Review

Biological Psychiatry

Risk in Relatives, Heritability, SNP-Based Heritability, and Genetic Correlations in Psychiatric Disorders: A Review

Bart M.L. Baselmans, Loïc Yengo, Wouter van Rheenen, and Naomi R. Wray

ABSTRACT

The genetic contribution to psychiatric disorders is observed through the increased rates of disorders in the relatives of those diagnosed with disorders. These increased rates are observed to be nonspecific; for example, children of those with schizophrenia have increased rates of schizophrenia but also a broad range of other psychiatric diagnoses. While many factors contribute to risk, epidemiological evidence suggests that the genetic contribution carries the highest risk burden. The patterns of inheritance are consistent with a polygenic architecture of many contributing risk loci. The genetic studies of the past decade have provided empirical evidence identifying thousands of DNA variants associated with psychiatric disorders. Here, we describe how these latest results are consistent with observations from epidemiology. We provide an R tool (CHARRGe) to calculate genetic parameters from epidemiological parameters and vice versa. We discuss how the single nucleotide polymorphism–based estimates of heritability and genetic correlation relate to those estimated from family records.

Keywords: Family register data, Genetic correlation, GWAS, Heritability, Psychiatric genetics, Risk in relatives

https://doi.org/10.1016/j.biopsych.2020.05.034

It has long been known that psychiatric disorders run in families [(1); cited in (2)] and that there is increased risk of the same disorder, and also of other psychiatric disorders, in relatives of those with a psychiatric diagnosis [(3); cited in (4)]. Epidemiological studies (5–7) continue to provide evidence-based reports of the genetic contribution to psychiatric disorders (8–11). However, genome-wide association studies (GWASs) of the past decade have allowed a whole new approach to understanding the genetic contribution to, and similarities between, psychiatric disorders (12–17). Here, we bring together evidence for the genetic contribution to risk of psychiatric disorders and the shared genetic contribution between them. We focus on the classical methods from genetic epidemiology and the new methods of the past decade that use single nucleotide polymorphism (SNP) GWAS data, and the relationship between these approaches. We emphasize the concepts in the main text but provide more detail (and analysis tools) in Supplement 1, which we hope will be useful as an aid to teaching and learning.

RISK IN RELATIVES, HERITABILITY, AND SNP-BASED HERITABILITY

Risk in Relatives

Genetic epidemiology is the study of disease patterns in families that are observed in population samples of many families. These studies provide direct measurements of disease rates in family members of those diagnosed with a disease, which can be compared with the rate of disease in the population as a whole. Across the full range of psychiatric disorders, increased rates are observed in family members of those diagnosed with a disorder (5–11,18–20) (Figure 1; Table S1 in Supplement 2).

Genetic epidemiological studies are easy to conceptualize but can be difficult to conduct. Consider a disease that affects one in a hundred individuals in the population in their lifetime—very large samples of families are needed to detect whether children or siblings of those affected by the disease have increased rates of disease above this baseline rate (8). An efficient study design is to select individuals with psychiatric disorders as probands and then track only family members of these probands, comparing rates of disease with those in a general population sample. As in any study design, ascertainment biases can occur. For example, age at onset of disease can lead to censorship of lifetime disease prevalence. This bias can be overcome in a study design that follows risk of disease in family members of both case probands and age- and sex-matched controls, to allow comparison of risk rates in samples ascertained under the same protocols. Notably, there are some large studies using national registry data (5,20) reporting risk in relatives, as well as reviews that provide meta-analyses of smaller studies (19,21) (Tables S2–S5 in Supplement 2).

Risk in First-Degree Relatives and Heritability

If population lifetime risk of disease (K) and the risk of disease in first-degree relatives (K_1) are directly measured, then the risk ratio (RR) for disease is $RR_1 = K_1/K$ (sometimes λ is used to

ISSN: 0006-3223

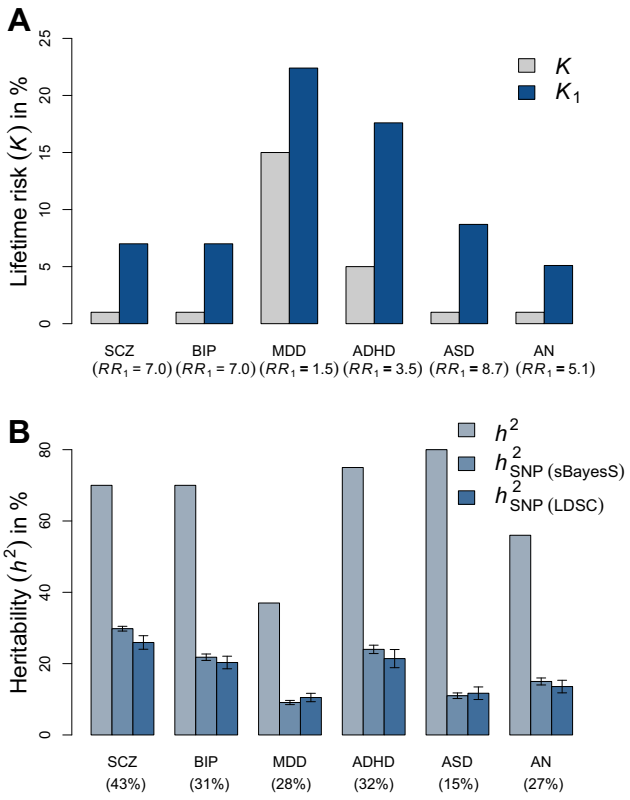

Figure 1. Different ways to express the genetic contribution to psychiatric disorders. **(A)** Lifetime risk of major psychiatric disorders in the population (K) and in those who have a first-degree relative affected (K_1). Also, the risk ratio (RR) for those with a first-degree relative affected ($RR_1 = K_1/K$) is shown under each disorder (x-axis). Note that part of the increased risk in relatives may reflect nongenetic factors shared by relatives. **(B)** Heritability (h^2), the proportion of variance in liability attributable to genetic factors, and single nucleotide polymorphism (SNP)–based heritability (h^2_{SNP}), the proportion of variance in liability associated with common SNPs genome-wide. Here, the h^2 values are consistent with the K and RR_1 reported in panel **(A)**, assuming the increased risk in relatives is attributable only to shared genetic factors. The parameter estimates used in this figure are composites justified from references cited in Table S1 in Supplement 2. SNP-based heritability was estimated from the genome-wide association study summary statistics using SBayesS and linkage disequilibrium score regression (LDSC) (46,48). Percentages under each label represent the ratio h^2_{SNP}/h^2. ADHD, attention-deficit/hyperactivity disorder; AN, anorexia nervosa; ASD, autism spectrum disorder; BIP, bipolar disorder; MDD, major depressive disorder; SCZ, schizophrenia.

denote RR). While all K_1 values are greater than K values, for psychiatric disorders (Figure 1A), they are much lower than those seen for "simple" Mendelian diseases such as Huntington disease, where on average 50% of individuals with a parent with the disease are also diagnosed with the disease, because a single dominant mutation is causal for disease (22). Note that these Mendelian disorders are generally less common [e.g., for Huntington disease $K = 2.7 \times 10^{-5}$ (24)] than psychiatric disorders, where K typically falls in the range of 0.005 to 0.15 (Figure 1; Table S1 in Supplement 2).

