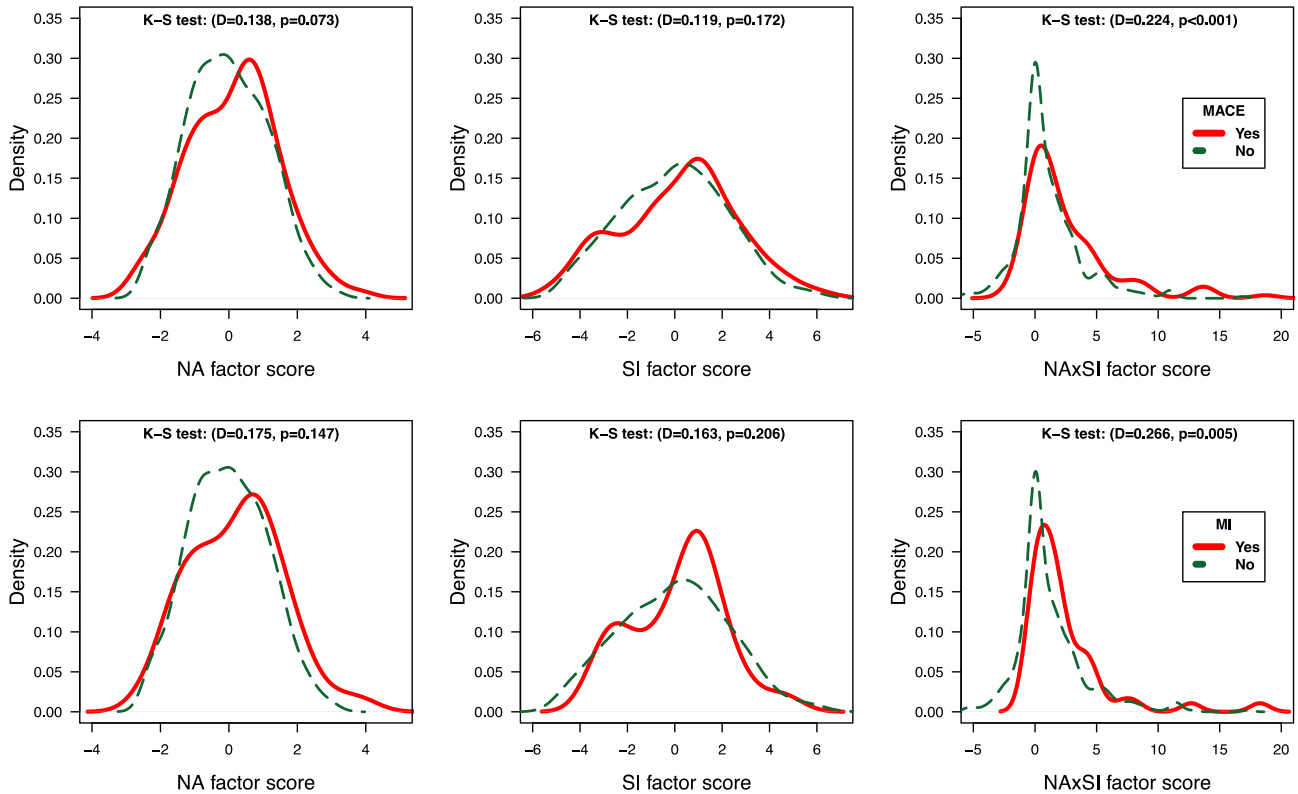


Figure 2. Factor score distributions of NA, SI, & NA x SI, separately for people who did (red/straight curve) or did not (green/broken curve) have cardiac events. The top row shows the results for the MACE endpoint and the bottom row for the MI endpoint. The top of each plot shows the result of the two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

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 et al., 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.

Study 2: Monte Carlo simulation

Method

Procedure

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 four 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 = .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.[1] 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, continuous item scores for each individual (i) and for each item (j) measuring the traits (t) NA or SI were obtained as follows:

$$Y_{ij} = \lambda_{jt}\xi_{ti} + \varepsilon_{ij} \tag{4}$$

As input we used a matrix with individual NA and SI trait scores ($\boldsymbol{\Xi}$), the factor loading matrix retrieved from Study 1 ($\boldsymbol{\Lambda}$), and residual error matrix ($\boldsymbol{\Theta}$) 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 B. Muthén and Kaplan (1985) .[2] 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:

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

In Equation (5), ρ_{Ci} denotes the probability of a cardiac event score of individual i. $\beta_0$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.

