Advances and Applications of Polygenic Scores for Coronary Artery Disease

Aniruddh P. Patel[1,2,3], **Amit V. Khera**[1,2,3,4]

[1]Division of Cardiology and Center for Genomic Medicine, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA

[2]Cardiovascular Disease Initiative, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA

[3]Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA

[4]Verve Therapeutics, Cambridge, Massachusetts, USA

Abstract

Polygenic scores quantify inherited risk by integrating information from many common sites of DNA variation into a single number. Rapid increases in the scale of genetic association studies and new statistical algorithms have enabled development of polygenic scores that meaningfully measure—as early as birth—risk of coronary artery disease. These newer-generation polygenic scores identify up to 8% of the population with triple the normal risk based on genetic variation alone, and these individuals cannot be identified on the basis of family history or clinical risk factors alone. For those identified with increased genetic risk, evidence supports risk reduction with at least two interventions, adherence to a healthy lifestyle and cholesterol-lowering therapies, that can substantially reduce risk. Alongside considerable enthusiasm for the potential of polygenic risk estimation to enable a new era of preventive clinical medicine is recognition of a need for ongoing research into how best to ensure equitable performance across diverse ancestries, how and in whom to assess the scores in clinical practice, as well as randomized trials to confirm clinical utility.

Keywords

INTRODUCTION

Identification of individuals at high risk for coronary artery disease (CAD)—the leading cause of death globally—to enable targeted prevention therapies remains a major public health need (1). The inherited component of risk of CAD was recognized as early as 1951, when close relatives of an afflicted family member were reported to have an approximately two-fold increase in risk (2-5). Subsequent studies of twins suggested that the inherited component of CAD ("heritability") may be up to 60% (6-8). Early investigation into the genetic basis of CAD involved careful family-based studies, which ultimately identified monogenic forms of the disease related to rare mutations in specific genes (9-11). Here, the prototypical example relates to mutations causal for familial hypercholesterolemia, a condition associated with impaired clearance of low-density lipoprotein (LDL) cholesterol from the circulation. Familial hypercholesterolemia variants are identified in approximately 2% of individuals with premature CAD and 0.4% of the general population (12-14).

Over the past two decades, the focus of study has shifted to the role of common genetic variants in predisposing to CAD. Genetic association studies conducted in more than one million individuals over the past decade have identified over 240 independent variants associated with risk of CAD (15-18). The variants discovered from these large-scale studies are common in the population and confer smaller differences in risk, typically associated with a <5% change in risk per inherited allele. Polygenic scores—which integrate information from many sites of common DNA variation—provide a quantitative metric of inherited risk that integrates information from many such variants into a single number. The concepts of polygenic prediction were first used in plant and animal breeding as early as the 1940s and continue to be used to maximize milk production in the dairy industry (19, 20). These scores have more recently been applied to human traits and diseases.

FUNDAMENTALS OF SCORE CONSTRUCTION

Results from genome-wide association studies (GWAS)—which aim to precisely quantify the relationship between each common DNA variant and disease risk—are the primary input data into the construction of polygenic scores. Common DNA variants are assessed in study participants across millions of sites in their genome. Investigators then compare frequencies of each variant occurring in cases versus controls, thereby generating an output of summary statistics that consists of a directional effect size per allele associated with the disease outcome and a measure of statistical significance (21) (Figure 1a).

Polygenic scores are calculated by summing the effect sizes of risk variants identified in an individual to get a single score, which generally forms a normal distribution when aggregated in the population. Starting with the most significant variants, sequentially higher p-value thresholds for inclusion of variants in the score are tested to identify the best-performing overall score (22). Initial efforts focused on only the variants with the strongest statistical associations; however, for polygenic diseases such as CAD, it was recognized that inclusion of variants with less significant associations was of value in improving polygenic score prediction (Figure 1b).

Linkage disequilibrium—the nonrandom association of variants at different locations in the genome—plays a key role in the application of GWAS data in polygenic score construction and needs to be accounted for to minimize noise. Briefly, the process of crossing over of chromosomes occurs during meiosis at recombination sites across the genome. This results in large blocks of genetic sequence patterns that segregate and are inherited together —haplotype blocks. Knowledge of these patterns is used to impute the sequences of DNA at certain sites within the vicinity of directly genotyped sites. Linkage disequilibrium patterns and haplotype blocks vary between ancestries. When constructing scores, it is important to prune out the variants within the same haplotype block as the strongest association to avoid double-counting risk conferred by a set of DNA variants that tend to be inherited together (Figure 1b).

The past decade has seen numerous advancements in the methods used to calculate polygenic scores from the traditional scoring approach and from pruning and thresholding (23). These methods use different statistical techniques to adjust the effect sizes measured from GWAS to reduce polygenic score error and improve performance (24-28). With a growing number of individuals included in GWAS for CAD over the past decade, the predictive power of the scores generated has continued to grow (12, 29-40). From 2010 to 2018, the odds ratio per standard deviation has increased from 1.23 to 1.72 due to a 3.5-fold increase in discovery cases and controls (Figure 1d). Recently developed methods that focus on refining the polygenic score weights based on fine mapping or the functional repercussions of certain variants have also been employed (41-47). Because functional annotations do not differ across ancestral backgrounds, polygenic scores based on this approach may have improved generalizability across ancestries. Other newly developed methods also incorporate GWAS data from multiple ancestries or multiple related traits to improve prediction (48-53).

The validation and testing of polygenic scores must occur in a data set independent from the data sets used in the discovery GWAS from which polygenic score weights are estimated. A subset of the target data is allocated for training, where polygenic score weights are optimized, calculated for a given trait, and benchmarked against other scores (54, 55). The best-performing score is chosen using measures of discrimination to then apply to the testing cohort for targeted risk prediction (Figure 1c,d). The Polygenic Score Catalog has been developed as a database of published scores which allows for cross-application and evaluation of different scores (56).

RISK PREDICTION

Polygenic scores hold considerable promise for enabling a tailored approach to clinical medicine. The advent of large biobanks that link genetic data with clinical outcomes has proven critical in the development and assessment of polygenic scores (57-62). An important example of such a resource is the UK Biobank, which is a prospective cohort study with extensive genetic, lifestyle, and linked clinical data in >500,000 individuals (63, 64).

