The Genetic Relationship Between Psychological Distress, Somatic Distress, Affective Disorders, and Substance Use in Young Australian Adults: A Multivariate Twin Study

Lun-Hsien Chang,[1,2] Baptiste Couvy-Duchesne,[3] Sarah E. Medland,[1] Nathan A. Gillespie,[4] Ian B. Hickie,[5] Richard Parker,[1] and Nicholas G. Martin[1]

[1]Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia
[2]Faculty of Medicine, University of Queensland, Brisbane, Queensland, Australia
[3]Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland, Australia
[4]Virginia Institute for Psychiatric and Behavioural Genetics, Virginia Commonwealth University, Richmond, VA, USA
[5]Brain and Mind Centre, University of Sydney, Sydney, New South Wales, Australia

Psychological distress (PSYCH), somatic distress (SOMA), affective disorders (AD), and substance use (SU) frequently co-occur. The genetic relationship between PSYCH and SOMA, however, remains understudied. We examined the genetic and environmental influences on these two disorders and their comorbid AD and SU using structural equation modeling. Self-reported PSYCH and SOMA were measured in 1,548 twins using the two subscales of a 12-item questionnaire, the Somatic and Psychological Health Report. Its reliability and psychometric properties were examined. Six ADs, involvement of licit and illicit substance, and two SU disorders were obtained from 1,663–2,132 twins using the World Mental Health Composite International Diagnostic Interview and/or from an online adaption of the same. SU phenotypes (heritability: 49–79%) were found to be more heritable than the affective disorder phenotypes (heritability: 32–42%), SOMA (heritability: 25%), and PSYCH (heritability: 23%). We fit separate non-parametric item response theory models for PSYCH, SOMA, AD, and SU. The IRT scores were used as the refined phenotypes for fitting multivariate genetic models. The best-fitting model showed the similar amount of genetic overlap between PSYCH–AD (genetic correlation $r_G = 0.49$) and SOMA–AD ($r_G = 0.53$), as well as between PSYCH–SU ($r_G = 0.23$) and SOMA–SU ($r_G = 0.25$). Unique environmental factors explained 53% to 76% of the variance in each of these four phenotypes, whereas additive genetic factors explained 17% to 46% of the variance. The covariance between the four phenotypes was largely explained by unique environmental factors. Common genetic factor had a significant influence on all the four phenotypes, but they explained a moderate portion of the covariance.

■ **Keywords:** twin studies, psychiatric disorders, addiction, genetics, social anxiety, psychometrics, item response theory, structural equation modeling

Psychological Distress and Somatic Distress

Psychological distress (PSYCH) and somatic distress (SOMA) commonly co-occur (Clarke et al., 2008; de Waal et al., 2004; Lee et al., 2015; Shidhaye et al., 2013). PSYCH can be described as a negative feeling that affects a person's life. This discomfort is usually manifested as anxiety, frustration, hopelessness, sadness, or worthlessness. SOMA, also known as somatoform disorder or somatization syndrome, can involve various bodily discomforts such as prolonged fatigue, pain, tiredness, or sleep disturbance (Hanel et al., 2009; Hiller et al., 2006). The prevalence of PSYCH and SOMA tends to vary across studies, likely due to the heterogeneous definition of the distress and the group from which the distress is measured. An analysis

of five Australian National Health Surveys found that a high level of PSYCH affected 11.1–13.4% of the population aged 18–65. This prevalence remained fairly stable from 2001 to 2014 (Harvey et al., 2017). In primary care settings, approximately 9.5% of Australian general practice attenders had a very high level (K10 ≥ 30) of PSYCH (Clarke et al., 2008). Somatization syndromes based on the Patient Health Questionnaire, on the other hand, occurred in 9.3% of the German general population (Kocalevent et al., 2013). A higher prevalence has been reported in clinical samples, with 16.1% of Dutch general practice participants reporting somatoform disorders as classified in the *Diagnostic and Statistical Manual of Mental Disorders*, *Fourth Edition* (*DSM-IV*; de Waal et al., 2004), and 18.5% of Australians being classified as somatisers (Clarke et al., 2008). Despite the varying prevalence, a common finding in the studies is a high comorbidity of PSYCH, anxiety/depressive disorder, and somatoform disorders (Barsky et al., 1992; Clarke et al., 2008; de Waal et al., 2004; Escobar et al., 1998; Eytan et al., 2011; Maier & Falkai, 1999; Ormel et al., 1994). A higher level of distress, as measured by the 10-item Kessler Psychological Distress Scale (K10; Kessler et al., 2002), is associated with higher probability of meeting criteria for a mental disorder or substance use (SU) disorder (Andrews & Slade, 2001; Slade et al., 2011).

SPHERE-12 for Measuring PSYCH and SOMA

The 12-item Somatic and Psychological Health Report (SPHERE-12) is shortened from the original 34-item SPHERE for screening mental illnesses that are commonly found in primary care settings and for use in research (Hickie et al., 2001). The SPHERE-12 questionnaire consists of two six-item sub-scales: the PSYCH-6, measuring PSYCH, and the SOMA-6, measuring SOMA.

Short questionnaires can be advantageous as short instruments with 12 or fewer items, such as the Beck Depression Inventory-Short Form (BDI-SF; Beck & Steer, 1993), Patient Health Questionnaire-9 (PHQ-9; Kroenke et al., 2001), and K10 (Kessler et al., 2002), have been shown to be as accurate as longer questionnaires that contain 15 or more items (Cheng & Chan, 2005) in detecting depression (Akena et al., 2012). In addition, they can be completed in a shorter time with lower participant burden and higher completion rates. Few studies have focused on PSYCH and SOMA measured with the SPHERE scale (Clarke & McKenzie, 2003; Couvy-Duchesne et al., 2017a).

Classical Twin Studies

The classical twin study provides a unique opportunity to understand the extent to which a trait or a group of related traits is governed by nature or nurture. This method is commonly used to estimate the relative importance of genetic or environmental influences on comorbid disorders (Agrawal et al., 2010; Han et al., 1999; Karkowski et al., 2000; Kendler et al., 2003; Kendler et al., 2007; Lynskey et al., 2004). Twin

studies typically employ standard error of the mean (SEM), which provides various models to partition the variation of a single phenotype, or the covariance of multiple phenotypes, into additive genetic component and environmental component. Multivariate twin studies typically find that psychiatric disorders are influenced by both genetic and environmental factors (Gelernter, 2015; Sullivan et al., 2000).

The Genetic Relationship Between PSYCH and SOMA

Few studies have examined the genetic etiology of PSYCH and SOMA (Ball et al., 2011; Gillespie et al., 2000; Hansell et al., 2012; Hickie et al., 1999; Ivkovic et al., 2007). Moderate heritability has been estimated for PSYCH (0.18–0.4) and SOMA (0.32–0.43). A high genetic correlation (0.84) has been found between PSYCH and SOMA in Australian adolescents and young adults (Hansell et al., 2012), suggesting a large overlapping genetic effect on these two phenotypes. Little is known about the genetic relationship between PSYCH, SOMA, and the more severe mental disorders and SU.

Aims

The objectives of this study were to (1) validate the PSYCH-6 and SOMA-6 sub-scales; (2) estimate the heritability for the PSYCH-6 and the SOMA-6, as well as CIDI-based assessment of affective disorders (AD) and SU; and (3) examine the overlapping genetic influence on the PSYCH-6, SOMA-6, AD, and SU. We examined the internal consistency and reliability of the PSYCH-6 and SOMA-6. We applied the item response theory (IRT) modeling to create refined phenotypes that estimate the normalized liability to PSYCH, SOMA, AD, and SU. We estimated the heritability for individual AD, SU phenotypes, as well as the IRT theta scores for the four scales (PSYCH-6–IRT, SOMA-6–IRT, AD–IRT, and SU–IRT). Finally, we investigated the genetic and environmental influences on the covariance between these four IRT scores.

Materials and Methods

The 19Up Study

The Brisbane Longitudinal Twin Study (BLTS; Wright & Martin, 2004) has collected a comprehensive array of neurobiological correlates, environmental risk factors, and endophenotypes for psychiatric disorders, since its inception in 1992 at QIMR Berghofer Medical Research Institute. The current study used data collected in one of the BLTS studies, 19Up (see Figure 1 for details of data collection; Couvy-Duchesne et al., 2018; Gillespie et al., 2013).

