

Systematic Review

Genetic risk scores used in cardiovascular disease prediction models: a systematic reviewHyunok Yun¹, Nan Iee Noh², Eun Young Lee^{1,*}¹Department of Nursing, Catholic Kkottongnae University, 28211 Cheongju, Republic of Korea²Department of Nursing, Donggang University, 61200 Gwangju, Republic of Korea*Correspondence: eylee@kkot.ac.kr (Eun Young Lee)

Academic Editors: Brian Tomlinson and Takatoshi Kasai

Submitted: 2 November 2021 Revised: 12 December 2021 Accepted: 17 December 2021 Published: 11 January 2022

Abstract

Background: Cardiovascular disease is caused by a combination of genetic and environmental risk factors. Some risk factors can change with age, but a genetic predisposition is permanent. Therefore, identifying the genotype of cardiovascular disease and using it alone or in combination with existing risk algorithms can improve risk prediction. This systematic review was conducted to examine existing studies on predictive models for cardiovascular disease using genetic risk score and to determine the clinical utility. **Methods:** An electronic database search was conducted to identify studies published from January 2005 to July 2020. The literature search was performed using the search terms “coronary artery disease”, “coronary heart disease”, “cardiovascular diseases”, “genetic risk score”, and “polygenic risk score”. **Results:** Through systematic review, 29 studies were identified. In most studies, genetic risk score was associated with the incidence of cardiovascular disease. In 23 studies, clinical utility was improved based on discrimination between or reclassification of subjects who did and did not experience an event, but the improvement was modest. **Conclusions:** The predictive model for cardiovascular disease using genetic risk score has limited usefulness in clinical practice due to methodological heterogeneity of genetic risk score constructs. Further research to develop a standardized protocol of genetic risk score constructs and validation studies with various cohorts from diverse populations are required.

Keywords: Genetic risk score; Cardiovascular disease; Prediction model; Systematic review**1. Introduction**

Cardiovascular disease (CVD) is the leading cause of mortality worldwide, causing an estimated 17.9 million deaths each year [1]. Early and accurate identification of individuals at high risk of CVD facilitates timely prevention and treatment and can lower public health costs by reducing unnecessary disease burden [2].

Conventional risk scores such as the Framingham Risk Score (FRS) [3], the American College of Cardiology/American Heart Association 2013 risk score (ACC/AHA13) [4], and QRESEARCH cardiovascular risk (QRISK1 and QRISK2) [5] have been developed and used in clinics. Conventional risk scores are useful for both the individual and the clinician by identifying individuals at increased risk of future cardiovascular events and helping them select appropriate lifestyle modifications and preventive medical treatment [6]. However, these risk scores focus on relatively short-term risk (5–10 years), which is insufficient to identify people with subclinical disease [7,8]. In particular, conventional risk scores might not identify individuals at a younger age who could likely attain long-term benefits [9].

The cause of CVD is a combination of genetic and environmental risk factors [10]. Some risk factors can change with age; however, genetic predisposition is permanent. Therefore, identifying genotypes for CVD and us-

ing them alone or in combination with existing risk algorithms can improve risk prediction [11]. To associate genotype and phenotype, genetic researchers have performed many genome-wide association studies (GWAS) and have made significant advances in identifying CVD-associated genetic variations/single nucleotide polymorphisms (SNPs) [12]. In 2007, the use of multi-location genetic risk scores (GRSs) was proposed to integrate the relatively small effects of individual genes and to better improve the accuracy of conventional risk scores [13]. Because GRSs allow analysis of high genetic risk at any age, people at higher risk for the disease can be identified before clinical signs appear [14]. Therefore, using GRSs for CVD prediction can help detect and prevent disease earlier.

The clinical utility of GRSs for CVD prediction depends not on the strength of their association with typical CVDs, but on their capability to predict future CVD events [15]. Although research on GRSs for CVD has been progressing, the effects of GRSs on clinical decision-making are unclear and the predictive power remains limited. Comparison of GRSs that were added to well-established and validated risk prediction models is insufficient [16] and has shown mixed results. Because information on the current development of CVD prediction models using the GRS and comprehensive evaluation based on the models are lacking, systematic efforts are needed. Therefore, in the present



study, the methodological characteristics of individual studies were identified, and the clinical utility of the GRS prediction model was evaluated by systematic review of the literature on CVD prediction models using GRSs.

