



# The clinical applicability of polygenic risk scores for LDL-cholesterol: considerations, current evidence and future perspectives

Arjen J. Cupido, Tycho R. Tromp, and G. Kees Hovingh

## Purpose of review

The current review describes the development, clinical relevance and potential caveats of polygenic risk scores (PRS) for LDL cholesterol (LDL-C).

## Recent findings

In recent years, a large number of common variants have been shown to have a small effect on LDL-C levels. The aggregate effect of all of these variants on LDL-C levels can be captured in a PRS and an elevated number of LDL-C increasing common variants is considered to be a cause of high LDL-C levels in patients with familial hypercholesterolemia (FH) without a large effect, rare mutation. PRS do not only serve as a tool in diagnostics, but are also helpful in cardiovascular disease (CVD) risk prediction. Moreover, PRS modulate CVD risk even in patients without a monogenic FH. However, future larger scale PRS directly aimed at CVD risk may serve as more sensitive tools to identify individuals with severely increased CVD risk.

## Summary

LDL-C PRS help explain part of hypercholesterolemia in a proportion of dyslipidemic patients that do not have monogenic FH. Nevertheless, the CVD risk conferred by current PRS does not appear to match that of monogenic FH. LDL-C PRS are currently not widely used in clinical care.

## Keywords

polygenic risk score, LDL-cholesterol, familial hypercholesterolemia, cardiovascular disease risk prediction, clinical relevance

## INTRODUCTION

Over the last 2 decades, a large number of genome-wide association studies (GWAS) have been conducted. These studies have resulted in a better understanding of the impact of common genetic variants on a large number of clinical phenotypes, including dyslipidemia. It is now widely established that the aggregate burden of many common small effect-size genetic variants can explain a proportion of the variation in plasma lipid levels at population level. This aggregated effect can be captured in a so-called polygenic risk score (PRS), which is generated by combining the effect of independent genetic variants that have been shown to be associated with the clinical trait of interest.

Several different PRS have been constructed for LDL-cholesterol (LDL-C) levels. In recent independent observational studies, it has been shown that individuals with an increased LDL-C PRS score are at increased risk for incident cardiovascular events, which has widely been considered as yet further proof that long-term exposure to high LDL-C levels

is an independent cause of atherosclerosis. Currently, the LDL-C PRS may be used as a diagnostic criterion, or as a tool to assess the individual's cardiovascular disease (CVD) risk. Moreover, a PRS may, in the future, also serve to tailor therapeutic choices.

In this review, we describe the development, clinical relevance and potential caveats of polygenic LDL-C risk scores. We also describe our reflections on the potential future perspective of the LDL-C PRS

Department of Vascular Medicine, Amsterdam UMC, Location AMC, Amsterdam, The Netherlands

Correspondence to G. Kees Hovingh, Department of Vascular Medicine, Amsterdam UMC, Location AMC, D3-223, Meibergdreef 9, 1100DD Amsterdam, The Netherlands. Tel: +31 20 5666612; fax: +31 20 5669343; e-mail: g.k.hovingh@amsterdamumc.nl

**Curr Opin Lipidol** 2021, 32:112–116

DOI:10.1097/MOL.0000000000000741

This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

## KEY POINTS

- Polygenic LDL-C risk scores capture information on the aggregate burden of many small effect genetic variants associated with LDL-cholesterol and may help explain severe hypercholesterolemia in some patients where no single large effect variant is found.
- A multitude of different PRS have been described and, currently there is no clear consensus on how to define a 'high PRS' nor on how the PRS could and should be of additive value in the clinical setting.
- In the setting of similar lipid levels, patients with a polygenic form of hypercholesterolemia based on the currently available PRS are at lower CVD risk compared with patients with monogenic familial hypercholesterolemia, and at higher CVD risk compared with patients who have neither a monogenic nor a polygenic form of dyslipidemia.
- A PRS in which effect sizes from many variants on a direct coronary artery disease phenotype are combined, could have a larger predictive value and may therefore be preferred to classify CVD risk and tailor therapy.

in relation to recently developed 'genome-wide' risk scores for CVD in general.