Within the UK Biobank, several studies have demonstrated striking gradients in risk of CAD according to polygenic scores. A score constructed using 6.6 million genetic variants and

well-powered GWAS summary statistics was able to predict CAD prevalence with an odds ratio of 1.72 per standard deviation in participants of European ancestry (12). This approach identified 8% of the population at threefold increased risk for CAD (Figure 2a). This prevalence is 20-fold higher than the carrier frequency of rare familial hypercholesterolemia variants conferring comparable risk. Other studies have stratified individuals by their cumulative risk for incident disease based on their decile of the score distribution (31) (Figure 2b). The individuals with high genetic risk identified by these methods are "flying under the radar" in current clinical practice, despite evidence that targeted screening or prevention efforts can markedly attenuate polygenic risk. Across three US healthcare systems, individuals in the top quintile of the polygenic score, when compared with the rest of the population, did not meet criteria for increased statin therapy per the American College of Cardiology/American Heart Association recommendations (46.2% versus 46.8%, respectively) or have higher statin prescription rates (25.0% versus 23.8%, respectively) (65).

Polygenic scores capture risk information that is largely independent of traditional risk factors (66). For example, the Pearson correlation coefficients between risk estimated by the Pooled Cohort Equations and polygenic score for CAD ranged from 0.01 to 0.03 across prior studies (66, 67) (Figure 2c). Of the genome-wide significant variants identified through large-scale CAD GWAS, only approximately 10% relate to known traditional risk factors such as lipid levels and blood pressure (68, 69). Although the remaining variants may not have understood mechanisms, they nonetheless offer additional predictive power when incorporated into genetic scoring. Polygenic scores have a higher discriminative capacity for incident CAD in baseline models including age and sex (change in C-statistic 0.045) when compared with 11 traditional risk factors (0.007 to 0.032) (66). Although there is a modest improvement in discrimination when polygenic scores are combined with risk estimates from the Pooled Cohort Equations, the combination of age, sex, and polygenic score alone demonstrates equivalent predictive power to this clinical risk estimator (70). Polygenic scores stratify CAD risk even among individuals with similar predicted risk according to current risk algorithms. Among individuals at borderline risk (5–7.5% 10-year risk estimate by Pooled Cohort Equations), polygenic scores stratified risk further, with lifetime risk ranging from 11.3% in the bottom quintile to 34.1% in the top quintile. Most, but not all, studies have shown improvement in risk prediction resulting from models that incorporate clinical variables and polygenic scores, as compared with clinical variables alone (31, 70-75). A key advantage of a polygenic score is that it can be assessed very early in life, before the onset of traditional risk factors. CAD polygenic scores perform better in younger individuals than in older individuals because at younger ages genetic determinants likely play a larger role in explaining variance with respect to other risk factors (71). An analysis of the polygenic basis of obesity and body weight across the lifespan illustrated the ability to stratify individuals into very different trajectories of cardiometabolic health. Analysis of a prospective cohort study that followed participants from birth to early adulthood noted minimal differences in birth weight across polygenic score deciles; however, this gradient widened through childhood so that by 18 years of age, a 26-pound average difference in weight was noted between the top and bottom deciles (76). Similarly, polygenic scores for

CAD assessed in young adulthood stratify individuals along a substantial gradient in risk of future coronary artery calcification and CAD events (77).

Polygenic scores offer an opportunity to refine genetic risk prediction even for those who inherit a familial hypercholesterolemia variant, who have long been recognized in clinical practice to have highly variable expression of disease. A portion of this variation is mediated by a polygenic background (78-80). For carriers of pathogenic variants for hypercholesterolemia, who on average exhibit triple the risk of CAD borne by noncarriers, substantial gradients in CAD risk exist, with the probability of disease by age 75 ranging from 17% to 78% (81). Therefore, familial hypercholesterolemia variant carriers with polygenic scores in the lowest quintile are at a near-population-average CAD risk. Moving forward, state-of-the-art genome interpretation is likely to involve integration of monogenic and polygenic risk into a single report for a complete and accurate picture of common disease risk (82, 83) (Figure 2d).

ACTIONABILITY OF POLYGENIC SCORES

Once a high-risk individual is identified based on their polygenic score, several studies have shown they may stand to benefit from preventive interventions. In observational studies, adherence to a healthy lifestyle—specified as no current smoking, no obesity, regular physical activity, and a healthy diet—was associated with an approximately 50% lower relative risk of CAD across the spectrum of polygenic scores compared to those with unfavorable lifestyle characteristics. Individuals with the highest polygenic risk showed the greatest absolute risk reduction (5.6%) from adherence to a healthy lifestyle so as to match the average population-level risk (84, 85) (Figure 3a). Similar to healthy lifestyle, post hoc analysis of trials of cholesterol-lowering therapies also noted increased absolute risk reduction among individuals with high polygenic scores (86-89). In a meta-analysis of primary CAD prevention trials, statin therapy was associated with a relative risk reduction of 46% and absolute risk reduction of 3.6% in the highest genetic risk group, compared to relative risk reduction of 26% and absolute risk reduction of 1.3% across all other groups (89) (Figure 3b).

Early efforts to study the impact of polygenic score disclosure have generally noted mixed results of clinical utility, providing an opportunity for additional investigation. In a prospective study of individuals randomized to receive CAD risk estimates with or without polygenic scores, individuals informed of their polygenic scores had higher rates of statin prescription and lower LDL cholesterol at 6 months of follow-up (90). Another small study did not find change in lipids or behavior after disclosure of a polygenic score result (91). Further study of polygenic score implementation in large randomized controlled trials is needed to guide treatment recommendations based on polygenic scores.

The ability to identify very high-risk individuals using a polygenic score suggests value for clinical trial development. Estimating a polygenic score for potential participants prior to enrollment and selection of individuals within the highest risk categories would enrich for CAD event incidence, reducing the needed sample size and maintaining the same statistical

power. This could contribute to substantial savings in clinical trial costs and ultimately in drug costs (92).

Polygenic scores can be used to identify causal pathways using the principles of Mendelian randomization. As genetic variants are randomly assorted during meiosis and are independent of confounders, their association with both a risk factor and a disease is consistent with a causal pathway between the risk factor and the disease. Multiple studies have shown that polygenic scores allow for cross-trait prediction and causal inference (93). For example, a polygenic score optimized for predicting body mass index is also able to detect significant risk for a range of other cardiometabolic diseases including CAD (76). Genetic scores constructed using variants in genes associated with plasma high-density lipoprotein (HDL) cholesterol levels demonstrate no causal association between certain HDL-raising mechanisms and CAD (94). By contrast, similar analyses using variants influencing triglyceride concentrations did suggest a causal relationship with CAD (95).