2. Methods

This review was conducted and reported in accordance with the Preferred Reporting Items for Systematic Reviews (PRISMA) guideline [17].

2.1 Search strategy

To retrieve published studies for this review, systematic searches were conducted using three electronic databases: PubMed, Embase, and SCOPUS. We attempted to reconcile the definition of CVD prior to selecting a search term. CVD is a general term for conditions affecting the heart or blood vessels. The major contributing factor to CVD is atherosclerosis, which is narrowing of the arteries resulting from subendothelial deposition of cholesterol, cholesterol esters, and calcium within the vessel walls. Rupture of atherosclerotic plaques yields blood clots that result in myocardial infarction or stroke [18,19]. Based on these mechanisms, CVD was defined as a composite of coronary heart death, stable or unstable angina, fatal or non-fatal myocardial infarction, coronary artery bypass grafting or percutaneous coronary intervention, and ischemic stroke events. Depending on the type of CVD in the included study, coronary artery disease (CAD) or coronary heart disease (CHD) was sometimes used instead of CVD, and all terms were included in the search. The following MeSH terms and keywords were used: (“coronary artery disease” OR “coronary heart disease” OR “cardiovascular diseases” OR “ischemic heart disease” OR “angina pectoris” OR “myocardial infarction” OR “stroke”) AND (“genetic risk score” OR “polygenic risk score” OR “genome-wide association study”). Studies were searched from January 2005, when a public database of common variations in the human genome was reported, to July 2020. To identify additional studies, the bibliography of each included study was searched manually.

2.2 Selection criteria

After elimination of duplicates, studies related to the subject were screened through titles and abstracts. Next, studies were selected by full-text review based on the inclusion and exclusion criteria. Three authors independently selected the studies, and inconsistencies were resolved through discussion.

The inclusion and exclusion criteria were as follows: (1) The study population was adults dwelling in the community. Studies in which subjects were recruited from hospitals or clinical trials were excluded. Studies focusing on animals, children, and diseased adults were excluded. Although some studies began when subjects were children, studies in which disease occurred in adulthood were in-

cluded. (2) After the observation period, only those outcomes referring to CVD were considered. However, hemorrhagic stroke was excluded due to differences in the underlying pathology. (3) Studies in which GRSs were used to predict CVD were included. All studies using GRSs consisting of direct and intermediate risk factors for CVD were included, and studies using only a single SNP or found only an association between GRS and CVD were excluded. (4) Studies were published in academic journals in English.

2.3 Data extraction

Two authors (HY and NIN) extracted data using a standardized form. To increase the accuracy of coding and data entry, the other author (EYL) independently verified all extracted data. The following items were extracted: Study characteristics (cohort name, ethnic group, sample size, age [mean], % of females, incident CVD, and follow-up period); development of a GRS (reference for SNP selection, selected phenotypes, number of SNPs used in GRS construction, and GRS calculation method); and evaluation of GRSs to predict CVD risk (base model for comparison with GRSs, whether family history was included in the base model, association between GRS and CVD, risk discrimination and reclassification to determine clinical utility of the GRS for CVD).

2.4 Assessment of risk bias

Each selected study was assessed for risk of bias using the Risk of Bias Assessment Tool for Non-randomized Studies (RoBANS 2.0) developed in 2011 by Health Insurance Review and Assessment Service (Wonju-Si, Republic of Korea) [20]. Evaluation items included target group comparability, target group selection, confounding variables, measurement of exposure, evaluator blinding, evaluation of results, incomplete data, and selective outcome reporting. Individual studies were rated as “high”, “low”, or “unclear” with respect to bias and were assessed independently by the three authors. Where inconsistency was noted, a consensus was reached through discussion among the authors.

3. Results

The search resulted in 27,485 studies retrieved; 28 studies were selected based on the inclusion and exclusion criteria and one study was added from the manual search. Fig. 1 shows the search and selection processes.

