

Predicting Cognitive Executive Functioning with Polygenic Risk Scores for Psychiatric Disorders

Chelsie E. Benca^{1,3} · Jaime L. Derringer² · Robin P. Corley³ · Susan E. Young⁴ · Matthew C. Keller^{1,3} · John K. Hewitt^{1,3} · Naomi P. Friedman^{1,3}

Received: 7 April 2016 / Accepted: 26 August 2016 / Published online: 14 October 2016
© Springer Science+Business Media New York 2016

Abstract Executive functions (EFs) have been proposed as an endophenotype for psychopathology because EF deficits are associated with most psychiatric disorders. To examine this hypothesis, we derived polygenic risk scores for autism, attention-deficit/hyperactive disorder (ADHD), bipolar disorder, major depression (MDD), and schizophrenia, using genome-wide data from the Psychiatric Genomics Consortium as discovery samples. We then examined the relationships between these polygenic risk scores and three separable EF components measured with a latent variable model. We also examined the relationship between genetic risk for ADHD and MDD and their respective symptom counts and lifetime diagnoses. We found no evidence for larger effect sizes for EFs as endophenotypes for psychiatric disorders. However, larger sample sizes will be important in examining this relationship further.

Keywords Cognitive control · Executive control · Polygenic risk scores · Intermediate phenotypes

Executive functions (EFs)—higher-order cognitive processes that regulate thoughts and actions during goal-directed behavior—are implicated in many types of psychopathology. Individuals with attention-deficit/hyperactivity disorder (ADHD), autism (AUT), schizophrenia (SCZ), major depressive disorder (MDD) and bipolar disorder (BP), as well as other psychiatric disorders, show EF deficits, and many of the symptoms for these disorders reflect EF dysfunction (Amann et al. 2012; Rosenthal et al. 2013; Snyder 2013; Snyder et al. 2015). In fact, researchers have hypothesized that EFs are endophenotypes—intermediate phenotypes on the pathway between genes and diagnosis—for these disorders (Nyden et al. 2011; Glahn et al. 2004; Hasler et al. 2004; Snitz et al. 2006; Willcutt et al. 2005). If so, then the genes that influence EFs also influence vulnerability to psychiatric disorders. In this study, we examine the hypothesis that genetic risk for psychiatric disorders predicts individual differences in EFs. We use large, publically available samples for ADHD, AUT, BP, MDD, and SCZ to find genetic risk variants and construct polygenic risk scores (PRSs) for each disorder, then test whether these risk scores predict EFs in an independent population-based sample that is smaller but more deeply phenotyped.

The EF framework we use is the unity/diversity model (Miyake and Friedman 2012), which was recently discussed by Snyder et al. (2015) as a particularly promising framework for gaining new insights into the relationship between EFs and psychopathology. This model examines nine tasks tapping three separable but correlated latent variable EFs (response inhibition, updating working

Electronic supplementary material The online version of this article (doi:10.1007/s10519-016-9814-2) contains supplementary material, which is available to authorized users.

Edited by Valerie Knopik.

✉ Chelsie E. Benca
chelsie.benca@colorado.edu

¹ Department of Psychology and Neuroscience, University of Colorado Boulder, Boulder, CO, USA

² Department of Psychology, University of Illinois at Urbana-Champaign, Urbana-Champaign, IL, USA

³ Institute for Behavioral Genetics, University of Colorado Boulder, Boulder, CO, USA

⁴ Division of Substance Dependence, Psychiatry Department, University of Colorado Denver, Denver, CO, USA

memory, and shifting sets). The covariances among the nine tasks are partitioned into three orthogonal factors: Common EF, which explains variance in all nine tasks, including the response inhibition tasks; Updating-Specific, which explains residual covariance among the updating working memory tasks (once the common factor is accounted for); and Shifting-Specific, which similarly explains residual covariance among the tasks designed to examine task shifting ability.

The Common EF latent factor is thought to reflect active goal maintenance and top-down biasing of lower-level cognitive processing (Miyake and Friedman 2012), which may be particularly important to avoid dominant or automatic responses. In fact, Common EF is isomorphic with response inhibition; in other words after accounting for Common EF, there is no Inhibition-Specific factor. The Shifting-Specific factor is thought to capture individual differences in the speed with which no-longer-relevant goals are cleared from working memory, and the Updating-Specific factor is thought to capture individual differences in gating information into working memory, as well as possibly memory-specific factors like retrieval (Miyake and Friedman 2012). Our prior work with this model (see Miyake and Friedman 2012), as well as existing meta-analyses and reviews (e.g., Snyder 2013; Snyder et al. 2015) suggests that the Common EF factor is the most closely related to multiple forms of psychopathology. There is less work examining specific variances in updating and shifting (i.e., after removing Common EF variance), but some prior research suggests that they show different relationships with psychopathology-relevant behavior (see summary in Herd et al. 2014). Given this body of research, we use this model as a candidate endophenotype.

A mediational endophenotype, also referred to as an *intermediate* phenotype, is assumed to be closer to the genetic risk factors for the disorder and the behavioral symptoms (Kendler and Neale 2010). Therefore, relevant genes should be more strongly associated with the endophenotype than the psychiatric disorder itself (Flint and Munafò 2007; Walters and Owen 2007). Proposed criteria for endophenotypes include the following: They are associated with the disorder, heritable, and found in unaffected family members at higher rates than in the general population (Gottesman and Gould 2003). Endophenotypes should also co-segregate in families, and be state-independent, or exist in probands even when they are not currently exhibiting the disorder (Gottesman and Gould 2003). Thus, one should be able to find an association between genetic risk for psychopathology and purported endophenotypes even in individuals who do not meet criteria for a disorder at the time they are measured on the endophenotypes.

Twin and family studies have shown that most complex psychiatric disorders are heritable (Shih et al. 2004), with

heritability estimates of 76 % for ADHD, 85–92 % for AUT, 59–87 % for BP, 37 % for MDD, and 81 % for SCZ (Faraone et al. 2005; Miles 2011; Smoller and Finn 2003; Sullivan et al. 2000, 2003). As relatively few to no single nucleotide polymorphisms (SNPs) have been identified at a genome-wide significance level for most of these psychiatric disorders (with the exception of SCZ; Ripke et al. 2014), and at best only a handful of SNPs have been identified for constructs related to EFs (Davis et al. 2010; Ibrahim-Verbaas et al. 2016; Plomin et al. 2013; Rietveld et al. 2014), it is difficult to assess whether the same genetic variants that predict EFs also predict these psychiatric disorders or vice versa. Even in cases for which a relatively large number of genome-wide significant variants have been identified, such as the 128 independent associations with SCZ identified by Ripke et al. (2014), the variants collectively explain very little of the phenotypic variance on a liability scale (3.4 %), with the individual SNPs explaining much less (by one estimate for genetic studies more generally, each SNP is typically associated with a 1.1 odds ratio; Dick et al. 2015).

One approach to increasing effect sizes is to use PRSs. PRSs aggregate the signals from multiple SNPs related to the disorder of interest, instead of testing the association of variants one by one (Dudbridge 2013; Morrison et al. 2007). To calculate a PRS, first a GWAS in a discovery sample is used to quantify the relations between all SNPs and a disorder. Then SNPs that meet a certain p -value threshold (for example, $p < .0005$) in the discovery sample are binned together. However, it is unclear what significance threshold is optimal, because adding SNPs can increase noise as well as signal; the threshold that results in the optimal signal to noise ratio likely varies depending on phenotype and sample size. Thus, studies commonly look at PRSs for SNPs at different p -value bins (e.g., all SNPs with $p < .10$, $.05$, $.005$, etc.). Then, for a given bin or collection of a priori chosen SNPs, in an independent testing sample, the PRS is computed as a summed count of whether or not each individual has 0, 1, or 2 copies of the risk variants. If the discovery sample is sufficiently large, then one might expect the estimates of the regression betas for each SNP to be stable and accurate, and each SNP can be weighted by its beta from the discovery sample (Dudbridge 2013). Finally, the PRSs can be used to predict the disorder or another phenotype in the testing sample.