## THE CONSTRUCTION OF LDL-CHOLESTEROL POLYGENIC RISK SCORES

At present, over 3100 common genetic variants have been shown to be associated with LDL-C levels or LDL-C-related traits such as statin-response [1]. Current large-scale studies such as the UK Biobank or Million Veterans Program [2] will probably result in an even larger number of variants, with smaller effect size. As such, we are getting to an even further refinement of our understanding of the genomic determinants of variation in LDL-C levels. The identified variants are located all across the genome, both in coding and noncoding regions of genes [3]. Ever since the first LDL-C PRS was constructed over a decade ago, more than 50 different lipid PRS have been used to assess the contribution of polygenic variation to lipid and lipoprotein levels in the population [4–6].

The statistical and technical aspects of combining variants in a PRS are beyond the scope this review, and have been extensively reviewed elsewhere [7–9]. In general, a PRS is constructed by summing the number of alleles from trait-affecting variants an individual has, weighted by their effect size as reported in the GWAS. This results in a single value quantifying the overall genetically defined

effect sizes. Since elevated LDL-C is associated with increased CVD risk, the LDL-raising allele at a locus is conventionally considered to be the 'risk' allele, irrespective of whether it is the major or minor allele in terms of frequency.

The selection and effect sizes of genetic variants are specific to the population and biochemical and statistical techniques used in the reference GWAS. Currently, the large majority of PRS are derived from cohorts with participants from European descent [6]. This may have ramifications for the performance of these PRS in individuals with a different ethnic background. Indeed, it has been shown that a PRS based on the effect sizes derived from a GWAS conducted in individuals from European descent, generated significant variation in the predictability for lipid levels among different ethnic subgroups from the MESA study [10]. The variance explained by the PRS (on top of age and sex) for any lipid trait was between 3 and 6% in whites and Hispanics, but only between 0.1 and 2% in African-Americans. Another study investigated the generalizability of lipid loci in European, Chinese, Japanese and Ugandan cohorts. After selecting all major lipid loci with a  $P$  value less than  $10^{-100}$  in a European discovery GWAS, they tested whether the loci were associated with lipid levels in other ethnic populations at a much less restrictive  $P$  value less than  $10^{-3}$ . Reproducibility for LDL-C associated variants only ranged between 62 and 77% [11<sup>■</sup>]. Possible explanations for this relatively poor generalizability could be sought in environmental factors or ethnic differences in gene expression [12].

## POLYGENIC RISK SCORES AS A TOOL TO REFINE DIAGNOSING FAMILIAL HYPERCHOLESTEROLEMIA

In the 1980s, studies on relatively few patients with extreme LDL-C phenotypes led to the discovery of the gene encoding the LDL receptor. Mutations in this single gene were identified as the causative defect in patients with familial hypercholesterolemia (FH) [13–15]. The CVD risk in carriers of an FH mutation is approximately two to three-fold higher when compared with individuals with similar LDL-C levels without such a pathogenic variant [16], underscoring the importance of cumulative lifelong exposure to high plasma LDL-C levels on progression of atherosclerosis in FH patients. Cascade screening is recommended in families of FH patients who carry a causal monogenic variant, as FH is a dominant inheritance disease.

However, a single pathogenic variant in the canonical FH-genes (*LDLR*, *APOB* or *PCSK9*) is identified in only 15–50% of phenotypical FH patients

(classified by clinical scoring systems) [17–19]. In a study by Wang *et al.* [17], no monogenic variant was found in 30% of the clinical FH patients with LDL-C levels above 7 mmol/l, and the authors suggest that the clinical FH phenotype may be caused by the cumulative effect of 10 small effect-size variants in roughly one-third of these patients. Already in 2013, a significantly higher PRS (containing 12 LDL-C associated variants) was observed in clinical FH patients without FH-mutation compared with both the general population and to FH patients with a confirmed monogenic mutation [20]. These results suggest that a substantial proportion of patients with mutation-negative FH could have a polygenic explanation for the observed high LDL-C levels.

