Using Polygenic Risk Scores to Establish Endophenotypes: Considerations and Current Constraints

David. C. Glahn[1,2], **Andrew M McIntosh**[3,4]

[1.]Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA

[2.]Olin Neuropsychiatry Research Center, Institute of Living, Hartford, CT, USA

[3.]Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh, UK

[4.]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK

Improvement in our understanding of the genetic architecture of mental illness has made it possible to rank individuals upon a theoretical liability distribution for a particular disorder using common single nucleotide polymorphisms. This ranking is based on selecting alleles from the results of a large-scale case-control genome wide association (GWA) study conducted in a "discovery" sample and designed to elucidate the genetic contributions to a particular disease. A weighted sum of these disease-associated alleles, referred to as a polygenic risk score, can then be constructed for any individual in an independent "target" sample. The polygenic risk score for a disorder reflects the individual-level genetic burden for this disorder which is attributable to common alleles and can be applied to assess associations with additional traits or endophenotypes. An association between a trait in the target sample (e.g. an illness or a quantitative measure) and the polygenic risk score implies that genetic signal associated with the illness studied in the discovery sample can be used for prediction of an individual's trait values in the target sample. In this issue of *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, Carey and colleagues (1) use such a polygenic score approach to link liability for attention-deficit/hyperactivity disorder (ADHD), alcohol misuse and activation within the ventral striatum during a reward processing task. While the various ADHD polygenic risk scores (different scores reflect different statistical thresholds from the discovery sample) correlated with ventral striatum activity, no significant association between the polygenic score for ADHD and subsyndromal alcohol misuse was observed. However, problematic alcohol use was phenotypically correlated with ventral striatum activity. Using a structural equation mode, Carey and colleagues suggest that genetic risk for ADHD has an indirect effect on problematic alcohol use through differences in ventral striatum activity to positive relative to negative feedback.

*Correspondence to David C Glahn Ph.D., Suite 6E, 40 Temple Street, New Haven, CT 06510, USA; david.glahn@yale.edu.

The article by Carey et al (1) represents an important and popular approach for characterizing the genetic liability for an illness that has the potential to uncover the contribution of genetic burden to different mechanisms related to the clinical expression of mental disorders. The large GWA studies tend to focus exclusively on defining the genetic contribution of diagnostic entities alone. Therefore, carefully executed follow up studies are needed to examine the influence of risk variants or polygenic scores on illness-associated behavioral, cognitive or brain imaging endophenotypes. These findings, in turn, could provide new insights into biological pathways, implicated by genetic factors, that give rise to psychiatric symptoms. Indeed, this is precisely the goal of Dr. Ahmad Hariri's ongoing Duke Neurogenetics Study (DNS), which provided the data presented by Carey et al (1). The DNS and similar studies have the potential to realize the clinical utility of findings from large-scale genetics analyses. Although polygenic scores do not, in themselves, point to a specific biological mechanism, they can be used to characterize the genetic relationship between an illness and related endophenotypes.

While we continue to make progress, there remain a number of important limitations to our knowledge of the genetic architecture of these debilitating disorders and to the commonly applied polygenic score techniques. Available evidence on the genetic architecture of mental illnesses suggests that both common and rare variants as well as rare structural changes (e.g. copy number variants, deletions, insertions) confer risk for these highly complex disorders. However, the relative importance of common vs. rare genetic variation may differ dramatically between illnesses. For example, numerous loci have been identified for schizophrenia where common variants may explain a significant proportion (>30%) of liability (2). In contrast, in the case of autism spectrum disorder, our current understanding of disease liability is much more strongly associated with rare *de novo* mutations (3) rather than common variation. While future work with larger sample sizes could shift this perception (4), to date relatively little of the genetic variance of autism is captured with common variance. The utility of the polygenic score methods as typically implemented is therefore dependent upon the absolute level of genetic variance explained by common variants. Thus, while a polygenic score approach may be useful for schizophrenia (5), the same analytic strategy may currently be less fruitful for autism spectrum disorder. Similarly, as the relative influence of common vs. rare variation is unknown for most psychiatric disorders, the effectiveness of a polygenic risk approach is unclear for these illnesses. Given additional experimentation, larger sample sizes and/or more informative cases, we anticipate that the genetic architecture of most mental illnesses will be better enumerated, either supporting or rejecting the use of the polygenic risk score for a particular disease.