Statistical methods

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

[1]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.

[2]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.

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 et al., 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 3–5 show for all 240 simulation conditions the relative bias in the estimated interaction effect averaged across all 500 replications. Tables 6–8 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. The online supplement contains six supplemental Tables. Tables S1, S2 and S3 present the bias in the estimated interaction effect in terms of the mean squared error (the squared difference between the true

Table 3. Relative bias in the estimated interaction effects for all simulation scenarios involving continuous item scores and skewness generated at the latent variable level. Bold faced cells indicate acceptable bias (<10%).

				Sum score regression		Factor score regression		SEM LMS continuous	
N	N/p ratio	Skewness	Interaction	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	18	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**
1000	18	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 4. Relative bias in the estimated interaction effects for all simulation scenarios involving ordinal item scores and skewness generated at the latent variable level. Bold faced cells indicate acceptable bias (<10%).

				Sum score regression		Factor score regression		SEM LMS continuous		SEM LMS categorical	
N	N/p ratio	Skewness	Interaction	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**	**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	18	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
1000	18	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

interaction effect and the estimated interaction, averaged across 500 replications). Tables S4, S5, and S6 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 Figure 3 (continuous item scores), Figure 4 (ordinal item scores with latent skewness), and Figure 5 (ordinal item scores with skewness generated on the item level). The online supplement contains similar figures for both the two 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 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 5–7 (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.

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.

Table 5. Relative bias in the estimated interaction effects for all simulation scenarios involving ordinal item scores and skewness generated at the item score level. Bold faced cells indicate acceptable bias (<10%).

				Sum score regression		Factor score regression		SEM LMS continuous		SEM LMS categorical	
N	N/p ratio	Skewness	Interaction	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	18	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**
1000	18	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 6. 95% confidence interval coverage rates of the estimated interaction effects for all simulation scenarios involving continuous item scores and skewness generated at the latent variable level. Bold faced cells indicate acceptable coverage.

				Sum scores		Factor scores		LMS (continuous)	
N	N/p ratio	Skewness	Interaction	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	18	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
1000	18	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