PSYCH and SOMA were assessed with the SPHERE-12 questionnaire when the participants were on average 26 years old. The SPHERE-12 data were available from all the three waves of the 19Up: NU1, NU2, and NU3. In addition, the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (*DSM-5*, American Psychiatric Association, 2013) diagnoses were made using an online version
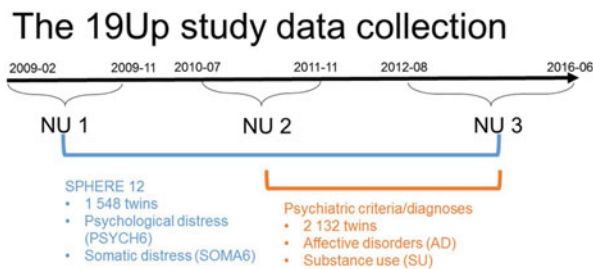
The 19Up study data collection

FIGURE 1

(Colour online) Timeline of data collection of psychological distress, somatic distress, and CIDI-based psychiatric diagnoses in the 19Up study.

of the Composite International Diagnostic Interview (CIDI; World Health Organization, 1990). The CIDI-based diagnoses were collected in NU2 and NU3. The 19Up study was approved by the QIMR Human Research Ethics Committee. Data were stored in compliance with national regulations regarding personal data protection. Informed consent was obtained from all the participants.

Participants

Detailed descriptions of participation rates, sample demographics, and the prevalence of mental or SU disorders in the 19Up cohort have been published (Couvy-Duchesne et al., 2018). The current study only used the data collected from twin individuals. The SPHERE-12 sample was comprised of 1,548 twin individuals. Their mean age was 25.3 years (age range: 18.4–38.4) and the sex ratio was 56.9% females. The CIDI sample was composed of measurements from 2,132 twin individuals. Their mean age was 26.1 years (age range: 18.7–38.6) and the sex ratio was 57.8% females. The number of complete twin pairs and twin individuals are shown for each study wave and zygosity group in Table 1.

Measures

Following standard administration procedures, respondents were asked to rate each SPHERE-12 item on the extent it had troubled them over the previous two weeks: sometimes or never having the problem (coded as 0), often (coded as 1), and most of the time (coded as 2). Internal consistency of the SPHERE-12 questionnaire was assessed using Cronbach's alpha (Cronbach, 1951) both item by item and for the whole scale without excluding any items. A small proportion of the participants answered the SPHERE-12 questions in two or three waves, enabling us to evaluate the test–retest reliability of the questionnaire between (1) NU1 and NU2, (2) NU2 and NU3, and (3) NU1 and NU3. There were 132 participants in both NU1 and NU2, 20 participants in both NU2 and NU3, and 22 participants in both NU1 and NU3. SPHERE data used for the item-by-item assessment of internal consistency consisted of 2,122–2,126 twins and siblings. Phenotypes used in our univari-

ate and multivariate genetic modeling were composed of measurements per twin individual. For twins with multiple measurements, their earliest wave was used.

The CIDI was used to assess various lifetime *DSM-5* psychiatric criteria to obtain diagnostic classifications relevant to AD or SU. The AD included agoraphobia, depressive episodes, major depressive disorder (MDD), panic attack (PA), panic disorder (PD), and social anxiety. The SU phenotypes included ever using alcoholic beverages, tobacco products, cannabis, non-medical use of prescription drugs (e.g., pain killers, stimulants), or illicit drugs, such as cocaine, amphetamine, inhalants, sedatives or sleeping pills, hallucinogens, opioids, and party drugs (e.g., MDMA, ketamine, or gamma-hydroxybutyrate), as well as alcohol-use disorder (AUD), and cannabis-use disorder (CUD). We selected six binary AD diagnoses and three ordinal SU diagnoses to construct the AD and SU scale.

Non-Parametric Item Response Theory Modeling

We employed the non-parametric IRT to evaluate item properties within each psychiatric domain or scale (i.e., PSYCH-6, SOMA-6, AD, and SU). One advantage of IRT is that item difficulty and item discriminability are used to weight the items when calculating the IRT score. In addition, IRT makes it possible to combine information from uniquely measured psychopathology (Thomas, 2011). Consequently, the IRT score can yield more precise estimation of the underlying trait than a summed score. This represents an improvement over the classical test theory, in which different items are usually assumed to have equal difficulty and discriminability and items' properties cannot be evaluated.

IRT models can be categorized as parametric IRT or non-parametric IRT. The parametric IRT assumes a logistic shape on item-response step functions (IRSFs). This more restricted method can lead to the rejection of useful items when the assumption is violated. Non-parametric IRT, however, does not impose a logistic shape on the IRSFs. This flexible method has become increasingly popular in modeling various scales, such as personality (Maydeu-Olivares, 2005), quality of life (Sijtsma et al., 2008), psychopathology (e.g., Khan et al., 2011; Meijer & Baneke, 2004), and mental health (Stochl et al., 2012).

In IRT, the family of graded response models (GRMs; Samejima, 1968) is suitable for analyzing polychotomous items collected by means of response scales (Hemker et al., 1997; Mellenbergh, 1995). IRT models the probability of answering each option in an item as a function of the underlying trait (e.g., level of depression or somatization). Such conditional probabilities are often referred to as IRSFs. Here, we used non-parametric GRM (np-GRM) that makes no assumptions about the underlying shape of the IRSF (Sijtsma et al., 2008). This model allows the checking of a fundamental requirement in IRT and scale construction — the probability of reporting a symptom is a strictly growing function of the latent score (hypothesis of monotonicity of

TABLE 1

Number of Complete Twin Pairs and Individual Twins (Parenthetical) in Each Zygosity and Sex Group and Across All the Groups (Sum) From Different Waves of (1) SPHERE-12 Questionnaire, (2) Diagnostic Interviews, and (3) Twins Who Are in Both Samples (Overlapped)

| | SPHERE-12 | | | | Diagnoses | | | Overlapped |
Zygosity	NU1	NU2	NU3	All waves	NU2	NU3	All waves	All waves
Monozygotic females	43 (98)	34 (109)	65 (178)	160 (385)	47 (144)	151 (380)	232 (524)	154 (376)
Monozygotic males	14 (46)	25 (86)	46 (146)	102 (278)	27 (96)	89 (265)	139 (361)	93 (265)
Dizygotic females	27 (76)	13 (54)	42 (125)	101 (255)	17 (64)	127 (318)	166 (382)	96 (248)
Dizygotic males	10 (41)	10 (48)	22 (97)	60 (186)	12 (57)	71 (216)	100 (273)	52 (170)
Dizygotic opposite sexes	43 (113)	20 (110)	64 (221)	160 (444)	26 (126)	155 (466)	219 (592)	144 (421)
Sum	137 (374)	102 (407)	239 (767)	583 (1,548)	129 (487)	596 (1,645)	860 (2,132)	539 (1,480)

Note: SPHERE-12 data were collected from three waves: NU1, NU2, and NU3. Diagnostic data were collected from two waves: NU2 and NU3.

IRSF). In practice, this ensures that the ordering of respondents on the score reflects the true ordering on the latent trait (Ark, 2005).

The np-GRM employs a kernel-smoothing technique to model the relationship between the level of underlying trait and the probability of choosing a particular option for a questionnaire (Ramsay, 1991). We visually examined the IRSF of every item for determining items that were included in the final scales. Items were included in the np-GRMs if they were monotonically increasing. The PSYCH-6 (questions 1–6 in Table 2) and the SOMA-6 (questions 7–12 in Table 2) included every item in their np-GRMs. The AD scale included the aforementioned six binary *DSM-5* diagnosis items. The SU scale included three ordinal diagnoses: AUD, CUD, and degree of drug use. The degree of drug use was derived from ever using any illicit substance aforementioned. This variable was coded as 0 for never using any of the nine drugs, 1 for ever using any one of them, 2 for ever using any two of them, and 3 for ever using 3 or more of them. The IRT scaling provided a better distributional shape, with reduced skewness and kurtosis relative to the scaling of the PSYCH-6 and the SOMA-6 summed scores (Figure S1).

We attempted various data transformations for PSYCH-6 and SOMA-6 measurements; however, a reverse J-shaped distribution was seen in all the transformed measurements. Even though IRT modeling attempts to put all data on a normal scale of liability with a limited number of items, it was not possible to remove all kurtosis and skewness, which is evident in the distribution of our data. We combined the PSYCH-6 and SOMA-6 measurements across the three waves, as we did not find significant differences in the IRT scores among the three study waves for the PSYCH-6 (Kruskal–Wallis chi-squared = 3.4,511, *df* = 2, *p* value = .1781) and the SOMA-6 (Kruskal–Wallis chi-squared = 0.69,418, *df* = 2, *p* value = .7067). The IRSFs for the four scales, PSYCH-6, SOMA-6, AD, and SU, are shown in Figures S2–S5.

Basic Assumption Tests

There are different zygosity and sex groups in the twin design (Evans et al., 1999). For binary phenotypes, the dif-

ferences in thresholds and covariances were examined between different zygosity and sex groups. For continuous phenotypes, the differences in means, variances, and covariances were examined across all the groups (Evans et al., 1999). When no significant differences are found between the groups, means and variances can then be combined across the groups to estimate the underlying population mean and variance (Evans et al., 1999).