The observed inheritance patterns of psychiatric disorders can be resolved as being consistent with the laws of genetic segregation if there are many DNA variants associated with risk of disease (i.e., polygenic), and if other risk factors also play a role (23). Other risk factors could include stress (25–27) or trauma (28,29); however, they also include de novo genetic events (30–32) (i.e., genetic mutations new in a child not present in their parents). Moreover, since each individual's life experience is different, nongenetic risk factors are also likely made up of many events. Therefore, psychiatric diseases are often termed "complex," in contrast to "simple" Mendelian diseases [although the latest results suggest that even Mendelian diseases are complex (33)]. When many factors contribute to the risk of disease, it can be helpful to consider a model of disease that describes a latent distribution of liability to disease (23,34). Since this latent liability is a sum of many genetic and other risks, it is reasonable to assume that the liability distribution is approximately normal (since many things added together will make a bell-shaped distribution in a population sample, the central limit theorem). Hence, those with disease have disease liability above a threshold that is required for disease to occur, and therefore this model is sometimes called the liability threshold model (Figures 2 and 3). On one hand, this description of disease is simplistic. On the other hand, this representation has proven to be useful, and no empirical data have demonstrated a reason to abandon this model. In other branches of medicine, the liability to disease, such as blood pressure, body mass index, hemoglobin levels, and bone mineral density, can be directly measured, where individuals at one end of the distribution (as appropriate) are labeled as having hypertension, obesity, anemia, or osteoporosis, respectively. The liability model can be viewed in different ways (Figure 3). There is a binary relationship between phenotypic liability and probability of disease (Figure 3A, B), and while the many genetic and other risks effects work additively to generate the approximately normal distribution of liability to disease in the population (Figure 2), the relationship between genetic liability and disease risk is very nonlinear (Figure 3C–G).

For children of those with disease, it is logical to assume that the threshold in liability associated with disease has the same value. As a result, the liability distribution in first-degree members must be shifted (in the direction of increased liability) compared with the population, to be consistent with the observed higher risk (Figure 2B, C). It is hard to appreciate the importance of the genetic contribution to disease from the measurable risk parameters K and K_1 directly (Figure 1A). Instead, the concept of heritability (Figure 2) is used, which describes how much of the variation between people in the normally distributed liability to disease must be attributed to genetic factors in order to be consistent with the observed risk rates of K and K_1. This concept takes into account the laws of segregation, where each parent passes exactly half of their DNA variants to a child and where each child receives a different sampled half from each of two parents, so that full siblings share, on average, half of their DNA variants, ranging from approximately 40% to approximately 60% (35). The theory and the equations used to calculate heritability from risk to relatives are presented in Supplement 1.

Risk in Other Relatives and Heritability

When epidemiological data are available on only one type of relative (e.g., exclusively parent–offspring pairs or exclusively

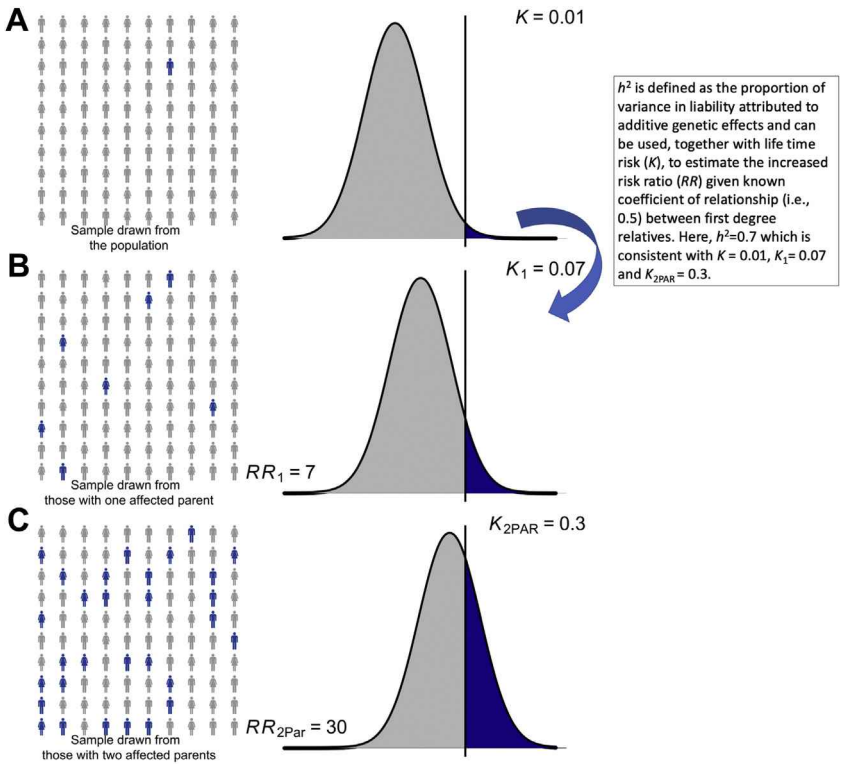

Figure 2. Risk, risk in relatives, and the liability threshold model of common polygenic disease. **(A)** Consider a population where 1 in 100 people are affected by disease (K = 0.01). If the liability of disease is normally distributed, the top 1% of the distribution (colored blue, bisected by the liability threshold) represents individuals with disease. **(B)** Consider the lifetime risk of disease in individuals who have one affected parent (K_1). In this example, 7 in 100 are affected. Hence, under the same diagnosis criteria, the liability distribution has shifted to the right. The top 7% of the distribution is colored blue. **(C)** Consider the lifetime risk of disease in individuals who have two affected parents (K_{2PAR}). In this example, 30 in 100 are affected. Hence, under the same diagnostic criteria as in the general population, the liability distribution has shifted far to the right. The top 30% of the distribution is colored blue. The risk ratio values for those with one parent affected (RR_1) and for those with two parents affected (RR_{2PAR}) were selected to be consistent with a heritability (h^2) of 0.7, so they are approximately representative of schizophrenia, bipolar disorder, and autism spectrum disorder. To generate similar graphs for other disorders, use the CHARRGe Shiny application (shiny.cnsgenomics.com/CHARRGe).

full-sibling pairs), it is possible that the estimates of heritability made from the two measurable pieces of information (i.e., K and K_1) could be inflated if some of the increased risk in relatives is attributable to nongenetic factors (such as family household, economic factors, stress) associated with having a close relative with a disorder. If epidemiological data are available on different types of family members, then these data points can be used to disentangle the contributions to variation in liability that might be attributed to genetic factors or to common environmental factors. This is a motivation for twin studies where probands are selected because they have a twin, who could be a monozygotic (MZ) or dizygotic (DZ) twin, from whom the risk rates (K_{MZ} and K_{DZ}) can be measured and compared to the baseline risk (K) calculated from a population sample. Since both types of twins likely grew up together in their shared households, and since MZ twins share all their DNA variants while DZ twins share only (on average) half of their DNA variants, in principle, the contribution from the shared genetic and household factors can be separated. In practice, it is difficult to achieve large enough samples of twins to make accurate estimates of heritability for disease traits (36,37) (Figure S1 in Supplement 2).