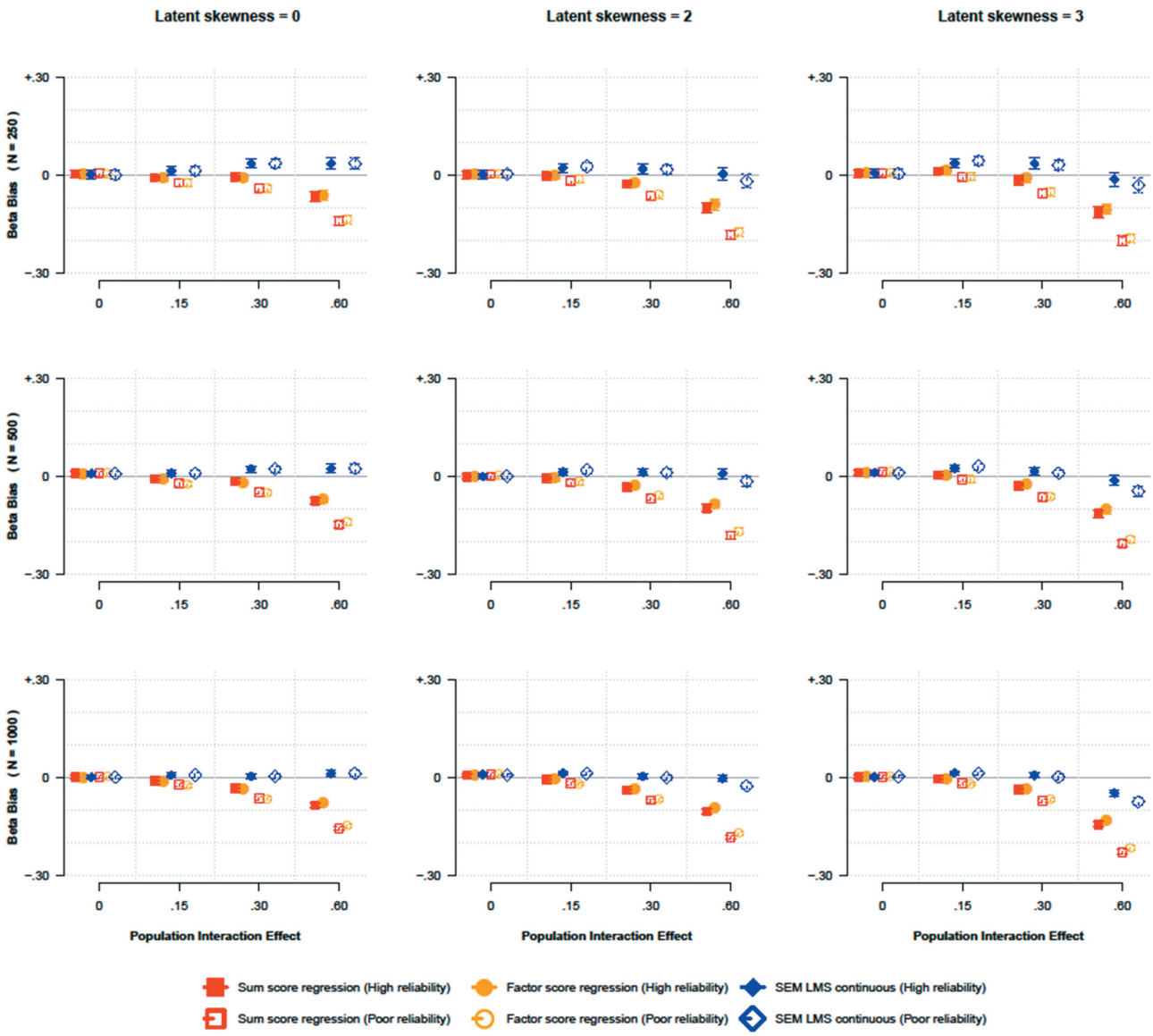


Figure 3. Comparison of three methods to estimate interaction effects between NA and SI on a binary outcome in scenarios with **continuous items**, varying over the true interaction size (x axis), reliability of the NA and SI scales, amount of **latent skewness** (columns) and sample size (rows). Each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