Table 1 (Ref. [9,13,15,21–46]) shows the characteristics of the 29 studies. The sample size of the cohort used in the analysis varied from 1306 to 482,629 (median 6041). Studies in which prediction of CVD using GRSs was conducted primarily included Caucasian or European-ancestry populations, and only three studies included Asians (included as a category in one study).

Table 2 (Ref. [9,13,15,21–46]) shows how GRSs were developed in the reviewed studies. To construct GRSs,

Table 1. Characteristics of included studies.

First author (year)	Cohort name	Ethnic group	Sample size	Age (mean)	% Female	Incidence of CVD	Follow up (year)
Morrison A.C. (2007) [13]	ARIC	Black and white	13,907	54.0	56.6	1452	13 (median)
Kathiresan S. (2008) [21]	Malmö Diet and Cancer Study	European	5414	58.6	58.5	238	10.6 (median)
Paynter N.P. (2010) [22]	WGHS	White	19,313	52.8 (median)	100.0	777	12.3 (median)
Ripatti S. (2010) [23]	FINRISK 1992, 1997, 2002, Health 2000, MDC	European	30,725	49.8	53.4	1264	10.7 (median)
Vaarhorst A.A.M. (2012) [24]	CAREMA	European	2559	46.7	46.8	646	12.1 (median)
	ARIC		8542	54.1	54.9	1110	18 (maximum)
Brautbar A. (2012) [25]	Rotterdam	non-Hispanic whites	2068	60.3	51.8	271	18 (maximum)
	Framingham Offspring Study		2339	40.4	56.4	215	35 (maximum)
	REGICOR		2351	53.9	52.2	107	9.7 (mean)
Lluis-Ganella C. (2012) [26]	Framingham Heart Study	European Caucasians	3537	56.0	56.5	429	13.3 (mean)
Thanassoulis G. (2012) [27]	Framingham Offspring Study	Multi-ethnic	3014	49.0	54.0	539	11 (median)
Hughes M.F. (2012) [28]	MORGAM project_9 cohorts	European	4818	45.2–62.7	0.0	1736	18 (median)
Ganna A. (2013) [29]	SATSA, OCTO-Twin, GENDER, HARMONY, ULSAM, WTCCC	Caucasian or European	10,612	67.5	50.0	781	4.3 (median)
Havulinna A.S. (2013) [30]	FINRISK 1992, 1997, 2002, 2007, Health 2000, OPERA, HBCS	European	32,669	49.1	54.7	2295	9.8 (median)
Tikkanen E. (2013) [31]	FINRISK, Health 2000	European	24,124	47.1	54.6	3376	12 (median)
	WHII		5059	36.4	26.4	300	10
Shah S. (2013) [32]	BWHHS	Caucasian	3414	68.8	100.0	342	8
Benjamin A.G. (2014) [33]	ARIC	White/Europeans	8491	54.0	55.0	620	10 (maximum)
Beaney K.E. (2015) [34]	NPHSII	European	2775	56.1	0.0	284	13.5 (median)
Kraru N.T. (2015) [35]	Inter99	European	6041	46.2	51.0	374	11.6 (median)
De Vries P.S. (2015) [36]	Rotterdam	European	5899	69.3	59.1	964	12.8 (mean)
Lu X. (2015) [37]	InterASIA, CCHS	Asian	26,262	50.9	55.3	1078	6.9 (mean)
Iribarren C. (2016) [38]	GERA	non-Hispanic whites	51,954	59.4	66.8	1864	5.9 (mean)
	FINRISK		12,676	46.0	54.0	757	13.6 (mean)
Abraham G. (2016) [9]	Framingham Heart Study	Multi-ethnic	3406	44.3	55.0	587	31.4 (mean)
Tada H. (2016) [39]	Malmö Diet and Cancer Study	European	23,595	58.1	62.0	2213	14.4 (median)
Antiochos P. (2016) [40]	CoLaus study	Caucasian	4283	53.4	53.0	133	5.6 (median)
Beaney K.E. (2017) [41]	NPHSII	European	2775	56.1	0.0	284	13.5 (median)
Chang X. (2017) [42]	SCHS	Asian	1306	64.1	36.0	395	8.6 (mean)
Iribarren C. (2018) [43]	GERA	non-White or Minority descent	11,242	55.8	70.4	450	8.7 (median)
Inouye M. (2018) [44]	UK Biobank	European	482,629	56.5	54.4	12,513	6.2 (mean)
Liu R. (2019) [45]	ARIC	European and African american	10,578	54.2	53.5	836	NA
	ARIC		4847	62.9	56.4	696	15.5 (median)
Mosley J.D. (2020) [15]	MESA	European ancestry	2390	61.8	52.2	227	14.2 (median)
Elliott J. (2020) [46]	UK Biobank	European ancestry	352,660	55.9	58.2	6272	8 (median)