The enrichment of common LDL-C raising variants in patients with severe hypercholesterolemia was independently confirmed multiple times thereafter. Recently, a study in healthy young women from the general population used the same 12-variant PRS and found that 21% of women with an LDL-C above the 99th percentile had a high PRS (defined as >90% percentile in the reference population) [21]. Similarly, a study investigating a different 29-variant PRS in a large Canadian cohort of clinical FH patients showed that 38% of all FH patients without a known mutation had a PRS in the top quintile compared with the reference population [19]. In a Portuguese FH-cohort, 41% of FH-mutation negative patients were considered to have ‘high PRS’ (defined as the top 25% of a 6-variant PRS) [22].

The abovementioned studies emphasize that both the number of variants used in the different PRS, as well as the cut-off threshold that defines a ‘high PRS’ vary greatly. There is currently no international consensus which PRS, and which cut-off-point defines a diagnosis of ‘polygenic hypercholesterolemia’ [23]. Regardless of the exact definition, identification of polygenic hypercholesterolemia could have an impact on the clinical value of cascade screening, because polygenic hypercholesterolemia does not follow the strict autosomal dominant pattern of inheritance of the monogenic form of FH [20]. The aggregation of many small effect variants means that only a portion of these variants will be inherited by the offspring, making it highly unlikely that they inherit a similar hypercholesterolemic phenotype as one of their parents.

### POLYGENIC RISK SCORES IN CARDIOVASCULAR RISK PREDICTION

Several studies have shown that the LDL-C PRS is independently associated with the risk for CVD. It is

important to realize that ‘polygenic hypercholesterolemia’, unlike monogenic hypercholesterolemia, is not a dichotomous diagnosis but rather a continuous scale that confers CVD risk in a dose-dependent manner. Moreover, hypercholesterolemia due to a polygenic or monogenic origin are not mutually exclusive but rather interacting entities. An individual’s predisposition to an LDL-C level based on polygenic background may aggravate CVD risk even in patients who already have extreme hypercholesterolemia due to a single FH-mutation.

Patients with monogenic FH and superimposed elevated LDL-C PRS are at greatest risk of CVD, which was shown by Trinder *et al.*, who investigated the occurrence of CVD in a cohort of 626 clinical FH patients of whom 274 (44%) had monogenic FH. The authors found that an elevated 28-variant LDL-C PRS (defined as >80th percentile) further increased CVD risk in patients with monogenic FH [hazard ratio 3.06 (95% confidence interval (CI) 1.56–5.99)] compared with patients with similar LDL-C levels without a genetic explanation for their hypercholesterolemia. In turn, patients with monogenic FH alone had a more ‘moderate’ cardiovascular risk increase [hazard ratio 1.97 (95% CI 1.09–3.56)] [19]. Using the same PRS, the authors later showed that among exclusively genetically confirmed FH patients ( $n = 1120$  from three separate cohorts) an elevated PRS was associated with increased CVD risk [hazard ratio 1.48 (95% CI 1.02–2.14,  $P = 0.04$ )] [24]. These studies show that polygenic contributions to LDL-C can still modulate CVD risk, even in a population already at extreme CVD risk such as FH patients.

In a different study, the same authors investigated the CVD risk associated with both monogenic and polygenic hypercholesterolemia in 47 841 individuals from the UK Biobank for whom whole exome sequencing data was available [25]. This allowed for the unbiased identification of 227 (0.57%, one in 176 individuals) carriers of a FH-mutation, as well as for the creation of a polygenic hypercholesterolemia group ( $n = 2379$ ) which was defined by a PRS more than 95th percentile using a novel 223-variant PRS. A ‘nongenetic hypercholesterolemia’ group was created for comparison, by matching the polygenic hypercholesterolemia participants 1:1 according to LDL-C, age, sex and genetic ancestry. The authors showed that risk for cardiovascular events increased with higher PRS in a dose-dependent manner, illustrated by a hazard ratio of 1.35 (1.30–1.40) when comparing the top PRS decile with the bottom decile. After selecting a subset of patients from each group to form three groups with similar LDL-C levels, it was found that