Univariate Twin Modeling

The observed variance of the phenotypes was partitioned into (A) additive genetic, (D) dominant genetic, (C) common environmental, and (E) unshared environmental variation using a univariate ACE or ADE model. The ACE models were fit for the phenotypes where MZ correlations (rMZ) were smaller than twice the DZ correlations (rDZ). The ADE models were fit for the phenotypes where rMZ were greater than twice the rDZ. The presence of D effect was not significant, given large sample sizes are required to detect a significant D.

Univariate twin models were fit separately for each continuous, binary, and ordinal phenotypes using the full-information maximum likelihood (FIML) method, which allows the use of data from all available individuals (see the number of twins in Table 3), including those without co-twins and those with co-twins who had missing measures. The log-likelihood ratio test was performed to assess the fit of nested models. We regressed the IRT scores (PSYCH-6–IRT, SOMA-6–IRT, AD–IRT, and SU–IRT) against covariates age of survey, sex, and study waves for obtaining their residuals. The four residualized IRT scores (PSYCH-6r, SOMA-6r, ADr, and SUr) were used to fit multivariate twin models.

Multivariate Twin Modeling

Multivariate twin models allow us to estimate the relative contribution of genes and environment to the covariation between measures (Neale & Cardon, 1992). Our analytical flow started with fitting the fully saturated Cholesky decomposition (CD) for PSYCH-6r, SOMA-6r, ADr, and SUr, followed by the more parsimonious independent pathway (IP) models, and common pathway (CP) models. To make

TABLE 2

Internal Consistency Assessed With Cronbach's Alpha Coefficients and Test–Retest Reliability Assessed With Intra-class Correlation Coefficients for Individual SPHERE-12 Questionnaire Items

| | | Internal consistency | | Test–retest reliability | | | | | | | | |
| | | | | NU1–NU2 ($N = 132$, $df = 131$) | | | NU2–NU3 ($N = 20$, $df = 19$) | | | NU1–NU3 ($N = 22$, $df = 21$) | | |
SPHERE-12 item	Number of items	Sample size	Cronbach's alpha	ICC	F	p value	ICC	F	p value	ICC	F	p value
1. Feeling nervous or tense	11	2,124	0.87	0.61 [0.49, 0.71]	4.16	2.1E-15	0.52 [0.12, 0.78]	3.17	.0069	0.20 [-0.22, 0.57]	1.50	.17
2. Feeling unhappy and depressed	11	2,122	0.87	0.60 [0.48, 0.70]	4.04	6.6E-15	0.57 [0.20, 0.81]	3.70	.0027	0.23 [-0.19, 0.59]	1.61	.14
3. Feeling constantly under strain	11	2,122	0.87	0.60 [0.48, 0.70]	4.01	8.8E-15	0.57 [0.19, 0.80]	3.68	.0028	0.27 [-0.15, 0.61]	1.75	.1
4. Everything getting on top of you	11	2,122	0.87	0.59 [0.47, 0.69]	3.92	2.2E-14	0.57 [0.18, 0.80]	3.60	.0032	0.26 [-0.17, 0.60]	1.69	.12
5. Losing confidence	11	2,122	0.87	0.61 [0.49, 0.71]	4.11	3.3E-15	0.56 [0.18, 0.80]	3.59	.0033	0.22 [-0.20, 0.58]	1.57	.15
6. Being unable to overcome difficulties	11	2,122	0.87	0.62 [0.50, 0.71]	4.26	7.2E-16	0.52 [0.12, 0.78]	3.17	.0068	0.22 [-0.20, 0.58]	1.57	.15
7. Muscle pain after activity	11	2,123	0.88	0.65 [0.54, 0.74]	4.70	1.1E-17	0.62 [0.27, 0.83]	4.33	.001	0.25 [-0.17, 0.60]	1.68	.12
8. Needing to sleep longer	11	2,124	0.87	0.63 [0.52, 0.73]	4.46	1.1E-16	0.55 [0.16, 0.79]	3.41	.0045	0.27 [-0.15, 0.62]	1.76	.099
9. Prolonged tiredness after activity	11	2,122	0.87	0.62 [0.50, 0.71]	4.26	7.6E-16	0.55 [0.17, 0.79]	3.49	.0039	0.19 [-0.23, 0.56]	1.48	.18
10. Poor sleep	11	2,122	0.88	0.61 [0.50, 0.71]	4.19	1.5E-15	0.50 [0.10, 0.77]	3.01	.0092	0.21 [-0.22, 0.57]	1.52	.17
11. Poor concentration	11	2,122	0.87	0.63 [0.51, 0.72]	4.36	2.8E-16	0.57 [0.19, 0.80]	3.64	.0031	0.15 [-0.27, 0.53]	1.36	.24
12. Tired muscles after activity	11	2,126	0.88	0.64 [0.53, 0.73]	4.62	2.6E-17	0.62 [0.26, 0.83]	4.24	.0012	0.25 [-0.18, 0.60]	1.65	.12
Total scores	12	2,122	0.88	0.62 [0.51, 0.72]	4.31	4.7E-16	0.57 [0.19, 0.80]	3.63	.0031	0.23 [-0.20, 0.58]	1.59	.14

Note: Psychological distress sub-scale included items 1–6. Somatic distress sub-scale included items 7–12. Cronbach's alpha coefficients for each item represent the effect of removing that item from the computation of the alpha coefficients (e.g., if item 1 is removed, the resulting value for the scale is 0.87, if item 2 is omitted, it is 0.87, and so on). 95% confidence intervals shown in square brackets.

TABLE 3

Maximum Likelihood Estimate Beta (Standard Error) of Age and Sex, Twin Correlations (Standard Errors), and Heritability (95% Confidence Interval) for Affective Disorders, Substance Use, PSYCH- 6, and SOMA- 6 Distress Sub-scales of the SPHERE-12

Scale	Phenotypes	N	Beta of covariates		Monozygotic twins		Dizygotic twins		Heritability
			Age	Sex (F vs. M)	Pairs (Ind)	rMZ	Pairs (Ind)	rDZ	
PSYCH-6	IRT score	1,548	**-0.04 (0.01)**	**-0.27 (0.11)**	266 (666)	0.27 (0.06)	322 (882)	0.06 (0.05)	**0.23 (0.13–0.33)**
SOMA-6	IRT score	1,548	**-0.05 (0.01)**	-0.13 (0.10)	266 (666)	0.29 (0.05)	322 (882)	0.04 (0.06)	**0.25 (0.15–0.35)**
AD	AD–IRT	2,132	0.01 (0.01)	**-0.54 (0.10)**	374 (885)	0.34 (0.04)	491 (1,247)	0.10 (0.05)	**0.32 (0.23–0.39)**
	Major depressive disorder	2,132	-0.01 (0.04)	**0.28 (1.54)**	374 (885)	0.43 (0.09)	491 (1,247)	0.17 (0.10)	**0.41 (0.24–0.57)**
	Agoraphobia	2,132	0.01	**0.64**	374 (885)	NC	491 (1,247)	NC	NC
	Depressive episodes	2,132	-0.01	**0.31**	374 (885)	0.43 (0.09)	491 (1,247)	0.16 (0.10)	**0.41 (0.24–0.56)**
	Panic attack	2,132	-0.01 (0.13)	**0.40 (7.01)**	374 (885)	0.43 (0.10)	491 (1,247)	0.04 (0.12)	**0.37 (0.17–0.55)**
	Panic disorder	2,132	-0.01	**0.47**	374 (885)	0.39 (0.27)	491 (1,247)	NC	0.23 (0.00–0.68)
	Social anxiety	2,132	0.01 (0.07)	**0.26 (1.64)**	374 (885)	0.44 (0.10)	491 (1,247)	0.15 (0.10)	**0.42 (0.23–0.58)**
SU	SU-IRT	2,132	**0.05 (0.01)**	**0.70 (0.09)**	374 (885)	0.49 (0.04)	491 (1,247)	0.26 (0.04)	**0.49 (0.42–0.56)**
	Alcohol use	2,132	**0.02 (0.37)**	-0.15 (2.62)	374 (885)	0.71 (0.06)	491 (1,247)	0.56 (0.08)	**0.75 (0.64–0.84)**
	Drug use ever	2,132	**-0.03 (0.31)**	-0.26 (2.87)	374 (885)	0.78 (0.04)	491 (1,247)	0.44 (0.06)	**0.79 (0.71–0.86)**
	Cannabis use ever	1,663	-0.03	-0.38	245 (648)	0.78 (0.05)	372 (1,015)	0.38 (0.08)	**0.78 (0.66–0.86)**
	Tobacco use ever	2,132	**-0.03 (0.26)**	-0.32 (2.63)	374 (885)	0.65 (0.06)	491 (1,247)	0.50 (0.06)	**0.70 (0.60–0.78)**
	Alcohol use disorder	2,132	**-0.03 (0.07)**	-0.43 (0.91)	374 (885)	0.43	491 (1,247)	0.23	**0.50 (0.40–0.59)**
	Cannabis use disorder	2,132	**-0.04**	**-0.42**	374 (885)	0.50	491 (1,247)	0.21	**0.74 (0.61–0.83)**