For schizophrenia and bipolar disorder, data sets have been collected that provide estimates of risk rates from different types of relative pairs (Figure 4). A Swedish registry (7) study (9 million individuals, 2 million nuclear families, approximately 36,000 individuals with schizophrenia, approximately 40,000 individuals with bipolar disorder) could identify parent–offspring, full and half siblings, and adoptees. Resulting estimates of the proportion of variance in liability attributable to shared environment of nuclear families were significant for both schizophrenia (4.5%, 95% confidence interval [CI] 4.4%–7.4%) and bipolar disorder (3.4%, 95% CI 2.3%–5.5%) but small relative to the genetic contribution, with heritabilities of 64% (95% CI 62%–68%) and 59% (95% CI 56%–62%), respectively (20). A greatly increased risk of a psychiatric disorder is reported for persons who have both parents affected by the disorder (38). Although environmental factors may contribute to the risk, very high rates of disease are expected, predicted from the liability threshold model. So, for a disease of lifetime risk 1% and heritability 70%, under a polygenic liability model of disease, while only 7% of the individuals who have one parent with the disease are expected to have the disease, 30% of the individuals who have two parents with the disease are expected to have the disease themselves (Figure 2B, C).

We have described the estimation of heritability through the measurable risk in relatives. Lichtenstein *et al.* estimated heritability in a linear mixed model framework directly using the disease-status observations (20). However, the methods can be shown to give similar estimates (39). We note that we focus on the narrow-sense heritability, which considers only additive genetic factors on the liability scale, which as discussed do act nonadditively for disease risk (i.e., observed scale) (Figure 3). Since the empirical evidence is that psychiatric disorders are highly polygenic, with individual variants having small effects, nonadditive genetic effects on the liability scale are unlikely [see (40,41) for further discussion].

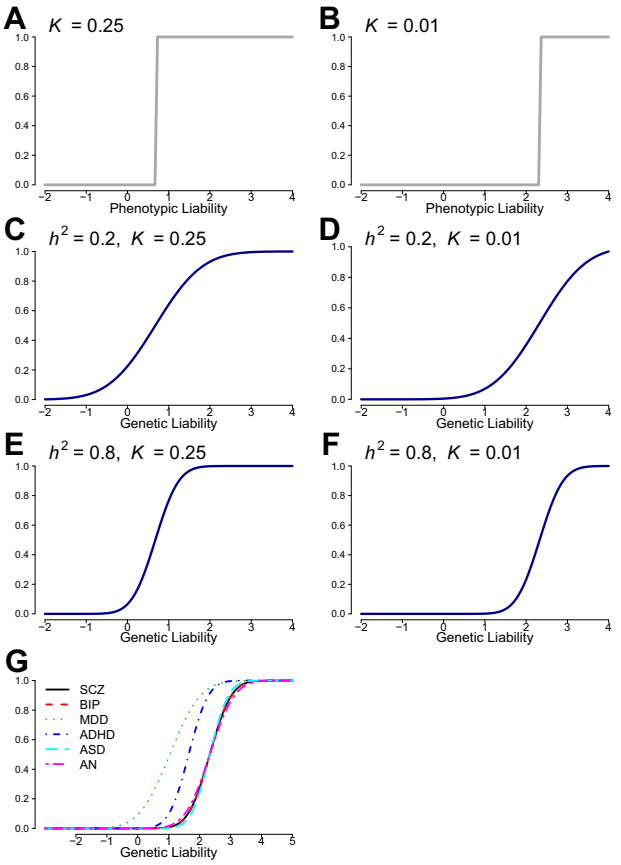

Figure 3. Different views of the liability threshold distribution. Although the liability threshold model is usually portrayed as a normal distribution of liability, with liability comprising many genetic and nongenetic risks (Figure 2), here the model is described by the relationship between liability (x-axis, standard deviation units) and probability of disease (y-axis) for disorder of lifetime risk of **(A, C, E)** $K = 0.25$ and **(B, D, F)** $K = 0.01$. Panels **(A, B)** show the relationship between phenotypic liability (x-axis) and risk/probability of disease. Individuals either have (probability of disease = 1) or do not have (probability of disease = 0) disease, and the vertical line corresponds to the threshold that bisects the x-axis at a point that corresponds to the normal distribution threshold of the risk of disease (as in Figure 2A) defined by K. Panels **(C–G)** show that the relationship between genetic liability (x-axis) and risk/probability of disease is very nonlinear. The gradient of the relationship is steeper when the heritability (h^2) is higher, as in panels **(E, F)** vs. panels **(C, D)**. A nonzero probability of disease is seen at a lower genetic liability for more common diseases (higher K), as in panels **(C, E)** vs. panels **(D, F)**. Hence, the liability threshold model describes risk distributions relative to genetic burden (liability) in only two parameters, K and h^2, and the normal distribution is mathematically very tractable to use for making predictions, hence its utility as a model. Panel **(G)** depicts risk curves using the estimates of K and h^2 (Figure 1, and Table S1 in Supplement 2) for six psychiatric disorders. ADHD, attention-deficit/hyperactivity disorder; AN, anorexia nervosa; ASD, autism spectrum disorder; BIP, bipolar disorder; MDD, major depressive disorder; SCZ, schizophrenia.

SNP-Based Heritability

Heritability is the proportion of the variation in liability to a disease that can be attributed to genetic factors and is calculated based on inference of genetic factors shared between relatives. Hence, genetic factors that are unique to an individual but that affect disease status for that individual (i.e., de novo mutations) would be partitioned into the residual (difference between total and genetic) variance and do not contribute to heritability. Other than these unique genetic factors, the genetic factors shared between relatives represent DNA variants that are both common and rare in the population. In contrast, SNP-based heritability (h^2_{SNP}) represents only the proportion of variance attributable to common DNA variants of frequency, typically $\geq 1\%$. Hence, h^2_{SNP} is, by definition, lower than heritability estimated from family and twin study designs. The ratio of SNP-based to total heritability provides a description of the relative importance of common variants to the genetic architecture, which could differ between disorders (42) (Figure 1B). For quantitative traits such as height, there is good evidence (43) that the difference between SNP-based heritability and heritability reflects contributions from rare and less common variants not assessed in or poorly correlated with (i.e., in low linkage disequilibrium [LD]) the common DNA variants used. For disease traits, additional factors such as phenotype definition or technical artifacts that are more likely to correlate with a binary trait, compared with a quantitative trait, may also play a role.