Polygenic scores for a variety of key conditions can be assessed using a single test with a genotyping array or low-coverage whole-genome sequencing at a technical cost of less than $50. A blood or saliva sample collected early in life would allow for a polygenic susceptibility "report card" including numerous polygenic scores for many diseases. As with CAD, polygenic score methodologies have proven highly generalizable across traits, identifying up to 6.1%, 3.5%, and 1.5% of European individuals at greater than threefold risk for atrial fibrillation, type 2 diabetes, and breast cancer, respectively (12). Polygenic scores may be cost-effective when applied to other diseases, although more work needs to be done to assess this for CAD (96).

CURRENT LIMITATIONS OF POLYGENIC SCORE GENERALIZABILITY

Alongside considerable enthusiasm for polygenic scores, limitations exist with regard to their generalizability to different populations. A key equity issue has arisen in that although current scores are associated with increased risk across different ancestries, they have diminished predictive power in non-European ancestries (13, 97-99). Approximately 80% of currently available GWAS data are from individuals of European ancestry, and these discovery data sets are not fully informative for generating polygenic scores in individuals of other ancestries due to differences in linkage disequilibrium patterns, allele frequencies, heritability, and genetic architecture (100). To address these challenges, important ongoing efforts aim to gather GWAS data in individuals of non-European ancestry as well as to develop polygenic scoring methods that incorporate multiancestry data or adjust variant weights to leverage shared genetic variation between ancestries (50, 51, 101-107).

FUTURE DIRECTIONS OF POLYGENIC SCORES

As polygenic scores continue to improve in portability and accuracy, several key steps remain before they can be incorporated into clinical practice (Figure 4). First, integration of genetic risk with clinical factors into a unified risk prediction model would harness the influence of both primordial and acquired risk factors. Initial efforts have combined information from established clinical risk calculators with polygenic scores to determine an

absolute 10-year CAD event rate; however, further work is needed in calibrating integrated risk models to target populations (74, 108, 109). Although polygenic risk is orthogonal and additive to conventional risk factors, characterizing gene-by-risk-factor interaction across subgroups of age and sex remains a challenge, as does grounding prediction in absolute risk estimates. Similarly, unifying reference distributions for ancestry-based comparisons will be key to accurately calculating risk in different populations. Additionally, as more genomic discovery data are acquired, polygenic scores for a given disease will change and need to be updated over time.

Second, moving forward, large-scale randomized controlled trials testing preventative therapies among individuals with elevated genetic risk may help prospectively quantify benefit of use and inform future prevention guidelines. Although most studies of polygenic scores to date have been observational, the evidence collected supporting a high polygenic score as a bona fide risk enhancer is comparable to that supporting other established risk-enhancing factors (Figure 5).

Third, proper interpretation and communication of polygenic scores to clinicians and patients are important additional steps in incorporating polygenic scores into clinical practice. Standards for studies reporting polygenic score results have been established to help facilitate translation into clinical care (54). Several groups have focused on improving the communication and interpretability of polygenic score reports for patients (82, 110). The EMERGE Network study recently launched several efforts to improve the optimal return, understanding, and clinical use of polygenic scores (111). Within the United Kingdom, similar efforts are under way in Our Future Health, which is a nationwide health research program with plans to recruit up to five million participants and with a focus on calculating and communicating polygenic and other risk scores. Direct-to-consumer testing companies have begun offering polygenic score reports for CAD and other diseases to subscribers. With the increase in availability of these data in healthcare system biobanks and patient-ordered testing, future efforts will need to focus on re-education of the clinician workforce (112).

Significant progress has been made from the first observations of inherited mechanisms of CAD to the calculation of polygenic scores to predict disease risk. Their predictive ability has steadily improved with the increase in availability of discovery data sets and more refined statistical genetics methods. Much work remains to be done to improve equity, portability, and integration with available models. Nevertheless, in the era of personalized medicine, polygenic scores are poised to have an important impact on the prediction and prevention of CAD.

LITERATURE CITED

1. Roth GA, Abate D, Abate KH, et al. 2018. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 392(10159):1736–88 [PubMed: 30496103]

2. White PD. 1957. Genes, the heart and destiny. N. Engl. J. Med 256(21):965–69 [PubMed: 13451977]

3. Gertler MM, Garn SM, White PD. 1951. Young candidates for coronary heart disease. JAMA 147(7):621–25

4. Russek HI, Zohman BL. 1958. Relative significance of heredity, diet and occupational stress in coronary heart disease of young adults; based on an analysis of 100 patients between the ages of 25 and 40 years and a similar group of 100 normal control subjects. Am. J. Med. Sci 235(3):266–77 [PubMed: 13520712]

5. Lloyd-Jones DM, Nam B-H, D'Agostino RB, et al. 2004. Parental cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults: a prospective study of parents and offspring. JAMA 291(18):2204–11 [PubMed: 15138242]

6. Marenberg ME, Risch N, Berkman LF, et al. 1994. Genetic susceptibility to death from coronary heart disease in a study of twins. N. Engl. J. Med 330(15):1041–46 [PubMed: 8127331]

7. Wienke A, Holm NV, Skytthe A, Yashin AI. 2001. The heritability of mortality due to heart diseases: a correlated frailty model applied to Danish twins. Twin Res. 4(4):266–74 [PubMed: 11665307]

8. Zdravkovic S, Wienke A, Pedersen NL, et al. 2002. Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins. J. Intern. Med 252(3):247–54 [PubMed: 12270005]

9. Lehrman MA, Schneider WJ, Südhof TC, et al. 1985. Mutation in LDL receptor: Alu-Alu recombination deletes exons encoding transmembrane and cytoplasmic domains. Science 227(4683):140–46 [PubMed: 3155573]

10. Soria LF, Ludwig EH, Clarke HR, et al. 1989. Association between a specific apolipoprotein B mutation and familial defective apolipoprotein B-100. PNAS 86(2):587–91 [PubMed: 2563166]

11. Brown MS, Goldstein JL. 1986. A receptor-mediated pathway for cholesterol homeostasis. Science 232(4746):34–47 [PubMed: 3513311]

12. Khera AV, Chaffin M, Aragam KG, et al. 2018. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat. Genet 50(9):1219–24 [PubMed: 30104762]