Note: Total number of subjects (*N*), number of complete MZ, or DZ twin pairs (individuals) are given. Significant betas are in bold-face. rMZ and rDZ are shown as NC when these correlations are too small. The heritability is bolded when the AE models are significantly different from the E models. The heritability is shown as NC when both rMZ and rDZ are NC. NC = not calculable.

it easier for researchers who are not familiar with twin studies and SEM, we provided conceptual path diagrams for the quadrivariate CD (Figure S6), IP (Figure S7), and CP (Figure S8) model. SEM enables us to quantify the effect of a single latent variable on an observed variable. High magnitudes of path coefficients indicate that latent variables have a large effect on the observed variables. The relative contribution of different latent variables to the variation of an observed variable can be quantified by squaring the standardized path coefficients. The CD model assumes one A, one C, and one E factor influences each of the four observed variables. As this model does not carry an assumption about the underlying genetic and environmental architecture, it was used as the base model against its nested sub-models (e.g., Cholesky AE, Cholesky CE, and Cholesky E) and the more parsimonious IP and CP models were compared.

The IP model hypothesizes that genes and environment exert differential influence on the covariance between phenotypes (Gillespie & Martin, 2005). This model estimates one set of shared genetic factor (Ac), common environmental factors (Cc), and unshared environmental factors (Ec), which influence the covariation between the four phenotypes via separate paths to each of them. In addition, this model also estimates genetic factor (As), common environmental factors (Cs), and unshared environmental factors (Es) that are specific to each phenotype, to explain the remaining phenotype-specific variance.

Nested within the IP model, the CP model is more parsimonious and restrictive than the IP model. This model hypothesizes that the covariation between the four phenotypes is influenced by common A, C, and E factors through a latent factor (Gillespie & Martin, 2005). Similar to the IP model, the CP model estimates phenotype-specific variance with A, C, and E factors that are specific to each phenotype. More complicated models, such as the IP with two or three common genetic factors, were also explored. We interpreted the parameter estimates, such as phenotypic correlations (r_P), genetic correlations (r_G), environmental correlations (r_E), and factor loading, from the best-fitting multivariate model.

Statistical Software

We used R (R Core Team, 2016) and SAS 9.4 (SAS Institute Inc., 2013) for statistical analyses, graph, and table generation. The following R packages were used: package psy (Falissard, 2012) for internal consistency (Cronbach's alpha coefficient), package irr (Gamer et al., 2012) for test–retest reliability, package KernSmoothIRT (Mazza et al., 2014) for the np-GRM models, and package OpenMx (Boker et al., 2017) for twin modeling. Path diagrams were generated using Ωnyx (Oertzen et al., 2015).

Results

Internal Consistency and Test–Retest Reliability of SPHERE-12

Good internal consistency (Cronbach's alpha: 0.87–0.88, Table 2) was seen in the SPHERE-12 scale, with the scale hardly varying as each item was dropped in turn,

suggesting a high level of homogeneity among these items. A normal distribution in the intervals, assessed with Shapiro–Wilk test, was found between NU2 and NU3 ($W = 0.94$, $Pr < W$: 0.27) and between NU1 and NU3 ($W = 0.94$, $Pr < W$: 0.16), but not between NU1 and NU2 ($W = 0.97$, $Pr < W$: 0.004). Moderate intra-class correlations (ICC; Table 2) were found between NU1 and NU2 (ICC: 0.59–0.65, interval: 1.6 ± 0.3 years, range: 0.2–2.5 years) and between NU2 and NU3 (ICC: 0.5–0.62, interval: 4.3 ± 0.8 years, range: 2.5–6.3 years). The ICC values were non-significant between NU1 and NU3 in each of the 12 items and the total scores.

Basic Assumption Tests

For binary phenotypes, no significant birth-order effect on the thresholds was found (Table S2: H1t). A zygosity effect (Table S2: H2t) was found only in alcohol use. Significant differences between correlations of MZ twins (rMZ) and DZ twins (rDZ) were found in drug use ever, and cannabis use ever (Table S2: H3c). We found a significant familiar aggregation effect (Table S2: H4c) in all the AD and SU phenotypes, except for agoraphobia and PD.

For continuous phenotypes, we did not find a significant birth-order effect (Table S3: H1m) or zygosity effect (Table S3: H2m) on the means. Significant differences between rMZ and rDZ (Table S3: H3c) were found in all the phenotypes, suggesting the presence of significant genetic effects in these phenotypes. A significant familiar aggregation effect on the covariance (Table S3: H4c) was found in all the phenotypes.

Univariate Genetic Analyses

As shown in Table S4, the results of the univariate genetic analyses revealed that the additive genetic (A) and unshared environmental (E) effect model best explained the variation in all the phenotypes (bolded AE models) except for AU and tobacco use ever. Table 3 shows the sample sizes, estimates of betas, twin correlations, and heritability from the best-fitting models for each of these phenotypes. We found a significant effect of age in all SU phenotypes, PSYCH-6-IRT, and SOMA-6-IRT, but not in any AD phenotypes. Sex had a significant effect on the variance of all the phenotypes except for SOMA-6-IRT. Significantly higher rMZ than rDZ was found in all the phenotypes except for agoraphobia and PD. Most AD phenotypes were moderately heritable, with heritability ranging from 37% for PA to 42% for social anxiety. The SU phenotypes were found to be more heritable than the AD phenotypes, with heritability ranging from 49% for AUD to 79% for ever using drugs. We were unable to estimate heritability for agoraphobia and PD as the low prevalence of these disorders resulted in a lack of discordant/concordant twin pairs, which resulted in insufficient information in the analyses.

Multivariate Genetic Analyses

Table S5 shows the model-fitting results for various multivariate models. Dropping the C component from the full Cholesky ACE model did not lead to a significant loss of model fit (Model 2). Dropping the A component (Model 3), or A and C component (Model 4), however, led to significant losses of fit. We then fit alternative IP models with one common genetic factor (Model 5), IP with two common genetic factors (IP2A; Model 6), IP with three common genetic factors (Model 7), as well as the CP model (Model 8). As none of these models led to significant loss of fit, we then explored the most parsimonious models from the IP model and IP2A model. To test the importance of the A, C, and E components, we fit a series of nested models by dropping these components from the common and the specific paths of the IP model (Models 10–17) and the IP2A model (Models 21–28). The C component could be dropped from both the common and specific paths of the IP model (Model 17), as well as the IP2A model (Model 28), without significantly loss of fit. Moreover, the IP_AE_AE model (Model 33) and the IP2A_AE_AE model (Model 34) had a lower AIC than the Cholesky AE model (Model 12). We present the parameter estimates from IP_AE_AE (Figure 2) as it had the lowest AIC.

Table 4 shows r_P, r_G, and r_E between any two of the four phenotypes. Significant r_P's were found between all the four phenotypes, with an average of 0.22 (range: 0.06–0.51). Significant and higher r_G, ranging from 0.15 to 0.85, was also found between the four phenotypes. The magnitude of r_G between the PSYCH-6r and ADr was very close to that between the SOMA-6r and ADr. The same pattern was also seen when comparing the r_G between PSYCH-6r and SUr with the r_G between the SOMA-6r and SUr. Significant r_E was only found between the PSYCH-6r and the SOMA-6r, the PSYCH-6r and ADr, and the PSYCH-6r and SUr.

The common unshared environmental factor (Ec in Figure 2) had a significant influence on the PSYCH-6r, SOMA-6r, and ADr, explaining 5–73% of their variation. The specific environmental factors (Es) had a significant effect on SOMA-6r, ADr, and SUr, explaining 53–64% of their variations. In phenotypes with significant Ec and Es, the Es had a much larger contribution than the Ec. A different pattern was seen in the additive genetic influence. The common additive genetic factors (Ac) had significant loading onto each set of the four phenotypes, explaining 3–22% of their variations. The specific additive genetic factors (As), however, were allowed to load only onto ADr and SUr. The relative contribution of the As was two to four times larger than the Ac.