NA, Not Available.

Table 2. Development of genetic risk scores.

First author (year)	Reference for SNPs	Phenotype	No. of SNPs	Method of GRS
Morrison A.C. (2007) [13]	GWAS	CHD	11	Unweighted
Kathiresan S. (2008) [21]	GWAS	Lipid	9	Unweighted
Paynter N.P. (2010) [22]	NHGRI catalog ^a	CVD/ Intermediate risk factors	Risk factors 101 CVD 12	Unweighted
Ripatti S. (2010) [23]	GWAS consortium ^b /GWAS	CHD or MI	13	Weighted
Vaarhorst A.A.M. (2012) [24]	NHGRI catalog	CHD/Intermediate risk factors	Overall 179 Risk factors 153 CHD 29	Unweighted/Weighted
Brautbar A. (2012) [25]	NHGRI catalog/GWAS	CHD	13	Unweighted/Weighted
Lluis-Ganella C. (2012) [26]	NHGRI catalog	CHD	8	Weighted
Thanassoulis G. (2012) [27]	GWAS consortium/GWAS	CHD or MI/Intermediate risk factors	CHD or MI 13 Risk factors 89 Overall 102	Weighted
Hughes M.F. (2012) [28]	GWAS consortium/GWAS	CHD or MI	13/15/8	Weighted
Ganna A. (2013) [29]	NHGRI catalog	CHD/Intermediate risk factors	Overall 395 CHD 46	Unweighted/Weighted
Havulinna A.S. (2013) [30]	GWAS consortium	BP	32	Weighted
Tikkanen E. (2013) [31]	GWAS consortium	CHD or MI	28	Weighted
Shah S. (2013) [32]	GWAS	Lipid	TC: 19 LDL: 21 HDL: 12 TG: 16	Unweighted
Benjamin A. G. (2014) [33]	Cplus4D consortium ^c	CHD/Intermediate risk factors	Overall 50 Risk factors 17 Non-risk factors 33	Weighted
Beane K.E. (2015) [34]	GWAS consortium/GWAS	CHD	13/19	Weighted
Kraru N.T. (2015) [35]	Cplus4D consortium	CAD	45	Unweighted
De Vries P.S. (2015) [36]	Cplus4D consortium	CHD	49/103/152	Weighted
Lu X. (2015) [37]	GWAS	BP	22	Weighted
Iribarren C. (2016) [38]	Cplus4D consortium	CAD/ Intermediate risk factors	CAD 8/12/36 Overall 51	Weighted
Abraham G. (2016) [9]	Cplus4D consortium	CHD	49,310	Weighted
Tada H. (2016) [39]	GWAS	CHD	27/50	Weighted
Antiochos P. (2016) [40]	Cplus4D consortium/GWAS	CHD	38/53/153	Unweighted/Weighted
Beane K.E. (2017) [41]	GWAS consortium/GWAS	CHD	19/21	Weighted
Chang X. (2017) [42]	NHGRI catalog	CHD/ Intermediate risk factors	156	Unweighted
Iribarren C. (2018) [43]	Cplus4D consortium	CAD/Intermediate risk factors	12/51	Weighted
Inouye M. (2018) [44]	GWAS consortium	CAD	1,745,180	Weighted
Liu R. (2019) [45]	GWAS	CAD	267	Weighted
Mosley J.D. (2020) [15]	GWAS consortium	CHD	6,630,149	Weighted
Elliott J. (2020) [46]	Cplus4D consortium	CAD	1,037,385	Weighted

^a NHGRI catalog: The NHGRI-EBI Catalog of human genome-wide association studies.