h^2_{SNP} is estimated from GWAS data, either directly from individual-level genotype data (44,45) or from GWAS summary statistics (46–48). In the approach that uses individual-level data, such as genomic relatedness–based restricted maximum-likelihood (GREML) (44,45), GWAS samples are selected so that there are no close family members either within or between the cases and controls (as including these could inflate the estimate of the contribution from common DNA variants). Hence, compared with genetic epidemiology studies (which have to focus on collection of

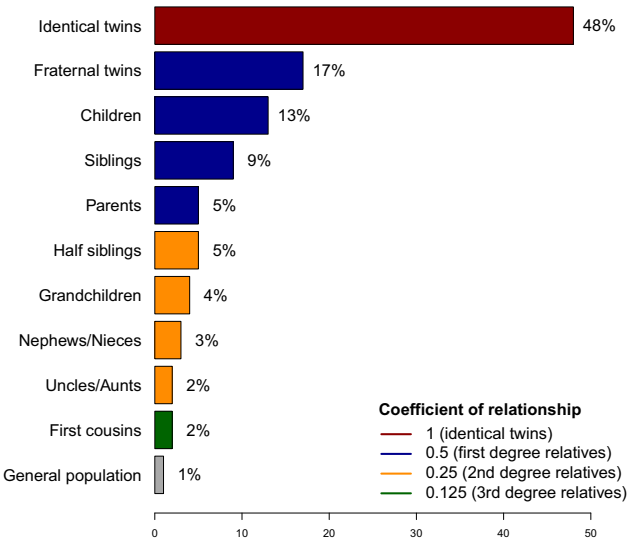

Figure 4. Estimate of risk to relatives for different types of family members. Results based on Gottesman (74). A disorder with lifetime risk of 1% indicates that roughly 1 in every 100 individuals will develop the disorder. However, increased risks are observed in relatives (of different degrees) of those affected. Risk is highest if a person has an identical twin with a disorder (indicated in red) and lowest in those with third-degree relatives with a disorder (indicated in green).

data from families), SNP-based heritability estimates are easy to obtain, given that they are a by-product of GWASs that have already collected large samples. In practice, owing to privacy concerns, it can be difficult to access individual-level genotype data; moreover, the computational burden of the methods is high and increases nonlinearly with the sample size. In order to maximize sample sizes for GWASs, most disease GWASs, including those for psychiatric disorders, are based on meta-analyses of GWAS summary statistics, as the summary statistics can be shared easily across sites. LD score regression (LDSC) (46) was the first [of now many methods (45,49–53)] to estimate h^2_{SNP} from GWAS summary statistics. The methods are described in more detail in Supplement 1. The methods for estimation of SNP-based heritability can differ in underlying assumptions made, and a comprehensive evaluation of methods (54) provides some guidelines. However, simulations show that the best choice of method depends on the simulated genetic architecture (54), which in real analyses is specific to a trait and often remains unknown. Notably, the simulations show that LDSC estimates of SNP-based heritability are generally biased downward; we recognize that the primary LDSC article (46) did not dwell on the estimation of SNP-based heritability (focusing instead on issues of residual population stratification), likely because bias in estimates was recognized. Nonetheless, it is an intensely applied method because of its light computational burden and the widespread availability of GWAS summary statistic data (55) and software (LDHub). Bayesian framework methods SBayesR (47) and SBayesS (48) [applied to GWAS summary statistics and published after, and hence not included in, the Evans *et al.* (54) comparison of methods] make fewer assumptions about the distribution of SNP effects. Instead, they use the GWAS results to infer the genetic architecture and so are optimized for a broader range of underlying genetic architectures. These methods applied to GWAS summary statistics provided estimates that agreed well with GREML estimates and had small standard errors (14,47,48). Since these methods are relatively new, we estimated the SNP-based heritability of all major psychiatric disorders using publicly available GWAS summary statistics (https://www.med.unc.edu/pgc/download-results/) using both LDSC and SBayesS (Figure 1; Table S6 in Supplement 2). As reported from simulated data, we found that for most traits SBayesS provided higher estimates, and for all traits the standard errors were lower, compared with those of LDSC. In theory, SNP-based heritability estimates should be unbiased, such that increases in sample size impact only the precision (i.e., the standard error of the estimates decrease with increasing sample size). However, a noteworthy empirical observation is that SNP-based heritability estimates tend to decrease as sample sizes increase, because of the inclusion of multiple cohorts (Figure S2 in Supplement 2). Careful consideration of the scale of estimates is required in genetic studies of disease traits, since transformations must be applied to the estimates made directly from data to achieve estimates on the liability scale (56–59), which is a scale that is interpretable across studies (see Supplements 1 and 2 for more details).

CROSS-DISORDER RISK IN RELATIVES, GENETIC CORRELATION, AND SNP-BASED CORRELATION

Genetic Relationships Between Psychiatric Disorders

Just as epidemiological studies can investigate the increased risk of disorder A in relatives of those with disorder A, they also collect the data to estimate the increased risk of disorder B in relatives of those with disorder A. However, since studies of a single disorder are difficult to conduct, those collecting data on two diseases in the same families are even more difficult to achieve. Bivariate extensions of the methods used to estimate heritability [see (23,39) and Supplement 1 for details] can be used to estimate genetic correlation (r_g) between traits. Given the complexities of collecting genetically informative epidemiological data, estimates of genetic correlations between psychiatric disorders are not commonly presented, although there are more estimates presented of cross-disorder risk in relatives. These cross-disorder risks can be difficult to benchmark. For example, the increased risk of major depressive disorder in first-degree relatives of those with schizophrenia estimated from a meta-analysis of 11 family studies seems modest at 1.5 (95% CI 1.2–1.8) (17), which is perceived as low. However, it can be shown (Supplement 1 and Figure 5) that this risk ratio implies a sizeable genetic correlation of about 0.34 (13). The national registry data of Sweden (20) and Denmark (38,39) provide cross-disorder risk rates between schizophrenia and bipolar disorder. Since genetic correlations require the estimation of three parameters, the standard errors of estimates are larger for the same sample size than estimates of heritability, i.e., the numerator is the estimate of the genetic covariance ($\widehat{\sigma}_{g_x.g_y}$) between the traits x and y, while the denominator includes estimates of the genetic variances ($\widehat{\sigma}^2_{g_x}$ and $\widehat{\sigma}^2_{g_y}$) of the two traits, $r_g = \widehat{\sigma}_{g_x.g_y} / \sqrt{\widehat{\sigma}^2_{g_x}\widehat{\sigma}^2_{g_y}}$; equivalently using the liability distribution (where phenotypic variances of both traits are 1) $r_g = h_{xy}/\sqrt{h^2_x h^2_y}$, where h^2_x and h^2_y are heritability estimates and h_{xy} the co-heritability. It can be informative to benchmark genetic correlation together with the trait heritabilities. For example, a high genetic correlation in the context of high heritabilities may be more meaningful than a high genetic correlation in the context of low heritabilities.

