# Polygenic epidemiology

Frank Dudbridge

 Department of Non-communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK.

Tel: +44 7927 2025

Correspondence: frank.dudbridge@lshtm.ac.uk

## Abstract

Much of the genetic basis of complex traits is present on current genotyping products, but the individual variants that affect the traits have largely not been identified. Several traditional problems in genetic epidemiology have recently been addressed by assuming a polygenic basis for disease and treating it as a single entity. Here I briefly review some of these applications, which collectively may be termed polygenic epidemiology. Methodologies in this area include polygenic scoring, linear mixed models, and linkage disequilibrium scoring. They have been used to establish a polygenic effect, estimate genetic correlation between traits, estimate how many variants affect a trait, stratify cases into sub-phenotypes, predict individual disease risks, and infer causal effects using Mendelian randomization. Polygenic epidemiology will continue to yield useful applications even while much of the specific variation underlying complex traits remains undiscovered.

## Main text

The completion of genome-wide association studies (GWAS) for hundreds of complex diseases and traits has created new challenges for genetic epidemiologists as we seek to understand the function of associated loci, using multiple emerging technologies [1](#_ENREF_1). At the same time, the realisation that most if not all complex traits are polygenic – that is, they are influenced by thousands of genetic variants each having a small effect – has spurred the development of methods that address traditional problems in genetic epidemiology by treating the entire polygenic basis as a single entity. By dealing with the genetic basis *en masse,* these methods access more of the heritable component of complex traits than is possible by single-variant approaches, and thus alleviate much of the missing heritability problem that has recently so exercised genetic epidemiologists [2](#_ENREF_2). Here I briefly review several such applications, which together form an emerging field of *polygenic epidemiology*.

The concept of a polygenic risk is well established in classical genetics and in humans has been applied for some time in, for example, segregation analysis of complex diseases [3](#_ENREF_3); [4](#_ENREF_4). But only with the development of genome-wide panels of SNP markers has it become possible to treat the polygenic basis explicitly. An early success was in the recalculation of twin-based heritabilities using observed, rather than expected, genetic similarity between dizygous pairs [5](#_ENREF_5). Two influential papers then introduced complementary methods for demonstrating evidence of a polygenic effect, and have each led to further developments across a range of applications.

The polygenic scoring method was applied by Purcell et al [6](#_ENREF_6) to argue that schizophrenia has a polygenic risk. Although their GWAS identified few individually significant SNPs, a small systematic increase in χ2 statistics was observed across the genome, and could not be reduced by control for population structure or genotyping error. The polygenic scoring method uses a GWAS dataset, called the *training sample*, to estimate effect sizes for each SNP, and to select SNPs according to a *P*-value threshold. In a second dataset, called the *target sample*, a polygenic risk score is calculated for each subject as the weighted sum of risk alleles at the selected SNPs, with the weights being the effects estimated in the training data. A polygenic risk would cause this score to be associated for very liberal selection *P*-values. Purcell et al observed significant association of the score with a selection threshold as high as *P*<0.5, and used simulation to argue that this was consistent with a polygenic effect; their argument has since been strengthened by theory [7](#_ENREF_7); [8](#_ENREF_8).

Linear mixed models were used by Yang et al [9](#_ENREF_9) to show that much of the heritability of height can be explained by current GWAS chips, even though very little could be explained by known associated variants. The approach derives from methods of quantitative genetics used in crop and livestock breeding, in which the genetic value (or breeding value) of an individual can be derived as a random effect by relating the phenotypes of the study subjects to their known pedigree structure. The innovation in human studies is that nominally unrelated subjects can be used, by estimating their (distant) relatedness from genome-wide marker data. By relating genetic to phenotypic similarity across all pairs of subjects, the variance of the genetic values can be estimated and taken as an estimate of narrow-sense heritability. Since this depends upon the markers used to estimate relatedness, especially the fact that they are usually selected to be common (eg >1% minor allele frequency), this estimate is known as *chip heritability*. A further advantage of using unrelated individuals is that shared environmental effects are minimal and therefore unlikely to bias the chip heritability.

