Different results on tetrachorical correlations in Mplus and Stata – Stata announces modified procedure

AGNES GÜNTHER,¹ MICHAEL HÖFLER¹

¹ Technische Universität Dresden, Institute of Clinical Psychology and Psychotherapy

Abstract

To identify the structure of mental disorders in large-scale epidemiological data sets, investigators frequently use tetrachoric correlations as a first step for subsequent application of latent class and factor analytic methods. It has been possible to do this with Stata since 2005, whereas the corresponding Mplus routine has been on the market for some years. Using an identical data set we observed considerable differences between the results of the packages. This paper illustrates the differences with several examples from the Early Developmental Stages of Psychopathology Study data set, which consists of 3021 subjects, with diagnostic information assessed by the CID-I. Results reveal that tetrachoric correlations resulting from Mplus were often considerably smaller than those computed with Stata. The results were dramatically different, especially where there were few observation per cell or even empty cells. These findings were put to Mplus and Stata, whose responses clarified the discrepancies by describing the different mathematical assumptions and procedures used. Stata announced that it intended to launch a modified procedure. Copyright © 2006 John Wiley & Sons, Ltd.

Introduction

We recently encountered a rather disturbing problem when computing tetrachoric correlations with two alternative statistical packages, Stata and Mplus. Using data from the Early Developmental Stages of Psychiatry (EDSP) study we examined tetrachoric correlations as a first step for subsequent application of latent class and factor analytic methods in an attempt to replicate and expand previous findings on the common structure of mental disorders (Kendler et al., 1995; Krueger et al., 1996; Krueger et al., 1998; Krueger, 1999; Wittchen et al., 1999; Vollebergh et al., 2001; Krueger and Finger, 2001; Krueger et al., 2001; Kendler et al., 2003; Hettema et al., 2005; Kessler et al., 2005).

We found considerably different results when using Stata's tetrachoric procedure (Stata Corp., 2006) and the MPlus type = basic procedure (Muthén and Muthén, 1998–2005). According to the respective manuals, MPlus

uses a numerical maximum likelihood approach to compute the correlation coefficient from the bivariate normal distribution function (Kirk, 1973; Divgi, 1979), whereas Stata applies an approximate explicit formula published by Edwards and Edwards (1984). In this short note we demonstrate the type and the degree of these discrepancies using several examples with this data set.

Methods

The data came from the baseline assessment and three follow-up assessments of the 10-year prospective-longitudinal EDSP study. The EDSP is based on a community sample of $N = 3021$ subjects aged 14 to 24 at the outset of the study (21 to 34 at the last assessment). Its aims are to examine the developmental patterns of mental and substance-use disorders and their risk and vulnerability factors over time. Sampling, weighting and other design features are presented in Wittchen

et al. (1998a), Wittchen et al. (1998b) and Lieb et al. (2000). The data presented in this paper are based on the DSM-IV version of the Composite International Diagnostic Interview (CIDI) (Wittchen et al., 1984) and its diagnostic algorithms. The information about the temporal patterns of disorders refers to the age-of-onset reports in each of the four assessments of the EDSP. Data from all respondents were aggregated across all four assessment waves to reflect all individual histories up to the age at the last completed assessment in a person $\times$ year data file.

The diagnosis of antisocial personality disorder (ASP) was only assessed for 2548 individuals who completed the third assessment wave. Thus, we had missing values for this diagnosis in 473 cases. Since Mplus only allows the computation of tetrachoric correlations with casewise deletion of missing data we had to apply casewise deletion when using Stata too, in order to keep the analyses comparable. Another 83 individuals were deleted from the data set because of missing age-of-onset information leaving a total of 2485 persons for the analyses.

The person-year data set for the total sample contained a total of 63 376 observations with complete

information on all disorders from 2485 persons. This data set was first split into two data sets: one for the younger birth cohort (aged 14 to 17 at the outset, $N = 25\,187$ observations from 1121 Persons; 1395 completed the baseline, 226 were lost due to not completing the second follow-up and 48 due to missing at of onset data), and one for the older birth cohort (aged 18 to 24 at baseline, $N = 38\,189$ observations from 1339; 1621 Persons completed the baseline, 247 did not complete the second follow-up and 35 were lost due to missing age of onset information). Note that the first follow-up assessment was only conducted in the younger cohort. Finally, the data set was split into three data sets covering different age periods (age 0 to 18, $N = 44\,465$; age 19 to 25, $N = 14\,034$; age 26 to 34, $N = 4867$). We examined 17 diagnostic variables.

Although the individuals were sampled with different probabilities according to age at baseline assessment (Wittchen et al, 1998a) we did not weight the data here because the tetrachoric procedure in Stata does not allow for sampling weights, and we aimed to run exactly the same analysis with both software packages.