^b GWAS consortium: combining two or more of “International Consortium for Blood Pressure GWAS” or “Global BPgen” or “The Cohorts for Heart and Aging Research in Genomic Epidemiology” or “Myocardial Infarction Genetics” or “Coronary Artery Disease”.

^c Cplus4D consortium: CARDIoGRAMplusC4D GWAS consortium.

SNP, Single nucleotide polymorphism; GRS, Genetic risk score; GWAS, Genome-wide association studies; CHD, Coronary heart disease; CVD, Cardiovascular disease; MI, Myocardial infarction; BP, Blood pressure; TC, Total cholesterol; LDL, Low density lipoproteins; HDL, High density lipoproteins; TG, Tri-glyceride; CAD, Coronary artery disease.

Table 3. Evaluation of genetic risk score for predicting cardiovascular disease risk.

First author (year)	Base model	Family history	Association between GRS and CVD (HR)	Discrimination and risk reclassification
Morrison A.C. (2007) [13]	ACRS	-	Whites 1.10*/Blacks 1.20*	Discrimination: improved
Kathiresan S. (2008) [21]	ATPIII	+	1.15*	Discrimination: not improved Risk reclassification: improved
Paynter N.P. (2010) [22]	ATPIII Reynolds	only Reynolds	NS	not improved
Ripatti S. (2010) [23]	TRF	+	CHD 1.66*/CVD 1.50*/MI 1.46*	not improved
Vaarhorst A.A.M. (2012) [24]	TRF	+	Weighted CHD GRS 1.12*	Discrimination: not improved Risk reclassification: improved in weighted CHD GRS only improved in ARIC
Bratbar A. (2012) [25]	ARIC, Rotterdam: - ACRS Framingham off spring: FRS	-	Unweighted GRS*/Weighted GRS*	
Lluis-Ganella C. (2012) [26]	TRF	+	REGICOR 1.21*/Framingham 1.07*	Discrimination: only improved in Framingham Risk reclassification: improved in the intermediate risk group
Thanassoulis G. (2012) [27]	TRF	+	13-GRS 1.05*/102-GRS 1.00 (NS)	not improved
Hughes M.F. (2012) [28]	FRS	subgroup (men NA aged 50–59)	NA	Discrimination: partially improved Risk reclassification: improved
Ganna A. (2013) [29]	FRS	-	Overall 1.54*/CHD-specific GRS 1.52*	Subgroup: improved discrimination and risk reclassification
Havulinna A.S. (2013) [30]	FRS	-	SBP GRS 1.23*/DBP GRS 1.26*	Discrimination: CHD-specific GRS improved Risk reclassification: improved
Tikkanen E. (2013) [31]	TRF	+	CHD 1.27*/CVD 1.18*/ACS 1.27*	Discrimination: not improved Risk reclassification: improved in the intermediate risk group
Shah S. (2013) [32]	FRS	-	WHII† all lipid GRS (NS) BWHHS† TC GRS 1.37*/LDL GRS 1.44*	not improved
Benjamin A. G. (2014) [33]	TRF	-	Full GRS‡ 1.28*/Non-risk GRS‡ 1.29*/Risk GRS‡ 1.05(NS)	Discrimination: improved
Beane K.E. (2015) [34]	FRS	-	19-GRS† 1.43 (NS)/13-GRS† 2.31*	Discrimination: not improved Risk reclassification: improved in the intermediate risk group
Krurup N.T. (2015) [35]	European SCORE risk factors	-	MI 1.06*/CAD 1.01 (NS)	Discrimination: improved Risk reclassification: not improved
De Vries P.S. (2015) [36]	TRF	+	49-GRS 1.11*/103-GRS 1.07*/152-GRS 1.13*	not improved
Lu X. (2015) [37]	TRF	+	CVD 1.11*	improved
Iribarren C. (2016) [38]	FRS	-	8-GRS 1.21*/12-GRS 1.20*/36-GRS 1.23*/51-GRS 1.23*	improved
Abraham G. (2016) [9]	FRS ACC/AHA	only FINRISK	FINRISK 1.74*/FHS 1.28*	Discrimination: improved Risk reclassification: improved in FINRISK
Tada H. (2016) [39]	TRF	+	27-GRS 1.70*/50-GRS 1.92*	Discrimination: improved Risk reclassification: not improved
Antiochos P. (2016) [40]	TRF	+	Unweighted 38-GRS 1.16 (NS)/53-GRS 1.25*/153-GRS 1.38*/Weighted 38-GRS 1.29*/53-GRS 1.39*/153-GRS 1.50*	Discrimination: improved Risk reclassification: weight 153-GRS improved
Beane K.E. (2017) [41]	QRISK2	+	19-GRS† 1.31*/21-GRS† 1.39*	Discrimination: 19-GRS improved Risk reclassification: 19-GRS improved
Chang X. (2017) [42]	ATPIII	-	Men 4.16 */Women 2.82 *	Discrimination: improved Risk reclassification: improved
Iribarren C. (2018) [43]	FRS	+	12-GRS African-Americans 1.86*/Latinos 1.52*/Asians 1.19 (NS)/All minority Groups 1.48*/51-GRS All minority Groups 1.43*	Discrimination: not improved Risk reclassification: improved partially
Inouye M. (2018) [44]	TRF	+	1.71*	Discrimination: improved
Liu R. (2019) [45]	TRF	-	European 1.32*/African 1.11 (NS)	Discrimination: improved in the European
Mosley J.D. (2020) [15]	Revised ACC/AHA	-	ARIC 1.24*/MESA 1.38*	not improved
Elliott J. (2020) [46]	TRF	+	1.32*	improved