SNP-Based Genetic Correlation

Bivariate extensions of both the GREML (60) and LDSC (61) methods allow estimation of a SNP-based genetic correlation from GWAS data sets that have been collected independently for the two traits (62). To estimate r_g from independently collected data sets, and as a by-product of GWASs, has been an important advance of the past decade. In essence, bivariate GREML detects whether cases of the two diseases are significantly more similar genetically than they are to controls (or significantly less similar in the case of negative correlation), and LDSC detects whether SNP associations from the two GWAS are nonrandom. Notably, genetic correlation estimates are scale independent (no transformations are needed) and are more robust to the assumptions made in different methods compared with heritability estimates (61,62). The first

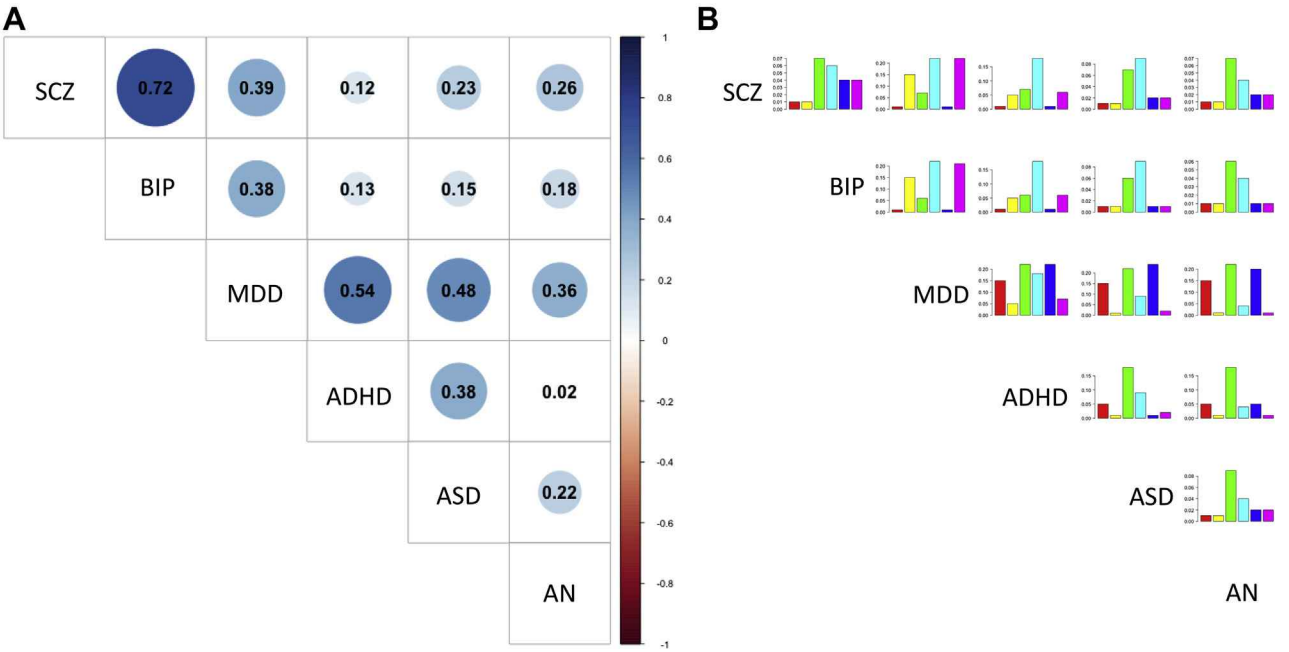

Figure 5. Genetic correlations and cross-disorder risk in relatives. **(A)** The genetic correlations (r_g) between two disorders, disorder *x* (row) and disorder *y* (column), are estimated from linkage disequilibrium score regression using genome-wide association study summary statistics (see Table S1 in Supplement 2 for references). We assume that genetic correlations are the same across the allelic spectrum. The diagonal elements are entered as 1, because this is the expectation for samples drawn from the same population. In real data analyses, these correlations have been estimated as <1 (see Figure S2 in Supplement 2). **(B)** We use lifetime risk (K_x and K_y) and total heritabilities (h^2_x and h^2_y) as reported in Table S1 in Supplement 2. We back-calculate the expected risks in first-degree relatives, within and between disorders from K_x, K_y, h^2_x, h^2_y, and r_g (see Supplement 1). Red represents lifetime risk of disorder *x* (K_x), yellow represents lifetime risk of disorder *y* (K_y), green represents lifetime risk of disorder *x* in first-degree relatives of those with disorder *x*, light blue represents lifetime risk of disorder *y* in first-degree relatives of those with disorder *y*, dark blue represents lifetime risk of disorder *y* in first-degree relatives of those with disorder *x*, and purple represents lifetime risk of disorder *x* in first-degree relatives of those with disorder *y*. ADHD, attention-deficit/hyperactivity disorder; AN, anorexia nervosa; ASD, autism spectrum disorder; BIP, bipolar disorder; MDD, major depressive disorder; SCZ, schizophrenia.

estimates of SNP-based genetic correlations between diseases was applied to psychiatric disorders (17). With the ever-increasing availability of GWAS data sets, r_g has now been reported for many trait-pair combinations, which would have been deemed impossible from traditional genetic epidemiology (63). It is open to debate whether the r_g estimated from the common SNPs in GWAS data is the same as the r_g when estimated across the full allele frequency spectrum. In the absence of contradicting data, it seems like a reasonable first assumption. Where r_g has been estimated from both family studies and GWAS data in the context of psychiatric disorders, estimates agree well. For example, a large-scale Swedish family and adoption study (20) estimated the genetic correlation between schizophrenia and bipolar disorder to be 0.6, which is in high concordance with what is found using genome-wide SNP data (13,17).

An important question is the impact of disease misdiagnosis on estimates of genetic correlation. This is particularly important in psychiatry, where the stable diagnosis of a person may differ from the diagnosis at first presentation. These concerns impact equally on genetic epidemiology studies estimates as on SNP-based estimates, but it is only in the GWAS era that estimation of genetic correlation has become commonplace. This question was studied prior to the publication of the first r_g estimates for psychiatric disorders and showed that misdiagnosis can indeed generate upwardly biased estimates of r_g (64). However, reported rates of change in diagnosis (e.g., approximately 15% of subjects initially diagnosed with bipolar disorder may later receive a stable diagnosis of schizophrenia) were found to inflate the genetic correlation by approximately 0.1 only and could not alone explain the higher genetic correlation that is generally observed (64). Moreover, in the context of a true high genetic correlation, people may be expected to present with symptoms consistent with both disorders. Other ascertainment biases are discussed in our genetic correlation review (62), but as a rule of thumb, ascertainment impacts both the genetic covariance (numerator) and genetic variances (denominator) of the r_g calculation, which makes them relatively robust to these type of biases.