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

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

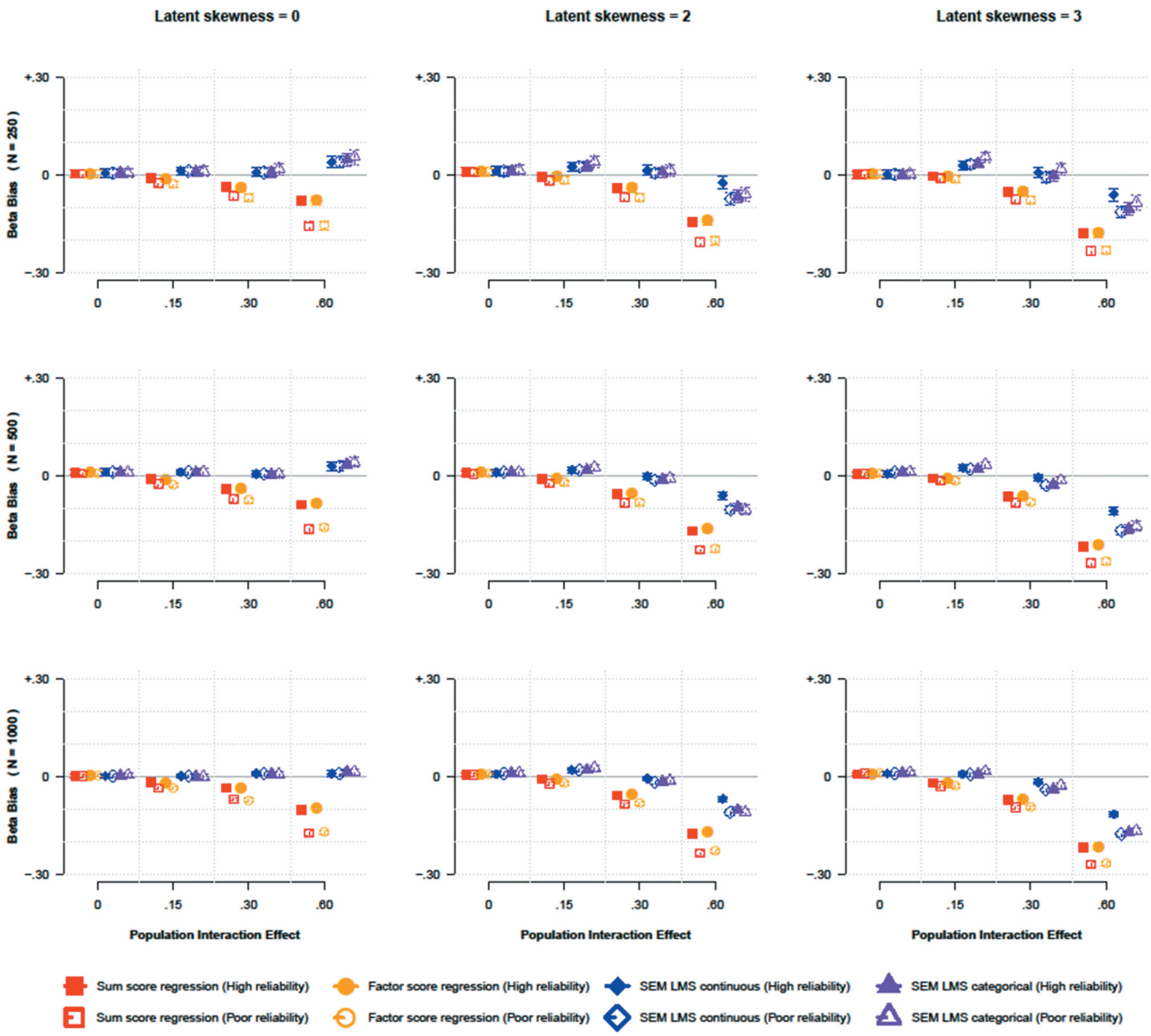


Figure 4. Comparison of four methods to estimate interaction effects between NA and SI on a binary outcome in scenarios with **ordinal items**, varying over the true interaction size (x axis), reliability of the NA and SI scales, amount of **latent skewness** (columns) and sample size (rows). Each data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