A benefit of using a PRS is the ability to use a large discovery set for one phenotype (i.e., a psychiatric disorder) to estimate genetic risk and then test for association in an independent, more deeply phenotyped sample. Both samples do not need to have both phenotypes, and a larger sample size is more important in the discovery sample for determining the risk variants and estimating the SNP effect sizes (Dudbridge 2013). So, a smaller testing sample, which in this case

has been assessed with great rigor, can be used to test the genetic association between the two phenotypes.

Genome-wide association studies (GWAS) for constructs related to EFs (such as intelligence test subscales, matrix reasoning, the Stroop test, the trail-making test, and educational attainment) have had varying success in identifying significant genetic variants (Davis et al. 2010; Ibrahim-Verbaas et al. 2016; Plomin et al. 2013; Rietveld et al. 2014). However, due to sample-size constraints, reverse-phenotyping is frequently employed. For example, the observed heritability of educational attainment is due in part to cognitive ability, but also reflects much more, such as work ethic, motivation, and behavioral problems (Krapohl et al. 2016). A recent GWAS by Rietveld et al. (2014) found that a PRS for educational attainment predicted general cognitive ability better than it did educational attainment in an independent sample. The authors suggested that the higher relation to general cognitive ability than to the originally investigated trait (educational attainment) arose because general cognitive ability is an endophenotype for educational attainment. The authors describe this phenomenon of using risk variants for the disorder of interest to try to predict a purported endophenotype as “reverse endophenotyping.” We utilized this approach because we have an extensive EF battery on a relatively small sample that would be inappropriate for risk score discovery. That is, even though it may be more logical to calculate a PRS for the endophenotype and test it with a psychiatric phenotype, we do the opposite because there are currently larger sample sizes for psychiatric disorders than for these EFs.

One recent study found associations between psychopathology and single cognitive measures (verbal-numerical reasoning, educational attainment, reaction time, and memory) in sample sizes of 36,035–112,067 individuals from the UK Biobank (Hagenaars et al. 2016). Associations were examined in two ways: genetic correlations from LD score regression, and PRSs. Schizophrenia was the only disorder consistently related to each measure, with genetic correlations ranging from 0.13 to -0.34 and betas from regressions with PRSs ranging from -0.062 to 0.025. This study shows the best-case scenario for effect sizes in large samples with single measures related to cognition. However, this study focused on individual cognitive tests that did not target particular EFs. In the current study, we use a similar approach to examine relations to multiple EFs, measured at the level of latent variables.

The current study

We used publicly available genome-wide summary data from five case–control samples (AUT, ADHD, BP, MDD, and SCZ) from the Psychiatric Genomics Consortium

(PGC) (The Psychiatric GWAS Consortium Steering Committee 2009). We calculated PRSs for each disorder at multiple p -value bins, and then used them to predict three separable EFs (Common EF, Updating-Specific, and Shifting-Specific latent variables) in an independent sample composed of unrelated individuals drawn from two Colorado twin studies ($n = 386$ with both genetic and EF data).

For our EF measures, we employed a latent variable model, which has two major advantages over individual tasks. First, because they only reflect variance that correlates across tasks, latent variables are free from measurement error due to unreliability (Bollen 1989). Second, particularly for EF constructs, latent variables are more valid measures, because they remove task impurity (Miyake et al. 2000). EFs are higher-level processes that act on lower-level processes; so individual EF tasks typically include a good deal of variance that is not related to the EF of interest (such as verbal or spatial ability). The EF model that we use includes measures that were selected to tap the same EFs but differ in these non-EF requirements so that this non-EF variance would be removed from the latent variables. The result is a purer measure of the EF, but the consequence is that standard errors for estimates of relations with these latent variables may be larger than those for individual tasks to the extent that the latent variable loadings are low (which they typically are for EF models). High reliability and validity is particularly important in evaluating endophenotypes, because poor measurement can outweigh the benefits gained by an endophenotype’s more proximal connections between genes and behavior. Prior research with a subset of the data used here demonstrates that these EF latent variables are highly heritable and show high stability across a 6-year time window (Friedman et al. 2016).

While many previous studies of general cognitive ability have larger samples, deep phenotyping by selecting highly heritable EF constructs based on a well-characterized model of EF should increase our ability to detect an association between psychopathology PRSs and EF, particularly in a smaller sample. Prior work with the data from the Colorado Longitudinal Twin Study sample (Friedman et al. 2016) indicates that these EF latent variables have heritabilities at age 17 of 98 % for Common EF, 100 % for Updating-Specific latent factor, and 76 % for Shifting-Specific. In the same sample, heritability for a general intelligence factor was estimated at 76 % (Friedman et al. 2008). These same latent variables are stable from ages 17 to 23 years, with correlations between the two ages of .86, 1.0, and .91 for Common EF, Updating-Specific, and Shifting-Specific abilities, respectively (Friedman et al. 2016). Moreover the Common EF factor is more strongly related than general cognitive ability to behavior that is

relevant to psychopathology, such as attention problems and self-restraint (e.g., Friedman et al. 2007, 2011). Thus, Common EF is a strong candidate for examination as an endophenotype for psychopathology.

For comparison purposes, we also included measures that were more similar to the psychiatric disorders on which the risk scores were based. Specifically, we used the ADHD and MDD PRSs to predict attention problems, depression symptoms, ADHD and MDD lifetime diagnosis, and a joint general anxiety disorder (GAD) and/or MDD lifetime diagnosis (given that depression and anxiety have a high genetic correlation; Kendler et al., 1992). These analyses enabled us to examine whether our effect sizes for EFs are larger or smaller than those for phenotypes that more closely match the original psychiatric phenotypes used to generate the PRSs.

Intelligence quotient (IQ), another proposed endophenotype for psychopathology and a construct related to EF (Friedman et al. 2006), has been previously linked to genetic risk for SCZ. As such, we tested whether the PRSs were correlated with IQ in our sample to see if we replicate this association and to better interpret the observed relationships between PRSs and EFs.

While we are interested in the relationship between all three latent factors in the EF model and psychopathology, we hypothesize that the PRSs will be negatively related to the Common EF factor, based on prior work suggesting that multiple forms of psychopathology are associated with broad EF deficits (e.g., Snyder et al. 2015). In addition, prior phenotypic and genetic models with one of these samples suggest a possible positive relationship between the Shifting-Specific factor and PRSs for psychopathology, reflecting a stability/flexibility tradeoff with the Common EF factor (Miyake and Friedman 2012); therefore, we hypothesize that PRSs will be positively related to the Shifting-Specific factor.

Method

Participants

Target sample

For genetic analyses participants were 452 individual twins (178 female; mean age at time of EF testing 19.6 [SD = 2.3]), a subset from 2935 twins recruited from the Colorado Longitudinal Twin Study (LTS) and the Colorado Community Twin Study (CTS) at the University of Colorado (Rhea et al., 2013). For all models we included all individuals who had phenotypic data in order to get a more robust estimation of phenotypic traits (distributions or thresholds); however, only a subset of 452–386 individuals who had both genotypic and phenotypic information, depending on the analysis,

contributed to the correlation between PRS and phenotype. For example, in the estimation of the EF models, we used all twins who had EF data ($n = 1543$) in order to get better, more stable estimates of the latent factor loadings; however, only one twin from a subset of those twin pairs was genotyped, and of those, only Caucasian samples were imputed to the 1000 genomes reference panel. Out of the 452 individuals with imputed genotype data, 386 also had EF data, 387 had IQ data, 452 had diagnostic information for ADHD, MDD, general anxiety disorder and/or major depression, or depression symptoms from the Center for Epidemiologic Studies-Depression scale (CES-D), and 257 also had Child Behavioral Checklist (CBCL; Achenbach 1991) data (see supplemental Table S1 for *ns*).