Discussion

This is the first study to investigate the overlapping genetic and environmental influences underlying the four
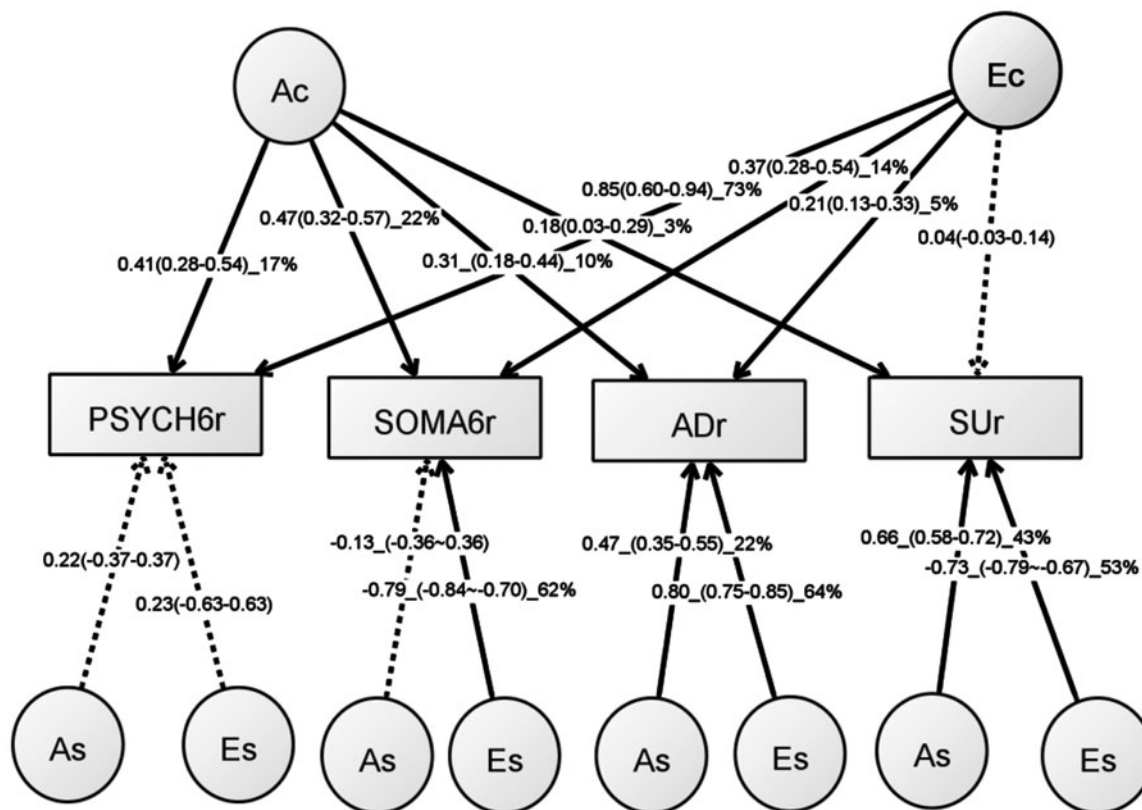
FIGURE 2

Path diagram of an independent pathway (IP) model with standardized parameter estimates. The IP model has an additive genetic (Ac) factor and a unique environmental factor (Ec) that are common to all four phenotypes, as well as genetic factor (As) and unshared environmental factor (Es) that are specific to each phenotype. All shared environmental influences have been dropped without worsening the model fit. Solid lines indicate significant paths and dotted lines indicate non-significant paths (95% confidence intervals include zero). Standardized path coefficients, their 95% confidence intervals (CIs), and percentage of variation of a phenotype explained by a factor are presented for significant paths. Non-significant paths are shown with path coefficients and 95% CIs.

comorbid disorders — PSYCH, SOMA, AD, and SU — in young Australian adults. We began our analytical flow with the validation of the SPHERE-12 instrument and then dissected the genetic and environmental influences using univariate and multivariate twin modeling. We found evidence for a small but significant genetic influence that was common to all the four phenotypes (Ac). The unshared environmental factors exerted greater influence on the four phenotypes; however, these influences were largely specific to SOMA, AD, and SU. This suggests that the genetic and environmental influences on the four phenotypes may be best explained by one common genetic factor (Ac), one unshared environmental factor that is common to the PSYCH-6, the SOMA-6, and AD (Ec), and genetic (As) and unshared environmental factors (Es) that were specific to some phenotypes.

Test–Retest Reliability of SPHERE-12

The SPHERE-12 has proved to be a highly consistent and moderately reliable instrument in our young Australian sample, which is consistent with longer SPHERE question-

naires such as the SPHERE-21 (Couvy-Duchesne et al., 2017) and the SPHERE-34 (Hansell et al., 2012). Internal consistency (Cronbach's alpha) was consistent between the PSYCH sub-scales (our PSYCH-6: 0.86; anxiety depression: 0.86–0.88; PSYCH-14: 0.87) and between SOMA sub-scales (our SOMA-6: 0.81; chronic fatigue: 0.78–0.79; SOMA-10: 0.7), suggesting that the item reduction did not lower the consistency of the SPHERE questionnaire.

Test–retest reliability, commonly assessed by ICC, reflects the variation in measurements taken during a test and a retest on the same subjects and under the same conditions (Koo & Li, 2016). The value of ICC can be influenced by factors such as variability among participants, sample sizes, and the time between the test and the retest. The ICC of our PSYCH-6, 0.47–0.48, was consistent with the anxiety-depression sub-scale, 0.47, reported by Couvy-Duchesne et al. (2017). However, the ICC in our SOMA-6 (ICC: 0.36–0.5) was lower than their chronic fatigue sub-scale (ICC: 0.57). The fact that our NU1–NU3 ICCs were non-significant and lower than NU1–NU2 ICCs and NU2–NU3 ICCs may be attributed to the long NU1–NU3 interval

TABLE 4

Estimates of Phenotypic, Genetic, and Environmental Correlations Between Psychological (PSYCH-6r) and somatic (SOMA-6r) Distress Sub-scales of the SPHERE-12, the Affective Disorders (ADr), and Substance Use (SUr) Scales

Phenotype	Phenotypic correlation				Genetic correlation				Environmental correlation			
	PSYCH-6r	SOMA-6r	ADr	SUr	PSYCH-6r	SOMA-6r	ADr	SUr	PSYCH-6r	SOMA-6r	ADr	SUr
PSYCH-6r	1				1				1			
SOMA-6r	**0.51 [0.47, 0.55]**	1			**0.85 [0.61, 1.0]**	1			**0.41 [0.33, 0.49]**	1		
ADr	**0.31 [0.26, 0.36]**	**0.23 [0.18, 0.27]**	1		**0.49 [0.27, 0.73]**	**0.53 [0.31, 0.74]**	1		**0.25 [0.16, 0.33]**	**0.11 [0.06, 0.21]**	1	
SUr	**0.10 [0.05, 0.15]**	**0.10 [0.05, 0.15]**	**0.06 [0.03, 0.11]**	1	**0.23 [0.04, 0.41]**	**0.25 [0.04, 0.42]**	**0.15 [0.02, 0.27]**	1	0.05 [-0.0, 0.15]	0.02 [-0.0, 0.11]	0.01 [-0.0, 0.06]	1

Note: 95% confidence intervals shown in square brackets. Significant correlations that did not include zero in their confidence intervals are bolded.

(mean interval: 5.4 ± 0.7 years, range: 4.4–6.8 years) or the stability of PSYCH-6 or SOMA-6 measurements over time. Longer test–retest intervals have been shown to increase inconsistency (Dareng et al., 2017).

Univariate Findings

Our heritability estimates were generally consistent with previous studies. Heritability estimates can vary between studies due to their variations in study designs, such as age groups, sample sizes, sexes, data collection methods (e.g., via self-reporting questionnaires or diagnostic interviews), and source of participants (e.g., community or clinical samples). Bearing in mind these differences, we checked to see whether our heritability estimates overlapped with the confidence intervals of heritability estimates reported from previous studies.

We found low but significant heritable influences on the PSYCH and the SOMA. Our heritability estimates for the PSYCH-6–IRT (0.23, 95% CI [0.13, 0.33]) and the SOMA-6–IRT (0.25, 95% CI [0.15, 0.35]) were consistent with heritability estimates for similar sub-scales in the oldest age groups (Couvy-Duchesne et al., 2017). They reported heritability 0.37 (95% CI [0.21, 0.51]) for anxiety-depression IRT scores at age 17–19 and 0.27 (95% CI [0.11, 0.41]) for chronic fatigue at the same ages.