Both polygenic scores and linear mixed models have been used to infer a polygenic basis for a wide range of traits [10-13](#_ENREF_10" \o "Bush, 2010 #3426), to the degree that it is now generally accepted that all complex traits are polygenic. Much attention is now directed at demonstrating and estimating the genetic correlation between pairs of traits. This can be done with polygenic scores, by training the score on one trait and testing it against another: if there is no shared genetic basis, the score will not be associated. A bivariate linear mixed model may also be used, modelling the genetic variance for two traits simultaneously with their covariance. Both approaches have shown a shared basis for schizophrenia and bipolar disorder, as well as for other pairs of psychiatric disorders [6](#_ENREF_6); [14](#_ENREF_14); [15](#_ENREF_15). This indicates a common molecular etiology for these conditions, which may prove useful for developing novel treatments or identifying individuals at risk, but may also suggest problems with current nosology and diagnosis. A shared basis has also been demonstrated for mammographic breast density and breast cancer, which may in part explain the former as a risk factor for the latter, by mediating some of the genetic risk of disease [16](#_ENREF_16" \o "Varghese, 2012 #5156). On the other hand, genetic correlation has not been observed for some pairs of traits for which it might be expected, especially in neurology [17](#_ENREF_17" \o "Goris, 2014 #5388). Autism appears to occupy a distinct position among psychiatric disorders in that it has no genetic correlation with the other major disorders studied, other than a low (16%) correlation with schizophrenia [14](#_ENREF_14" \o "Lee, 2013 #5330).

Linear mixed models give immediate estimates of chip heritability and genetic correlation: inference of the polygenic effect follows by testing these effects against their standard errors [18](#_ENREF_18" \o "Visscher, 2014 #5391). By contrast, polygenic scoring gives a hypothesis test for the polygenic effect, without an immediate estimate of chip heritability. However, methods have been developed to infer chip heritability and genetic correlation from the result of a polygenic score test [19](#_ENREF_19); [20](#_ENREF_20), by estimating under what degree of chip heritability would the observed result be expected. These models also allow for a proportion of variants with no effect on the trait; linear mixed models have also been implemented with this feature [21-23](#_ENREF_21" \o "Meuwissen, 2001 #5434). A notable finding has been that the proportion of variants affecting a complex trait rarely exceeds 5%, even in heterogeneous samples [20](#_ENREF_20). Thus the classical infinitesimal model may not hold in truth, although the traits remain highly polygenic. This finding affects the interpretation of genetic correlation, since the overall genetic correlation is determined by the proportion of variants with effects on both traits and the correlation of effects among those variants. A given genetic correlation may be concentrated among a few variants with highly concordant effects on both traits, or dispersed among more variants with only weakly concordant effects. Although the genetic correlation is informative at the whole subject level, it is not in itself very informative at the single variant level. Chip heritability itself is only interpretable in the context of an assumed model, usually (for disease traits) a liability threshold model, and the population-specific environment [24](#_ENREF_24); however it is generally useful to estimate the broad degree of genetic variance and correlation, even if it is challenging to provide or interpret a precise estimate.

A third approach to assessing chip heritability is LD-scoring [25](#_ENREF_25" \o "Bulik-Sullivan, 2015 #5370). Its rationale is that the more variants a given marker is in linkage disequilibrium (LD) with, the higher is its (marginal) association statistic likely to be. Regression of χ2 statistics for genome-wide markers on their LD “scores” gives an estimate of chip heritability (the slope of the regression) as well as of systematic bias due to population structure (the intercept). An extension allows the estimation of genetic covariance between traits [26](#_ENREF_26" \o "Bulik-Sullivan, 2015 #5442).