Discovery samples

We used publicly available summary statistics from GWAS to obtain the sets of SNPs (and associated beta weights) to be included in the PRSs for each disorder. The discovery data came from the PGC (Sullivan 2010) and included an AUT sample from the Autism Disorder Working Group (March 2015 Release; URL:PGC) with 10,610 individuals (5305 ASD cases and 5305 pseudo-controls), an ADHD sample (Neale et al. 2010) with 9543 individuals (896 cases, 2455 controls, 2064 trios), a BP sample (Sklar et al. 2011) with 16,731 individuals (7481 cases, 9250 controls), a MDD sample (Ripke et al. 2013) with 76,237 individuals (16,023 cases, 60,214 controls), and a SCZ sample (Ripke et al. 2014) with 150,064 individuals (36,989 cases, 113,075 controls). For more details on the discovery samples' characteristics, preprocessing procedures, and analysis methods used by the PGC, see the papers associated with each dataset.

Materials

Attention problem symptoms

Attention problems were assessed by the attention problems subscale of the Child Behavior Checklist (CBCL; Achenbach 1991). This subscale had 11 symptoms that could be endorsed as not true (0), somewhat true (1), or very true (2), for a maximum score of 22 points. For the LTS, we used multiple waves of parent (either mother or father) ratings from age 7 until age 16 years. After taking the square root of the raw score to help normalize the distribution, we regressed out age separately within each sex at each time point, then averaged the standardized residuals across time. We followed the same procedure for the CTS sample, however we only had parent ratings (mother, father, or both) at one time point. Across both the LTS and CTS samples, mothers' ratings were more

common than fathers'; only mothers answered approximately 77 % of the time, only fathers answered 9 % of the time, and both parents answered approximately 13 % of the time. When both were available, we averaged the parents' ratings at that time point, and then averaged the combined rating with the other time points. Descriptive statistics for raw scores are provided in Supplemental Table S1.

Depression symptoms

Participants completed the Center for Epidemiologic Studies-Depression scale (CES-D; Radloff 1977) at three waves: wave 1 (ages 11.33–15.99 years), wave 2 (ages 15.75–27.45 years), and wave 3 (ages 21.10–34.37 years). This 20-question scale assesses how often a person experiences depressive symptoms on a scale of 0 (rarely or none of the time) to 4 (most or all of the time). At each wave, after reverse-scoring appropriate questions, if an individual answered at least 16 questions, we took the mean of those questions and multiplied it by 20 in order to get a sum score.¹ We used a square root transformation to help normalize the distribution and regressed out age, sex, and their interaction, then averaged the standardized residuals across waves to get a single score for each participant.

Lifetime diagnoses

We examined three lifetime diagnoses: ADHD, MDD, and GAD and/or MDD. Adult case-control status was assessed by the DSM-IV diagnostic criteria, or the DSM-III-R adjusted to be equivalent with the DSM-IV diagnostic criteria if data were collected before 2002. We used the Diagnostic Interview Schedule (DIS; Robins et al. 2000) for participants 18 or older, and the Diagnostic Interview Schedule for Children (DISC; Shaffer, et al. 2000) for participants younger than 18. We had three waves of data available (see *Depression symptoms* section) and used all of the data to create our measures. Age at time of psychopathology assessment ranged from 12 to 34 with a mean age of 24.4 (SD = 3.7). When there were multiple assessments, the age in supplemental Table S1 is from the most recent wave of available data. Our final variables were dichotomous variables for each disorder, where if the participant had ever met criteria for diagnosis at any wave, he or she was considered a case. Out of the 452 participants who had genetic data and information on lifetime diagnosis, 43 (9.5 %) had a lifetime diagnosis of ADHD, 107 (23.6 %) had a lifetime diagnosis of MDD, 45 (10 %) had a lifetime diagnosis of GAD, and 120 (26.5 %) had a lifetime diagnosis of MDD and/or GAD.

¹ Across all waves, only 4 scores were not computed because the participant did not answer at least 16 questions.

Full-scale intelligence

IQ was measured using the Wechsler Adult Intelligence Scale, third edition (WAIS-III; Wechsler 1997) in the LTS sample, and the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler 1999) in the CTS sample. The WAIS-III was collected at a mean age 16.58 (SD = 0.79), with a mean score of 102.2 (range 70–142). The WASI was collected at a mean age of 21.09 (SD = 1.72), with a mean score of 106 (range 53–135). Scores were regressed on age, sex, and their interaction within sample, and the standardized residuals were then concatenated.

EF tasks

Nine EF tasks were used to construct EF latent variables. The inhibition tasks (antisaccade, stop-signal, and Stroop) required stopping a prepotent behavioral response (eye movements, categorization, or word reading, respectively). The dependent measures were antisaccade accuracy, estimated stop-signal reaction time in the stop-signal task, and Stroop response time interference (for incongruent minus asterisks stimuli). The updating working-memory tasks (keep track, letter memory, and spatial 2-back) required monitoring incoming stimuli (words, letters, or spatial locations, respectively) updating working-memory with new relevant information (deleting no longer relevant information) when appropriate. The dependent measures were accuracy. The set-shifting tasks (number-letter, color-shape, and category-switch) required participants to switch between two subtasks (categorizing numbers or letters, colors or shapes, or animacy or size, respectively) on the basis of cues that appeared before each trial. The dependent measures were local switch costs, or the difference in reaction time on switch trials minus repeat trials. Additional information is provided in Table 1; see Friedman et al. (2008) for full details. Tasks were administered in the LTS sample at mean age 17.25 years (SD = 0.65) and in the CTS sample at mean age 21.01 years (SD = 1.68). The CTS and LTS samples were combined and then age, sex, and their interaction were regressed out of each EF task score. Standardized factor loadings for the three orthogonal EF latent variables are provided in Table 1 for the combined sample. See supplemental Table S1 for task descriptive statistics for the sample with genetic data.

Procedures

Genotyping: discovery sample

The AUT2 and SCZ2 sample were part of a second phase and were imputed to the 1000 Genome reference panel

Table 1 Descriptions and Factor Loadings of the Executive Function Tasks

Measure	Description	Standardized factor loading		
		Common EF	Updating-specific	Shifting-specific
Inhibiting				
Antisaccade	Avoid the prepotent response to saccade to a cue and instead look in the opposite direction to view a briefly displayed target	.54	–	–
Stop-signal	Stop a dominant categorization response on infrequent trials in which an auditory signal sounds	.50	–	–
Stroop	Avoid the prepotent tendency to read a word and instead name the color of the font in which the word is printed	.41	–	–
Updating				
Keep-track	From a series of 15 words, remember the most recently presented exemplar of 2-4 specified categories	.38	.63	–
Letter-memory	During a series of letters, continuously rehearse the last three letters and recall them at the end	.38	.47	–
Spatial 2-back	Respond whether an indicated location is the same as that two trials back	.40	.17	–
Shifting				
Number-letter	Categorize whether the number in a letter-number pair is odd or even, or whether the letter is a consonant or vowel, depending on the location of stimuli (top or bottom of screen)	.42	–	.45
Color-shape	Categorize whether a colored shape is a circle or triangle, or red or green, depending on a cue letter (C or S) appearing above the stimulus	.39	–	.43
Category-switch	Categorize a word as living or nonliving, or small or big, depending on a cue symbol appearing above the word	.45	–	.59

Standardized factor loadings (all $p < .05$) from a model with no genetic risk score or principal components included. Models included the full sample ($n = 1549$) although only a subset of 389 individuals contributed to the correlation with the genetic risk scores. The model showed an acceptable fit, $\chi^2(21) = 97.22$, $p < .001$; CFI = .959; RMSEA = .048. EF executive function

(The 1000 Genomes Project Consortium, 2010). The BP sample was imputed to HapMap phase 2; the ADHD and MDD samples were imputed to HapMap phase 3 (Thorisson et al. 2005; The International HapMap Consortium 2003). After quality control through PGC (see individual references for more information) all results files were downloaded to our servers. All discovery samples went through a clumping procedure in PLINK (Purcell et al. 2007) to account for linkage disequilibrium (LD). Clumping accounts for LD by taking the most significant SNPs in a GWAS, then grouping SNPs that meet an LD threshold with this most significant index SNP, resulting in only one signal per LD block. We used an LD threshold of $R^2 < 0.2$, with no SNPs excluded based on p -values for association with the disorder. The resulting SNPs were then put into R (R Core Team 2013) and the list of SNP names were matched to the imputed SNPs in the testing sample for PRS generation in the testing sample.