We found moderate to high heritability estimates on the liability to AD and SU phenotypes (Table 3). For AD diagnostic phenotypes, our heritability estimates were in accordance with previous studies. Sullivan et al. (2000) attributed on average 37% (range: 31–42%) of the variation in major depression to additive genetic factors by reviewing four community studies (Bierut et al., 1999; Kendler et al., 1995; Kendler & Prescott, 1999; Lyons et al., 1998) and two clinical studies (Kendler et al., 1995; McGuffin et al., 1996). These reviewed studies employed various data collection methods, such as semi-structured interviews, mailed questionnaires, telephone interviews, or diagnostic assessments on probands with MDD. Rijsdijk et al. (2003) reported an estimate of 0.39–0.42 for severe depression, measured with GHQ, in female twins aged 18–79 (mean age: 47.7). Our heritability estimates for MDD (0.41, 95% CI [0.24, 0.57]) and depressive episodes (0.41, 95% CI [0.24, 0.56]) fell within the range reported from these studies. Unfortunately, heritability for agoraphobia and PD was not estimable, given insufficient cases in MZ and DZ twins (Table S1). We therefore compared our heritability of PA with that of PD, as PA is one of the defining features of PD. Previous studies reported a considerable variation in the heritability of PD, ranging from 0.28 in female twins (Hettema et al., 2005) to 0.48 from a meta-analysis (Hettema et al., 2001) that combined data from five family studies and five twin studies. Our heritability of PA (0.37; 95% CI [0.17, 0.55]) was within this range. To our knowledge, the heritability of *DSM-5* social anxiety has not been reported. However, studies that assessed a relevant symptom, social

anxiety disorder (SAD), with diagnostic interviews on twin samples have reported a considerable variation in their heritability estimates, ranging from 0.1 to 0.55 (Czajkowski et al., 2011; Hettema et al., 2006; Kendler et al., 2001; Nelson et al., 2000; Reichborn-Kjennerud et al., 2007). Our heritability estimate for social anxiety 0.42 (95% CI [0.23, 0.58]) was close to the upper range.

For SU phenotypes, we found that 50–79% of the variation was attributable to the additive genetic influence, which is comparable with heritability estimates for SU. To our knowledge, this is the first twin study that reports heritability estimate (0.78, 95% CI [0.66, 0.86]) for CUD using *DSM-5* diagnosis. A meta-analysis (Verweij et al., 2010) reviewed 28 studies for heritability estimates for cannabis use initiation and problematic cannabis use; most of these phenotypes were defined by *DSM-IV-TR* (American Psychiatric Association, 2000). They reported heritability estimates of 48% in male twins and of 40% in female twins for cannabis use initiation. Higher heritability was found in problematic cannabis use: 51% in men and 59% in women. Another study reported heritability of 72% for the *DSM-IV* (American Psychiatric Association, 1994) definition of cannabis abuse or dependence in young Australian adults (Lynskey et al., 2012). A recent study found that 21% of the phenotypic variation in the *DSM-5* CUD was attributable to the genomic variation in common SNPs, but this estimate was found to be non-significant (Agrawal et al., 2014). It is worth noting that the diagnostic criteria for SU disorder differ between the *DSM-IV* and the *DSM-5*. The distinction between substance abuse and substance dependence in the *DSM-IV* has been replaced with a single SU disorder in the *DSM-5* (Hasin et al., 2013). Our heritability for AUD, 0.50 (95% CI [0.4, 0.59]), is similar to the best-fit estimate of 0.49 (95% CI [0.43, 0.53]) reported by Verhulst et al. (2015). Another study reported a wide range of heritability (49–64%) for alcoholism (McGue, 1999). For AU and tobacco use, dropping the A component did not lead to losses of fit, whereas dropping the C component led to losses of fit. With the lowest AIC, the resulting CE models best explained the variance of the two phenotypes.

Multivariate Findings

We found a strong genetic correlation ($r_G = 0.85$) between the PSYCH-6r and the SOMA-6r, which is consistent with longer SPHERE sub-scales. Couvy-Duchesne et al. (2017) found high genetic correlations of 0.85–1.00 between anxiety-depression and chronic fatigue IRT scores in four younger age groups (<13, 13–15, 15–17, and 17–19 years). Hansell et al. (2012) reported a high r_G of 0.87 between the PSYCH-14 and the SOMA-10 at mean age 15.5 years (age range: 12.0–25.6). Our results, together with these findings, suggest that the genetic overlap between PSYCH and SOMA remained consistently high in adolescents and young adults. It is interesting to note that the two SPHERE sub-scales had a similar genetic overlap with ADr

($r_G = 0.49, 0.53$) and SUr ($r_G = 0.23, 0.25$). Although the main objectives of this study were not to compare summed scores with IRT scores, we found a high r_P between them in both the PSYCH-6 (r_P: 0.81, 95% CI [0.79, 0.82]) and the SOMA-6 (r_P: 0.82, [95% CI: 0.8, 0.84]).

An important finding from our independent pathway model was that the genetic influence on the four phenotypes was through the actions of three sets of genetic factors. The first set, the common genetic factor (Ac), explained 3–22% of the variance of the four phenotypes. The second and third sets were genetic factors (As) that were specific to ADr and SUr. For these two phenotypes, these As explained a larger proportion of their variance than the Ac.

What are the implications of the common genetic influence underlying comorbid psychiatric disorders? Given the genetic factors derived from multivariate models are statistically defined as a hypothetical predictive construct (Carey, 1988), the genetic covariation does not directly map onto an inference about shared loci. However, the shared genetic construct suggests the potential utility of treatment approaches that target comorbid conditions in the absence of manifest comorbidity.

Limitations

The use of a genetically informative sample and comprehensive sets of phenotypes were the strength of this study; however, our results should be interpreted considering the following limitations. First, the low prevalence rates of agoraphobia and PD did not allow the estimation of their heritability. The estimation of heritability requires calculable rMZ and rDZ (i.e., neither zero nor negative). There were limited pairs of twins who were concordant for these disorders, meaning that we were unable to estimate the heritability for these phenotypes (Table S1). However, we included these items in our IRT modeling as they showed acceptable IRSF (Figure S4). Low prevalence is a common problem in these binary traits (Goodwin et al., 2005; Mosing et al., 2009). Second, our use of a cross-sectional design could limit the generalizability of our results beyond the age group 18–38 years. Third, we combined the PSYCH-6 and SOMA-6 measurement from all the study waves and sexes in order to obtain their genetic estimates in a sample as large as possible. This may have diluted the difference in variance and covariance components between the waves and between the sexes. Fourth, our study was well powered to reject the C component in favor of an AE source model, but was insufficiently powered to distinguish the IP and CP structural models within the AE source model. Finally, although we found that unique environmental factors and additive genetic factors explained on average 67.5% and 29.5% of the covariance between the four phenotypes, we do not know how these influences may change over age. For personality traits, genetic contribution tends to be stable over time, whereas new environmental effects emerge over time (Read et al., 2006). Similar results have also been

found on the genetic and environmental influences on subjective well-being (Nes et al., 2006). To study the stability of genetic influence on common psychiatric disorders, future work needs to use a longitudinal twin design.

Conclusions

The present study shows that both nature and nurture play an important role in the liability to PSYCH, SOMA, AD, and SU in young adults. We found consistently high internal consistency between the 12-item SPHERE and the longer SPHERE instruments. We examined the item properties for each of these four illnesses using non-parametric item response theory models. We found a high genetic overlap between PSYCH and SOMA. These shared genetic effects suggest the potential utility of treatment approaches that target comorbid conditions in the absence of manifest comorbidity.

Disclosure of Interests

None of the authors had a conflict of interest to disclose.

Supplementary material

To view supplementary material for this article, please visit https://doi.org/10.1017/thg.2018.33

References

Agrawal, A., Lynskey, M. T., Bucholz, K. K., Kapoor, M., Almasy, L., Dick, D. M., … Bierut, L. J. (2014). DSM-5 cannabis use disorder: A phenotypic and genomic perspective. *Drug and Alcohol Dependence*, *134*, 362–369.

Agrawal, A., Silberg, J. L., Lynskey, M. T., Maes, H. H., & Eaves, L. J. (2010). Mechanisms underlying the lifetime co-occurrence of tobacco and cannabis use in adolescent and young adult twins. *Drug and Alcohol Dependence*, *108*, 49–55.

Akena, D., Joska, J., Obuku, E. A., Amos, T., Musisi, S., & Stein, D. J. (2012). Comparing the accuracy of brief versus long depression screening instruments which have been validated in low and middle income countries: A systematic review. *BMC Psychiatry*, *12*, 187.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text revision). Washington, DC: Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: American Psychiatric Publishing.

Andrews, G., & Slade, T. (2001). Interpreting scores on the Kessler Psychological Distress Scale (K10). *Australian and New Zealand Journal of Public Health*, *25*, 494–497.

Ark, L. A. van der. (2005). Stochastic ordering of the latent trait by the sum score under various polytomous IRT models. *Psychometrika*, *70*, 283–304.