A further application is the estimation of chip heritability within sub-groups of markers. Such “genome partitioning” analyses have demonstrated that polygenic effects are spread uniformly across the chromosomes, but variants with known functional effects tend to explain more variation than others [27](#_ENREF_27); [28](#_ENREF_28).

Polygenic scores show promise for patient stratification and sub-phenotyping. Hamshere et al [29](#_ENREF_29) showed that, among bipolar disorder cases, polygenic scores for schizophrenia risk could distinguish schizo-affective cases from others, while not distinguishing psychotic cases from non-psychotic. In imflammatory bowel disease, polygenic scores can distinguish cases with colonic from ileal Crohn’s disease, and from those with ulcerative colitis [30](#_ENREF_30" \o "Cleynen, 2015 #5449). Prostate cancer screening targeted to men with high polygenic risk could reduce the rate of overdiagnosis [31](#_ENREF_31" \o "Pashayan, 2015 #5463).

Ultimately we might hope for individual risk prediction from polygenic analysis, which indeed first motivated the polygenic scoring method [32](#_ENREF_32). The performance of genetic prediction is bounded by the heritability [33](#_ENREF_33), but early attempts fell well short of that limit [34](#_ENREF_34). Furthermore, family history is an informative marker for genetic risk, and any predictor based on measured genotypes should exceed that benchmark to be deemed worthwhile [35](#_ENREF_35" \o "Aulchenko, 2009 #5464). It is now clear that much larger samples are required to attain accurate prediction from polygenic models, of the order of 105 subjects [8](#_ENREF_8); [36](#_ENREF_36); [37](#_ENREF_37). The reason is simply that as more variants enter the prediction model (as is necessary to explain the genetic risk), the greater is the sampling error in the total score, so the latter must be kept extremely small at the single variant level. National biobanks and disease consortia are now beginning to approach this scale [38-41](#_ENREF_38" \o "Michailidou, 2013 #5148), and the outlook for genetic prediction is becoming more promising than has recently appeared.

A common practice is to perform genetic prediction using only variants that are robustly associated at genome-wide significance [42-44](#_ENREF_42" \o "Szulkin, 2015 #5481). This is based on intuition that the risk score contains genuine predictors and no “noise”, and could therefore be easily conveyed to clinicians and policy makers. However, precisely in line with the missing heritability problem, such predictors explain very little variation in disease risk, and therefore have little predictive accuracy. Analytic results [8](#_ENREF_8) suggest that, even at very large sample sizes, prediction is optimised by selecting variants with *P-*values as high as 0.001 into a risk score (figure 1). Perhaps counter-intuitively, it is possible to predict with accuracy close to the theoretical maximum even while including many neutral variants in the model. By the same token, as sample sizes continue to grow it should be possible to achieve useful levels of prediction even while many individually associated variants remain to be discovered. In translating such predictors into practice, it will be important to present the polygenic risk as a single entity in order to allay concerns that it contains many irrelevant variants. This is naturally achieved by linear mixed models, which estimate a single genetic value for each individual as a random effect, and these models are now being extended to allow for subgroups of variants with no or little effect [23](#_ENREF_23); [45](#_ENREF_45).

As a final aspect of polygenic epidemiology discussed here, Mendelian randomisation studies are increasingly using composite genetic scores to draw causal inferences from observational data [46-49](#_ENREF_46" \o "Voight, 2012 #5505). The advantages are similar to those of individual prediction, in that a composite score can predict the intermediate trait to a greater accuracy than can single variants. A particular difficulty however is that Mendelian randomisation assumes that the genetic instrument only affects the outcome through the exposure of interest. This assumption is increasingly likely to be violated as more variants enter the model [50](#_ENREF_50); [51](#_ENREF_51). There is a stronger argument for restricting gene scores to robustly associated loci, but pleiotropy can still violate the assumption [52](#_ENREF_52" \o "Holmes, 2015 #5411). In a highly polygenic score, variants with effects on the outcome but not on the exposure can create substantial bias [53](#_ENREF_53" \o "Evans, 2013 #5499). Methods are now available to adjust Mendelian randomisation analyses for known [54](#_ENREF_54); [55](#_ENREF_55) and unknown [56](#_ENREF_56) pleiotropic effects. Concerns about pleiotropy have so far discouraged the application of linear mixed models to Mendelian randomisation, although in principle they could improve precision over polygenic scores.