Genotyping: testing sample

Individuals were genotyped on the Affymetrix 6.0 platform (Affymetrix, Inc., Santa Clara CA) and called by

BEAGLECALL 1.0.1 (Browning and Yu 2009). See the description for the “Center on Antisocial Drug Dependence (CADD)” sample in Derringer et al. (2015) for full details of the cleaning and quality control procedures before imputation.

Caucasians were identified by visual inspection of the first 10 components from a principal components analysis calculated in PLINK using the full, unrelated CADD sample (described in Derringer et al. 2015). Cut-offs for the first 3 PCs were applied, and then the remaining subjects were imputed to the 1000 Genome reference panel using IMPUTE2 (Howie et al. 2009).² The 10 ancestry components were also used as covariates in the analyses.

SHAPEIT was used for the prephasing process (Delaneau et al. 2012). A cut-off info score of $> .4$ was used to ensure good quality imputed SNPs, resulting in approximately 14.9 million SNPs. After restricting imputed SNPs to those also identified in the discovery sample (see

² Visual inspection involved comparing the self-reported ancestry to the places in the distribution that showed breakpoints (or drop-offs) between the sample’s ancestry groups. This resulted in identification of European ancestry participants by component 1 > 0.014 , 0 $<$ component 2 < 0.013 , and component 3 > -0.006 .

Table S3 in supplemental materials for number of SNPs in each PRS), the beta weights for those SNPs were used to calculate weighted risk scores in the testing sample by multiplying 0, 1, or 2 (for copies of the risk allele), or dosages for imputed SNPs, by the beta weight for those SNPs, and summing across SNPs in each *p*-value bin.

Analyses

Analyses were run in Mplus 7.3 (Muthén and Muthén 1998–2012) to allow for estimation of the EF latent variables. Models used all available phenotypic data when possible; however, only individuals who also had genetic information contributed to the correlation between PRSs and phenotype (i.e., individuals with EF data, but without genetic data were included in the models to obtain the best estimates of the factor loadings, but the covariance with the PRS was only based on the subset with both genetic and phenotypic data). For models with categorical diagnoses, mean and variance adjusted weighted least squares (WLSMV) estimation (delta parameterization) was used, which models the underlying liability as a normal distribution using a probit model; for models with only continuous data, robust maximum likelihood (MLR) was used. Non-independence (due to including both twins) was corrected for with the `type = COMPLEX` option, which clusters by family. In all analyses, all individual indicators (e.g., all nine EF tasks) as well as the PRS were regressed on 10 ethnicity PCs.³

As described earlier, all continuous phenotypic variables were age, sex, and age by sex⁴ regressed before analysis. Age (of last diagnostic assessment) and sex were included as covariates for models including diagnoses. PRSs were not regressed on age and sex.

PRSs and EFs

We used structural equation modeling to estimate the three EF latent variables: a Common EF latent variable, representing what is shared between all of the tasks (with loadings from all nine tasks), an Updating-Specific latent variable capturing additional variance specific to Updating tasks (with loadings from three updating tasks), and a Shifting-Specific latent variable capturing additional

variance unique to the shifting tasks (with loadings from the three shifting tasks). The latent factors in the EF model are orthogonal, where Common EF explains covariance across all nine tasks, and the Updating- and Shifting-Specific factors explain additional covariance among the updating and shifting tasks, respectively, that is not explained by the Common EF factor.

To examine the relations of these EF latent variables to each PRS, we correlated them with the residual of the PRS (after removing the PCs from the PRS). Thus, the correlations we present are actually partial correlations controlling for ethnicity, because the 10 PCs were regressed out of both the PRS and the individual EF tasks (and the EF tasks were also residualized on age and sex).

PRSs and psychopathological symptoms, diagnoses, and IQ

We used the five PRSs to predict IQ, ADHD symptom scores, ADHD lifetime diagnosis, depression symptom scores, depression lifetime diagnosis, and MDD and GAD lifetime diagnoses. IQ was correlated with all five PRSs, however ADHD symptom scores and diagnosis, and depression symptom scores and diagnosis, were correlated only with the ADHD PRSs and the MDD PRSs respectively. As with the EF model, the correlations we present are actually partial correlations controlling for ethnicity, because the 10 PCs were regressed out of both the PRS and the phenotype (and the phenotype was also regressed on age and sex).

Permutation

PRSs for higher *p*-value bins include the same SNPs as lower threshold bins for PRSs based on the same disorder. Due to high correlations between *p*-value bins within each risk score (see supplemental Table S2), correlations with the same phenotype across bins of the same risk score are not independent. Therefore, we used permutation to correct for multiple testing. For each permutation, we retained the relatedness of the *p*-value bins within PRSs for each disorder. The association between the independent and dependent variables was broken by randomly shuffling scores for the dependent variables 1000 times and constructing a distribution of statistical coefficients under this null. For example, for the EF model, we shuffled the rows of the nine EF task scores (residualized on age and sex), so that the correlations among the nine EF tasks were retained, but their associations with the PRSs were broken. Because the PRSs were not shuffled, the associations among *p*-value bins remained intact. For each shuffle, we then ran the same model (including ethnicity PCs, which were not shuffled), and obtained the newly estimated correlations between PRSs and EFs. We constructed the

³ To include individuals without genetic data in the estimation of the EF latent variables (Mplus will exclude individuals missing on covariates) and other phenotypic measures, we imputed missing PCs as the average for that self-identified ethnicity in our genetic sample. The number of individuals who contributed to each phenotype was as follows: EFs = 1543; CES-D = 2875; CBCL = 1684; IQ = 1571; DIS diagnoses = 2875.

⁴ This interaction term was included even though it was not significant in any models.

empirical distribution of correlation coefficients for each disorder in this way, and used it to calculate empirical p -values for the correlations we obtained in our unpermuted models (i.e., a correlation would be significant if it was more extreme than 95 % of the empirical correlation values in the distribution of permuted correlations). This is ultimately less stringent than a Bonferroni correction (Carmargo et al. 2008) for multiple testing, but does not correct for the multiple testing due to examining multiple phenotypes, for which we divided our alpha of .05 by the number of phenotypes tests (nine) examined, for a new alpha of .006.

Results

PRSs with cognitive measures

EFs

To examine the relationship between EFs and genetic risk for psychopathology, we correlated the PRSs (residualized on PCs) with the EF latent variables (individual tasks regressed on PCs). Correlations are shown in Fig. 1. Common EF was positively correlated with the MDD $p < .05$ bin PRS, but did not significantly relate to the other PRSs at any p -value bin, and this correlation did not survive multiple testing correction.

The Updating-Specific latent variable significantly positively correlated with the ADHD $p < .0005$ and SCZ $p < 1$ bins. While the former survived permutation, neither of these results were significant after correcting for multiple-testing of the nine phenotypes. Updating-Specific abilities did not appear to be related to any of the other three disorders. Likewise, Shifting-Specific abilities were not related to genetic risk for any of the five disorders.