13. Khera AV, Chaffin M, Zekavat SM, et al. 2019. Whole-genome sequencing to characterize monogenic and polygenic contributions in patients hospitalized with early-onset myocardial infarction. Circulation 139(13):1593–602 [PubMed: 30586733]

14. Patel AP, Wang M, Fahed AC, et al. 2020. Association of rare pathogenic DNA variants for familial hypercholesterolemia, hereditary breast and ovarian cancer syndrome, and Lynch syndrome with disease risk in adults according to family history. JAMA Netw. Open 3(4):e203959 [PubMed: 32347951]

15. Aragam KG, Jiang T, Goel A, et al. 2021. Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants. medRxiv 2021.05.24.21257377

16. Samani NJ, Erdmann J, Hall AS, et al. 2007. Genomewide association analysis of coronary artery disease. N. Engl. J. Med 357(5):443–53 [PubMed: 17634449]

17. Schunkert H, König IR, Kathiresan S, et al. 2011. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. Nat. Genet 43(4):333–38 [PubMed: 21378990]

18. CARDIoGRAMplusC4D Consortium, Deloukas P, Kanoni S, et al. 2013. Large-scale association analysis identifies new risk loci for coronary artery disease. Nat. Genet 45(1):25–33 [PubMed: 23202125]

19. Henderson CR. 1975. Use of relationships among sires to increase accuracy of sire evaluation. J. Dairy Sci 58(11):1731–38

20. Hazel LN. 1943. The genetic basis for constructing selection indexes. Genetics 28(6):476–90 [PubMed: 17247099]

21. Altshuler D, Daly MJ, Lander ES. 2008. Genetic mapping in human disease. Science 322(5903):881–88 [PubMed: 18988837]

22. Euesden J, Lewis CM, O'Reilly PF. 2015. PRSice: polygenic risk score software. Bioinformatics 31(9):1466–68 [PubMed: 25550326]

23. Choi SW, Mak TS-H, O'Reilly PF. 2020. Tutorial: a guide to performing polygenic risk score analyses. Nat. Protoc 15(9):2759–72 [PubMed: 32709988]

24. Vilhjálmsson BJ, Yang J, Finucane HK, et al. 2015. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. Am. J. Hum. Genet 97(4):576–92 [PubMed: 26430803]

25. Mak TSH, Porsch RM, Choi SW, et al. 2017. Polygenic scores via penalized regression on summary statistics. Genet. Epidemiol 41(6):469–80 [PubMed: 28480976]

26. Ge T, Chen C-Y, Ni Y, et al. 2019. Polygenic prediction via Bayesian regression and continuous shrinkage priors. Nat. Commun 10(1):1776 [PubMed: 30992449]

27. Newcombe PJ, Nelson CP, Samani NJ, et al. 2019. A flexible and parallelizable approach to genome-wide polygenic risk scores. Genet. Epidemiol 43(7):730–41 [PubMed: 31328830]

28. Lloyd-Jones LR, Zeng J, Sidorenko J, et al. 2019. Improved polygenic prediction by Bayesian multiple regression on summary statistics. Nat. Commun 10(1):5086 [PubMed: 31704910]

29. Abraham G, Havulinna AS, Bhalala OG, et al. 2016. Genomic prediction of coronary heart disease. Eur. Heart J 37(43):3267–78 [PubMed: 27655226]

30. Ripatti S, Tikkanen E, Orho-Melander M, et al. 2010. A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. Lancet 376(9750):1393–400 [PubMed: 20971364]

31. Inouye M, Abraham G, Nelson CP, et al. 2018. Genomic risk prediction of coronary artery disease in 480,000 adults. J. Am. Coll. Cardiol 72(16):1883–93 [PubMed: 30309464]

32. Ganna A, Magnusson PKE, Pedersen NL, et al. 2013. Multilocus genetic risk scores for coronary heart disease prediction. Arterioscler. Thromb. Vasc. Biol 33(9):2267–72 [PubMed: 23685553]

33. Tada H, Melander O, Louie JZ, et al. 2016. Risk prediction by genetic risk scores for coronary heart disease is independent of self-reported family history. Eur. Heart J 37(6):561–67 [PubMed: 26392438]

34. Patel RS, Sun YV, Hartiala J, et al. 2012. Association of a genetic risk score with prevalent and incident myocardial infarction in subjects undergoing coronary angiography. Circ. Cardiovasc. Genet 5(4):441–49 [PubMed: 22767652]

35. Tikkanen E, Havulinna AS, Palotie A, et al. 2013. Genetic risk prediction and a 2-stage risk screening strategy for coronary heart disease. Arterioscler. Thromb. Vasc. Biol 33(9):2261–66 [PubMed: 23599444]

36. Thanassoulis G. 2019. Using genetics to plan future randomized trials of lipoprotein(a) lowering —how much reduction, for how long, and in whom? JAMA Cardiol. 4(6):513–14 [PubMed: 31017615]

37. Hughes MF, Saarela O, Stritzke J, et al. 2012. Genetic markers enhance coronary risk prediction in men: the MORGAM prospective cohorts. PLOS ONE 7(7):e40922 [PubMed: 22848412]

38. de Vries PS, Kavousi M, Ligthart S, et al. 2015. Incremental predictive value of 152 single nucleotide polymorphisms in the 10-year risk prediction of incident coronary heart disease: the Rotterdam Study. Int. J. Epidemiol 44(2):682–88 [PubMed: 25953786]

39. Krarup NT, Borglykke A, Allin KH, et al. 2015. A genetic risk score of 45 coronary artery disease risk variants associates with increased risk of myocardial infarction in 6041 Danish individuals. Atherosclerosis 240(2):305–10 [PubMed: 25864160]

40. Weijmans M, de Bakker PIW, van der Graaf Y, et al. 2015. Incremental value of a genetic risk score for the prediction of new vascular events in patients with clinically manifest vascular disease. Atherosclerosis 239(2):451–58 [PubMed: 25687271]

41. Hu Y, Lu Q, Liu W, et al. 2017. Joint modeling of genetically correlated diseases and functional annotations increases accuracy of polygenic risk prediction. PLOS Genet. 13(6):e1006836 [PubMed: 28598966]

42. Hu Y, Lu Q, Powles R, et al. 2017. Leveraging functional annotations in genetic risk prediction for human complex diseases. PLOS Comput. Biol 13(6):e1005589 [PubMed: 28594818]

43. Márquez-Luna C, Gazal S, Loh P-R, et al. 2021. Incorporating functional priors improves polygenic prediction accuracy in UK Biobank and 23andMe data sets. Nat. Commun 12(1):6052 [PubMed: 34663819]