The small effects of individual variants have caused the value of GWAS to be questioned [57](#_ENREF_57). But bearing in mind that most complex traits are strongly heritable, the polygenic risk is highly informative taken *en masse* as a single risk factor. Confirmation is appearing through the applications of polygenic epidemiology discussed here, and further applications – such as identifying interactions, or mediation analysis – are sure to be developed in the near future. So far, the field has mainly developed from psychiatric genetics, in which progress in identifying and following up GWAS associations has been slower than in other areas. However, as seen here, applications are becoming common in cardiovascular, cancer and immunological genetics. Other areas will add further insights: for example, population geneticists have used GWAS signals to demonstrate polygenic adaptation [58](#_ENREF_58" \o "Berg, 2014 #5517) . Linear mixed models are in principle the most powerful and accurate class of methods, with the principal challenges being the correct specification of the random effects distribution, and computational issues in large samples [59](#_ENREF_59" \o "Ge, 2015 #5512). Polygenic risk scores, and LD scoring, offer simpler and faster approaches, requiring only summary statistics from completed GWAS, and generally incurring only a moderate loss of precision compared to linear mixed models. The variety of methods and applications now emerging in polygenic epidemiology will make this field a fertile ground for development in the coming years.

## Acknowledgements

This work was funded by the MRC (MR/K006215/1). The author declares no conflict of interest.



Figure 1. P-values (-log10 scale) for selecting variants into a polygenic score such that the area under the ROC curve (AUC) is maximised. A binary trait with prevalence 10% is assumed, with variants selected from a case/control study with equal number of cases and controls. Chip heritability of 40% (solid line) and 20% (dashed line) is distributed among 100,000 independent variants, of which 5% have normally-distributed effects and the rest have no effect. The vertical line is at 50,000 cases and 50,000 controls, at which point over 95% of the maximum AUC is achieved.

## References

1. Ziegler, A., and Konig, I.R. (2014). Celebrating the 30th anniversary of genetic epidemiology: how to define our scope? Genet Epidemiol 38, 379-380.

2. Maher, B. (2008). Personal genomes: The case of the missing heritability. Nature 456, 18-21.

3. Risch, N. (1990). Linkage strategies for genetically complex traits. I. Multilocus models. Am J Hum Genet 46, 222-228.

4. Antoniou, A.C., Pharoah, P.D., McMullan, G., Day, N.E., Stratton, M.R., Peto, J., Ponder, B.J., and Easton, D.F. (2002). A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. Br J Cancer 86, 76-83.

5. Visscher, P.M., Hill, W.G., and Wray, N.R. (2008). Heritability in the genomics era--concepts and misconceptions. Nat Rev Genet 9, 255-266.

6. Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F., and Sklar, P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 460, 748-752.

7. Yang, J., Weedon, M.N., Purcell, S., Lettre, G., Estrada, K., Willer, C.J., Smith, A.V., Ingelsson, E., O'Connell, J.R., Mangino, M., et al. (2011). Genomic inflation factors under polygenic inheritance. Eur J Hum Genet 19, 807-812.

8. Dudbridge, F. (2013). Power and predictive accuracy of polygenic risk scores. PLoS Genet 9, e1003348.

9. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42, 565-569.

10. Bush, W.S., Sawcer, S.J., de Jager, P.L., Oksenberg, J.R., McCauley, J.L., Pericak-Vance, M.A., and Haines, J.L. (2010). Evidence for polygenic susceptibility to multiple sclerosis--the shape of things to come. Am J Hum Genet 86, 621-625.