Table 1. Matrix of tetrachoric correlations among 17 lifetime mental disorders (person $\times$ year data, $N = 63\,376$, total sample)

	Percentage affected	Alcohol SUD*		Ill. drug SUD		Nicotine SUD		Eating disorder		Panic disorder		Agora-phobia		Specific phobia/ NOS		Social phobia	
		Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata
Alcohol SUD	10.46	1.00	1.00														
Ill. drug SUD	3.61	0.69	0.82	1.00	1.00												
Nicotine SUD	9.79	0.62	0.71	0.68	0.81	1.00	1.00										
Eating disorder	1.53	0.22	0.38	0.27	0.49	0.39	0.59	1.00	1.00								
Panic disorder	1.21	0.34	0.54	0.29	0.52	0.42	0.64	0.37	0.66	1.00	1.00						
Agoraphobia	1.82	0.29	0.45	0.32	0.55	0.47	0.66	0.39	0.67	0.84	0.95	1.00	1.00				
Specific phobia/NOS	12.86	0.32	0.41	0.33	0.47	0.45	0.55	0.48	0.68	0.53	0.74	0.54	0.73	1.00	1.00		
Social phobia	4.76	0.34	0.47	0.28	0.45	0.41	0.56	0.37	0.60	0.50	0.73	0.56	0.77	0.45	0.58	1.00	1.00
GAD	1.17	0.28	0.47	0.39	0.64	0.53	0.74	0.39	0.69	0.58	0.84	0.62	0.85	0.46	0.68	0.56	0.79
OCD	0.68	0.21	0.39	0.33	0.61	0.42	0.66	0.49	0.79	0.62	0.88	0.62	0.87	0.46	0.70	0.47	0.73
PTSD	0.83	0.24	0.42	0.42	0.69	0.45	0.68	0.39	0.70	0.57	0.85	0.56	0.83	0.46	0.69	0.47	0.72
Mania/2.57 hypomania	0.39	0.57	0.40	0.62	0.46	0.63	0.29	0.53	0.34	0.60	0.48	0.72	0.42	0.59	0.32	0.52	0.44
MDD/8.30 dysthymia	0.40	0.51	0.42	0.59	0.53	0.64	0.55	0.75	0.55	0.76	0.54	0.74	0.55	0.65	0.55	0.69	0.68
Pain disorder	4.52	0.15	0.23	0.26	0.42	0.37	0.52	0.34	0.57	0.41	0.66	0.41	0.63	0.45	0.59	0.34	0.52
SSI 4/6	1.23	0.27	0.45	0.38	0.63	0.44	0.65	0.49	0.77	0.49	0.78	0.56	0.81	0.37	0.57	0.43	0.67
Any psychotic	0.17	0.45	0.76	0.37	0.71	0.28	0.55	0.26	0.61	-0.10	-1.00	0.34	0.70	0.31	0.59	0.28	0.58
ASP	4.42	0.61	0.74	0.62	0.78	0.53	0.67	0.07	0.15	0.28	0.50	0.24	0.44	0.29	0.41	0.29	0.45

*SUD: substance abuse disorder.

Results

Table 1 shows the tetrachoric correlations in the overall sample as well as the frequency of each disorder in the person-year data set. In the text we mention the raw 2×2 tables of those associations that we highlight because of their strikingly different results when using Stata versus Mplus.

In each case the tetrachoric correlation as calculated with Stata was higher in absolute value than the tetrachoric correlation as calculated with Mplus. Several discrepancies were substantial, for example that between any psychotic disorder and mania/hypomania (Stata: 0.70 versus Mplus: 0.35; no psychotic disorder/no mania/hypomania $N = 61\,661$, no psychotic disorder/mania/hypomania $N = 1610$, psychotic disorder/no mania/hypomania $N = 85$, psychotic disorder/mania/hypomania $N = 20$). An extremely high difference emerged for the association between any psychotic disorder and panic disorder (-1.00 versus -0.10 ; no psychotic disorder/no panic disorder $N = 62\,503$, psychotic disorder/no panic disorder $N = 105$, no psychotic disorder/panic disorder $N = 768$, psychotic disorder/panic disorder $N = 0$).

Tables 2a and b show the results in the younger cohort (Table 2a) and the older cohort (Table 2b) respectively.

Again, Stata revealed consistently higher tetrachoric correlations. In the younger cohort, the tetrachoric correlation between any psychotic disorder and SSI 4/6 was even twofold in Stata (0.76 versus 0.38, no psychotic disorder/no SSI 4/6 $N = 24\,853$, no psychotic disorder/SSI 4/6 $N = 287$, no psychotic disorder/no SSI 4/6 $N = 41$, no psychotic disorder/no SSI 4/6 $N = 6$). The correlation between any psychotic disorder and panic disorder equalled -1.00 in Stata and 0.00 in Mplus (no psychotic disorder/no panic $N = 24\,875$, no psychotic disorder/panic $N = 265$, psychotic disorder/no panic $N = 47$, psychotic disorder/panic $N = 0$). In the older cohort a similar pattern was found for the tetrachoric correlation between any psychotic disorder on the one hand and GAD and OCD on the other hand.

