

Polygenic Scores in Epidemiology: Risk Prediction, Etiology, and Clinical Utility

Brion S. Maher¹

Published online: 28 September 2015
© Springer International Publishing AG 2015

Abstract Genes account for a significant proportion of the risk for most common diseases. The genome-wide association scan (GWAS) era of genetic epidemiology has generated a massive amount of data, revolutionized our thinking on the genetic architecture of common diseases and positioned the field to realistically consider risk prediction for common polygenic diseases, such as non-familial cancers, and autoimmune, cardiovascular, and psychiatric diseases. Polygenic scoring is an approach that shows promise for understanding the polygenic contribution to common human diseases. This is an approach typically relying on genome-wide SNP data, where a set of SNPs identified in a discovery GWAS are used to construct composite polygenic scores. These scores are then used in additional samples for association testing or risk prediction. This review summarizes the extant literature on the use, power, and accuracy of polygenic scores in studies of the etiology of disease and the promise for disease risk prediction.

Keywords Polygenic · Genetic risk · Polygenic risk · Polygenic scores · GWAS

Introduction

Roughly a century ago, RA Fisher proved that familial resemblance could be due to many factors inherited in a Mendelian

fashion [1]. Fisher's publication not only resolved a key difference between the Mendelians and biometricians, who believed that discrete genes could not underlie quantitative traits, but also led to many core statistical concepts used today. In the decades that followed that publication, it became widely accepted that the substantial heritability observed for many human traits was likely due to the composite influence of many genes of very small effect. Despite this, even among the most ardent proponents of Fisher's work, there was belief that Mendelian analysis (e.g., linkage analysis, segregation analysis) was a "preferable" form of scientific inquiry especially where variation was not continuous and "individually traceable genes" were available [2]. The notion that the Mendelian approach, which looks at cosegregation of a single specific measurable genetic locus and a discrete trait, was more powerful than a biometric approach or that the biometric approach would be "more tedious and cumbersome to use than Mendelian analysis" for discrete traits stood for many years [2]. The advances in the computer implementation of Mendelian statistical methods [3, 4], the growing availability of polymorphic markers across the genome [5], and the focus of human genetics on clinically defined diseases all further encouraged the use of single locus methods. These approaches have experienced a recent resurgence, again largely due to the availability of novel technology that allows for the interrogation of the whole genome [6]. While linkage analysis was very successful in identifying single gene contributions to Mendelian disorders, the approach has been much less successful in disentangling the genetic risk for more common genetically complex disorders [7]. Consequently, the history of human disease gene mapping has best been characterized as a series of fits and starts largely focused on attempting to identify single detectable genetic contributions.

The perceived general lack of success of linkage analysis in identifying complex disease risk variants, coupled with the

This article is part of the Topical Collection on *Genetic Epidemiology*

✉ Brion S. Maher
brion@jhu.edu

¹ Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, 850 Hampton House 624 North Broadway, Baltimore, MD 21205, USA

technological advances that allowed for the development of large genotyping arrays have made genome-wide association studies (GWAS) commonplace in disease gene mapping over the past decade. GWAS have identified thousands of variants for hundreds of diseases that while detectable and replicable at stringent multiple test corrected levels explain only a very small amount of the genetic or phenotypic variance [8]. Initial work on genetic risk prediction showed that these loci, even when taken together within a given disease, appeared to provide very little improvement in risk prediction over known clinical indicators [9] and account for a very modest amount of disease or trait heritability. While this “lost” heritability has left some puzzled [10], a reasonable explanation for this is the very small true effect size of each gene contributing to complex disease. The stringent statistical significance criterion of GWAS has resulted in relatively few of the modest effect size polymorphisms being deemed “significant” and “replicated.” It has been shown that the expected rank of realistic effect sizes would not likely be among the top findings in a typical GWAS [11]. Thus, many associated loci will not be deemed “significant” based on a single, large genome-wide study. However, there is an expectation that the SNPs in the upper tail of the distribution of test statistics from a given GWAS will be enriched with true signals.