11. Speliotes, E.K., Willer, C.J., Berndt, S.I., Monda, K.L., Thorleifsson, G., Jackson, A.U., Allen, H.L., Lindgren, C.M., Luan, J., Magi, R., et al. (2010). Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet 42, 937-948.

12. Lu, Y., Ek, W.E., Whiteman, D., Vaughan, T.L., Spurdle, A.B., Easton, D.F., Pharoah, P.D., Thompson, D.J., Dunning, A.M., Hayward, N.K., et al. (2014). Most common 'sporadic' cancers have a significant germline genetic component. Hum Mol Genet 23, 6112-6118.

13. Visscher, P.M., Brown, M.A., McCarthy, M.I., and Yang, J. (2012). Five years of GWAS discovery. Am J Hum Genet 90, 7-24.

14. Lee, S.H., Ripke, S., Neale, B.M., Faraone, S.V., Purcell, S.M., Perlis, R.H., Mowry, B.J., Thapar, A., Goddard, M.E., Witte, J.S., et al. (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat Genet 45, 984-994.

15. Cross-Disorder Group of the Psychiatric Genomics Consortium. (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet 381, 1371-1379.

16. Varghese, J.S., Thompson, D.J., Michailidou, K., Lindstrom, S., Turnbull, C., Brown, J., Leyland, J., Warren, R.M., Luben, R.N., Loos, R.J., et al. (2012). Mammographic breast density and breast cancer: evidence of a shared genetic basis. Cancer Res 72, 1478-1484.

17. Goris, A., van Setten, J., Diekstra, F., Ripke, S., Patsopoulos, N.A., Sawcer, S.J., van Es, M., Andersen, P.M., Melki, J., Meininger, V., et al. (2014). No evidence for shared genetic basis of common variants in multiple sclerosis and amyotrophic lateral sclerosis. Hum Mol Genet 23, 1916-1922.

18. Visscher, P.M., Hemani, G., Vinkhuyzen, A.A., Chen, G.B., Lee, S.H., Wray, N.R., Goddard, M.E., and Yang, J. (2014). Statistical power to detect genetic (co)variance of complex traits using SNP data in unrelated samples. PLoS Genet 10, e1004269.

19. Stahl, E.A., Wegmann, D., Trynka, G., Gutierrez-Achury, J., Do, R., Voight, B.F., Kraft, P., Chen, R., Kallberg, H.J., Kurreeman, F.A., et al. (2012). Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. Nat Genet 44, 483-489.

20. Palla, L., and Dudbridge, F. (2015). A 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 Hum Genet 97, 250-259.

21. Meuwissen, T.H., Hayes, B.J., and Goddard, M.E. (2001). Prediction of total genetic value using genome-wide dense marker maps. Genetics 157, 1819-1829.

22. Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. Nat Genet 44, 821-824.

23. Moser, G., Lee, S.H., Hayes, B.J., Goddard, M.E., Wray, N.R., and Visscher, P.M. (2015). Simultaneous discovery, estimation and prediction analysis of complex traits using a bayesian mixture model. PLoS Genet 11, e1004969.

24. Hopper, J.L., and Mack, T.M. (2015). The heritability of prostate cancer-letter. Cancer Epidemiol Biomarkers Prev 24, 878.

25. Bulik-Sullivan, B.K., Loh, P.R., Finucane, H.K., Ripke, S., Yang, J., Patterson, N., Daly, M.J., Price, A.L., and Neale, B.M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet 47, 291-295.

26. Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.R., Duncan, L., Perry, J.R., Patterson, N., Robinson, E.B., et al. (2015). An atlas of genetic correlations across human diseases and traits. Nat Genet.

27. Yang, J., Manolio, T.A., Pasquale, L.R., Boerwinkle, E., Caporaso, N., Cunningham, J.M., de Andrade, M., Feenstra, B., Feingold, E., Hayes, M.G., et al. (2011). Genome partitioning of genetic variation for complex traits using common SNPs. Nat Genet 43, 519-525.