Tables 3 a, b and c show results in the three age periods considered.

The tetrachoric correlations were almost always higher in absolute value or equal in Stata. However, in the 0 to 18 age period we found four pairs of variables

GAD		OCD		PTSD		Mania/ hypomania		MDD/ dysthymia		Pain disorder		SSI 4/6		Any psychotic		ASP	
Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata
1.00	1.00																
0.64	0.89	1.00	1.00														
0.56	0.84	0.56	0.86	1.00	1.00												
0.71	0.52	0.80	0.40	0.69	1.00	1.00											
0.87	0.53	0.78	0.60	0.82	0.34	0.51	1.00	1.00									
0.39	0.64	0.37	0.64	0.38	0.65	0.40	0.61	0.43	0.59	1.00	1.00						
0.48	0.78	0.60	0.87	0.63	0.88	0.47	0.73	0.49	0.71	0.63	0.83	1.00	1.00				
0.25	0.61	0.32	0.73	0.48	0.85	0.35	0.70	0.36	0.66	0.21	0.48	0.50	0.86	1.00	1.00		
0.33	0.56	0.35	0.61	0.33	0.58	0.36	0.57	0.28	0.41	0.21	0.35	0.26	0.47	0.25	0.55	1.00	1.00

Table 2a. Matrix of tetrachoric correlations among 17 lifetime mental disorders (person $\times$ year data, $N = 25\,187$, 14–17 years at baseline)

	Percentage affected	Alcohol SUD*		Ill. drug SUD		Nicotine SUD		Eating disorder		Panic disorder		Agora-phobia		Specific phobia/ NOS		Social phobia	
		Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata
Alcohol SUD	9.62	1.00	1.00														
Ill. drug SUD	4.39	0.74	0.84	1.00	1.00												
Nicotine SUD	9.22	0.64	0.73	0.69	0.80	1.00	1.00										
Eating disorder	1.10	0.16	0.30	0.14	0.28	0.47	0.69	1.00	1.00								
Panic disorder	1.05	0.22	0.38	0.27	0.50	0.32	0.53	0.52	0.81	1.00	1.00						
Agoraphobia	1.64	0.15	0.27	0.17	0.33	0.39	0.59	0.47	0.76	0.74	0.92	1.00	1.00				
Specific phobia/NOS	14.88	0.28	0.35	0.32	0.44	0.39	0.48	0.37	0.57	0.54	0.75	0.52	0.72	1.00	1.00		
Social phobia	4.31	0.28	0.41	0.27	0.43	0.27	0.41	0.34	0.58	0.43	0.68	0.49	0.71	0.40	0.53	1.00	1.00
GAD	0.60	0.27	0.49	0.42	0.70	0.36	0.60	0.26	0.57	0.61	0.88	0.62	0.88	0.44	0.68	0.48	0.75
OCD	0.65	0.25	0.45	0.33	0.59	0.35	0.59	0.34	0.68	0.57	0.86	0.52	0.81	0.38	0.61	0.48	0.75
PTSD	0.45	0.24	0.45	0.51	0.79	0.35	0.60	0.58	0.87	0.63	0.90	0.46	0.78	0.37	0.61	0.50	0.78
Mania/hypomania	2.69	0.49	0.67	0.48	0.69	0.48	0.65	0.25	0.49	0.37	0.64	0.46	0.71	0.40	0.56	0.23	0.41
MDD/7.41 dysthymia	0.43	0.55	0.44	0.60	0.53	0.65	0.47	0.70	0.62	0.82	0.44	0.65	0.52	0.62	0.45	0.61	0.65
Pain disorder	5.25	0.12	0.19	0.21	0.35	0.35	0.48	0.24	0.45	0.50	0.74	0.43	0.66	0.39	0.51	0.38	0.55
SSI 4/6	1.16	0.30	0.50	0.37	0.62	0.45	0.67	0.41	0.72	0.47	0.77	0.56	0.82	0.32	0.51	0.33	0.57
Any psychotic ASP	0.19	0.35	0.64	0.30	0.60	0.24	0.48	0.31	0.69	0.00	−1.00	0.39	0.76	0.30	0.56	0.22	0.49
	4.83	0.61	0.73	0.54	0.71	0.52	0.66	0.09	0.19	0.28	0.51	0.15	0.28	0.28	0.39	0.28	0.44

*SUD: substance abuse disorder.