carriers of a monogenic FH-mutation were at greatest CVD risk [hazard ratio 1.93 (95% CI 1.34–2.77)], followed by the polygenic hypercholesterolemia group [hazard ratio 1.26 (95% CI 1.03–1.55)] compared with the nongenetic hypercholesterolemic group (reference group). This suggests that a larger cumulative LDL-C exposure may cause the risk of atherosclerosis in the former despite similar LDL-C levels at the time of inclusion. These observations also show that when hypercholesterolemia is attributable to a high PRS, it only explains a modestly increased CVD risk compared with patients with similar LDL-C levels without polygenic explanation and that the CVD risk in these patients does not come close to that of their ‘monogenic FH’ counterparts with similar LDL-C levels. This is likely due to the fact that the PRS in this study only explains 10% of the total LDL-C as being genetic and thus lifelong additive, which is modest compared with the LDL-C increase explained by an FH mutation.

The studies described above all exclusively used PRS limited to LDL-C associated variants to predict CVD risk. To put these results into perspective it is noteworthy that other PRS models have been generated that incorporate many more variants whose effect on CVD risk is not only explained by changes in LDL-C.

In 2018, Khera *et al.* [26] used the UK Biobank to derive and test a genome-wide PRS comprising over 6 million variants associated with coronary artery disease (CAD). They showed dose-dependent increase in the prevalence of CAD with increasing CAD-PRS. Moreover, participants within the highest 8% of this CAD-PRS had a three-fold increased risk of CAD compared with the mean of the rest of the population. This risk is comparable with the reported risk in patients with a monogenic form of FH. The latter is relatively uncommon (prevalence approximately one in 250 individuals; 0.4% [13,14]), while the high CAD-PRS was defined as a score above the 92nd percentile. This implicates that a 20-fold higher number of patients of risk could be identified by this CAD-PRS compared with genomic sequencing of the FH genes. Early assessment of CVD risk based on a large-scale CAD-PRS may thus help in identifying many more patients at high CVD risk to start preventive interventions.

## THE USE OF POLYGENIC RISK SCORES IN CLINICAL MANAGEMENT

Current clinical guidelines provide no recommendations on whether and how to use an LDL-C PRS in the clinical decision-making process about therapeutic interventions. This is in contrast to patients carrying a monogenic FH mutation who are by

default classified as being at high risk for atherosclerosis, warranting early and aggressive treatment of all CVD risk factors [23,27].

The improvement of current risk models by quantification of the lifelong cumulative burden of LDL-C as a result of high LDL-C PRS could possibly inform which patients are likely to benefit most from early interventions with – for example – statin therapy. This is of importance because preventive measures (both lifestyle and pharmacological) in those patients are likely to result in greater risk reductions than in patients with lower genetic risk. Although not yet investigated for LDL-C PRS, a study from 2017 found that patients in the top quintile of a 57-variant CAD-PRS achieved higher relative risk reduction from statin therapy compared with the rest of the study population (46 versus 26%,  $P=0.05$ ), despite similar LDL-C lowering [28]. Another recent study used a CAD-PRS to predict the benefit of LDL-C lowering by evolocumab in patients with atherosclerotic disease from the FOURIER trial [29<sup>†</sup>]. The clinical benefit of evolocumab was almost two-fold higher in patients with high genetic risk compared with the overall trial population.

Besides improved risk stratification, PRS may also be used to more adequately inform patients about their polygenetic risk, which could motivate patients to adhere to medication and a healthy lifestyle. While this has never specifically been investigated for LDL-C PRS, a CAD-PRS was used to investigate this [30]. A total of 203 patients at intermediate risk for CAD, who were not taking statin therapy were randomly assigned 1:1 to receive information about their individual CAD risk, either based on a conventional risk score or a score supplemented with information on their CAD-PRS. It was found that statins were initiated more often in the group that received information on their polygenetic risk (39 versus 22%,  $P < 0.01$ ) which helped to achieve slightly lower LDL-C levels after 6 months ( $96.5 \pm 32.7$  versus  $105.9 \pm 33.3$  mg/dl;  $P = 0.04$ ).