SNP-Based Genetic Correlations Between Data Sets of the Same Disorder Including Across Ancestry

SNP-based r_g can be calculated between two independently collected data sets for the same disorder. In this case, the r_g is expected to be 1 because the expected values of the two SNP-based heritability estimates, as well as those of the SNP-based co-heritability, are the same. In practice, the r_g estimates are usually found to be less than 1, and the SNP-based heritability estimates are observed to differ more than expected by the standard error estimates. Together these results

imply that the nature of the GWAS data does not meet the assumptions of random samples from the same population. This finding could reflect differences in phenotype definition, population differences, or differences in technical factors. This is important to note since estimation of r_g has been used as a strategy to investigate the genetic relationship between different phenotype definitions, a topic of importance in psychiatry (65). While appealing in principle, such approaches need to be benchmarked against genetic correlations estimated between different data sets that have measured the same phenotype (Figure S2 in Supplement 2).

SNP-based correlations can also be estimated from the same disorder but from samples collected from different ancestries. The Popcorn method (66) is an extension of LDSC that attempts to model the differences in allele frequency and LD between ancestries to estimate cross-ancestry r_g. For instance, the genetic correlation for schizophrenia between East Asian and European ancestries was estimated as r_g = .98, SE .03 (67). When estimates of r_g between ancestries are estimated to be less than 1, it is important that the estimate is benchmarked against that calculated from two data sets of the same disorder from the same ancestry to ensure that r_g values are interpreted with recognition that factors other than ancestry can generate r_g values less than 1.

CONCLUSIONS

In this capstone narrative, we bring together the methods and results that summarize the genetic contribution to psychiatric disorders and the genetic relationship between them. We note that we use the common assumption that psychiatric disorder diagnosis definitions are underpinned by a consistent polygenic biology. If this is not true—for example, if a single clinical diagnosis is allocated to one or more independent or correlated biological diseases—then further thought is needed to interpret the estimates of heritability and genetic correlation. Such a scenario could explain (41), in part, the large difference between heritability and SNP-based heritability (Figure 1) in addition to contributions from rare variants and low LD between genotyped and causal variants. Previously, we concluded that only with large GWAS sample sizes and extensive clinical data (40,41) would we have the information needed to examine this interesting question. Despite this caveat, multiple results from GWAS data confirm that individuals allocated a specific diagnosis are genetically more similar, on average, than those allocated other diagnoses (i.e., heritabilities of individual disorders are greater than co-heritabilities between disorders) (Figure S2 in Supplement 2).

Understanding the genetic contribution to common disease is a foundation for many other research directions. It is outside the scope of this review to focus on the utility of the estimates of heritability and genetic correlation in detail. Estimates of SNP-based heritability help to guide whether efforts to increase GWAS sample sizes should continue, as they provide an upper limit on the combined effects of individual associated loci. Estimates of heritability and SNP-based heritability provide guidelines of maximum future accuracy of risk prediction applied to people whose disease status is not yet known. Genetic correlations can be used to determine how much the accuracy of the risk prediction can be improved by drawing on information from correlated traits, which perhaps are available in much larger samples than for the primary disorder itself (68). Here, we have focused on genetic correlations between psychiatric disorders, an approach that is likely to reflect pleiotropy (same causal variants affecting more than one disorder). However, genetic correlations can also be estimated between psychiatric disorders and other common diseases, or between psychiatric disorders and traits measurable in the population (such as educational attainment or smoking status), and these estimates could reflect causal relationships, which have been long-discussed in the psychiatric epidemiology literature (69). In the past 5 years, results from GWASs have allowed causal relationships using putative exposure traits and psychiatric disorders to be explored, as well as those between psychiatric disorders and subsequent metabolic disease, using the Mendelian randomization approach. The application of Mendelian randomization to psychiatric disorders has been discussed elsewhere (70) and is an exciting tool in psychiatry (as long as studies are well powered) to investigate putative causal relationships that are impossible or unethical to address through clinical trials. As an example, we recently showed that although there is considerable pleiotropy between genetic variants for vitamin D and psychiatric disorders, there is no evidence of a causal relationship (71). Such analyses contribute hard data to a long discussion in psychiatric epidemiology (72,73). Finally, we hope that our Supplementary materials, including Rmarkdown script and CHARRGe Shiny application (https://shiny.cnsgenomics.com/CHARRGe/), are useful to others both in research and as teaching and learning aids.

ACKNOWLEDGMENTS AND DISCLOSURES

This work was supported by the National Health and Medical Research Council (Grant Nos. 1078901, 1087889, 1113400, and 1151854 [to NRW]), ZonMw Rubicon (Grant No. 45219101 [to BMLB]), and the European Union Joint Programme Neurodegenerative Disease Biological Resource Analysis to Identify New Mechanisms and Phenotypes in Neurodegenerative Diseases (JPND BRAIN-MEND) project (to NRW and WvR). The fundamentals of Supplement 1 were generated as teaching material for a course given at the Summer Institute of Statistical Genetics, Seattle, WA, 2015–2018, and improved and built upon by BMLB and WvR.

The authors report no biomedical financial interests or potential conflicts of interest.

ARTICLE INFORMATION

From the Institute for Molecular Bioscience (BMLB, LY, NRW) and Queensland Brain Institute (NRW), The University of Queensland, Brisbane, Queensland, Australia; and Department of Neurology (WvR), Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, Netherlands.

Address correspondence to Naomi R. Wray, Ph.D., University of Queensland, Room 432, Carmody Road, Brisbane 4072, Australia; E-mail: naomi.wray@uq.edu.au.

Received Mar 6, 2020; revised May 14, 2020; accepted May 27, 2020.

Supplementary material cited in this article is available online at https://doi.org/10.1016/j.biopsych.2020.05.034.

The CHARRGe Shiny application is available at https://shiny.cnsgenomics.com/CHARRGe/.

Code for Rmarkdown and CHARRGe is shared at https://github.com/BartBaselmans/CHARRGe.

REFERENCES

1. Bleuler E (1917): Mendelismus bei Psychosen, speziell bei der Schizophrenie [Mendelism in the psychoses, especially schizophrenia]. Arch Neurol Psychiatr 1:19–40.