Table 2b. Matrix of tetrachoric correlations among 17 lifetime mental disorders (person $\times$ year data, $N = 38\,189$, 18–24 years at baseline)

	Percentage affected	Alcohol SUD		Ill. drug SUD		Nicotine SUD		Eating disorder		Panic disorder		Agora-phobia		Specific phobia/ NOS		Social phobia	
		Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata
Alcohol SUD	11.01																
Ill. drug SUD	3.10	0.66	0.80	1.00	1.00												
Nicotine SUD	10.16	0.60	0.69	0.68	0.82	1.00	1.00										
Eating disorder	1.82	0.24	0.40	0.34	0.58	0.35	0.53	1.00	1.00								
Panic disorder	1.32	0.39	0.60	0.30	0.55	0.47	0.68	0.27	0.53	1.00	1.00						
Agoraphobia	1.94	0.35	0.52	0.41	0.65	0.51	0.69	0.35	0.61	0.88	0.96	1.00	1.00				
Specific phobia/NOS	11.53	0.36	0.45	0.33	0.48	0.50	0.59	0.54	0.73	0.53	0.74	0.56	0.74	1.00	1.00		
Social phobia	5.05	0.36	0.49	0.30	0.48	0.48	0.62	0.38	0.60	0.53	0.75	0.59	0.79	0.49	0.62	1.00	1.00
GAD	1.55	0.28	0.45	0.40	0.65	0.57	0.77	0.41	0.68	0.56	0.82	0.63	0.85	0.50	0.70	0.58	0.79
OCD	0.70	0.19	0.35	0.34	0.62	0.45	0.69	0.54	0.82	0.65	0.89	0.68	0.89	0.51	0.75	0.46	0.72
PTSD	1.08	0.24	0.41	0.40	0.67	0.48	0.70	0.29	0.57	0.55	0.83	0.59	0.84	0.51	0.73	0.45	0.70
Mania/2.50 hypomania	0.32	0.48	0.33	0.55	0.44	0.62	0.31	0.55	0.32	0.58	0.49	0.73	0.43	0.60	0.37	0.57	0.44
MDD/8.88 dysthymia	0.39	0.49	0.42	0.59	0.53	0.63	0.59	0.77	0.50	0.72	0.59	0.77	0.59	0.68	0.59	0.72	
Pain disorder	4.04	0.17	0.27	0.29	0.47	0.40	0.54	0.39	0.62	0.36	0.61	0.39	0.62	0.49	0.64	0.33	0.50
SSI 4/6	1.27	0.24	0.41	0.39	0.65	0.42	0.64	0.53	0.79	0.50	0.78	0.56	0.81	0.41	0.61	0.48	0.71
Any psychotic ASP	0.15	0.53	0.83	0.42	0.77	0.32	0.60	0.24	0.57	−0.05	−1.00	0.29	0.64	0.32	0.60	0.32	0.63
	4.15	0.61	0.74	0.67	0.82	0.54	0.69	0.07	0.14	0.28	0.50	0.30	0.51	0.30	0.43	0.30	0.46

GAD		OCD		PTSD		Mania/ hypomania		MDD/ dysthymia		Pain disorder		SSI 4/6		Any psychotic		ASP	
Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata
1.00	1.00																
0.65	0.91	1.00	1.00														
0.53	0.86	0.59	0.89	1.00	1.00												
0.46	0.75	0.41	0.70	0.46	0.76	1.00	1.00										
0.87	0.49	0.75	0.55	0.81	0.36	0.54	1.00	1.00									
0.39	0.66	0.03	0.07	0.35	0.63	0.28	0.46	0.34	0.48	1.00	1.00						
0.50	0.82	0.50	0.81	0.66	0.91	0.36	0.63	0.43	0.65	0.48	0.71	1.00	1.00				
0.50	0.87	0.49	0.86	0.47	0.86	0.22	0.52	0.35	0.65	0.08	0.20	0.38	0.76	1.00	1.00		
0.14	0.30	0.31	0.57	0.04	0.10	0.42	0.62	0.23	0.35	0.21	0.35	0.22	0.42	0.09	0.23	1.00	(1.00)

GAD		OCD		PTSD		Mania/ hypomania		MDD/ dysthymia		Pain disorder		SSI 4/6		Any psychotic		ASP	
Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata
1.00	1.00																
0.65	0.89	1.00	1.00														
0.55	0.82	0.56	0.85	1.00	1.00												
0.70	0.58	0.83	0.39	0.67	1.00	1.00											
0.70	0.86	0.56	0.79	0.61	0.82	0.32	0.49	1.00	1.00								
0.42	0.66	0.51	0.78	0.41	0.67	0.48	0.69	0.50	0.65	1.00	1.00						
0.48	0.76	0.65	0.89	0.63	0.87	0.53	0.78	0.53	0.74	0.71	0.88	1.00	1.00				
−0.07	−1.00	0.02	−1.00	0.50	0.86	0.42	0.77	0.37	0.67	0.30	0.62	0.57	0.89	1.00	1.00		
0.39	0.62	0.37	0.64	0.40	0.66	0.32	0.52	0.31	0.45	0.20	0.34	0.28	0.50	0.35	0.68	1.00	(1.00)

Table 3a. Matrix of tetrachoric correlations among 17 lifetime mental disorders (person $\times$ year data, $N = 44\,475$, age span 0–18 years)