## FUTURE PERSPECTIVES

Our understanding of the aggregate effect of a large number of variants on both individual LDL-C levels and CVD risk is evolving. While there is no consensus on how to use the LDL-C PRS in the setting of diagnosing FH yet, such PRS may hold great value in risk prediction and tailoring CVD directed therapy in the near future. It is to be anticipated that the decreasing cost of genome wide analysis, combined with an ever-increasing understanding of the impact of models including PRS that contain more than a million variants will result in a broader application of PRS in the clinic.

## Acknowledgements

None.

## Financial support and sponsorship

A.J.C. and T.R.T. have nothing to disclose. G.K.H. reports research grants from the Netherlands Organization for Scientific Research (Vidi 016.156.445), Cardiovascular Research Initiative, and European Union; institutional research support from Aegerion, Amgen, AstraZeneca, Eli Lilly, Genzyme, Ionis, Kowa, Pfizer, Regeneron, Roche, Sanofi and The Medicines Company; speaker's bureau and consulting fees from Amgen, Aegerion, Sanofi and Regeneron (fees paid to the academic institution); and part-time employment at Novo Nordisk.

## Conflicts of interest

There are no conflicts of interest.

## REFERENCES AND RECOMMENDED READING

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Buniello A, MacArthur JAL, Cerezo M, *et al*. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* 2019; 47:D1005–D1012.
2. Klarin D, Damrauer SM, Cho K, *et al*. Genetics of blood lipids among ~300,000 multiethnic participants of the Million Veteran Program. *Nat Genet* 2018; 50:1514–1523.
3. Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: from polygenic to omnigenic. *Cell* 2017; 169:1177–1186.
4. Kathiresan S, Melander O, Guiducci C, *et al*. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet* 2008; 40:189–197.
5. Kathiresan S, Willer CJ, Peloso GM, *et al*. Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet* 2009; 41:56–65.
6. Dron JS, Hegele RA. The evolution of genetic-based risk scores for lipids and cardiovascular disease. *Curr Opin Lipidol* 2019; 30:71–81.
7. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. *Genome Med* 2020; 12:1–11.
8. Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. *Nat Rev Genet* 2018; 19:581–590.
9. Lambert SA, Abraham G, Inouye M. Towards clinical utility of polygenic risk scores. *Hum Mol Genet* 2019; 28:R133–R142.
10. Johnson L, Zhu J, Scott ER, Wineinger NE. An examination of the relationship between lipid levels and associated genetic markers across racial/ethnic populations in the multiethnic study of atherosclerosis. *PLoS One* 2015; 10:1–12.
11. Kuchenbaecker K, Telkar N, Reiker T, *et al*. The transferability of lipid loci across African, Asian and European cohorts. *Nat Commun* 2019; 10:4330. The study investigates the generalizability of a European genome-wide association studies on lipid levels in cohorts from Asia and Uganda. Of all results, it is especially striking that only 10% of all associations with TG could be replicated in the Ugandan cohort.
12. Keys KL, Mak ACY, White MJ, *et al*. On the cross-population generalizability of gene expression prediction models. *PLoS Genet* 2020; 16:1–28.
13. Nordestgaard BG, Chapman MJ, Humphries SE, *et al*. Familial hypercholesterolaemia is underdiagnosed and undertreated in the general population: guidance for clinicians to prevent coronary heart disease Consensus Statement of the European Atherosclerosis Society. *Eur Heart J* 2013; 34:3478–3490.
14. Hu P, Dharmayat KI, Stevens CAT, *et al*. Prevalence of familial hypercholesterolemia among the general population and patients with atherosclerotic cardiovascular disease: a systematic review and meta-analysis. *Circulation* 2020; 141:1742–1759.
15. Goldstein JL, Brown MS. The LDL receptor. *Arterioscler Thromb Vasc Biol* 2009; 29:431–438.