Purcell and colleagues [12] used an approach in an early GWAS of schizophrenia, a moderately rare disease (estimated lifetime prevalence, K , of 1 %) of high heritability (80 %) [13], that steps well beyond single locus testing. In this approach, two-stage GWAS data were used to select a set of “independent” SNPs in linkage equilibrium that generated p values below some arbitrary threshold (P_T) in one sample as a discovery stage. Those SNPs were then used to create polygenic sum scores, with each allele weighted by the logarithm of the odds ratio from the discovery sample to be tested in a second sample. Case-control differences in polygenic score were then tested using logistic regression. They found highly significant differences in polygenic sum score between schizophrenic cases and controls at P_T across a range from 0.05 to 0.5. In addition, they found that SNPs selected from a stage 1 schizophrenia sample yielded significant case-control polygenic score differences in a bipolar disorder sample, possibly indicating a high degree of genetic overlap between the two diseases. This method approaches a true Fisherian approach by attempting to index all possible influential variation across the genome and literally allowing the statistically non-significant true signals to stand up and be counted.

The terms “polygenic scores” (PGS), “genetic risk scores” (GRS), and “polygenic risk scores” (PRS) are used interchangeably to describe metrics comprising a large number of SNPs pooled together to represent a measured set of variants underlying a particular trait or disease. The discussion herein

will make no distinction between application to diseases and normally distributed traits (e.g., height), and hence, the term “polygenic scores” will be used.

The application of PGS can be grouped into two broad categories: (1) exploration of the genetic contribution to the etiology and (2) prediction of individual disease risk or trait outcome.

Etiology

Dudbridge [14•] examined the power and predictive accuracy of polygenic scores for discrete and continuous traits, deriving analytic expressions as a function of the total heritability of the trait, the proportion of that variance explained by the measured marker panel, the total number of markers, the proportion of those markers with no effect on the trait, the total sample size, the proportions of the sample used for training and testing, and the p value threshold for selection of markers from the training sample for subsequent association testing or risk prediction. The primary finding was that a roughly even split between training and testing samples yielded the best power for association testing but the use of a larger training set increased the precision of prediction. In addition, fewer null markers at a fixed total marker number and heritability led to worse predictive power. While this result seems counterintuitive, fewer null markers lead to a decrease in the average effect per true marker, thus leading to poorer power to discriminate true from null effects in the training sample.

An important point while considering or comparing the findings of investigations of the power or precision of PGS is the choice of genetic models. A detailed discussion can be found elsewhere [15]. The generalizability of all theoretical work is limited by the degree to which the assumed distribution of effects reflects reality. Most theoretical studies tend to assume something close to additivity. The key variables influencing the power and precision of PGS are the number and size of effects. That is, the greater the ability to separate true from null effects, the greater the power and precision of PGS.

Marker-Based Heritability

In addition to approaches that create polygenic sum scores, multiple additional approaches have been developed to examine multi- or polygenic contributions to disease. These methods attempt to index what could accurately be termed “marker-based” or “molecular” heritability. The predecessor to this method used genome-wide markers in pedigrees to examine deviation from expected genetic similarity and phenotypic similarity [16]. This method was eventually extended to general populations using genome-wide SNP data and termed “genome-wide complex trait analysis (GCTA)” [17]. In this method, a population-adjusted genetic relationship

matrix is derived from all available SNPs and used in a mixed linear model to estimate via restricted maximum likelihood (REML) the proportion of phenotypic variation accounted for by the relationship matrix. While this method can be applied at the genome, chromosome, or regional level to determine the amount of variance accounted for, it does not identify specifically which variants account for the heritability nor does it specifically facilitate risk prediction. A recent competing method is a generalization of Haseman-Elston regression to population samples, termed phenotype correlation-genotype correlation (PCGC) regression. PCGC generates markedly higher genome-wide heritability estimates than REML-based approaches [18•]. Palla and Dudbridge [19] also extended their previous work [14•] to jointly estimate the number of variants influencing a trait and the variance accounted for by those variants. This eliminates the step of estimating P_T and, since the method uses only the distribution of test statistics, can be applied to GWAS results, without reanalysis of individual-level data.

Recent methods demonstrate that correcting for linkage disequilibrium, as opposed to pruning out correlated (but possibly true independent effect) SNPs increases the heritability estimates from PGS [20]. The authors examined the performance in two large, published GWAS datasets and found modest increases in heritability for schizophrenia (20.1 to 25.3 %) and multiple sclerosis (9.8 to 12.0 %).

An additional recent extension of polygenic scoring relies on haploSNPs [21]. HaploSNPs are large shared segments disrupted only by recombination. After identification, haploSNPs are recoded to 0,1,2 copies and treated as “SNPs” in PGS assessment via the aforementioned PCGC regression approach. HaploSNP heritability estimates jump from 32 to 64 % and from 20 to 67 % versus SNP alone estimates for schizophrenia and MS, respectively.