	Percentage affected	Alcohol SUD*		Ill. drug SUD		Nicotine SUD		Eating disorder		Panic disorder		Agora-phobia		Specific phobia/ NOS		Social phobia	
		Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata
Alcohol SUD	3.07	1.00	1.00														
Ill. drug SUD	1.01	0.74	0.91	1.00	1.00												
Nicotine SUD	3.09	0.70	0.85	0.74	0.91	1.00	1.00										
Eating disorder	0.43	0.32	0.61	0.35	0.70	0.55	0.83	1.00	1.00								
Panic disorder	0.38	0.37	0.68	0.43	0.78	0.47	0.78	0.54	0.88	1.00	1.00						
Agoraphobia	0.68	0.30	0.58	0.37	0.70	0.51	0.79	0.50	0.85	0.84	0.97	1.00	1.00				
Specific phobia/NOS	6.69	0.41	0.59	0.41	0.65	0.51	0.69	0.49	0.77	0.60	0.85	0.61	0.84	1.00	1.00		
Social phobia	1.98	0.41	0.65	0.36	0.64	0.40	0.64	0.40	0.72	0.52	0.82	0.58	0.84	0.52	0.72	1.00	1.00
GAD	0.27	0.33	0.65	0.37	0.74	0.51	0.82	0.48	0.86	0.71	0.95	0.64	0.92	0.49	0.78	0.52	0.84
OCD	0.27	0.31	0.62	0.47	0.83	0.38	0.70	0.42	0.82	0.66	0.94	0.61	0.91	0.52	0.81	0.54	0.85
PTSD	0.20	0.25	0.56	0.50	0.86	0.52	0.84	0.60	0.92	0.62	0.93	0.59	0.91	0.48	0.78	0.43	0.78
Mania/hypomania	0.96	0.54	0.79	0.53	0.82	0.53	0.79	0.39	0.75	0.41	0.77	0.53	0.84	0.51	0.75	0.35	0.64
MDD/dysthymia	2.57	0.50	0.72	0.54	0.80	0.60	0.80	0.59	0.86	0.64	0.89	0.56	0.83	0.61	0.78	0.57	0.79
Pain disorder	2.04	0.24	0.44	0.28	0.55	0.43	0.67	0.35	0.67	0.55	0.85	0.49	0.78	0.53	0.73	0.44	0.70
SSI 4/6	0.54	0.44	0.74	0.51	0.83	0.52	0.80	0.53	0.87	0.59	0.90	0.64	0.90	0.44	0.71	0.40	0.72
Any psychotic	0.03	0.11	0.34	0.15	−1.00	0.30	0.69	0.24	−1.00	0.25	−1.00	0.48	0.90	0.45	0.82	0.36	0.77
ASP	2.43	0.63	0.82	0.62	0.85	0.57	0.78	0.08	0.22	0.38	0.70	0.34	0.63	0.40	0.59	0.39	0.64

*SUD: substance abuse disorder.

Table 3b. Matrix of tetrachoric correlations among 17 lifetime mental disorders (person $\times$ year data, $N = 14\,034$, age span 19–25 years)

	Percentage affected	Alcohol SUD		Ill. drug SUD		Nicotine SUD		Eating disorder		Panic disorder		Agora-phobia		Specific phobia/ NOS		Social phobia	
		Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata
Alcohol SUD	27.24	1.00	1.00														
Ill. drug SUD	10.13	0.53	0.61	1.00	1.00												
Nicotine SUD	25.12	0.36	0.39	0.50	0.58	1.00	1.00										
Eating disorder	3.79	−0.05	−0.08	0.06	0.10	0.16	0.23	1.00	1.00								
Panic disorder	2.96	0.14	0.21	0.12	0.19	0.24	0.34	0.22	0.39	1.00	1.00						
Agoraphobia	3.98	0.05	0.07	0.15	0.24	0.29	0.39	0.29	0.46	0.79	0.90	1.00	1.00				
Specific phobia/NOS	27.01	0.01	0.01	0.10	0.12	0.18	0.20	0.31	0.42	0.38	0.53	0.37	0.49	1.00	1.00		
Social phobia	10.78	0.08	0.10	0.10	0.14	0.21	0.26	0.19	0.30	0.37	0.53	0.46	0.61	0.24	0.29	1.00	1.00
GAD	2.84	0.04	0.07	0.25	0.39	0.35	0.49	0.24	0.42	0.41	0.64	0.51	0.71	0.30	0.43	0.43	0.60
OCD	1.59	0.02	0.04	0.18	0.31	0.28	0.43	0.39	0.63	0.57	0.79	0.59	0.80	0.29	0.44	0.36	0.56
PTSD	2.07	0.00	0.01	0.27	0.43	0.23	0.35	0.30	0.51	0.48	0.71	0.46	0.68	0.33	0.48	0.34	0.51
Amania/hypomania	6.18	0.19	0.25	0.25	0.36	0.26	0.34	0.13	0.23	0.23	0.37	0.36	0.53	0.24	0.31	0.17	0.25
MDD/dysthymia	20.05	0.08	0.08	0.19	0.24	0.26	0.29	0.38	0.51	0.41	0.56	0.38	0.50	0.34	0.37	0.37	0.44
Pain disorder	10.25	−0.14	−0.18	0.08	0.11	0.16	0.20	0.20	0.30	0.28	0.43	0.28	0.41	0.25	0.31	0.18	0.24
SSI 4/6	2.79	0.03	0.05	0.24	0.37	0.27	0.38	0.38	0.59	0.37	0.59	0.46	0.67	0.21	0.31	0.30	0.45
Any psychotic	0.38	0.27	0.48	0.22	0.43	0.15	0.29	0.20	0.43	−0.15	−1.00	0.31	0.59	0.17	0.31	0.15	0.30
ASP	9.31	0.51	0.59	0.51	0.62	0.39	0.46	−0.06	−0.11	0.16	0.26	0.04	0.07	0.06	0.08	0.11	0.16