44. Weissbrod O, Hormozdiari F, Benner C, et al. 2020. Functionally informed fine-mapping and polygenic localization of complex trait heritability. Nat. Genet 52(12):1355–63 [PubMed: 33199916]

45. Chen W, McDonnell SK, Thibodeau SN, et al. 2016. Incorporating functional annotations for fine-mapping causal variants in a Bayesian framework using summary statistics. Genetics 204(3):933–58 [PubMed: 27655946]

46. Finucane HK, Bulik-Sullivan B, Gusev A, et al. 2015. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet 47(11):1228–35 [PubMed: 26414678]

47. Kichaev G, Yang W-Y, Lindstrom S, et al. 2014. Integrating functional data to prioritize causal variants in statistical fine-mapping studies. PLOS Genet. 10(10):e1004722 [PubMed: 25357204]

48. Chung W, Chen J, Turman C, et al. 2019. Efficient cross-trait penalized regression increases prediction accuracy in large cohorts using secondary phenotypes. Nat. Commun 10(1):569 [PubMed: 30718517]

49. Turley P, Walters RK, Maghzian O, et al. 2018. Multi-trait analysis of genome-wide association summary statistics using MTAG. Nat. Genet 50(2):229–37 [PubMed: 29292387]

50. Márquez-Luna C, Loh P-R, South Asian Type 2 Diabetes (SAT2D) Consort., et al. 2017. Multiethnic polygenic risk scores improve risk prediction in diverse populations. Genet. Epidemiol 41(8):811–23 [PubMed: 29110330]

51. Amariuta T, Ishigaki K, Sugishita H, et al. 2020. Improving the trans-ancestry portability of polygenic risk scores by prioritizing variants in predicted cell-type-specific regulatory elements. Nat. Genet 52(12):1346–54 [PubMed: 33257898]

52. Maier RM, Zhu Z, Lee SH, et al. 2018. Improving genetic prediction by leveraging genetic correlations among human diseases and traits. Nat. Commun 9(1):989 [PubMed: 29515099]

53. Weissbrod O, Kanai M, Shi H, et al. 2022. Leveraging fine-mapping and non-European training data to improve trans-ethnic polygenic risk scores. Nat. Genet 54(4):450–58 [PubMed: 35393596]

54. Wand H, Lambert SA, Tamburro C, et al. 2021. Improving reporting standards for polygenic scores in risk prediction studies. Nature 591(7849):211–19 [PubMed: 33692554]

55. Adeyemo A, Balaconis MK, Darnes DR, et al. 2021. Responsible use of polygenic risk scores in the clinic: potential benefits, risks and gaps. Nat. Med 27(11):1876–84 [PubMed: 34782789]

56. Lambert SA, Gil L, Jupp S, et al. 2021. The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. Nat. Genet 53(4):420–25 [PubMed: 33692568]

57. Carey DJ, Fetterolf SN, Davis FD, et al. 2016. The Geisinger MyCode Community Health Initiative: an electronic health record-linked biobank for precision medicine research. Genet. Med 18(9):906–13 [PubMed: 26866580]

58. Chen Z, Chen J, Collins R, et al. 2011. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. Int. J. Epidemiol 40(6):1652–66 [PubMed: 22158673]

59. Gaziano JM, Concato J, Brophy M, et al. 2016. Million Veteran Program: a mega-biobank to study genetic influences on health and disease. J. Clin. Epidemiol 70:214–23 [PubMed: 26441289]

60. Karlson EW, Boutin NT, Hoffnagle AG, et al. 2016. Building the Partners HealthCare Biobank at Partners Personalized Medicine: informed consent, return of research results, recruitment lessons and operational considerations. J. Pers. Med 6(1):2 [PubMed: 26784234]

61. Leitsalu L, Haller T, Esko T, et al. 2015. Cohort profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. Int. J. Epidemiol 44(4):1137–47 [PubMed: 24518929]

62. Pulley J, Clayton E, Bernard GR, et al. 2010. Principles of human subjects protections applied in an opt-out, de-identified biobank. Clin. Transl. Sci 3(1):42–48 [PubMed: 20443953]

63. Bycroft C, Freeman C, Petkova D, et al. 2018. The UK Biobank resource with deep phenotyping and genomic data. Nature 562(7726):203–9 [PubMed: 30305743]

64. Sudlow C, Gallacher J, Allen N, et al. 2015. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLOS Med. 12(3):e1001779 [PubMed: 25826379]

65. Aragam KG, Dobbyn A, Judy R, et al. 2020. Limitations of contemporary guidelines for managing patients at high genetic risk of coronary artery disease. J. Am. Coll. Cardiol 75(22):2769–80 [PubMed: 32498804]

66. Hindy G, Aragam KG, Ng K, et al. 2020. Genome-wide polygenic score, clinical risk factors, and long-term trajectories of coronary artery disease. Arterioscler. Thromb. Vasc. Biol 40(11):2738–46 [PubMed: 32957805]

67. Goff DC Jr., Lloyd-Jones DM, Bennett G, et al. 2014. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart

Association Task Force on Practice Guidelines. Circulation 129(25 Suppl. 2):S49–73 [PubMed: 24222018]

68. Nurnberg ST, Zhang H, Hand NJ, et al. 2016. From loci to biology: functional genomics of genomewide association for coronary disease. Circ. Res 118(4):586–606 [PubMed: 26892960]

69. Chen Z, Schunkert H. 2021. Genetics of coronary artery disease in the post-GWAS era. J. Int. Med 290(5):980–92

70. Elliott J, Bodinier B, Bond TA, et al. 2020. Predictive accuracy of a polygenic risk score-enhanced prediction model versus a clinical risk score for coronary artery disease. JAMA 323(7):636–45 [PubMed: 32068818]

71. Manikpurage HD, Eslami A, Perrot N, et al. 2021. Polygenic risk score for coronary artery disease improves the prediction of early-onset myocardial infarction and mortality in men. Circ. Genom. Precis. Med 14(6):e003452 [PubMed: 34670400]

72. Mosley JD, Gupta DK, Tan J, et al. 2020. Predictive accuracy of a polygenic risk score compared with a clinical risk score for incident coronary heart disease. JAMA 323(7):627–35 [PubMed: 32068817]