Risk Prediction

As GWAS initially yielded new genetic signals, Kraft and colleagues discussed quantification of genetic risk in terms of clinical utility [22]. While p values and odds ratios are useful for discovery of novel SNPs influencing disease risk, they are inadequate for describing the potential clinical utility of PGS. Metrics such as the sensitivity, or true positive rate, and the specificity, or true negative rate, provide more useful information regarding the potential clinical utility of a given measure. Functions of sensitivity and specificity are often plotted across the range of a given predictor with the area under that curve (AUC) as a common measure of the potential diagnostic utility of that predictor. AUC, in the context of polygenic scores, can best be interpreted as the probability that the PGS value for a randomly selected case will be higher than that for a randomly selected control. An AUC of 0.5 would indicate a predictor with no better than random

predictive utility. An AUC >0.8 is frequently cited as the standard for a clinically useful diagnostic indicator. Wray and colleagues specifically address the meaning and interpretability of AUC in the context of specific multi-locus models [23].

In early empirical work on genetic risk prediction, Janssens and colleagues [9] demonstrated that inclusion of a modest number of genetic risk loci in predictive models of type 2 diabetes and cardiovascular disease did not improve risk prediction over known environmental risk factors and family history. Jakobsdottir and colleagues [24] found similar results using a handful of replicating GWAS results. These authors conclude that predictive modeling for personalized medicine, which would require sensitive and specific measures of risk, was not yet realistic for most common disorders and diseases. In retrospect, this conclusion was premature due to the relatively small proportion of the genetic variance explained by the SNPs in the predictive models. Using large numbers of loci in simulated data, Wray and colleagues [25] conclude that large screening samples ($>10,000$) will provide adequate power for selection of loci that account for a significant proportion of the genetic variance. Moreover, singling out the segment of the population sample in the upper 5 % of risk allele count identifies a segment of the sample with three to seven times the relative risk for a disease of modest (10–20 %) heritability. One would expect the predictive ability to increase with an increase in heritability. Evans and colleagues [26] subsequently applied a genome-wide approach to the Wellcome Trust Case Control Consortium data sets and found that the approach yielded some predictive value beyond chance although the authors concede that this result could be due to technical artifact or stratification.

Chatterjee et al. [27•] examined the potential predictive utility of PGS derived from genome-wide SNP data. Using the predictive correlation coefficient to describe predictive accuracy, which has a known relationship with AUC ($AUC = \Phi(\sqrt{0.5} R_n)$, where R_n is the predictive correlation coefficient), the authors examine the impact of several factors on prediction accuracy. Overall, the authors' findings are similar to previous work [14•]. That is, factors allowing for greater resolution of true from null effects (e.g., increased heritability at a fixed number of true markers, larger sample sizes) have an impact on the prediction utility of derived models. A noteworthy addition was the examination of multiple effect distributions. Predictive power was greatest when the underlying true SNP effects followed an exponential distribution versus other distributions with more SNPs of smaller effect. An additional important conclusion of this work is the discussion of the inherent limitation of building predictive models from a host of univariate comparisons when the true underlying genetic architecture of most traits or common diseases is likely to be much more complicated. The authors argue for development of methods that better capture complex genetic architectures.

Aschard and colleagues [28] explored the impact of more complicated architectures in simulated models of gene-gene and gene-environment interactions on predictive accuracy of polygenic risk scores (although the number of SNPs included in the models ranged from 15–31). They simulated two to ten interactions (one to five gene-gene and gene-environment interactions, respectively) and found that as sample size, and the number and size of interactions increased, the predictive power also increased. However, these increases were very modest most likely due to the relatively small set of effects (SNPs, environments and interactions) accounting for a modest overall heritability.

Kong and colleagues explored real life variants, extracting 161 risk alleles from the NHGRI GWAS catalog previously implicated in traits or diseases commonly seen in primary care settings or commonly used in indexing disease risk [29]. Polygenic risk was calculated as follows: (1) the simple sum of the risk alleles present, (b) product of the odds ratio of the present risk present, and (c) the total population attributable risk of the alleles present. Under each of the three methods, the risk score was normalized against the score distribution in the control population. Risk scores were calculated for existing Wellcome Trust Case Control Consortium (WTCCC) data. They found a significant positive correlation between the methods, indicating that choice of weighting scheme is probably less critical compared to other factors. In addition, the scores were only modestly predictive with AUCs in the 0.5–0.7 range. It is important to note the relatively small number of SNPs used in this analysis. An important outcome of these analyses is the examination of PGS by decile, with consistent findings of an increase in proportion of cases by increasing risk decile, an approach worth further discussion and consideration.