IQ

We also examined the relationship between IQ and the PRSs, because IQ is phenotypically associated with EFs (Friedman et al., 2006; 2008) and has been related to PRSs for SCZ (Lencz et al. 2014; McIntosh et al. 2013). As shown in Fig. 1d, IQ was negatively correlated with the SCZ $p < 5 \times 10^{-5}$ bin, but this result did not survive correction for multiple testing.

PRSs with measures of psychopathology

Given the relatively small effects we observed with the proposed endophenotypes (EFs and IQ), we wondered if we would get similarly small effects with phenotypes that were arguably more closely related to the phenotypes used

to construct the PRSs. So, we examined how the ADHD and MDD PRSs related to attention and depression symptoms and lifetime diagnoses. The magnitude of effects found for relevant phenotypes within our sample allows for a better understanding of the magnitude of relationship observed with EF and IQ.

The relationships between ADHD and MDD symptom scores and their respective PRSs were assessed with correlational analyses of the residuals of PRS and phenotype after each was regressed on the PCs for ethnicity. As shown in Fig. 2, PRSs for MDD were not significantly related to any psychopathological phenotypes in our sample. While genetic risk for ADHD was not related to ADHD symptom scores, it was correlated with lifetime diagnosis for ADHD at one bin ($p < .05$), but this result did not survive multiple testing correction. Because we did not find significant results with either EFs or psychopathology measures after correcting for multiple testing, we did not test whether the magnitudes of effects were significantly larger for EFs.

Power

To better interpret our results, we conducted power analyses for the EF measures in our sample. As can be seen in Table 2, we would have enough power with our sample size used in this study ($N = 386$) if there were a moderate to large correlation ($r = .20$ – $.50$) between Common EF and a PRS. However, the observed effect sizes were smaller than this, and therefore we were underpowered with our sample size. We examined what sample size would be necessary for a power of .80 with a smaller correlation estimate ($r = .10$). Alpha levels were varied because we examined nine different phenotypes that are not fully independent of one another; for example, ADHD symptoms are correlated with ADHD lifetime diagnosis ($r = .25$) and MDD symptoms at a lower level ($r = .15$), so our adjusted alpha should be somewhere between .05 and .006. As shown in Table 2, for 80 % power to detect an effect with a correlation of .10 or smaller, larger sample sizes, on the order of 1510–2500 or more, are necessary. In summary, if latent EFs were strong endophenotypes for psychopathology and we observed stronger relations between EFs and these PRSs than previously seen with other cognitive measures, we would have been adequately powered. However with a correlation of .10 or smaller, we would need many more subjects to have adequate power.

Discussion

To understand the potential of EFs as endophenotypes for psychiatric disorders, we used large discovery datasets to generate PRSs for five disorders (AUT, ADHD, MDD, BP,

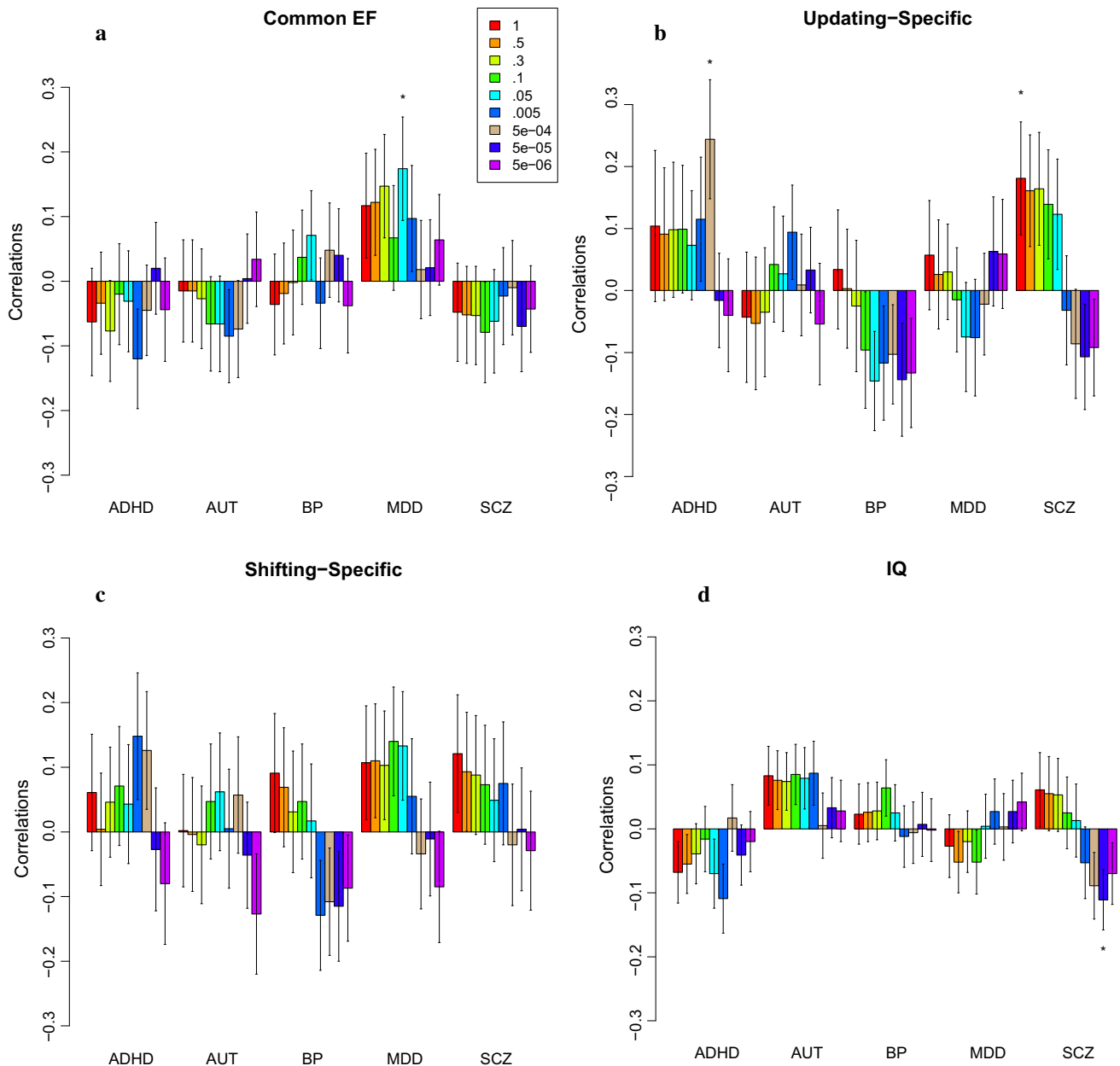


Fig. 1 Correlations between cognitive measures and psychopathological polygenic risk scores (PRSs). Bars represent standard errors. Legend shows colors corresponding to p -value threshold bins for each disorder. **a** Correlations between PRSs and the Common EF latent factor. **b** Correlations between PRSs and the Updating-Specific latent

factor. **c** Correlations between PRSs and the Shifting-Specific latent factor. **d** Correlations between PRSs and IQ. EF executive function, IQ intelligence quotient, ADHD Attention Deficit Hyperactive Disorder, AUT Autism, BP Bipolar Disorder, MDD Major Depressive Disorder, SCZ Schizophrenia. $*p < .05$ uncorrected

and SCZ) and related those PRSs to EF latent variables in an independent dataset. We found little evidence for stronger effect sizes for the EFs than measures more similar to these psychopathologies. The general pattern of results indicated that EFs might be related to psychopathology, but they may not lead us to find more genetic variants than symptom or diagnosis measures unless we have significantly larger sample sizes.