28. Visscher, P.M., Macgregor, S., Benyamin, B., Zhu, G., Gordon, S., Medland, S., Hill, W.G., Hottenga, J.J., Willemsen, G., Boomsma, D.I., et al. (2007). Genome partitioning of genetic variation for height from 11,214 sibling pairs. Am J Hum Genet 81, 1104-1110.

29. Hamshere, M.L., O'Donovan, M.C., Jones, I.R., Jones, L., Kirov, G., Green, E.K., Moskvina, V., Grozeva, D., Bass, N., McQuillin, A., et al. (2011). Polygenic dissection of the bipolar phenotype. Br J Psychiatry 198, 284-288.

30. Cleynen, I., Boucher, G., Jostins, L., Schumm, L.P., Zeissig, S., Ahmad, T., Andersen, V., Andrews, J.M., Annese, V., Brand, S., et al. (2015). Inherited determinants of Crohn's disease and ulcerative colitis phenotypes: a genetic association study. Lancet.

31. Pashayan, N., Duffy, S.W., Neal, D.E., Hamdy, F.C., Donovan, J.L., Martin, R.M., Harrington, P., Benlloch, S., Amin Al Olama, A., Shah, M., et al. (2015). Implications of polygenic risk-stratified screening for prostate cancer on overdiagnosis. Genet Med 17, 789-795.

32. Wray, N.R., Goddard, M.E., and Visscher, P.M. (2007). Prediction of individual genetic risk to disease from genome-wide association studies. Genome Res 17, 1520-1528.

33. Clayton, D.G. (2009). Prediction and interaction in complex disease genetics: experience in type 1 diabetes. PLoS Genet 5, e1000540.

34. Evans, D.M., Visscher, P.M., and Wray, N.R. (2009). Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk. Hum Mol Genet 18, 3525-3531.

35. Aulchenko, Y.S., Struchalin, M.V., Belonogova, N.M., Axenovich, T.I., Weedon, M.N., Hofman, A., Uitterlinden, A.G., Kayser, M., Oostra, B.A., van Duijn, C.M., et al. (2009). Predicting human height by Victorian and genomic methods. Eur J Hum Genet 17, 1070-1075.

36. Daetwyler, H.D., Villanueva, B., and Woolliams, J.A. (2008). Accuracy of predicting the genetic risk of disease using a genome-wide approach. PLoS One 3, e3395.

37. Chatterjee, N., Wheeler, B., Sampson, J., Hartge, P., Chanock, S.J., and Park, J.H. (2013). Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. Nat Genet 45, 400-405, 405e401-403.

38. Michailidou, K., Hall, P., Gonzalez-Neira, A., Ghoussaini, M., Dennis, J., Milne, R.L., Schmidt, M.K., Chang-Claude, J., Bojesen, S.E., Bolla, M.K., et al. (2013). Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat Genet 45, 353-361, 361e351-352.

39. Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature 511, 421-427.

40. CardiogramPlusC4D Consortium. (2015). A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet 47, 1121-1130.

41. Locke, A.E., Kahali, B., Berndt, S.I., Justice, A.E., Pers, T.H., Day, F.R., Powell, C., Vedantam, S., Buchkovich, M.L., Yang, J., et al. (2015). Genetic studies of body mass index yield new insights for obesity biology. Nature 518, 197-206.

42. Szulkin, R., Whitington, T., Eklund, M., Aly, M., Eeles, R.A., Easton, D., Kote-Jarai, Z.S., Amin Al Olama, A., Benlloch, S., Muir, K., et al. (2015). Prediction of individual genetic risk to prostate cancer using a polygenic score. The Prostate 75, 1467-1474.

43. Mavaddat, N., Pharoah, P.D., Michailidou, K., Tyrer, J., Brook, M.N., Bolla, M.K., Wang, Q., Dennis, J., Dunning, A.M., Shah, M., et al. (2015). Prediction of breast cancer risk based on profiling with common genetic variants. J Natl Cancer Inst 107.