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

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

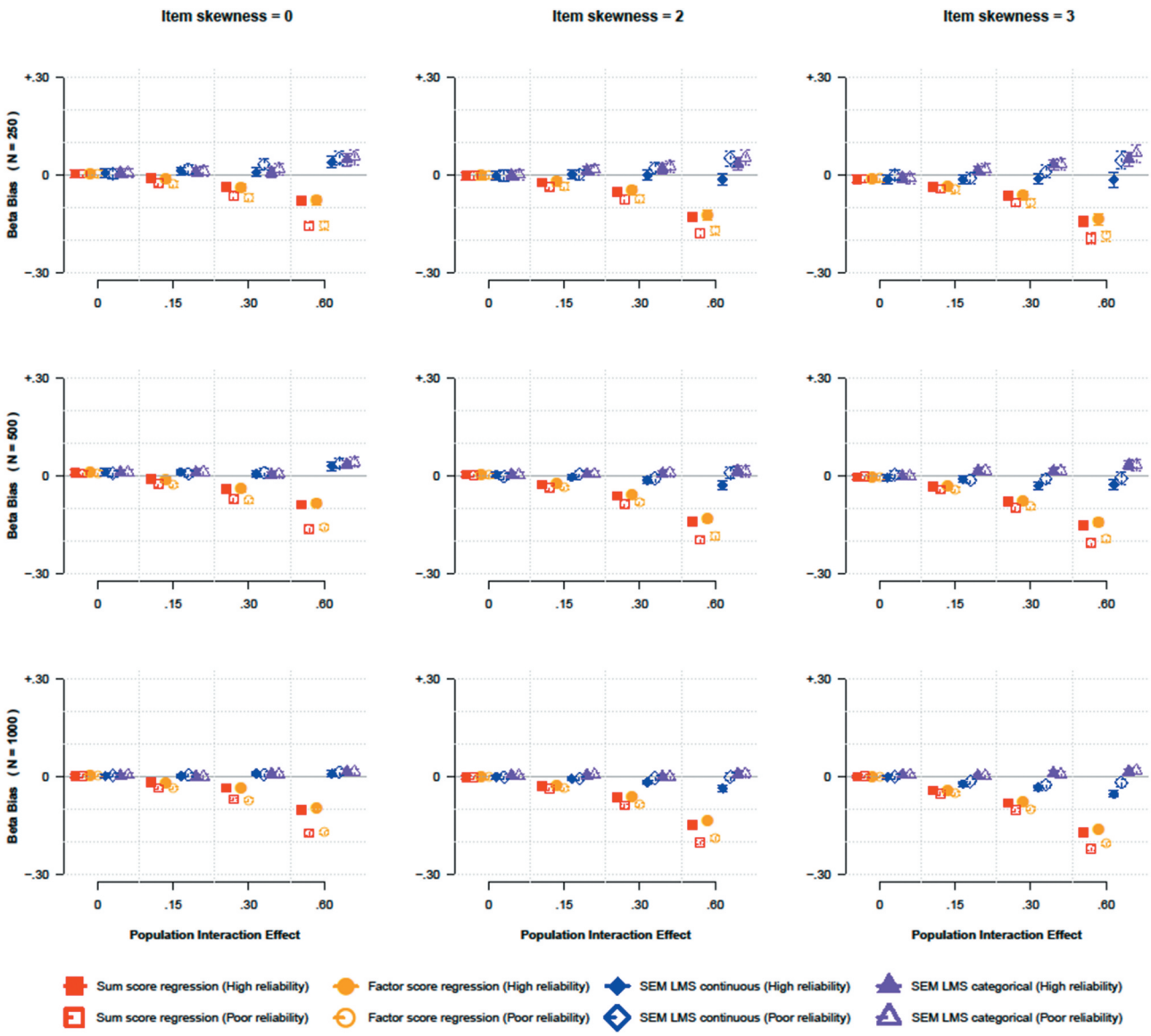


Figure 5. Comparison of four methods to estimate interaction effects between NA and SI on a binary outcome in scenarios with **ordinal items**, varying over the true interaction size (x axis), reliability of the NA and SI scales, amount **item skewness** (columns) and sample size (rows). Each data point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

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 four 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 four 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

Table 7. 95% confidence interval coverage rates of the estimated interaction effects for all simulation scenarios involving ordinal item scores and skewness generated at the latent variable level. Bold faced cells indicate acceptable coverage.

				Sum scores		Factor scores		LMS (continuous)		LMS (categorical)	
N	N/p ratio	Skewness	Interaction	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	18	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
1000	18	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

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 (Cham et al., 2012; Kelava & Nagengast, 2012; Kelava et al., 2014). 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 et al., 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

Table 8. 95% confidence interval coverage rates of the estimated interaction effects for all simulation scenarios involving ordinal item scores and skewness generated at the item score level. Bold faced cells indicate acceptable coverage.

				Sum scores		Factor scores		LMS (continuous)		LMS (categorical)	
N	N/p ratio	Skewness	Interaction	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	18	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**
1000	18	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**

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 a 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 ≥ 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 (Jöreskog & Yang, 1996; Kenny & Judd, 1984; Marsh et al., 2004), the two stage least squares approach (Bollen & Paxton, 1998), or mixture modeling (Kelava & Nagengast, 2012; Kelava et al., 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 et al. (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, 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 cutoff values and do therefore likely not reflect real differences between the groups (MacCallum et al., 2002; Royston et al., 2006). However, it is not necessary to classify people based on arbitrary cutoff 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 β = .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.

Funding

The research of J.M. Wicherts was funded by the European Research Council (Grant nr: 726361) as part of the programme: H2020-EU.1.1. - EXCELLENT SCIENCE.

ORCID

Paul Lodder http://orcid.org/0000-0003-1300-7853
Wilco H.M. Emons http://orcid.org/0000-0001-7576-0037

Johan Denollet http://orcid.org/0000-0001-8986-3256
Jelte M. Wicherts http://orcid.org/0000-0003-2415-2933

References

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

Asparouhov, T., & Muthén, B. (2010). *Multiple imputation with Mplus*. MPlus Web Notes.