At a nominally significant level, a Common EF latent variable was positively related to genetic risk for depression; however, this effect was in the opposite direction than expected, with higher genetic risk for depression indicating better Common EF in a general population sample. Higher genetic risk for ADHD was nominally related to better Updating-Specific abilities; this association was also not in the expected direction. The amount of variance explained

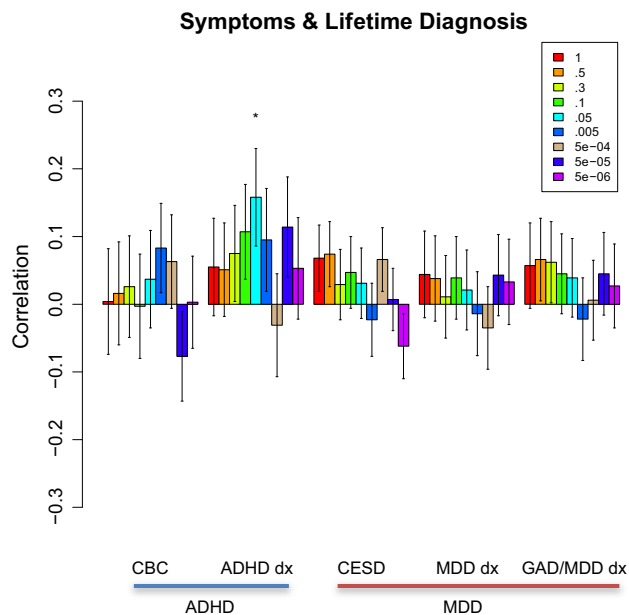


Fig. 2 Correlations between ADHD, MDD, and GAD/MDD symptoms and lifetime diagnosis, and ADHD and MDD polygenic risk scores (PRSs). Correlations are partial correlations after the 10 principal components for ethnicity have been regressed out of the PRSs and the phenotypic measures and age and sex have been regressed out of the phenotypic measures. Bars represent standard errors. Legend shows colors corresponding to p -value threshold bins for each disorder. *ADHD* Attention Deficit Hyperactive Disorder, *MDD* Major Depressive Disorder, *CBC* Child Behavioral Checklist, *ADHD dx* lifetime diagnosis of ADHD, *CESD* Center for Epidemiologic Studies-Depression Scale, *MDD dx* MDD lifetime diagnosis, *GAD/MDD dx* General Anxiety Disorder or MDD lifetime diagnosis. * $p < .05$ uncorrected

by the PRS for each latent factor was $R^2 = 0.03$ for Common EF and $R^2 = 0.06$ for Updating-Specific. However, these results did not survive correction for multiple testing, so they would need to be replicated to determine if they are real effects that are simply underpowered.

Likewise, we examined a measure of general cognitive ability, IQ, which has also been proposed as an endophenotype for psychopathology (Burdick et al. 2009).

Although our results did not survive correction for multiple testing, the directionality and variance explained was comparable to what has been observed in previous studies. Lencz et al. (2014) linked a PRS for general cognitive ability to case-control status of SCZ, and McIntosh et al. (2013) linked a PRS for SCZ to increased cognitive decline between the ages of 11 and 70. We also found that increased genetic risk for SCZ predicted lower IQ, with the amount of variance explained ($R^2 = 0.01$) comparable to that found by (Lencz et al. 2014; $R^2 = 0.000$ – 0.019) and (McIntosh et al. 2013; $R^2 = 0.006$ – 0.009). The replication of this association between increased genetic risk for SCZ and cognitive ability suggests that we may be seeing real, but underpowered, effects.

A mediation model of an endophenotype (Kendler and Neale 2010) assumes that the endophenotype is more proximal to genes that influence the psychiatric disorder. If the mediation assumption is incorrect and phenotypes related to disorders of interest, such as depression symptoms, are equally or more strongly related to the PRSs, then EFs as endophenotypes might not be as useful for PRS research. To address this assumption, we also used the PRSs to predict relevant phenotypes more similar to the psychopathologies used to generate the PRSs. Again, we found few associations. A relationship between increased risk for ADHD and lifetime diagnosis for ADHD emerged, where greater genetic risk was related to higher rates of lifetime diagnosis, but it did not survive correction for multiple testing. However, the amount of variance explained ($R^2 = 0.02$) is similar to what we observed for EF and IQ, suggesting that in a small testing sample, psychopathology phenotypes do not have a weaker relationship with PRSs than candidate endophenotypes.

Recently, a few studies have addressed the assumption that endophenotypes will elicit larger effect sizes with respect to genetic variants. A meta-analysis by Flint and Munafò (2007) concluded that endophenotypes were not necessarily showing larger effect sizes than the disorders of interest. These results could have occurred because the

Table 2 Power for Executive Function Analyses

Simulated correlations	$1-\beta$; $\alpha = .05$ and $N = 386$	Required N ; $\alpha = .05$ and $1-\beta = .8$	Required N ; $\alpha = .01$ and $1-\beta = .8$	Required N ; $\alpha = .006$ and $1-\beta = .8$
.50	1	57	84	94
.40	.999	91	135	151
.30	.990	165	245	274
.20	.811	375	558	625
.10	.294	1510	2247	2519

Power analysis for the executive functions (EFs) latent-variable model where simulated correlations represent a theoretical correlation between common EF and the polygenic risk score

$1-\beta$ = power; α = alpha; N number of participants included in parameter estimates

studies were not using appropriate endophenotypes or because the assumption that endophenotypes have larger effect sizes is incorrect. In particular, if the endophenotypes were not mediators between the genes and phenotypes as often assumed, but instead indices of liability, where the same genes influence both the endophenotypes and the phenotypes of interest (Kendler and Neale 2010), then one might not expect larger effect sizes for the endophenotypes.

The largest GWAS study to date found no significant hits for EF tasks (Stroop, trail-making, and fluency tests; Ibrahim-Verbaas et al. 2016), despite discovery sample sizes ranging from 5429 to 32,070. Thus, EF tasks, like other measures, seem to have relatively small effect sizes for individual variants. However, another meta-analysis by Rose and Donohoe (2013) found different effect sizes for two different classes of endophenotypes for SCZ, with larger effect sizes for cognitive neuroimaging endophenotypes than lab-based cognitive measures. More research is needed to establish good estimates of expected effect sizes for different types of endophenotypes.

Another emerging debate focuses on issues of sample size and phenotype specificity when testing for genotype-phenotype associations. Many studies have shown that with the small effect sizes for individual SNPs, large samples will be necessary to detect significant associations with the phenotypes of interest. However, when combining data sets or using large publicly available datasets, often only rudimentary phenotypic assessment is available (e.g., case-control status, without information on which symptoms were endorsed or degree of severity of illness). This thin phenotyping allows for the inclusion of more subjects, but potentially dilutes statistical power and the strength of association (Tracy 2008). While this trade-off holds in this study with regard to PRS generation, the deep phenotyping of a candidate endophenotype could possibly help in the testing sample. We had hoped that our deep phenotyping with the latent variable model of EF, which reduces measurement error and extracts highly heritable latent factors that are more stable across time than single measures (Friedman et al. 2016), would enable us to detect a larger effect. We were well powered to detect effects that explained 4 % or more of the variance, but the effects we obtained were smaller than that.

Despite being underpowered, there is still useful information to gain pertaining to the effect sizes we can reasonably expect from endophenotypes compared to more direct measures of psychopathology with a small testing sample size. Lab-based measures of EF, even at a highly heritable latent variable level, do not seem to generate substantially larger effect sizes for genes related to risk for psychopathology than measures of symptoms, at least in a population-based sample.

Limitations

In addition to the previously discussed power issues, a limitation of this study is that our sample was population-based with low levels of psychopathology; hence, the genetic variance related to psychopathology was likely restricted compared to a clinical sample. Although endophenotypes are present in individuals without the disorder of interest, particularly in family members of a proband, the use of a population sample might have limited the variance in the endophenotype as well. Thus, a stronger effect would perhaps be seen in a clinical sample.