† OR, Odds Ratio; ‡ RR, Relative risk; * significant.

GRS, Genetic risk score; CVD, Cardiovascular disease; HR, Hazard ratios; ACRS, ARIC CHD risk score; ATPIII, Adult treatment panel III; NS, not significant; TRF, Traditional risk factor; CHD, Coronary heart disease; MI, Myocardial infarction; FRS, Framingham heart study risk score; NA, not available; SBP, systolic blood pressure; DBP, diastolic blood pressure; ACS, Acute coronary syndrome; TC, Total cholesterol; LDL, Low density lipoproteins; CAD, Coronary artery disease; ACC/AHA, American College of Cardiology/American Heart Association risk score; QRISK2, QRESEARCH cardiovascular risk 2014 version.

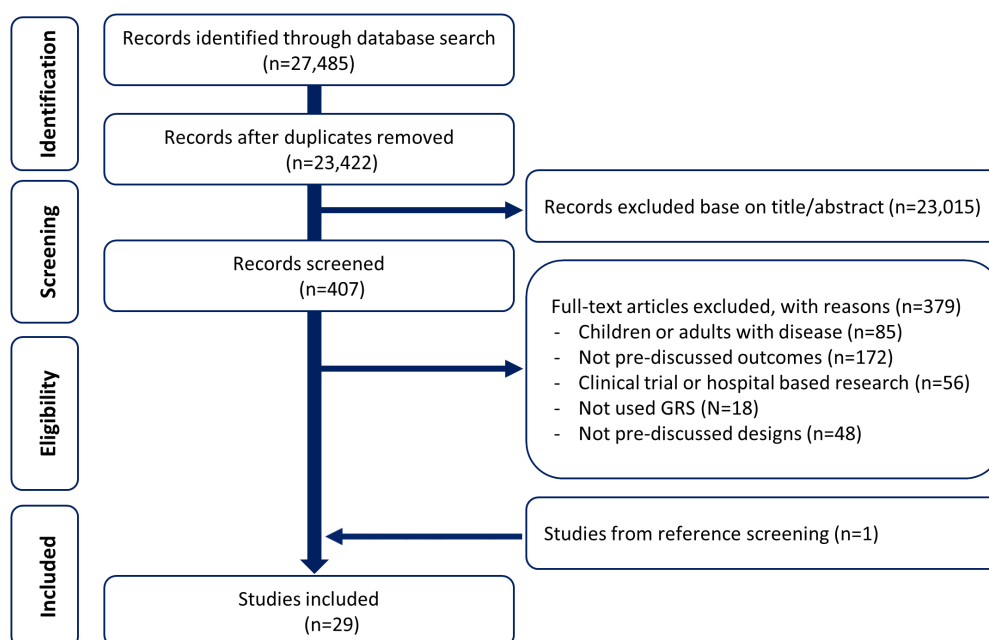


Fig. 1. PRISMA flow chart for inclusion in the systematic review. PRISMA, Preferred Reporting Items for systematic reviews and Meta-analysis.

most studies have used large-scale data such as the National Human Genome Research Institute (NHGRI) catalog or large international consortia, such as CARDIoGRAM-plusC4D (Coronary ARtery Disease Genome-wide Replication and Meta-analysis plus The Coronary Artery Disease). The number of SNPs used in GRSs ranged from 8–6,630,149. In the past, only a limited number of SNPs associated with CVD or intermediate risk factors were used. In recent years, however, millions of SNPs have been used for GRS calculations. In 19 studies, GRSs were calculated by assigning weights based on the effect of each SNP, and in six studies, a simple count of the total number of risk alleles was used. Three studies analyzed both weighted and unweighted values.

Table 3 (Ref. [9,13,15,21–46]) shows the comparison between base models and models with added GRS. In most base models, covariates included age, systolic blood pressure, total cholesterol, high-density lipoproteins, diabetes, and smoking status. In addition, sex, body mass index, diastolic blood pressure, lipid-lowering and antihypertensive agents, and serum markers were added or omitted depending on the existing risk score used (data not shown). In 17 studies, family history was included in the base model. An association between incidence of CVD and GRS was apparent in 27 studies. In a number of studies, C-statistic, the net reclassification index (NRI), or integrated discrimination index (IDI) was calculated to assess improvement between the base models and models with added GRS. In examination of C-statistic, discrimination was improved in some or all models in 18 studies. NRI results showed im-