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

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.

Bollen, K. A. (1989). *Structural equations with latent variables*. John Wiley & Sons.

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.

Busemeyer, J. R., & Jones, L. E. (1983). Analysis of multiplicative combination rules when the causal variables are measured with error. *Psychological Bulletin*, *93*, 549–562. https://doi.org/10.1037/0033-2909.93.3.549

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

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

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

Denollet, J. (2005). DS14: Standard assessment of negative affectivity, social inhibition, and Type D personality. *Psychosomatic Medicine*, *67*, 89–97. https://doi.org/10.1097/01.psy.0000149256.81953.49

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., Vrints, C. J., & Conraads, V. M. (2013). Predictive value of social inhibition and negative affectivity for cardiovascular 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., 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., 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., 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

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

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

Du, J., Zhang, D., Yin, Y., Zhang, X., Li, J., Liu, D., ... & Chen, W. (2016). The personality and psychological stress predict major adverse cardiovascular events in patients with coronary heart disease after percutaneous coronary intervention for five years. *Medicine, 95*. https://doi.org/10.1097/MD.0000000000003364

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

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

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*, 299–310. https://doi.org/10.1007/s12160-011-9339-0

Hallquist, M., & Wiley, J. (2011). *MplusAutomation: Automating Mplus model estimation and interpretation*. R package version 0.6–3.

Hayduk, L. A., & Littvay, L. (2012). Should researchers use single indicators, best indicators, or multiple indicators in structural equation models? *BMC Medical Research Methodology*, *12*, 159. https://doi.org/10.1186/1471-2288-12-159

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

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

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

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

Kenny, D. A., & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin*, *96*, 201–210. https://doi.org/10.1037/0033-2909.96.1.201

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.

Kupper, N., & Denollet, J. (2016). Explaining heterogeneity in the predictive 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

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

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. (2020a). Modeling synergy: How to assess a Type D personality effect. *Journal of Psychosomatic Research*, *132*, 109990. https://doi.org/10.1016/j.jpsychores.2020.109990

Lodder, P. (2020b). A re-evaluation of the Type D personality effect. *Personality and Individual Differences*, *167*, 110254. https://doi.org/10.1016/j.paid.2020.110254

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*, *54*, 637–665. https://doi.org/10.1080/00273171.2018.1562863

Lomax, R. G., & Schumacker, R. E. (2004). *A beginner's guide to structural equation modeling*. Psychology Press.

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

MacCallum, R. C., & Mar, C. M. (1995). Distinguishing between moderator 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

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

Muthén, B., & 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*, 171–189. https://doi.org/10.1111/j.2044-8317.1985.tb00832.x

Muthén, L. K., & Muthén, B. O. (1998–2011). *Mplus user's guide* (6th ed.). Muthén & Muthén.

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

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-analytic review. *Journal of Psychosomatic Research*, *71*, 199–206. https://doi.org/10.1016/j.jpsychores.2011.01.009

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

Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Copenhagen, Denmark: Paedagogiske Institut

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

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

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 continuous predictors in multiple regression: A bad idea. *Statistics in Medicine*, *25*, 127–141. https://doi.org/10.1002/sim.2331

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., & 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

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*, 461–464. https://doi.org/10.1214/aos/1176344136

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. Crc Press.

Smith, T. W. (2011). Toward a more systematic, cumulative, and applicable science of personality and health: Lessons from type D personality. *Psychosomatic Medicine*, *73*, 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

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*. Allyn & Bacon/Pearson Education.

Thurstone, L. (1935). *The vectors of mind*. Chicago, IL: University of Chicago Press.

Vale, C. D., & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika*, *48*, 465–471. https://doi.org/10.1007/BF02293687

Waller, N. G. (2016). *Fungible coefficients and Monte Carlo functions*. R-package. https://cran.r-project.org/web/packages/fungible/index.html

Appendix A: DS14 scale information for Study 1

DS14	*Negative Affectivity*							*Social Inhibition*						
Correlation	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