73. Khan SS, Page C, Wojdyla DM, et al. 2022. Predictive utility of a validated polygenic risk score for long-term risk of coronary heart disease in young and middle-aged adults. Circulation 146(8):587–96 [PubMed: 35880530]

74. Mars N, Koskela JT, Ripatti P, et al. 2020. Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers. Nat. Med 26(4):549–57 [PubMed: 32273609]

75. Sun L, Pennells L, Kaptoge S, et al. 2021. Polygenic risk scores in cardiovascular risk prediction: a cohort study and modelling analyses. PLOS Med. 18(1):e1003498 [PubMed: 33444330]

76. Khera AV, Chaffin M, Wade KH, et al. 2019. Polygenic prediction of weight and obesity trajectories from birth to adulthood. Cell 177(3):587–96.e9 [PubMed: 31002795]

77. Emdin CA, Xia R, Agrawal S, et al. 2022. Polygenic score assessed in young adulthood and onset of subclinical atherosclerosis and coronary heart disease. J. Am. Coll. Cardiol 80(3):280–82 [PubMed: 35835500]

78. Trinder M, Li X, DeCastro ML, et al. 2019. Risk of premature atherosclerotic disease in patients with monogenic versus polygenic familial hypercholesterolemia. J. Am. Coll. Cardiol 74(4):512–22 [PubMed: 31345425]

79. Oetjens MT, Kelly MA, Sturm AC, et al. 2019. Quantifying the polygenic contribution to variable expressivity in eleven rare genetic disorders. Nat. Commun 10(1):4897 [PubMed: 31653860]

80. Nauffal V, Morrill VN, Jurgens SJ, et al. 2022. Monogenic and polygenic contributions to QTc prolongation in the population. Circulation 145(20):1524–33 [PubMed: 35389749]

81. Fahed AC, Wang M, Homburger JR, et al. 2020. Polygenic background modifies penetrance of monogenic variants for tier 1 genomic conditions. Nat. Commun 11(1):3635 [PubMed: 32820175]

82. Brockman DG, Petronio L, Dron JS, et al. 2021. Design and user experience testing of a polygenic score report: a qualitative study of prospective users. BMC Med. Genom 14(1):238

83. Maamari DJ, Brockman DG, Aragam K, et al. 2022. Clinical implementation of a combined monogenic and polygenic risk disclosure for coronary artery disease. JACC Advances 3:1–11

84. Khera AV, Emdin CA, Drake I, et al. 2016. Genetic risk, adherence to a healthy lifestyle, and coronary disease. N. Engl. J. Med 375(24):2349–58 [PubMed: 27959714]

85. Hasbani NR, Ligthart S, Brown MR, et al. 2022. American Heart Association's Life's Simple 7: lifestyle recommendations, polygenic risk, and lifetime risk of coronary heart disease. Circulation 145(11):808–18 [PubMed: 35094551]

86. Mega J, Stitziel N, Smith J, et al. 2015. Genetic risk, coronary heart disease events, and the clinical benefit of statin therapy. Lancet 385(9984):2264–71 [PubMed: 25748612]

87. Damask A, Steg PG, Schwartz GG, et al. 2020. Patients with high genome-wide polygenic risk scores for coronary artery disease may receive greater clinical benefit from alirocumab treatment in the ODYSSEY OUTCOMES trial. Circulation 141(8):624–36 [PubMed: 31707832]

88. Marston NA, Kamanu FK, Nordio F, et al. 2020. Predicting benefit from evolocumab therapy in patients with atherosclerotic disease using a genetic risk score: results from the FOURIER TRIAL. Circulation 141(8):616–23 [PubMed: 31707849]

89. Natarajan P, Young R, Stitziel NO, et al. 2017. Polygenic risk score identifies subgroup with higher burden of atherosclerosis and greater relative benefit from statin therapy in the primary prevention setting. Circulation 135(22):2091–101 [PubMed: 28223407]

90. Kullo IJ, Jouni H, Austin EE, et al. 2016. Incorporating a genetic risk score into coronary heart disease risk estimates: effect on low-density lipoprotein cholesterol levels (the MI-GENES clinical trial). Circulation 133(12):1181–88 [PubMed: 26915630]

91. Knowles JW, Zarafshar S, Pavlovic A, et al. 2017. Impact of a genetic risk score for coronary artery disease on reducing cardiovascular risk: a pilot randomized controlled study. Front. Cardiovasc. Med 4:53 [PubMed: 28856136]

92. Fahed AC, Philippakis AA, Khera AV. 2022. The potential of polygenic scores to improve cost and efficiency of clinical trials. Nat. Commun 13(1):2922 [PubMed: 35614072]

93. Ritchie SC, Lambert SA, Arnold M, et al. 2021. Integrative analysis of the plasma proteome and polygenic risk of cardiometabolic diseases. Nat. Metab 3(11):1476–83 [PubMed: 34750571]

94. Voight BF, Peloso GM, Orho-Melander M, et al. 2012. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. Lancet 380(9841):572–80 [PubMed: 22607825]

95. Do R, Willer CJ, Schmidt EM, et al. 2013. Common variants associated with plasma triglycerides and risk for coronary artery disease. Nat. Genet 45(11):1345–52 [PubMed: 24097064]

96. Sharp SA, Jones SE, Kimmitt RA, et al. 2020. A single nucleotide polymorphism genetic risk score to aid diagnosis of coeliac disease: a pilot study in clinical care. Aliment. Pharmacol. Ther 52(7):1165–73 [PubMed: 32790217]

97. Martin AR, Kanai M, Kamatani Y, et al. 2019. Current clinical use of polygenic scores will risk exacerbating health disparities. Nat. Genet 51(4):584–91 [PubMed: 30926966]

98. Dikilitas O, Schaid DJ, Kosel ML, et al. 2020. Predictive utility of polygenic risk scores for coronary heart disease in three major racial and ethnic groups. Am. J. Hum. Genet 106(5):707–16 [PubMed: 32386537]

99. Kim MS, Patel KP, Teng AK, et al. 2018. Genetic disease risks can be misestimated across global populations. Genome Biol. 19(1):179 [PubMed: 30424772]

100. Martin AR, Gignoux CR, Walters RK, et al. 2017. Human demographic history impacts genetic risk prediction across diverse populations. Am. J. Hum. Genet 100(4):635–49 [PubMed: 28366442]

101. Ruan Y, Feng Y-CA, Chen C-Y, et al. 2022. Improving polygenic prediction in ancestrally diverse populations. Nat. Genet 54(5):573–80 [PubMed: 35513724]