GAD		OCD		PTSD		Mania/ hypomania		MDD/ dysthymia		Pain disorder		SSI 4/6		Any psychotic		ASP	
Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata
1.00	1.00																
0.70	0.95	1.00	1.00														
0.57	0.92	0.57	0.92	1.00	1.00												
0.57	0.88	0.47	0.83	0.50	0.85	1.00	1.00										
0.77	0.95	0.59	0.87	0.61	0.89	0.43	0.71	1.00	1.00								
0.40	0.75	0.26	0.58	0.48	0.82	0.43	0.72	0.43	0.68	1.00	1.00						
0.57	0.90	0.51	0.87	0.65	0.93	0.40	0.75	0.54	0.82	0.65	0.89	1.00	1.00				
0.58	0.95	0.58	0.95	0.61	0.96	0.15	−1.00	0.49	0.87	0.07	−1.00	0.43	0.87	1.00	1.00		
0.43	0.76	0.40	0.74	0.27	0.59	0.45	0.73	0.37	0.60	0.34	0.58	0.35	0.65	0.05	−1.00	1.00	1.00

GAD		OCD		PTSD		Mania/ hypomania		MDD/ dysthymia		Pain disorder		SSI 4/6		Any psychotic		ASP	
Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata
1.00	1.00																
0.55	0.78	1.00	1.00														
0.40	0.64	0.47	0.73	1.00	1.00												
0.28	0.44	0.45	0.67	0.27	0.45	1.00	1.00										
0.54	0.70	0.37	0.54	0.46	0.64	0.11	0.15	1.00	1.00								
0.24	0.38	0.25	0.42	0.22	0.36	0.25	0.35	0.26	0.32	1.00	1.00						
0.36	0.59	0.54	0.77	0.57	0.79	0.38	0.57	0.35	0.49	0.54	0.70	1.00	1.00				
0.16	0.38	0.25	0.57	0.36	0.69	0.25	0.50	0.27	0.49	0.18	0.36	0.51	0.81	1.00	1.00		
0.19	0.31	0.23	0.38	0.23	0.38	0.23	0.33	0.07	0.10	0.02	0.02	0.12	0.20	0.04	0.09	1.00	1.00

Table 3c. Matrix of tetrachoric correlations among 17 lifetime mental disorders (person $\times$ year data, $N = 4867$, age span 26–34 years)

	Percentage affected	Alcohol SUD		Ill. drug SUD		Nicotine SUD		Eating disorder		Panic disorder		Agora-phobia		Specific phobia/ NOS		Social phobia	
		Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata
Alcohol SUD	29.63	1.00	1.00														
Ill. drug SUD	8.65	0.53	0.61	1.00	1.00												
Nicotine SUD	26.73	0.32	0.34	0.54	0.62	1.00	1.00										
Eating disorder	5.03	0.01	0.02	0.16	0.25	0.10	0.13	1.00	1.00								
Panic disorder	3.76	0.10	0.14	0.06	0.11	0.23	0.33	0.14	0.24	1.00	1.00						
Agoraphobia	5.94	0.12	0.16	0.21	0.31	0.28	0.36	0.12	0.20	0.82	0.90	1.00	1.00				
Spec. phobia/NOS	28.44	0.01	0.01	0.11	0.14	0.22	0.24	0.41	0.52	0.35	0.47	0.42	0.52	1.00	1.00		
Social phobia	12.80	0.08	0.09	0.03	0.05	0.23	0.27	0.27	0.39	0.41	0.55	0.43	0.55	0.23	0.27	1.00	1.00
GAD	4.62	0.04	0.05	0.24	0.37	0.39	0.51	0.24	0.38	0.50	0.68	0.58	0.73	0.33	0.43	0.51	0.65
OCD	1.77	−0.05	−0.08	0.12	0.21	0.31	0.46	0.48	0.70	0.49	0.72	0.52	0.73	0.38	0.55	0.26	0.41
PTSD	3.00	0.04	0.06	0.25	0.39	0.31	0.44	−0.04	−0.09	0.46	0.67	0.48	0.67	0.29	0.41	0.39	0.54
Mania/hypomania	6.86	0.01	0.01	0.10	0.15	0.22	0.29	0.10	0.16	0.11	0.18	0.33	0.47	0.15	0.20	0.14	0.20
MDD/dysthymia	26.77	0.05	0.05	0.15	0.20	0.25	0.26	0.42	0.54	0.28	0.38	0.44	0.55	0.36	0.38	0.39	0.45
Pain disorder	10.70	−0.11	−0.14	0.08	0.12	0.13	0.16	0.22	0.32	0.14	0.22	0.21	0.31	0.21	0.25	0.10	0.13
SSI 4/6	3.02	−0.03	−0.05	0.13	0.23	0.22	0.32	0.44	0.63	0.35	0.56	0.43	0.62	0.17	0.24	0.42	0.58
Any psychotic	0.78	0.46	0.70	0.36	0.59	0.02	0.04	0.09	0.19	−0.15	−1.00	−0.02	−0.05	0.11	0.19	0.12	0.23
ASP	8.51	0.46	0.55	0.60	0.71	0.38	0.46	−0.10	−0.17	0.05	0.09	0.19	0.28	0.08	0.11	0.09	0.13