44. Talmud, P.J., Cooper, J.A., Morris, R.W., Dudbridge, F., Shah, T., Engmann, J., Dale, C., White, J., McLachlan, S., Zabaneh, D., et al. (2015). Sixty-five common genetic variants and prediction of type 2 diabetes. Diabetes 64, 1830-1840.

45. Speed, D., and Balding, D.J. (2014). MultiBLUP: improved SNP-based prediction for complex traits. Genome Res 24, 1550-1557.

46. Voight, B.F., Peloso, G.M., Orho-Melander, M., Frikke-Schmidt, R., Barbalic, M., Jensen, M.K., Hindy, G., Holm, H., Ding, E.L., Johnson, T., et al. (2012). Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. Lancet 380, 572-580.

47. Nuesch, E., Dale, C., Palmer, T.M., White, J., Keating, B.J., van Iperen, E.P., Goel, A., Padmanabhan, S., Asselbergs, F.W., Verschuren, W.M., et al. (2015). Adult height, coronary heart disease and stroke: a multi-locus Mendelian randomization meta-analysis. Int J Epidemiol.

48. Zhang, B., Shu, X.O., Delahanty, R.J., Zeng, C., Michailidou, K., Bolla, M.K., Wang, Q., Dennis, J., Wen, W., Long, J., et al. (2015). Height and Breast Cancer Risk: Evidence From Prospective Studies and Mendelian Randomization. J Natl Cancer Inst 107.

49. Thrift, A.P., Gong, J., Peters, U., Chang-Claude, J., Rudolph, A., Slattery, M.L., Chan, A.T., Locke, A.E., Kahali, B., Justice, A.E., et al. (2015). Mendelian Randomization Study of Body Mass Index and Colorectal Cancer Risk. Cancer Epidemiol Biomarkers Prev 24, 1024-1031.

50. Palmer, T.M., Lawlor, D.A., Harbord, R.M., Sheehan, N.A., Tobias, J.H., Timpson, N.J., Davey Smith, G., and Sterne, J.A. (2012). Using multiple genetic variants as instrumental variables for modifiable risk factors. Statistical methods in medical research 21, 223-242.

51. Burgess, S., and Thompson, S.G. (2013). Use of allele scores as instrumental variables for Mendelian randomization. Int J Epidemiol 42, 1134-1144.

52. Holmes, M.V., Asselbergs, F.W., Palmer, T.M., Drenos, F., Lanktree, M.B., Nelson, C.P., Dale, C.E., Padmanabhan, S., Finan, C., Swerdlow, D.I., et al. (2015). Mendelian randomization of blood lipids for coronary heart disease. Eur Heart J 36, 539-550.

53. Evans, D.M., Brion, M.J., Paternoster, L., Kemp, J.P., McMahon, G., Munafo, M., Whitfield, J.B., Medland, S.E., Montgomery, G.W., Timpson, N.J., et al. (2013). Mining the human phenome using allelic scores that index biological intermediates. PLoS Genet 9, e1003919.

54. Burgess, S., Dudbridge, F., and Thompson, S.G. (2015). Re: "Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects". Am J Epidemiol 181, 290-291.

55. Burgess, S., and Thompson, S.G. (2015). Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. Am J Epidemiol 181, 251-260.

56. Bowden, J., Davey Smith, G., and Burgess, S. (2015). Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol 44, 512-525.

57. Manolio, T.A. (2013). Bringing genome-wide association findings into clinical use. Nat Rev Genet 14, 549-558.

58. Berg, J.J., and Coop, G. (2014). A population genetic signal of polygenic adaptation. PLoS Genet 10, e1004412.

59. Ge, T., Nichols, T.E., Lee, P.H., Holmes, A.J., Roffman, J.L., Buckner, R.L., Sabuncu, M.R., and Smoller, J.W. (2015). Massively expedited genome-wide heritability analysis (MEGHA). Proc Natl Acad Sci U S A 112, 2479-2484.