102. Ishigaki K, Akiyama M, Kanai M, et al. 2020. Large-scale genome-wide association study in a Japanese population identifies novel susceptibility loci across different diseases. Nat. Genet 52(7):669–79 [PubMed: 32514122]

103. Gurdasani D, Carstensen T, Fatumo S, et al. 2019. Uganda genome resource enables insights into population history and genomic discovery in Africa. Cell 179(4):984–1002.e36 [PubMed: 31675503]

104. All of Us Research Program Investig., Denny JC, Rutter JL, et al. 2019. The "All of Us" research program. N. Engl. J. Med 381(7):668–76 [PubMed: 31412182]

105. Taliun D, Harris DN, Kessler MD, et al. 2021. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. Nature 590(7845):290–99 [PubMed: 33568819]

106. Wall JD, Stawiski EW, Ratan A, et al. 2019. The GenomeAsia 100K Project enables genetic discoveries across Asia. Nature 576(7785):106–11 [PubMed: 31802016]

107. Wang M, Menon R, Mishra S, et al. 2020. Validation of a genome-wide polygenic score for coronary artery disease in South Asians. J. Am. Coll. Cardiol 76(6):703–14 [PubMed: 32762905]

108. Riveros-Mckay F, Weale ME, Moore R, et al. 2021. Integrated polygenic tool substantially enhances coronary artery disease prediction. Circ. Genom. Precis. Med 14(2):e003304 [PubMed: 33651632]

109. Weale ME, Riveros-Mckay F, Selzam S, et al. 2021. Validation of an integrated risk tool, including polygenic risk score, for atherosclerotic cardiovascular disease in multiple ethnicities and ancestries. Am. J. Cardiol 148:157–64 [PubMed: 33675770]

110. Widén E, Junna N, Ruotsalainen S, et al. 2022. How communicating polygenic and clinical risk for atherosclerotic cardiovascular disease impacts health behavior: an observational follow-up study. Circ. Genom. Precis. Med 15(2):e003459 [PubMed: 35130028]

111. McCarty CA, Chisholm RL, Chute CG, et al. 2011. The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. BMC Med. Genom 4:13

112. Hao L, Kraft P, Berriz GF, et al. 2022. Development of a clinical polygenic risk score assay and reporting workflow. Nat. Med 28(5):1006–13 [PubMed: 35437332]
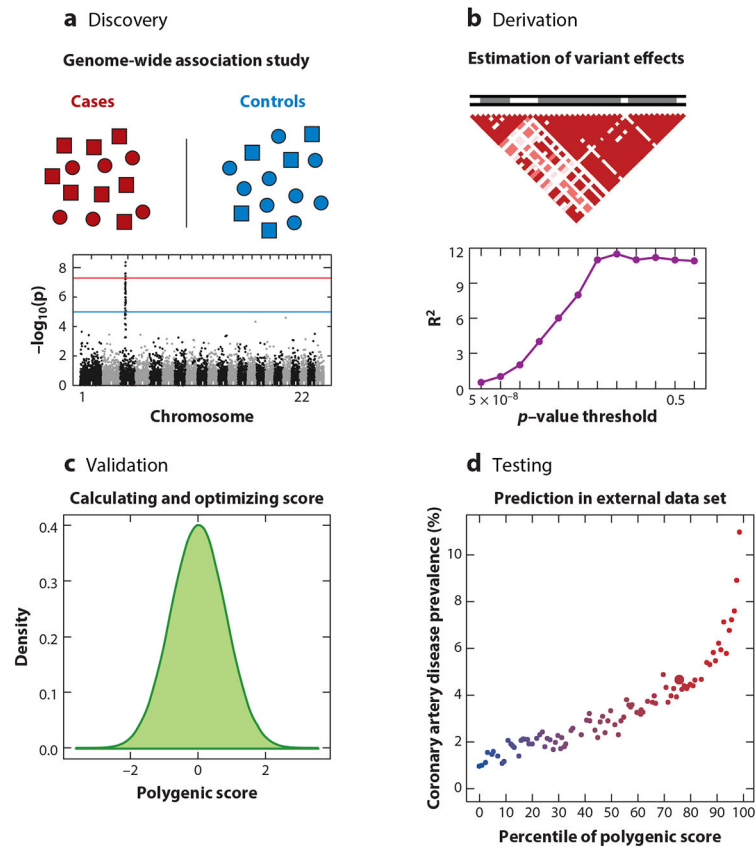
Figure 1.

Polygenic score construction. (*a*) Genome-wide association studies generate summary statistics to be used in polygenic score development. (*b*) Linkage disequilibrium patterns and significance of variant association are taken into account when selecting variants for score calculation. (*c*) Summing of effect sizes of each variant across the genome for each individual results in a normal distribution of scores in a population. (*d*) Polygenic scores stratify risk for coronary artery disease, with prevalence of disease in 100 groups binned according to percentile of polygenic score. Figure adapted with permission from Reference 12.
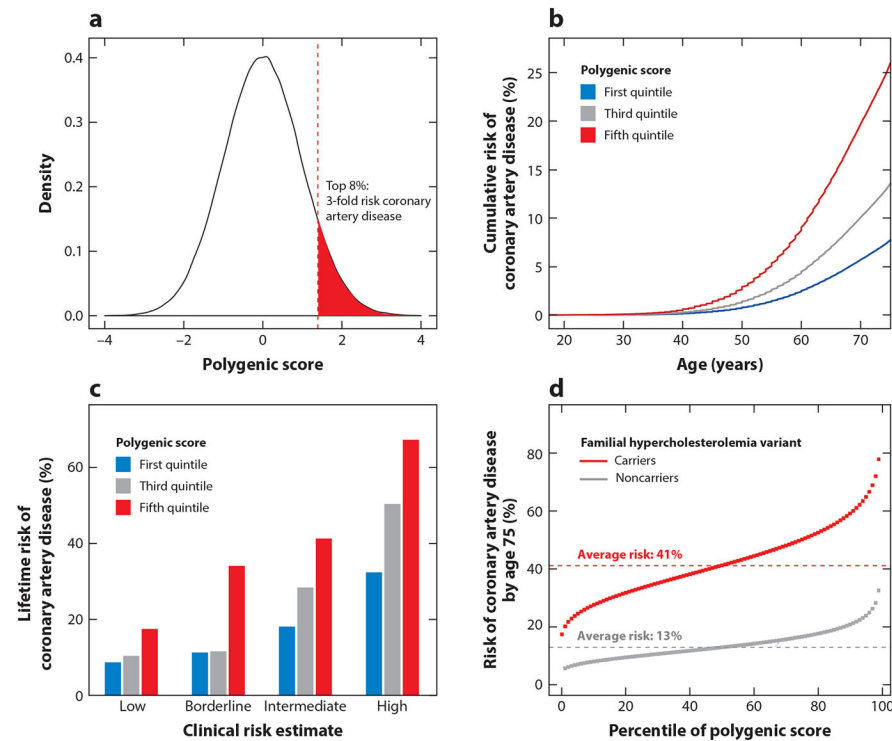
Figure 2.