16. Khera AV, Won HH, Peloso GM, *et al*. Diagnostic yield and clinical utility of sequencing familial hypercholesterolemia genes in patients with severe hypercholesterolemia. *J Am Coll Cardiol* 2016; 67:2578–2589.
17. Wang J, Dron JS, Ban MR, *et al*. Polygenic versus monogenic causes of hypercholesterolemia ascertained clinically. *Arterioscler Thromb Vasc Biol* 2016; 36:2439–2445.
18. Reeskamp LF, Tromp TR, Defesche JC, *et al*. Next-generation sequencing to confirm clinical familial hypercholesterolemia. *Eur J Prev Cardiol* 2020; 2047487320942996. [Epub ahead of print]
19. Trinder M, Li X, DeCastro ML, *et al*. Risk of premature atherosclerotic disease in patients with monogenic versus polygenic familial hypercholesterolemia. *J Am Coll Cardiol* 2019; 74:512–522.
20. Talmud PJ, Shah S, Whittall R, *et al*. Use of low-density lipoprotein cholesterol gene score to distinguish patients with polygenic and monogenic familial hypercholesterolaemia: a case-control study. *Lancet* 2013; 381:1293–1301.
21. Balder JW, Rimbert A, Zhang X, *et al*. Genetics, lifestyle, and low-density lipoprotein cholesterol in young and apparently healthy women. *Circulation* 2018; 137:820–831.
22. Mariano C, Alves AC, Medeiros AM, *et al*. The familial hypercholesterolaemia phenotype: monogenic familial hypercholesterolaemia, polygenic hypercholesterolaemia and other causes. *Clin Genet* 2020; 97:457–466.
23. Mach F, Baigent C, Catapano AL, *et al*. 2019 ESC/EAS guidelines for the management of dyslipidaemias: lipid modification to reduce cardiovascular risk. *Atherosclerosis* 2019; 290:140–205.
24. Trinder M, Paquette M, Cermakova L, *et al*. Polygenic contribution to low-density lipoprotein cholesterol levels and cardiovascular risk in monogenic familial hypercholesterolemia. *Circ Genomic Precis Med* 2020; 13:515–523. In this study, the authors combined three separate cohorts to amass 1120 monogenic familial hypercholesterolemia patients in whom an LDL cholesterol (LDL-C) polygenic risk scores (PRS) was calculated. It was shown that patients who had high polygenic risk superimposed upon their familial hypercholesterolemia mutation had greatest cardiovascular disease (CVD) risk.
25. Trinder M, Francis GA, Brunham LR. Association of monogenic vs polygenic hypercholesterolemia with risk of atherosclerotic cardiovascular disease. *JAMA Cardiol* 2020; 5:390–399. In the UK Biobank the authors identified monogenic and polygenic familial hypercholesterolemia patients as well as patients with high cholesterol without identifiable genetic cause. Patients with monogenic familial hypercholesterolemia were at highest CVD risk, followed by those with polygenic hypercholesterolemia.
26. Khera AV, Chaffin M, Aragam KG, *et al*. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet* 2018; 50:1219–1224.
27. Wiegman A, Gidding SS, Watts GF, *et al*. Familial hypercholesterolaemia in children and adolescents: gaining decades of life by optimizing detection and treatment. *Eur Heart J* 2015; 36:2425–2437.
28. Natarajan P, Young R, Stitzel NO, *et al*. Polygenic risk score identifies subgroup with higher burden of atherosclerosis and greater relative benefit from statin therapy in the primary prevention setting. *Circulation* 2017; 135:2091–2101.
29. Marston NA, Kamanu FK, Nordio F, *et al*. Predicting benefit from evolocumab therapy in patients with atherosclerotic disease using a genetic risk score: results from the FOURIER trial. *Circulation* 2019; 141:616–623. Using data from the landmark FOURIER trial, the authors show that patients with high polygenic CVD risk (measured by two separate PRS) derive greater absolute and relative risk reduction from LDL-C lowering by a mAb targeting PCSK9.
30. Kullo IJ, Jouni H, Austin EE, *et al*. Incorporating a genetic risk score into coronary heart disease risk estimates: effect on low-density lipoprotein cholesterol levels (the MI-GENES Clinical Trial). *Circulation* 2016; 133:1181–1188.