2. Kendler KS (2020): Eugen Bleuler's views on the genetics of schizophrenia in 1917. Schizophr Bull 1–7.
3. Rüdin E (1916): Studien uber Vererbung und entstehung geistiger Störungen. I. Zur vererbung und neuentstehung der Dementia praecox [Studies on the inheritance and origin of mental illness. I. The problem of the inheritance and primary origin of dementia praecox] Monographien aus dem Gesamtgebiet der Neurologie und Psychiatrie, vol. 12. Berlin: Springer.
4. Kendler KS, Zerbin-Rüdin E (1996): Abstract and review of "Studien über Vererbung und Entstehung Geistiger Störungen. I. Zur Vererbung und Neuentstehung der Dementia Praecox." (Studies on the inheritance and origin of mental illness: I. To the problem of the inheritance and primary origin of dementia praecox). Am J Med Genet 67:338–342.
5. Mortensen PB, Pedersen MG, Pedersen CB (2010): Psychiatric family history and schizophrenia risk in Denmark: Which mental disorders are relevant? Psychol Med 40:201–210.
6. Bøcker C, Pedersen C (2006): The Danish Civil Registration System. A cohort of eight million persons. Dan Med Bull 53:441–449.
7. Ekbom A (2011): The Swedish multi-generation register. Methods Mol Biol 675:215–220.
8. Hilker R, Helenius D, Fagerlund B, Skytthe A, Christensen K, Werge TM, *et al.* (2018): Heritability of schizophrenia and schizophrenia spectrum based on the nationwide Danish Twin Register. Biol Psychiatry 83:492–498.
9. Nordsletten AE, Brander G, Larsson H, Lichtenstein P, Crowley JJ, Sullivan PF, *et al.* (2019): Evaluating the impact of nonrandom mating: Psychiatric outcomes among the offspring of pairs diagnosed with schizophrenia and bipolar disorder. Biol Psychiatry 87:253–262.
10. Cederlöf M, Thornton LM, Baker J, Lichtenstein P, Larsson H, Rück C, *et al.* (2015): Etiological overlap between obsessive-compulsive disorder and anorexia nervosa: A longitudinal cohort, multigenerational family and twin study. World Psychiatry 14:333–338.
11. Lichtenstein P, Carlström E, Råstam M, Gillberg C, Anckarsäter H (2010): The genetics of autism spectrum disorders and related neuropsychiatric disorders in childhood. Am J Psychiatry 167:1357–1363.
12. Stahl EA, Breen G, Forstner AJ, McQuillin A, Ripke S, Trubetskoy V, *et al.* (2019): Genome-wide association study identifies 30 loci associated with bipolar disorder. Nat Genet 51:793–803.
13. Cross-Disorder Group of the Psychiatric Genomics Consortium (2019): Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. Cell 179:1469–1482, e11.
14. Zeng J, De Vlaming R, Wu Y, Robinson MR, Lloyd-Jones LR, Yengo L, *et al.* (2018): Signatures of negative selection in the genetic architecture of human complex traits. Nat Genet 50:746–753.
15. Ripke S, Neale BM, Corvin A, Walters JTR, Farh KH, Holmans PA, *et al.* (2014): Biological insights from 108 schizophrenia-associated genetic loci. Nature 511:421–427.
16. Baker JH, Schaumberg K, Munn-Chernoff MA (2017): Genetics of anorexia nervosa. Curr Psychiatry Rep 19:84.
17. Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM, Perlis RH, *et al.* (2013): Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat Genet 45:984–994.
18. Nordsletten AE, Larsson H, Crowley JJ, Almqvist C, Lichtenstein P, Mataix-Cols D (2016): Patterns of nonrandom mating within and across 11 major psychiatric disorders. JAMA Psychiatry 73:354–361.
19. Van Snellenberg JX, De Candia T (2009): Meta-analytic evidence for familial coaggregation of schizophrenia and bipolar disorder. Arch Gen Psychiatry 66:748–755.
20. Lichtenstein P, Yip BH, Björk C, Pawitan Y, Cannon TD, Sullivan PF, Hultman CM (2009): Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: A population-based study. Lancet 373:234–239.
21. Sullivan PF, Magnusson C, Reichenberg A, Boman M, Dalman C, Davidson M, *et al.* (2012): Family history of schizophrenia and bipolar disorder as risk factors for autism. Arch Gen Psychiatry 69:1099–1103.
22. Harper PS, Newcombe RG (1992): Age at onset and life table risks in genetic counselling for Huntington's disease. J Med Genet 29:239–242.
23. Falconer DS (1965): The inheritance of liability to certain diseases, estimated from the incidence among relatives. Ann Hum Genet 29:51–76.
24. Pringsheim T, Wiltshire K, Day L, Dykeman J, Steeves T, Jette N (2012): The incidence and prevalence of Huntington's disease: A systematic review and meta-analysis. Mov Disord 27:1083–1091.
25. Holtzman CW, Trotman HD, Goulding SM, Ryan AT, MacDonald AN, Shapiro DI, *et al.* (2013): Stress and neurodevelopmental processes in the emergence of psychosis. Neuroscience 249:172–191.
26. Agnew-Blais J, Danese A (2016): Childhood maltreatment and unfavourable clinical outcomes in bipolar disorder: A systematic review and meta-analysis. Lancet Psychiatry 3:342–349.
27. Kendler KS, Karkowski LM, Prescott CA (1999): Causal relationship between stressful life events and the onset of major depression. Am J Psychiatry 156:837–841.
28. Kessler RC, Aguilar-Gaxiola S, Alonso J, Bromet EJ, Gureje O, Karam EG, *et al.* (2018): The associations of earlier trauma exposures and history of mental disorders with PTSD after subsequent traumas. Mol Psychiatry 23:1892–1899.
29. Coleman JRI, Peyrot WJ, Purves KL, Davis KAS, Rayner C, Choi SW, *et al.* (2020): Genome-wide gene-environment analyses of major depressive disorder and reported lifetime traumatic experiences in UK Biobank. Mol Psychiatry 25:1430–1446.
30. Rees E, Han J, Morgan J, Carrera N, Escott-Price V, Pocklington AJ, *et al.* (2020): De novo mutations identified by exome sequencing implicate rare missense variants in SLC6A1 in schizophrenia. Nat Neurosci 23:179–184.
31. Yuen RKC, Merico D, Cao H, Pellecchia G, Alipanahi B, Thiruvahindrapuram B, *et al.* (2016): Genome-wide characteristics of de novo mutations in autism. NPJ Genomic Med 1:160271–1692710.
32. Li J, Cai T, Jiang Y, Chen H, He X, Chen C, *et al.* (2016): Genes with de novo mutations are shared by four neuropsychiatric disorders discovered from NPdenovo database. Mol Psychiatry 21:290–297.
33. Lee JM, Wheeler VC, Chao MJ, Vonsattel JPG, Pinto RM, Lucente D, *et al.* (2015): Identification of genetic factors that modify clinical onset of Huntington's disease. Cell 162:516–526.
34. Reich T, James JW, Morris CA (1972): The use of multiple thresholds in determining the mode of transmission of semi-continuous traits. Ann Hum Genet 36:163–184.
35. Visscher PM, Medland SE, Ferreira MAR, Morley KI, Zhu G, Cornes BK, *et al.* (2006): Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. PLoS Genet 2:0316–0325.
36. Visscher PM, Gordon S, Neale MC (2008): Power of the classical twin design revisited: II detection of common environmental variance. Twin Res Hum Genet 11:48–54.
37. Visscher PM (2004): Power of the classical twin design revisited. Twin Res 7:505–512.
38. Gottesman II, Laursen TM, Bertelsen A, Mortensen PB (2010): Severe mental disorders in offspring with 2 psychiatrically ill parents. Arch Gen Psychiatry 67:252–257.
39. Wray NR, Gottesman II (2012): Using summary data from the Danish National Registers to estimate heritabilities for schizophrenia, bipolar disorder, and major depressive disorder. Front Genet 3:118.
40. Stringer S, Derks EM, Kahn RS, Hill WG, Wray NR (2013): Assumptions and properties of limiting pathway models for analysis of epistasis in complex traits. PLoS One 8:e68913.
41. Wray NR, Maier R (2014): Genetic basis of complex genetic disease: The contribution of disease heterogeneity to missing heritability. Curr Epidemiol Rep 1:220–227.
42. Yang J, Zeng J, Goddard ME, Wray NR, Visscher PM (2017): Concepts, estimation and interpretation of SNP-based heritability. Nat Genet 49:1304–1310.
43. Wainschtein P, Jain DP, Yengo L, Zheng Z, Cupples LA, Shadyab AH, *et al.* (2019): Recovery of trait heritability from whole genome sequence data. bioRxiv. https://doi.org/10.1101/588020.
44. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, *et al.* (2010): Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42:565–569.
45. Yang J, Lee SH, Goddard ME, Visscher PM (2011): GCTA: A tool for genome-wide complex trait analysis. Am J Hum Genet 88:76–82.