Ball, H. A., Siribaddana, S. H., Sumathipala, A., Kovas, Y., Glozier, N., Rijsdijk, F., … Hotopf, M. (2011). Genetic and environmental contributions to the overlap between psychological, fatigue and somatic symptoms: A twin study in Sri Lanka. *Twin Research and Human Genetics*, *14*, 53–63.

Barsky, A. J., Wyshak, G., & Klerman, G. L. (1992). Psychiatric comorbidity in DSM-III-R hypochondriasis. *Archives of General Psychiatry*, *49*, 101–108.

Beck, A. T., & Steer, R. A. (1993). *Beck depression inventory manual*. San Antonio, TX: Psychological Corporation.

Bierut, L. J., Heath, A. C., Bucholz, K. K., Dinwiddie, S. H., Madden, P. A. F., Statham, D. J., … Martin, N. G. (1999). Major depressive disorder in a community-based twin sample: Are there different genetic and environmental contributions for men and women? *Archives of General Psychiatry*, *56*, 557–563.

Boker, S. M., Neale, M. C., Maes, H. H., Wilde, M. J., Spiegel, M., Brick, T. R., … Wilhelm, S. (2017). *OpenMx 2.7.9 user guide*. Retrieved from https://openmx.ssri.psu.edu/documentation

Carey, G. (1988). Inference about genetic correlations. *Behavior Genetics*, *18*, 329–338.

Cheng, S.-T., & Chan, A. C. M. (2005). The center for epidemiologic studies depression scale in older Chinese: Thresholds for long and short forms. *International Journal of Geriatric Psychiatry*, *20*, 465–470.

Clarke, D. M., & McKenzie, D. P. (2003). An examination of the efficiency of the 12-item SPHERE questionnaire as a screening instrument for common mental disorders in primary care. *Australian & New Zealand Journal of Psychiatry*, *37*, 236–239.

Clarke, D. M., Piterman, L., Byrne, C. J., & Austin, D. W. (2008). Somatic symptoms, hypochondriasis and psychological distress: A study of somatisation in Australian general practice. *The Medical Journal of Australia*, *189*, 560–564.

Couvy-Duchesne, B., Davenport, T. A., Martin, N. G., Wright, M. J., & Hickie, I. B. (2017). Validation and psychometric properties of the Somatic and Psychological HEalth REport (SPHERE) in a young Australian-based population sample using non-parametric item response theory. *BMC Psychiatry*, *17*, 279.

Couvy-Duchesne, B., O'Callaghan, V., Parker, R., Mills, N., Kirk, K. M., Scott, J., … Gillespie, N. A. (2018). Nineteen and Up study (19Up): Understanding pathways to mental health disorders in young Australian twins. *BMJ Open*, *8*, e018959.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334.

Czajkowski, N., Kendler, K. S., Tambs, K., Røysamb, E., & Reichborn-Kjennerud, T. (2011). The structure of genetic and environmental risk factors for phobias in women. *Psychological Medicine*, *41*, 1987–1995.

Dareng, E. O., Adebamowo, S. N., Eseyin, O. R., Odutola, M. K., Pharoah, P. P., & Adebamowo, C. A. (2017). Test–retest reliability of self-reported sexual behavior history in urbanized Nigerian women. *Frontiers in Public Health*, *5*, 172.

de Waal, M. W. M., Arnold, I. A., Eekhof, J. A. H., & van Hemert, A. M. (2004). Somatoform disorders in general practice: Prevalence, functional impairment and comorbidity with anxiety and depressive disorders. *The British Journal of Psychiatry*, *184*, 470–476.

Escobar, J. I., Gara, M., Silver, R. C., Waitzkin, H., Holman, A., & Compton, W. (1998). Somatisation disorder in primary care. *The British Journal of Psychiatry*, *173*, 262–266.

Evans, D. M., Frazer, I. H., & Martin, N. G. (1999). Genetic and environmental causes of variation in basal levels of blood cells. *Twin Research*, *2*, 250–257.

Eytan, A., Haller, D. M., Wolff, H., Cerutti, B., Sebo, P., Bertrand, D., & Niveau, G. (2011). Psychiatric symptoms, psychological distress and somatic comorbidity among remand prisoners in Switzerland. *International Journal of Law and Psychiatry*, *34*, 13–19.

Falissard, B. (2012). *Psy: Various procedures used in psychometry*. R package version 1.1. Retrieved from https://cran.r-project.org/web/packages/psy/index.html

Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2012). *Irr: Various coefficients of interrater reliability and agreement*. R package version 0.84. Retrieved from http://cran.us.r-project.org/web/packages/irr/index.html

Gelernter, J. (2015). Genetics of complex traits in psychiatry. *Biological Psychiatry*, *77*, 36–42.

Gillespie, N. A., Henders, A. K., Davenport, T. A., Hermens, D. F., Wright, M. J., Martin, N. G., & Hickie, I. B. (2013). The Brisbane longitudinal twin study: Pathways to cannabis use, abuse, and dependence project-Current status, preliminary results, and future directions. *Twin Research and Human Genetics*, *16*, 21–33.

Gillespie, N. A., & Martin, N. G. (2005). Multivariate genetic analysis. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (pp. 1363–1370). Hoboken, NJ: John Wiley & Sons.

Gillespie, N. A., Zhu, G., Heath, A. C., Hickie, I. B., & Martin, N. G. (2000). The genetic aetiology of somatic distress. *Psychological Medicine*, *30*, 1051–1061.

Goodwin, R. D., Faravelli, C., Rosi, S., Cosci, F., Truglia, E., de Graaf, R., & Wittchen, H. U. (2005). The epidemiology of panic disorder and agoraphobia in Europe. *European Neuropsychopharmacology*, *15*, 435–443.

Han, C., McGue, M. K., & Iacono, W. G. (1999). Lifetime tobacco, alcohol and other substance use in adolescent Minnesota twins: Univariate and multivariate behavioral genetic analyses. *Addiction*, *94*, 981–993.

Hanel, G., Henningsen, P., Herzog, W., Sauer, N., Schaefert, R., Szecsenyi, J., & Löwe, B. (2009). Depression, anxiety, and somatoform disorders: Vague or distinct categories in primary care? Results from a large cross-sectional study. *Journal of Psychosomatic Research*, *67*, 189–197.

Hansell, N. K., Wright, M. J., Medland, S. E., Davenport, T. A., Wray, N. R., Martin, N. G., & Hickie, I. B. (2012). Genetic co-morbidity between neuroticism, anxiety/depression and somatic distress in a population sample of adolescent and young adult twins. *Psychological Medicine*, *42*, 1249–1260.

Harvey, S. B., Deady, M., Wang, M.-J., Mykletun, A., Butterworth, P., Christensen, H., & Mitchell, P. B. (2017). Is the prevalence of mental illness increasing in Australia? Evidence from national health surveys and administrative data, 2001–2014. *The Medical Journal of Australia*, *206*, 490–493.

Hasin, D. S., O'Brien, C. P., Auriacombe, M., Borges, G., Bucholz, K., Budney, A., … Grant, B. F. (2013). DSM-5 criteria for substance use disorders: Eecommendations and eationale. *The American Journal of Psychiatry*, *170*, 834–851.

Hemker, B. T., Sijtsma, K., Molenaar, I. W., & Junker, B. W. (1997). Stochastic ordering using the latent trait and the sum score in polytomous IRT models. *Psychometrika*, *62*, 331–347.

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

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

Hettema, J. M., Prescott, C. A., Myers, J. M., Neale, M. C., & Kendler, K. S. (2005). The structure of genetic and environmental risk factors for anxiety disorders in men and women. *Archives of General Psychiatry*, *62*, 182–189.

Hickie, I. B., Davenport, T. A., Hadzi-Pavlovic, D., Koschera, A., Naismith, S. L., Scott, E. M., & Wilhelm, K. A. (2001). Development of a simple screening tool for common mental disorders in general practice. *The Medical Journal of Australia*, *175*, (Suppl.), S10–S17.

Hickie, I., Kirk, K., & Martin, N. (1999). Unique genetic and environmental determinants of prolonged fatigue: A twin study. *Psychological Medicine*, *29*, 259–268.

Hiller, W., Rief, W., & Brähler, E. (2006). Somatization in the population: From mild bodily misperceptions to disabling symptoms. *Social Psychiatry and Psychiatric Epidemiology*, *41*, 704–712.

Ivkovic, V., Vitart, V., Rudan, I., Janicijevic, B., Smolej-Narancic, N., Skaric-Juric, T., … Deary, I. J. (2007). The

Eysenck personality factors: Psychometric structure, reliability, heritability and phenotypic and genetic correlations with psychological distress in an isolated Croatian population. *Personality and Individual Differences*, *42*, 123–133.