Conclusions—Future Directions

Overall, PGS approaches hold promise for capturing the contribution of genome-wide common variation to complex disease. The method is generally most powerful when the discovery sample is large. Parameters that influence the number of risk loci, including the heritability of the disease, the effect size and distribution of the true loci, and their frequency have a dramatic impact on the power of the approach. This result is not surprising since the empirically derived sum score is always a mix of true and null SNPs. Thus, the power maximizes at P_T ranges where the PGS is most enriched with true signals. Consequently, any variable that enriches the polygenic score with disease loci will increase the power of the approach.

It is also clear that different factors drive discovery versus prediction, and frequently, investigators compromise between the two. While a larger discovery sample and a much smaller number of clearly associated SNPs maximize prediction

accuracy, an equal sample split maximizes power to detect association between PGSs and a trait or disease.

While the PGS approach appears very promising, there are a number of unanswered questions regarding the method that our review has not explored, most notably, optimal effect size or p value thresholds for discovery, deciling or dichotomizing of PGS, optimal weighting strategies, accommodation of variable ancestry, and inclusion of environmental variables.

A commonly used approach to selecting a SNP during the discovery process for subsequent inclusion in PGS simply continues to add SNPs until the proportion of variance plateaus. Frequently, in approaches using log(OR) weighting, this happens as the OR approaches unity (i.e., a weight=0) and as the proportion of true effects diminishes. This point in the distribution varies with the power of the sample and the genetic architecture of the disease or trait. Diseases with thousands of underlying small effects will generate inflated p -values leading to the inclusion of tens of thousands of SNPs, certainly a mix of true and null effects that will diminish the predictive power of the PGS. While a priori decisions regarding optimal P_T would be difficult since the genetic architecture of any given disease is largely unknown, it may be worth scanning a range of P_T values in applied data to get a picture of the “shape” of the PGS signal. Alternatively, a PGS could be limited to the relatively small number of replicated SNPs, again limiting the predictive power of the PGS. Clearly, more work needs to be done.

Another issue is the possibility of dichotomizing or quantiling PGS to identify particular high risk groups for prevention/intervention. If we consider a best case scenario (i.e., all true and no null SNPs) and use the probit model described by Wray [15], assuming a total trait heritability of 80 %, a per allele effect (a) of 0.05, and a prevalence of 3 %; the number of total alleles underlying the trait can be estimated at 762. Given the distribution of PGS under the probit model, the consequent risk at several quantiles can be estimated. For example, at the 90th, 95th, and 99th percentiles the risk of disease is 3.73, 8.1, and 24 %, respectively. This not only gives an idea of the upper limit of PGS risk prediction but also highlights the possibility that individuals at five- to tenfold increased risk may be identified using refined PGS.

The usual approach to dealing with population stratification by ancestry is to create ancestry principal components from random genome-wide SNP data for adjustment in subsequent testing. In the discovery stage, where simple association testing is used to select loci for the PGS, usual methods [30, 31] of controlling for estimated genome-wide ancestry can be used. In the PGS testing or prediction stage, where these loci are summed, measures of ancestry can also be included as covariates, where the impact of sum score on case-control status is modeled. This approach, while potentially viable, assumes a linear impact of stratification on the component parts (the SNPs) of the sum score. This may not be a

realistic assumption since different polymorphisms may display different signatures of stratification. However, we do concede that it is the norm of the field to use composite, or genome-average, measures of ancestry to correct for stratification in GWAS. While not an ideal approach, this may not be a fatal flaw. Furthermore, population stratification is commingled with genome-wide true effects. In fact, initial work highlighted the difficulty in disentangling population stratification from polygenic signature [32]. If, as we assume, Fisher's model is true for complex disease, then some degree of average genomic difference will be due to true case-control differences and not confounding by ancestry. Thus, while correcting for ancestry is a vital step to getting valid results, it is also important to not overcorrect for it. However, recent work has demonstrated that the inflation in test statistics commonly seen even in adjusted GWAS results is likely due to polygenicity and not stratification [33]. This is an area of active research and novel methods to assess and address the impact of population stratification on PGS are needed.