Although we chose to calculate PRSs from psychiatric disorders and test them with EFs because larger sample sizes are available for the former than the latter, and because its utility has been previously demonstrated in other studies (Lencz et al. 2014; Rietveld et al. 2014), the reverse endophenotype approach could also be considered a limitation. The relationship between purported endophenotypes and genetic risk for psychopathology is likely a complicated matter (Cannon and Keller 2006). If an endophenotype is only related to a portion of the genes influencing a given disorder, the strength of the relationship between all genes that affect the disorder and the endophenotype is unclear. Conversely, if the endophenotype is a complex trait itself, such as EF, there are likely unique genetic contributions to EF that do not overlap with the more distal phenotype of interest, such as psychopathology. Due to the unclear genetic relationship between endophenotypes and the more distal phenotype, it is difficult to estimate an expected effect size. However, the genetic architecture of both psychopathology and EF are important for the interpretation of our results.

Multiple testing could also be considered a complication of this study. Associations between five disorders and four phenotypes (Common EF, Updating-Specific, Shifting-Specific and IQ) were tested, as well as one disorder (ADHD) with two phenotypes (ADHD diagnosis and ADHD symptoms), and one disorder (MDD) with three phenotypes (MDD diagnosis, combined GAD/MDD diagnosis, and MDD symptoms), all of these at nine bins. In total, we conducted 225 tests. While this number is not remarkable for those working with GWAS data, it is greater than is typically done in PRS studies. How to adequately correct for multiple testing is complicated by the fact that the nine bins are not independent from each other, the phenotypes are not independent (e.g., ADHD symptom count is correlated with ADHD diagnosis), and the different disorders are also not independent of one another due to comorbidity. We chose to use permutation testing and then use a Bonferroni correction for the number of bins; however, there is no clear best way to correct for

multiple testing in this scenario. One suggestion for future studies would be to reduce the number of bins tested, particularly if testing several phenotypes. However, our initial thorough approach in the exploratory analyses presented here will guide future investigations of relationships between common and specific EFs and a range of psychopathology outcomes.

Conclusion

In this study, we examined the relationship between PRSs for psychopathology and EFs with highly heritable EF latent variables. Despite large sample sizes for deriving PRSs for psychopathology and deeply phenotyped candidate endophenotypes, we did not see substantial effects. The highest observed relations between PRSs for psychopathology and EFs ranged from an R^2 of .03–.06, which are smaller than we needed for adequate power with our sample size. The highest R^2 for non-EF phenotypes with PRSs was .03, in a similar range as our EF measures. Overall, our results are similar to what was found by Flint and Munafò (2007) and provide little evidence for EFs as endophenotypes that will give significantly larger estimates than psychiatric phenotypes such as lifetime diagnosis. However, even if EFs do not necessarily show larger genetic effect sizes than psychiatric measures, their transdiagnostic associations with psychopathology (Snyder et al. 2015) suggests that increasing understanding of their genetic influences can provide a window into disease mechanisms and pathways.

Funding This research was supported by National Institute of Health Grants MH063207, MH001865, MH016880, MH079485, DA011015, DA035804, AG046938, and HD010333.

Compliance with ethical standards

Conflict of interest Chelsie E. Benca, Jaime L. Derringer, Robin P. Corley, Susan E. Young, Matthew C. Keller, John K. Hewitt, and Naomi P. Friedman declares that they have no conflict of interest.

Human and Animal Rights All procedures performed in studies involving human participants were in accordance with the ethical standards of the University of Colorado Institutional Review Board and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Parental permission and informed consent or assent were obtained from each subject at each assessment.

References

Achenbach TM (1991) Manual for the Child Behavior Checklist 4-18 and 1991 Profile. University of Vermont Department of Psychiatry, Burlington

- Amann B, Gomar JJ, Ortiz-Gil J, McKenna P, Sans-Sansa B, Sarró S, Pomarol-Clotet E (2012) Executive dysfunction and memory impairment in schizoaffective disorder: a comparison with bipolar disorder, schizophrenia and healthy controls. *Psychol Med* 42:2127–2135
- Bollen KA (1989) Structural equations with latent variables. Wiley, New York
- Browning B, Yu Z (2009) Simultaneous genotype calling and haplotype phase inference improves genotype accuracy and reduces false positive associations for genome-wide association studies. *Am J Hum Genet* 85:847–861
- Burdick KE, Gunawardane N, Woodberry K, Malhotra AK (2009) The role of general intelligence as an intermediate phenotype for neuropsychiatric disorders. *Cognit Neuropsychiatr* 12:299–311
- Camargo A, Azuaje F, Wang H, Zheng H (2008) Permutation-based statistical tests for multiple hypotheses. *Source Code Biol Med* 3:15
- Cannon TD, Keller MC (2006) Endophenotypes in the genetic analyses of mental disorders. *Annu Rev Clin Psychol* 2:267–290
- Davis OSP, Butcher LM, Docherty SJ, Meaburn EL, Curtis CJC, Simpson MA, Plomin R (2010) A three-stage genome-wide association study of general cognitive ability: hunting the small effects. *Behav Genet* 40:759–767
- Delaneau O, Marchini J, Zagury JF (2012) A linear complexity phasing method for thousands of genomes. *Nat Methods* 9:179–181
- Derringer J, Corley RP, Haberstick BC, Young SE, Demmitt BA, Howrigan DP, McQueen MB (2015) Genome-wide association study of behavioral disinhibition in a selected adolescent sample. *Behav Genet* 45:375–381
- Dick DM, Agrawal A, Keller MC, Adkins A, Aliev F, Monroe S, Sher KJ (2015) Candidate gene-environment interaction research: reflections and recommendations. *Perspect Psychol Sci* 10:37–59
- Dudbridge F (2013) Power and predictive accuracy of polygenic risk scores. *PLoS Genet* 9:e1003348
- Faraone SV, Perlis RH, Doyle AE, Smoller JW, Goralnick JJ, Holmgren MA, Sklar P (2005) Molecular genetics of attention-deficit/hyperactivity disorder. *Biol Psychiatry* 57:1313–1323
- Flint J, Munafò MR (2007) The endophenotype concept in psychiatric genetics. *Psychol Med* 37:163–180
- Friedman NP, Haberstick BC, Willcutt EG, Miyake A, Young SE, Corley RP, Hewitt JK (2007) Greater attention problems during childhood predict poorer executive functioning in late adolescence. *Psychol Sci* 18:893–900
- Friedman NP, Miyake A, Corley RP, Young SE, Defries JC, Hewitt JK (2006) Not all executive functions are related to intelligence. *Psychol Sci* 17:172–179
- Friedman NP, Miyake A, Young SE, Defries JC, Corley RP, Hewitt JK (2008) Individual differences in executive function are almost entirely genetic in origin. *J Exp Psychol Gen* 137:201–225
- Friedman NP, Miyake A, Robinson JL, Hewitt JK (2011) Developmental trajectories in toddler's self-restraint predict individual differences in executive functions 14 years later: a behavioral genetic analysis. *Dev Psychol* 47:1410–1430
- Friedman NP, Miyake A, Altamirano LJ, Corley RP, Young SE, Rhea SA, Hewitt JK (2016) Stability and change in executive function abilities from late adolescence to early adulthood: a longitudinal twin study. *Dev Psychol* 52:326–340
- Glahn DC, Bearden CE, Niendam TA, Escamilla MA (2004) The feasibility of neuropsychological endophenotypes in the search for genes associated with bipolar affective disorder. *Bipolar Disord* 6:171–182
- Gottesman II, Gould TD (2003) The endophenotype concept in psychiatry: etymology and strategic intentions. *Am J Psychiatry* 160:636–645