for which Stata revealed a tetrachoric correlations of −1.00 whereas Mplus yielded positive correlations. In two of these cases Mplus revealed correlations greater than 0.2. These were the associations between any psychotic disorder and panic disorder (−1.00 versus 0.24, no psychotic disorder/no panic disorder $N = 44\,292$, no psychotic disorder/panic disorder $N = 169$, psychotic disorder/no panic disorder $N = 14$, psychotic disorder/panic disorder $N = 0$) and the one between any psychotic disorder and any eating disorder (−1.00 versus 0.24, no psychotic disorder/no eating disorder $N = 44\,268$, no psychotic disorder/eating disorder $N = 193$, psychotic disorder/eating disorder $N = 14$, psychotic disorder/panic disorder $N = 0$).

Discussion

The substantial discrepancies found in the results for tetrachoric correlations computed with Stata and Mplus require explanation. Particularly in light of the fact that both programs have been used extensively in recent publications, a clarification of this issue seems to be imperative. Assisted by the editors of the IJMPR a draft version of this manuscript was sent to STATA and

Mplus with a plea for comments. Answers were received in June 2006 and clarify the issues sufficiently.

Reply by Linda Muthen from Mplus (6 January 2006)

According to the respective manuals Mplus uses a numerical maximum likelihood approach to compute the correlation coefficient from the bivariate normal distribution function (Kirk, 1973; Divgi, 1979), Stata applies an approximate explicit formula published by Edwards and Edwards (1984).

Reply by Vince Wiggins from Stata Corporation (6 January 2006)

Günther and Höfler find that the different estimators of tetrachoric correlations used by Stata and Mplus produce substantially different estimates when applied to many of the variables in a psychopathology study. Mplus uses an iterative technique to compute maximum likelihood (ML) estimates, whereas in the interest of speed Stata uses the closed-form approximation due to Edwards (1984). The documentation for each package clearly describes the estimator used. Stata's manuals

GAD		OCD		PTSD		Mania/ hypomania		MDD/ dysthymia		Pain disorder		SSI 4/6		Any psychotic		ASP	
Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata	Mplus	Stata
1.00	1.00																
0.62	0.81	1.00	1.00														
0.55	0.73	0.55	0.77	1.00	1.00												
0.29	0.44	0.45	0.66	0.25	0.40	1.00	1.00										
0.55	0.68	0.48	0.67	0.47	0.63	0.13	0.17	1.00	1.00								
0.31	0.44	0.43	0.62	0.27	0.42	0.31	0.42	0.27	0.32	1.00	1.00						
0.38	0.58	0.65	0.84	0.57	0.77	0.45	0.63	0.36	0.49	0.58	0.73	1.00	1.00				
-0.18	-1.00	-0.04	-1.00	0.38	0.66	0.29	0.52	-0.01	-0.01	-0.00	-0.01	0.34	0.62	1.00	1.00		
0.19	0.29	0.24	0.40	0.25	0.40	0.16	0.24	0.05	0.07	-0.04	-0.06	0.08	0.14	0.36	0.60	1.00	1.00

even provide the formula for the Edwards approximation. Both packages are correctly computing what they claim to compute.

Despite that, the Edwards approximation used by Stata does not perform well for a subclass of problems. The tetrachoric correlation measures the correlation of a latent bivariate normal distribution by using the 2×2 tabulation of two observed binary outcomes, each of which takes the value 1 when some threshold is exceeded in its dimension of the bivariate distribution. When these thresholds are in the tails of the distribution, the Edwards approximation is indefensibly poor. The approximation works well when the marginals in both dimensions are above 10% of the observations; it works poorly when they are well below 10%.