While the work of Aschard and colleagues showed a limited impact by including gene-gene and gene-environment interactions in prediction models, there are clear instances where environments of large effect will modify risk calculated using genes alone and ultimately increase prediction precision. Using models described by Wray [15], we can examine the impact of inclusion of a measured environmental variable on the predictive power of a given PGS. If we assume a disease with 40 % heritability accounted for by 100 loci of equal effect with an average minor allele frequency of 20 % and disease prevalence of 10 %, we can estimate the AUC of such a model (including only true loci) of 0.78. However, under the same genetic model but with inclusion of a single main effect environmental variable with a case-control difference of 1 or 2 standard deviations, the estimated AUC increases to 0.85 and 0.95, respectively. Thus, in diseases or traits with a known environmental component, risk prediction will benefit from the joint consideration of genes and environment.

Another issue is the selection of an optimal weighting strategy. Simulations show that simple (unweighted) sum scores provide adequate power in situations where the number of loci is small. Purcell and colleagues [12] used an approach where sum scores were weighted by the logOR resulting from the case-control differences in the discovery sample. The logarithm of the odds ratio is frequently used since it is normally distributed with a mean of 0 and variance proportional to the inverse of the sample size. The use of logOR weighting leads to the creation of sum scores where the most extreme results are disproportionately up-weighted. This approach may provide different results from an approach that uses no weights. Kong and colleagues showed a strong correlation between multiple weighting approaches [29]. In addition, the inclusion and weighting of rare variant with relatively large effects in PGS is a subject that has not been adequately addressed.

The polygenic score approach will continue to provide some insight into the genetic architecture of human complex disease. This approach holds incredible promise for aiding in the categorizing of specific phenotypic or risk subgroups, for example, among those with PGS in upper and lower quantiles. Recently, these methods have been applied to identify individuals at high risk for early disease onset [34, 35], those likely to suffer from comorbid disorders [36, 37], those with treatment resistant disease [38], those likely to benefit from intervention [39], and those who are more likely to suffer from more severe chronic disease [40]. PGS methods will continue to be actively used in research settings to identify phenotypically interesting or unique case subsets. As PGS methods are further refined, sample sizes increase and research continues to demonstrate the clinical utility of the approach, and PGS will likely become a routine part of clinical risk assessment.

Compliance with Ethics Guidelines

Conflict of Interest Brion S. Maher declares grants from the National Institutes of Health/National Institute on Drug Abuse (NIH grants DA035693, DA036525, and DA039408).

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Papers of particular interest, published recently, have been highlighted as:

- Of importance

1. Fisher, R. A. *The correlation between relatives on the supposition of Mendelian inheritance*. Trans Royal Soc Edinburgh. 1918.
2. Mather K, Jinks JL. Biometrical genetics. *Biom Gen*. 1971;56:445–61.
3. Ott, J. *Analysis of human linkage*. The Johns Hopkins 1991.
4. Ott J. A computer program for linkage analysis of general human pedigrees. *Am J Hum Genet*. 1976;28:528–9.
5. Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet*. 1980;32:314–31.
6. Chong JX et al. The genetic basis of Mendelian phenotypes: discoveries, challenges, and opportunities. *Am J Human Gen*. 2015;97:199–215. **The ability to identify rare variants for Mendelian diseases has recently been bolstered by advances in whole genome sequencing technology. This paper discusses the promise for identification of clinically actionable rare variants.**
7. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science*. 1996;273:1516–7.
8. Visscher PM. Sizing up human height variation. *Nat Genet*. 2008;40:489–90.
9. Janssens AC, Gwinn M, Khoury MJ, Subramonia-Iyer S. Does genetic testing really improve the prediction of future type 2 diabetes? *PLoS Med*. 2006;3, e114.