- Hagenaars SP, Harris SE, Davies G, Hill WD, Liewald DCM, Ritchie SJ, Deary IJ (2016) Shared genetic aetiology between cognitive functions and physical and mental health in UK Biobank (N = 112151) and 24 GWAS consortia. *Mol Psychiatry*. doi:10.1038/mp.2015.225
- Hasler G, Drevets WC, Manji HK, Charney DS (2004) Discovering endophenotypes for major depression. *Neuropsychopharmacology* 29:1765–1781
- Herd SA, O'Reilly RC, Hazy TE, Chatham CH, Brant AM, Friedman NP (2014) A neural network model of individual differences in task switching abilities. *Neuropsychologia* 62:375–389
- Howie BN, Donnelly P, Marchini J (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 5:e1000529
- Ibrahim-Verbaas CA, Bressler J, Debette S, Schuur M, Smith AV, Bis JC, Mosley TH (2016) GWAS for executive function and processing speed suggests involvement of the *CADM2* gene. *Mol Psychiatry* 21:189–197
- Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ (1992) Major depression and generalized anxiety disorder: same genes, (partly) different environments? *Arch Gen Psychiatry* 49:716–722
- Kendler KS, Neale MC (2010) Endophenotype: a conceptual analysis. *Mol Psychiatry* 15:789–797
- Krapohl E, Euesden J, Zabaneh D, Pingault J-B, Rimfeld K, von Stumm S, Dale PS, Breen G, O'Reilly PF, Plomin R (2016) Phenome-wide analysis of genome-wide polygenic scores. *Mol Psychiatry* 21:1188–1193
- Lenz T, Knowles E, Davies G, Guha S, Liewald DC, Starr JM, Malhotra AK (2014) Molecular genetic evidence for overlap between general cognitive ability and risk for schizophrenia: a report from the cognitive genomics consortium (COGENT). *Mol Psychiatry* 19:168–174
- McIntosh AM, Gow A, Luciano M, Davies G, Liewald DC, Harris SE, Deary IJ (2013) Polygenic risk for schizophrenia is associated with cognitive change between childhood and old age. *Biol Psychiatry* 73:938–943
- Miles JH (2011) Autism spectrum disorders: a genetics review. *Genet Med* 13:278–294
- Miyake A, Friedman NP, Emerson MJ, Witzki AH, Howerter A, Wager TD (2000) The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: a latent variable analysis. *Cogn Psychol* 41:49–100
- Miyake A, Friedman NP (2012) The nature and organization of individual differences in executive functions: four general conclusions. *Curr Dir Psychol Sci* 21:8–14
- Morrison AC, Bare LA, Chambless LE, Ellis SG, Malloy M, Kane JP, Boerwinkle E (2007) Prediction of coronary heart disease risk using a genetic risk score: the atherosclerosis risk in communities study. *Am J Epidemiol* 166:28–35
- Muthén LK, Muthén BO (1998-2012) *Mplus User's Guide*. Seventh Edition. Los Angeles, CA: Muthén & Muthén
- Neale BM, Medland SE, Ripke S, Asherson P, Franke B, Lesch KP, Nelson S (2010) Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* 49:884–897
- Nyden A, Hagberg B, Gousse V, Rastam M (2011) A cognitive endophenotype of autism in families with multiple incidence. *Res Autism Spectr Disord* 5:191–200
- Plomin R, Haworth CMA, Meaburn EL, Price TS, Davis OSP (2013) Common DNA markers can account for more than half of the genetic influence on cognitive abilities. *Psychol Sci* 24:562–568
- Purcell S, Neale BM, Todd-brown K, Thomas L, Ferreira MAR, Bender D, Sham PC (2007) REPORT PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575
- R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. ISBN 3-900051-07-0
- Radloff LS (1977) The CES-D scale: a self-report depression scale for research in the general population. *Appl Psychol Meas* 1:385–401
- Rhea SA, Gross AA, Haberstick BC, Corley RP (2013) Colorado twin registry: an update. *Twin Res Hum Genet* 16:351–357
- Rietveld CA, Esko T, Davies G, Pers TH, Turley P, Benyamin B, Philipp D (2014) Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proc Natl Acad Sci* 111:13790–13794
- Ripke S, Neale BM, Corvin A, Walters JTR, Farh KH, Holmans PA, O'Donovan MC (2014) Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511:421–427
- Ripke S, Wray NR, Lewis CM, Hamilton SP, Weissman MM, Breen G, Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium (2013) A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry* 18:497–511
- Robins LN, Cottler LB, Buchholz KK, Compton WM, North CS, Rourke KM (2000) Diagnostic interview schedule for the DSM-IV (DIS-IV). Washington University School of Medicine, St Louis
- Rose EJ, Donohoe G (2013) Brain vs behavior: an effect size comparison of neuroimaging and cognitive studies of genetic risk for schizophrenia. *Schizophr Bull* 39:518–526
- Rosenthal M, Wallace GL, Lawson R, Wills MC, Dixon E, Yerys BE, Kenworthy L (2013) Impairments in real-world executive function increase from childhood to adolescence in autism spectrum disorders. *Neuropsychology* 27:13–18
- Shaffer D, Fisher P, Lucas CP, Dulcan MK, Schwab-Stone ME (2000) NIMH diagnostic interview schedule for children version IV (NIMH DISC-IV): description, differences from previous versions, and reliability of some common diagnoses. *J Am Acad Child Adolesc Psychiatry* 39:28–38
- Shih RA, Belmonte PL, Zandi PP (2004) A review of the evidence from family, twin and adoption studies for a genetic contribution to adult psychiatric disorders. *Int Rev Psychiatry* 16:260–283
- Sklar P, Ripke S, Scott LJ, Andreassen OA, Chichon S, Craddock N, Corvin A (2011) Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near *ODZ4*. *Nat Genet* 43:977–983
- Smoller JW, Finn CT (2003) Family, twin, and adoption studies of bipolar disorder. *Am J Med Genet* 123C:48–58
- Snitz BE, MacDonald III AW, Carter SC (2006) Cognitive deficits in unaffected first-degree relatives of schizophrenia patients: a meta-analytic review of putative endophenotypes. *Schizophr Bull* 32:179–194
- Snyder HR (2013) Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: a meta-analysis and review. *Psychol Bull* 139:81–132
- Snyder HR, Miyake A, Hankin BL (2015) Advancing understanding of executive function impairments and psychopathology: bridging the gap between clinical and cognitive approaches. *Front Psychol* 6:328
- Sullivan PF (2010) The psychiatric GWAS consortium: big science comes to psychiatry. *Neuron* 68:182–186
- Sullivan PF, Kendler KS, Neale MC (2003) Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry* 60:1187–1192
- Sullivan PF, Neale MC, Kendler KS (2000) Genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry* 157:1552–1562
- The 1000 Genomes Project Consortium (2010) A map of human genome variation from population-scale sequencing. *Nature* 467:1061–1073

- The International HapMap Consortium (2003) The International HapMap project. *Nature* 426:789–796
- The Psychiatric GWAS Consortium Steering Committee (2009) A framework for interpreting genome-wide association studies of psychiatric disorders. *Mol Psychiatry* 14:10–17
- Thorisson GA, Smith AV, Krishnan L, Stein LD (2005) The International HapMap Project Web site. *Genome Res* 15(11):1592–1593
- Tracy RP (2008) ‘Deep phenotyping’: characterizing populations in the era of genomics and systems biology. *Curr Opin Lipidol* 19:151–157
- Walters JTR, Owen MJ (2007) Endophenotypes in psychiatric genetics. *Mol Psychiatry* 12:886–890
- Wechsler D (1997) Wechsler adult intelligence scale, 3rd edn. Psychol Corp, San Antonio
- Wechsler D (1999) Wechsler abbreviated scale of intelligence (WASI). Psychological Corporation, San Antonio
- Willcutt EG, Doyle AE, Nigg JT, Faraone SV, Pennington BF (2005) Validity of the executive function theory of attention-deficit/hyperactivity disorder: a meta-analytic review. *Biol Psychiatry* 57:1336–1346

URLs

PGC: <http://www.med.unc.edu/pgc/downloads>

R: <http://www.R-project.org/>