46. Bulik-Sullivan B, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, *et al.* (2015): LD score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet 47:291–295.
47. Lloyd-Jones LR, Zeng J, Sidorenko J, Yengo L, Moser G, Kemper KE, *et al.* (2019): Improved polygenic prediction by Bayesian multiple regression on summary statistics. Nat Commun 10:5086.
48. Zeng J, Xue A, Jiang L, Lloyd-Jones LR, Wu Y, Wang H, *et al.* (2019): Bayesian analysis of GWAS summary data reveals differential signatures of natural selection across human complex traits and functional genomic categories. bioRxiv. https://doi.org/10.1101/752527.
49. Yang J, Bakshi A, Zhu Z, Hemani G, Vinkhuyzen AAE, Lee SH, *et al.* (2015): Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat Genet 47:1114–1120.
50. Speed D, Hemani G, Johnson MR, Balding DJ (2012): Improved heritability estimation from genome-wide SNPs. Am J Hum Genet 91:1011–1021.
51. Yang J, Manolio TA, Pasquale LR, Boerwinkle E, Caporaso N, Cunningham JM, *et al.* (2011): Genome partitioning of genetic variation for complex traits using common SNPs. Nat Genet 43:519–525.
52. Speed D, Cai N, Johnson MR, Nejentsev S, Balding DJ (2017): Reevaluation of SNP heritability in complex human traits. Nat Genet 49:986–992.
53. Zaitlen N, Kraft P, Patterson N, Pasaniuc B, Bhatia G, Pollack S, Price AL (2013): Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. PLoS Genet 9: e1003520.
54. Evans LM, Tahmasbi R, Vrieze SI, Abecasis GR, Das S, Gazal S, *et al.* (2018): Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits. Nat Genet 50:737–745.
55. Zheng J, Erzurumluoglu AM, Elsworth BL, Kemp JP, Howe L, Haycock PC, *et al.* (2017): LD Hub: A centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. Bioinformatics 33:272–279.
56. Lee SH, Wray NR, Goddard ME, Visscher PM (2011): Estimating missing heritability for disease from genome-wide association studies. Am J Hum Genet 88:294–305.
57. Yap CX, Sirodenko J, Marioni RE, Yengo L, Wray NR, Visscher PM (2018): Misestimation of heritability and prediction accuracy of male-pattern baldness. Nat Commun 9:2537.
58. Gianola D (1979): Heritability of polychotomous characters. Genetics 93:1051–1055.
59. Peyrot WJ, Boomsma DI, Penninx BWJH, Wray NR (2016): Disease and polygenic architecture: avoid trio design and appropriately account for unscreened control subjects for common disease. Am J Hum Genet 98:382–391.
60. Lee SH, Yang J, Goddard ME, Visscher PM, Wray NR (2012): Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. Bioinformatics 28:2540–2542.
61. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, *et al.* (2015): An atlas of genetic correlations across human diseases and traits. Nat Genet 47:1236–1241.
62. van Rheenen W, Peyrot WJ, Schork AJ, Lee SH, Wray NR (2019): Genetic correlations of polygenic disease traits: From theory to practice. Nat Rev Genet 20:567–581.
63. The Brainstorm Consortium (2018): Analysis of shared heritability in common disorders of the brain. Science 360:eaap8757.
64. Wray NR, Lee SH, Kendler KS (2012): Impact of diagnostic misclassification on estimation of genetic correlations using genome-wide genotypes. Eur J Hum Genet 20:668–674.
65. Cai N, Revez JA, Adams MJ, Andlauer TFM, Breen G, Byrne EM, *et al.* (2020): Minimal phenotyping yields genome-wide association signals of low specificity for major depression. Nat Genet 52:437–447.
66. Brown BC, Ye CJ, Price AL, Zaitlen N (2016): Transethnic genetic-correlation estimates from summary statistics. Am J Hum Genet 99:76–88.
67. Lam M, Chen CY, Li Z, Martin AR, Bryois J, Ma X, *et al.* (2019): Comparative genetic architectures of schizophrenia in East Asian and European populations. Nat Genet 51:1670–1678.
68. Maier R, Moser G, Chen GB, Ripke S, Coryell W, Potash JB, *et al.* (2015): Joint analysis of psychiatric disorders increases accuracy of risk prediction for schizophrenia, bipolar disorder, and major depressive disorder. Am J Hum Genet 96:283–294.
69. Stilo SA, Murray RM (2019): Non-genetic factors in schizophrenia. Curr Psychiatry Rep 21.
70. Smoller JW, Andreassen OA, Edenberg HJ, Faraone SV, Glatt SJ, Kendler KS (2019): Psychiatric genetics and the structure of psychopathology. Mol Psychiatry 24:409–420.
71. Revez JA, Lin T, Qiao Z, Xue A, Holtz Y, Zhu Z, *et al.* (2020): Genome-wide association study identifies 143 loci associated with 25 hydroxyvitamin D concentration. Nat Commun 11:1647.
72. McGrath JJ, Eyles DW, Pedersen CB, Anderson C, Ko P, Burne T, Mortensen P (2010): Neonatal vitamin D status and risk of schizophrenia: A population-based case-control study. Arch Gen Psychiatry 67:889–894.
73. McGrath J, Saari K, Hakko H, Jokelainen J, Jones P, Järvelin MR, *et al.* (2004): Vitamin D supplementation during the first year of life and risk of schizophrenia: A Finnish birth cohort study. Schizophr Res 67:237–245.
74. Gottesman II (1991): A Series of Books in Psychology. Schizophrenia Genesis: The Origins of Madness. New York: WH Freeman/Times Books/Henry Holt & Co.