10. Maher B. Personal genomes: the case of the missing heritability. *Nat News*. 2008;456:18–21.
11. Zaykin DV, Zhivotovsky LA. Ranks of genuine associations in whole-genome scans. *Genetics*. 2005;171:813–23.
12. Purcell SM et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009;460:748–52.
13. Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry*. 2003;60:1187–92.
14. • Dudbridge F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet*. 2013;9:e1003348. **This was the first paper to create analytic power calculations for the power to detect case-control differences and of the relationship between specific genetic models and risk prediction accuracy.**
15. Wray NR et al. Multi-locus models of genetic risk of disease. *Genome Med*. 2010;2:10.
16. Visscher PM et al. Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genet*. 2006;2, e41.
17. Yang J et al. Genome partitioning of genetic variation for complex traits using common SNPs. *Nat Genet*. 2011;43:519–25.
18. • Golan D, Lander ES, Rosset S. Measuring missing heritability: inferring the contribution of common variants. *Proc Natl Acad Sci U S A*. 2014;111:E5272–81. **The PGC regression approach to calculating molecular heritability is introduced. It provides an apparent dramatic jump in heritability estimated from genome-wide data over existing approaches.**
19. Palla L, Dudbridge FA. Fast method that uses polygenic scores to estimate the variance explained by genome-wide marker panels and the proportion of variants affecting a trait. *Am J Human Genet*. 2015;97(2):250–9.
20. Vilhjalmsón, B. et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *bioRxiv* 2015;015859.
21. Bhatia, G. et al. Haplotypes of common SNPs can explain missing heritability of complex diseases. *bioRxiv*. 2015;022418
22. Kraft P et al. Beyond odds ratios—communicating disease risk based on genetic profiles. *Nat Rev Genet*. 2009;10:264–9.
23. Wray NR, Yang J, Goddard ME, Visscher PM. The genetic interpretation of area under the ROC curve in genomic profiling. *PLoS Genet*. 2010;6, e1000864.
24. Jakobsdóttir J, Gorin MB, Conley YP, Ferrell RE, Weeks DE. Interpretation of genetic association studies: markers with replicated highly significant odds ratios may be poor classifiers. *PLoS Genet*. 2009;5, e1000337.
25. Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res*. 2007;17:1520–8.
26. Evans DM, Visscher PM, Wray NR. Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk. *Hum Mol Genet*. 2009;18:3525–31.
27. • Chatterjee N et al. Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. *Nat Genet*. 2013;45:400–5. **The predictive correlation coefficient is discussed. The distribution of genetic effects impacts predictive ability indicating that a smaller number of large effects yields better predictive power than a large number of small effects.**
28. Aschard H et al. Inclusion of gene-gene and gene-environment interactions unlikely to dramatically improve risk prediction for complex diseases. *Am J Human Genet*. 2012;90:962–72.
29. • Kong, S. W. et al. Summarizing polygenic risks for complex diseases in a clinical whole-genome report. *Genet Med*. 2014; *Real, replicated variants from the NHGRI GWAS database were studied in the context of clinical prediction. While predictive power was modest, the approach used to construct scores had minimal impact on accuracy.*
30. Devlin B, Roeder K. Genomic control for association studies. *Biometrics*. 1999;55:997–1004.
31. Price AL et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38: 904–9.
32. Yang J et al. Genomic inflation factors under polygenic inheritance. *Eur J Human Genet*. 2011;19:807–12.
33. • Bulik-Sullivan BK et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47:291–5. **LD score regression is introduced as a method to distinguish polygenic effects from other sources of test statistic inflation, such as population stratification, and estimate genome-wide heritability and co-heritability between traits. The authors demonstrate that inflation in published GWAS is due to polygenicity and not stratification.**
34. Ahn K, An S, Shugart Y, Rapoport J. Common polygenic variation and risk for childhood-onset schizophrenia. *Mol Psychiatry*. 2014. doi:10.1038/mp.2014.158.
35. Escott-Price V et al. Polygenic risk of Parkinson disease is correlated with disease age at onset. *Ann Neurol*. 2015;77:582–91.
36. Hamshere ML et al. High loading of polygenic risk for ADHD in children with comorbid aggression. *Am J Psychiatry*. 2014;170: 909–16.
37. Wiste A et al. Bipolar polygenic loading and bipolar spectrum features in major depressive disorder. *Bipolar Disord*. 2014;16:608–16.
38. Frank J et al. Identification of increased genetic risk scores for schizophrenia in treatment-resistant patients. *Mol Psychiatry*. 2014;20:150–1.
39. Musci RJ et al. Polygenic score × intervention moderation: an application of discrete-time survival analysis to modeling the timing of first tobacco use among urban youth. *Dev Psychopathol*. 2015;27: 111–22.
40. Meier, S., Mattheisen, M., Ripke, S. & Wray, N. R. Polygenic risk score, parental socioeconomic status, family history of psychiatric disorders, and the risk for schizophrenia: a Danish population-based study and meta-analysis. 2015.