Stata's **biprobit** command will compute the tetrachoric correlation by ML, but Stata's **tetrachoric** command uses the approximation. We will update the **tetrachoric** command to use the ML estimator. By the time you read this in print, **tetrachoric** will have already been issued in a free update.

Acknowledgement

We wish to thank the representatives of Stata and Mplus for their replies. We appreciate that the problem is solved and that Stata shall launch a modified procedure.

This work is further part of the Early Developmental Stages of Psychopathology (EDSP) Study and is funded by the German Federal Ministry of Education and Research (BMBF) project no. 01EB9405/6, 01EB 9901/6, EB01016200, 01EB0140, and 01EB0440. Part of the field work and analyses were also supported by grants of the Deutsche Forschungsgemeinschaft (DFG) LA1148/1-1, WI2246/1-1, WI 709/7-1, and WI 709/8-1.

References

- American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 4 edn. Washington DC: APA, 1994.
- Divgi DR. Calculation of the tetrachoric correlation coefficient. *Psychometrika* 1979; 44(2): 169–72.
- Edwards JH, Edwards AWF. Approximating the tetrachoric correlation coefficient. *Biometrics* 1984; 40: 563.
- Hettema JM, Prescott CA, Myers JM, Neale MC, Kendler KS. The structure of genetic and environmental risk

- factors for anxiety disorders in men and women. *Arch Gen Psychiatr* 2005; 62: 182–9.
- Kendler KS, Prescott CA, Myers J, Neale MC. The structure of genetic and environmental risk factors for common psychiatric and substance use disorders in men and women. *Arch Gen Psychiatr* 2003; 60: 929–37.
- Kendler KS, Walters EJ, Neale MS, Kessler RC, Heath AC, Eaves LJ. The structure of genetic and environmental risk factors for six major psychiatric disorders in women. *Arch Gen Psychiatr* 1995; 52: 374–83.
- Kessler RC, Chiu WT, Demler O, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey. *Arch Gen Psychiatr* 2005; 62: 617–27.
- Kirk DB. On the numerical approximation of the bivariate normal (tetrachoric) correlation coefficient. *Psychometrika* 1973; 38(2): 259–68.
- Krueger RF. The structure of common mental disorders. *Arch Gen Psychiatr* 1999; 56: 921–31.
- Krueger RF, Caspi A, Moffitt TE, Silva PA. The structure and stability of common mental disorders (DSM-III-R): a longitudinal – epidemiological study. *J Abnorm Psychol* 1998; 107: 216–27.
- Krueger RF, Caspi A, Moffitt TE, Silva PA, McGee. Personality traits are differentially linked to mental disorders: A multitrait-multidiagnosis study of an adolescent birth cohort. *J Abnorm Psychol* 1996; 105: 299–312.
- Krueger RF, Finger MS. Using item response theory to understand comorbidity among anxiety and unipolar mood disorders. *Psychol Assess* 2001; 13: 140–51.
- Krueger RF, McGue M, Iacono WG. The higher-order structure of common DSM mental disorders: internalization, externalization, and their connections to personality. *Pers Indiv Differ* 2001; 30: 1245–59.
- Lieb R, Isensee B, Sydow von K, Wittchen HU. The Early Developmental Stages of the Psychopathology Study (EDSP): A methodological update. *Eur Addiction Res* 2000; 6: 170–82.
- Muthén LK, Muthén BO. *Mplus User's Guide*. Third Edition. Los Angeles CA: Muthén & Muthén, 1998–2005.
- StataCorp. *Stata Statistical Software: Release 9.2*. College Station, TX: Stata Corporation, 2006.
- Vollebergh WAM, Iedema J, Bijl R, De Graaf R, Smit F, Ormel J. The structure and stability of common mental disorders. The Nemesis Study. *Arch Gen Psychiatr* 2001; 58: 597–603.
- Wittchen HU, Höfler M, Merikangas KR. Toward the identification of core psychopathological processes? *Arch Gen Psychiatr* 1999; 56: 929–31.
- Wittchen HU, Nelson CB, Lachner G. Prevalence of mental disorders and psychosocial impairments in adolescents and young adults. *Psychol Med* 1998b; 28: 109–26.
- Wittchen HU, Perkonig A, Lachner G, Nelson CB. Early Developmental Stages of Psychopathology Study (EDSP): objectives and design. *Eur Addiction Res* 1998a; 4: 18–27.
- Wittchen HU, Robins LN, Wing J. *Composite International Diagnostic Interview (CIDI)*. Munich: Max-Planck-Institut für Psychiatrie, Klinisches Institut, 1984.

Correspondence: Michael Höfler, Institute of Clinical Psychology and Psychotherapy Technische Universität Dresden, Chemnitz Str. 46a, 01187 Dresden, Germany.

Telephone (+49) 351-463-36921

Fax (+49) 351-463-3698

Email: hoefler@psychologie.tu-dresden.de