Utility of polygenic scores. (*a*) Distribution of a polygenic score with shading reflecting the proportion of the population with threefold increased risk for prevalent coronary artery disease (CAD) versus the remainder of the population. (*b*) Cumulative risk of CAD by age 75 stratified by quintiles of the polygenic score distribution. (*c*) Cumulative risk of CAD by age 90 stratified by Pooled Cohort Equations risk category. (*d*) Predicted probability of CAD by age 75 in each percentile of the polygenic score distribution stratified by carrier status for a familial hypercholesterolemia variant. Horizontal dashed lines show the probability of disease for people with average polygenic score. Panel (*a*) from Reference 12, (*b*) from Reference 31, (*c*) from Reference 66, (*d*) from Reference 81. All adapted with permission.
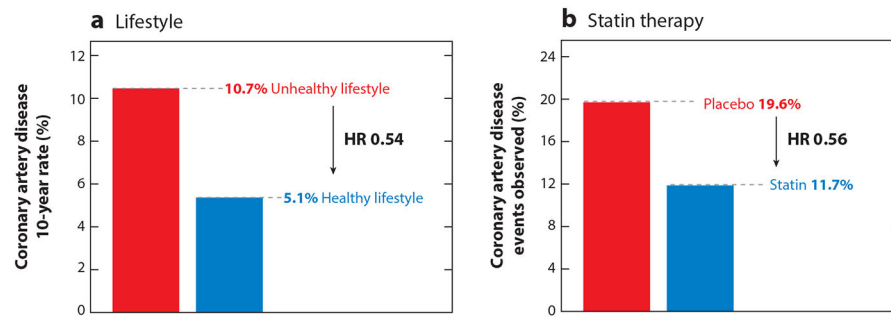
Figure 3.

Benefit of lifestyle and therapeutic interventions on genetic risk. Abbreviation: HR, hazard ratio. (*a*) Standardized 10-year cumulative incidence rates for coronary artery disease (CAD) events in individuals with the highest quintile of the polygenic score, stratified according to lifestyle (84). (*b*) Incident CAD event rates in clinical trial participants with the highest quintile of the polygenic score, stratified by statin therapy (89).
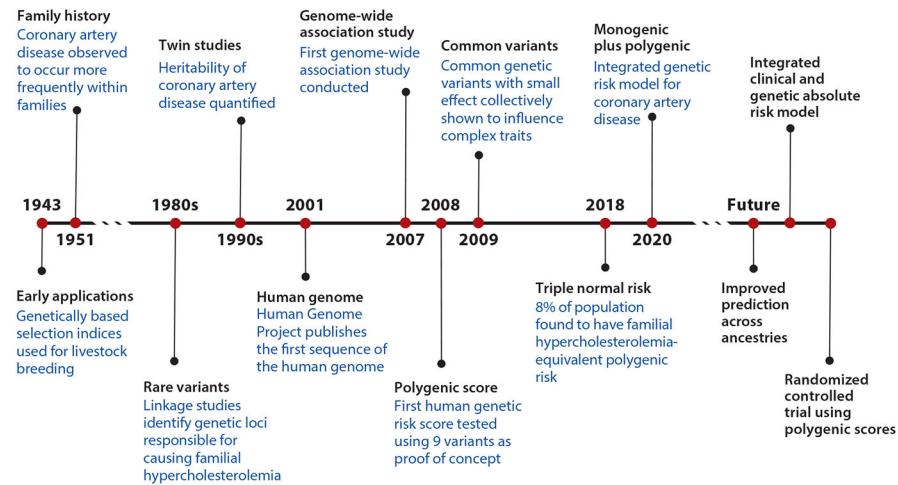
Figure 4.

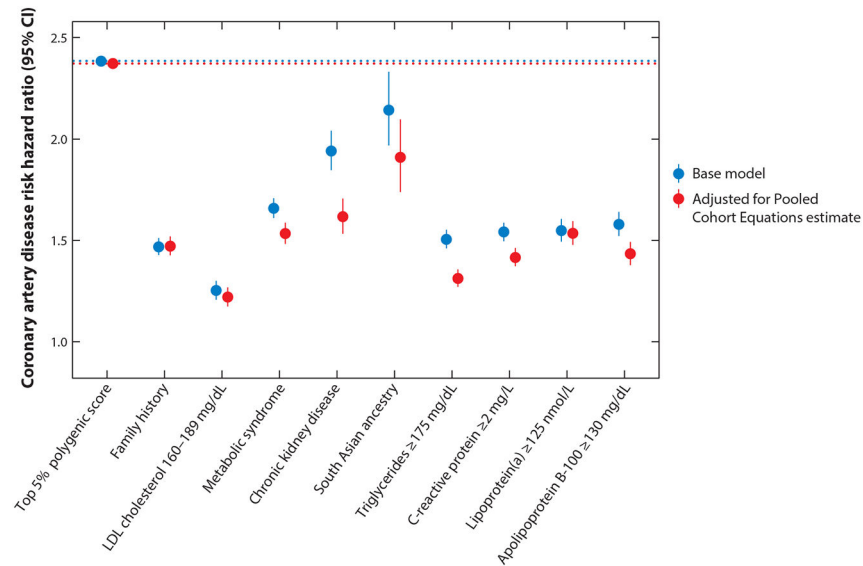Timeline of advances in the development of polygenic scores.

Figure 5.
Associations of polygenic scores and risk-enhancing factors with coronary artery disease (CAD). Hazard ratios with corresponding 95% confidence intervals are shown for incident CAD associated with risk-enhancing factors in the UK Biobank, calculated using Cox proportional-hazard models with covariates of enrollment age and sex in base model (*red*), or enrollment age, sex, and Pooled Cohort Equations 10-year risk estimate (*blue*).