A complete treatment of polygenic score methods is beyond the scope of this commentary and several recent articles on this topic are highly recommended (6, 7). Nonetheless, there are two issues of fundamental importance for interpreting results from polygenic score analyses; namely the level of genetic variance explained in the discovery sample and the size of the target sample. Because individual common alleles explain a very small amount of disease liability, the discovery of risk-conferring alleles requires very large samples. For example, the proportion of the genetic liability to schizophrenia explained by common variants dramatically increased with the number of schizophrenia cases as the sample size increased from 3,322 (5) to 36,989 (2). The change in the number and often the nature of the

risk-conferring alleles by necessity impacts polygenic risk score calculations, introducing a degree of uncertainty about its reliability. In addition, even when a definitive polygenic risk score might be available, its usefulness will ultimately depend on the proportion of the disease-associated genetic variance it explains. For example, if only 3% of the genetic variance of an illness is explained by the polygenic risk score, its explanatory power in subsequent analyses focusing on disease-related traits or mechanisms will be quite constrained. In contrast, if 50% of the genetic variance is explained by the common variants, then the polygenic score could be quite informative.

Dudbridge (7) identified further constraints associated with the size of the target sample. He convincingly demonstrated that for maximal power in a polygenic score analysis, the target and discovery samples should be of approximately equal size. Unfortunately, this dramatically reduces the usefulness of a polygenic score approach in samples with exhaustive or expensive phenotyping. In response, Wray and colleagues (8), using a number of assumptions, suggested that a target sample of ~2,000 individuals may provide sufficient statistical power to detect a significant proportion of the variance. While collecting a target sample of ~2,000 individuals is certainly more tractable than a sample of >30,000, polygenic score analysis using imaging data is likely to be atypical and generally unfeasible for all but a few large epidemiological studies (e.g. UK Biobank, Personalized Medicine Initiative).

Carney and colleagues (1) derived their polygenic risk score from the discovery sample reported by the Cross-Disorder Group of the Psychiatric Genomics Consortium (9), which included 1,947 trio cases, 1,947 trio pseudo-controls, 840 unrelated cases and 688 unrelated controls. While the results from the ADHD case-control analyses were not independently reported by the Cross-Disorder Group, it appears that these data represent a subset of individuals included in an analysis reported earlier by Neale and colleagues (10), including 2,064 trios, 896 cases, and 2,455 controls. No genome-wide significant associations were found by Neale and colleagues, who estimated that only 0.51% of the genetic variance was captured by the analysis. While the approach used by Carney and colleagues is conceptually useful, the results may be subject to revision as more data on the genetic risk factors of ADHD become available.

In conclusion, the polygenic risk score is currently a pragmatic approach for assessing the endophenotypic status of disease-related traits and improving our understanding of the brain systems associated with illness liability. Genomic prediction will likely be a core component of preventative and personalized medicine. These current findings should be viewed as a useful starting point for more detailed planning and larger studies.

Acknowledgments

References

1. Carey C, Knodt A, Drabant Conley E, Hariri A, Bogdan R (2017): Reward-related ventral striatum activity links polygenic risk for attention-deficit/hyperactivity disorder to problematic alcohol use in young adulthood. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging.

2. Schizophrenia Working Group of the Psychiatric Genomics C (2014): Biological insights from 108 schizophrenia-associated genetic loci. Nature. 511:421–427. [PubMed: 25056061]

3. Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, et al. (2014): The contribution of de novo coding mutations to autism spectrum disorder. Nature. 515:216–221. [PubMed: 25363768]

4. Gaugler T, Klei L, Sanders SJ, Bodea CA, Goldberg AP, Lee AB, et al. (2014): Most genetic risk for autism resides with common variation. Nat Genet. 46:881–885. [PubMed: 25038753]

5. Purcell S, Wray N, Stone J, Visscher P, O'Donovan M, Sullivan P, et al. (2009): Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 460:748–752. [PubMed: 19571811]

6. Wray NR, Yang J, Hayes BJ, Price AL, Goddard ME, Visscher PM (2013): Pitfalls of predicting complex traits from SNPs. Nat Rev Genet. 14:507–515. [PubMed: 23774735]

7. Dudbridge F (2013): Power and predictive accuracy of polygenic risk scores. PLoS Genet. 9:e1003348. [PubMed: 23555274]

8. Wray NR, Lee SH, Mehta D, Vinkhuyzen AA, Dudbridge F, Middeldorp CM (2014): Research review: Polygenic methods and their application to psychiatric traits. J Child Psychol Psychiatry. 55:1068–1087. [PubMed: 25132410]

9. Cross-Disorder Group of the Psychiatric Genomics C (2013): Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet. 381:1371–1379. [PubMed: 23453885]

10. Neale BM, Medland SE, Ripke S, Asherson P, Franke B, Lesch KP, et al. (2010): Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. J Am Acad Child Adolesc Psychiatry. 49:884–897. [PubMed: 20732625]