Karkowski, L. M., Prescott, C. A., & Kendler, K. S. (2000). Multivariate assessment of factors influencing illicit substance use in twins from female-female pairs. *American Journal of Medical Genetics*, *96*, 665–670.

Kendler, K. S., Myers, J., & Prescott, C. A. (2007). Specificity of genetic and environmental risk factors for symptoms of cannabis, cocaine, alcohol, caffeine, and nicotine dependence. *Archives of General Psychiatry*, *64*, 1313–1320.

Kendler, K. S., Myers, J., Prescott, C. A., & Neale, M. C. (2001). The genetic epidemiology of irrational fears and phobias in men. *Archives of General Psychiatry*, *58*, 257–265.

Kendler, K. S., Pedersen, N. L., Neale, M. C., & Mathé, A. A. (1995). A pilot Swedish twin study of affective illness including hospital- and population-ascertained subsamples: Results of model fitting. *Behavior Genetics*, *25*, 217–232.

Kendler, K. S., & Prescott, C. A. (1999). A population-based twin study of lifetime major depression in men and women. *Archives of General Psychiatry*, *56*, 39–44.

Kendler, K. S., Prescott, C. A., Myers, J., & Neale, M. C. (2003). The structure of genetic and environmental risk factors for common psychiatric and substance use disorders in men and women. *Archives of General Psychiatry*, *60*, 929–937.

Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S. L. T., … Zaslavsky, A. M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine*, *32*, 959–976.

Khan, A., Lewis, C., & Lindenmayer, J.-P. (2011). Use of non-parametric item response theory to develop a shortened version of the positive and negative syndrome scale (PANSS). *BMC Psychiatry*, *11*, 178.

Kocalevent, R.-D., Hinz, A., & Brähler, E. (2013). Standardization of a screening instrument (PHQ-15) for somatization syndromes in the general population. *BMC Psychiatry*, *13*, 91.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*, 155–163.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*, 606–613.

Lee, S., Leung, C. M., Kwok, K. P., & Ng, K. L. (2015). A community-based study of the relationship between somatic and psychological distress in Hong Kong. *Transcultural Psychiatry*, *52*, 594–615.

Lynskey, M. T., Agrawal, A., Henders, A., Nelson, E. C., Madden, P. A. F., & Martin, N. G. (2012). An Australian twin study of cannabis and other illicit drug use and misuse, and other psychopathology. *Twin Research and Human Genetics*, *15*, 631–641.

Lynskey, M. T., Glowinski, A. L., Todorov, A. A., Bucholz, K. K., Madden, P. A. F., Nelson, E. C., … Heath, A. C. (2004). Major depressive disorder, suicidal ideation, and suicide attempt in twins discordant for cannabis dependence and early-onset cannabis use. *Archives of General Psychiatry*, *61*, 1026–1032.

Lyons, M. J., Eisen, S. A., Goldberg, J., True, W., Lin, N., Meyer, J. M., … Tsuang, M. T. (1998). A registry-based twin study of depression in men. *Archives of General Psychiatry*, *55*, 468–472.

Maier, W., & Falkai, P. (1999). The epidemiology of comorbidity between depression, anxiety disorders and somatic diseases. *International Clinical Psychopharmacology*, *14*, S1–S6.

Maydeu-Olivares, A. (2005). Further empirical results on parametric versus non-parametric IRT modeling of Likert-type personality data. *Multivariate Behavioral Research*, *40*, 261–279.

Mazza, A., Punzo, A., & McGuire, B. (2014). KernSmoothIRT: An R package for kernel smoothing in item response theory. *Journal of Statistical Software*, *58*, 1–34.

McGue, M. (1999). The behavioral genetics of alcoholism. *Current Directions in Psychological Science*, *8*, 109–115.

McGuffin, P., Katz, R., Watkins, S., & Rutherford, J. (1996). A hospital-based twin register of the heritability of DSM-IV unipolar depression. *Archives of General Psychiatry*, *53*, 129–136.

Meijer, R. R., & Baneke, J. J. (2004). Analyzing psychopathology items: A case for nonparametric item response theory modeling. *Psychological Methods*, *9*, 354–368.

Mellenbergh, G. J. (1995). Conceptual notes on models for discrete polytomous item responses. *Applied Psychological Measurement*, *19*, 91–100.

Mosing, M. A., Gordon, S. D., Medland, S. E., Statham, D. J., Nelson, E. C., Heath, A. C., … Wray, N. R. (2009). Genetic and environmental influences on the co-morbidity between depression, panic disorder, agoraphobia, and social phobia: A twin study. *Depression and Anxiety*, *26*, 1004–1011.

Neale, M., & Cardon, L. (1992). *Methodology for genetic studies of twins and families*. Dordrecht, Netherlands: Springer.

Nelson, E. C., Grant, J. D., Bucholz, K. K., Glowinski, A., Madden, P. A. F., null, Reich, W., & Heath, A. C. (2000). Social phobia in a population-based female adolescent twin sample: Co-morbidity and associated suicide-related symptoms. *Psychological Medicine*, *30*, 797–804.

Nes, R. B., Røysamb, E., Tambs, K., Harris, J. R., & Reichborn-Kjennerud, T. (2006). Subjective well-being: Genetic and environmental contributions to stability and change. *Psychological Medicine*, *36*, 1033–1042.

Oertzen, T. von, Brandmaier, A. M., & Tsang, S. (2015). Structural equation modeling with Ωnyx. *Structural Equation Modeling*, *22*, 148–161.

Ormel, J., VonKorff, M., Ustun, T. B., Pini, S., Korten, A., & Oldehinkel, T. (1994). Common mental disorders and disability across cultures: Results from the WHO collaborative study on psychological problems in general health care. *JAMA*, *272*, 1741–1748.

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ramsay, J. O. (1991). Kernel smoothing approaches to non-parametric item characteristic curve estimation. *Psychometrika*, *56*, 611–630.

Read, S., Vogler, G. P., Pedersen, N. L., & Johansson, B. (2006). Stability and change in genetic and environmental components of personality in old age. *Personality and Individual Differences*, *40*, 1637–1647.

Reichborn-Kjennerud, T., Czajkowski, N., Torgersen, S., Neale, M. C., Ørstavik, R. E., Tambs, K., & Kendler, K. S. (2007). The relationship between avoidant personality disorder and social phobia: A population-based twin study. *American Journal of Psychiatry*, *164*, 1722–1728.

Rijsdijk, F. V., Snieder, H., Ormel, J., Sham, P., Goldberg, D. P., & Spector, T. D. (2003). Genetic and environmental influences on psychological distress in the population: General health questionnaire analyses in UK twins. *Psychological Medicine, 33*, 793–801.

Samejima, F. (1968). Estimation of latent ability using a response pattern of graded scores. *ETS Research Bulletin Series*, *1968*, i–169.

SAS Institute Inc. (2013). *SAS 9.4 programming documentation*. Cary, NC: Author.

Shidhaye, R., Mendenhall, E., Sumathipala, K., Sumathipala, A., & Patel, V. (2013). Association of somatoform disorders with anxiety and depression in women in low- and middle-income countries: A systematic review. *International Review of Psychiatry*, *25*, 65–76.

Sijtsma, K., Emons, W. H. M., Bouwmeester, S., Nyklíček, I., & Roorda, L. D. (2008). Nonparametric IRT analysis of quality-of-life scales and its application to the world health organization quality-of-life scale (WHOQOL-Bref). *Quality of Life Research*, *17*, 275–290.

Slade, T., Grove, R., & Burgess, P. (2011). Kessler psychological distress scale: Normative data from the 2007 Australian national survey of mental health and wellbeing. *Australian & New Zealand Journal of Psychiatry*, *45*, 308–316.

Stochl, J., Jones, P. B., & Croudace, T. J. (2012). Mokken scale analysis of mental health and well-being questionnaire item responses: A non-parametric IRT method in empirical research for applied health researchers. *BMC Medical Research Methodology*, *12*, 74.

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

Thomas, M. L. (2011). The value of item response theory in clinical assessment: A review. *Assessment*, *18*, 291–307.

Verhulst, B., Neale, M. C., & Kendler, K. S. (2015). The heritability of alcohol use disorders: A meta-analysis of twin and adoption studies. *Psychological Medicine*, *45*, 1061–1072.

Verweij, K. J. H., Zietsch, B. P., Lynskey, M. T., Medland, S. E., Neale, M. C., Martin, N. G., … Vink, J. M. (2010). Genetic and environmental influences on cannabis use initiation and problematic use: A meta-analysis of twin studies. *Addiction*, *105*, 417–430.

World Health Organization. (1990). *Composite international diagnostic interview*. Geneva, Switzerland: Author.

Wright, M. J., & Martin, N. G. (2004). Brisbane adolescent twin study: Outline of study methods and research projects. *Australian Journal of Psychology*, *56*, 65–78.