proved risk reclassification in some or all of the 17 studies. In two of those, classification improved only in the intermediate risk group. In 11 studies, clinical utility was confirmed by showing improvements in both discrimination and reclassification. The C-statistic value that discriminated CVD was 0.650 to 0.880 (mean \pm SD]: 0.751 \pm 0.057], median: 0.747) in the base models and 0.640 to 0.881 (mean \pm SD]: 0.756 \pm 0.057], median: 0.753) in models with added GRS. Increments range from –0.030 to 0.043 (mean \pm SD]: 0.006 \pm 0.010)].

On assessment of risk bias in each selected study, one study showed high bias in the ‘incomplete data’ category and another study showed high bias in the ‘selective outcome’ category. However, we included the two studies because the bias was insufficient to question the quality of the study results.

4. Discussion

This study was conducted to systematically review existing studies in which GRSs were used for CVD prediction and to determine the clinical relevance. Based on a systematic review process, 29 studies were identified. The GRSs developed in the reviewed studies were associated with incidence of CVD. A total of 23 studies showed clinical utility by improving discrimination or reclassification between subjects who did and did not experience an event.

The association between incidence of CVD and GRS implied that a genetic signal was present among the selected markers, and GRSs can be used for predicting of individual trait. Although individual SNPs have minimal effect

on CVD prediction, GRSs containing multiple SNPs potentially can be a strong predictor of disease [47].

In the present study, the ability of GRS predictive models to discriminate was improved, but the improvement was modest. The first procedure in generating a GRS is 'variable selection' to determine which SNPs should be included in the model [48]. In GWAS for CVD, 163 loci were reported through 2018, after the discovery of chromosome 9p21 risk locus in 2007. In addition, over 300 additional loci with false discovery rate values $<5\%$ indicate CAD risk and might be useful for improving CAD risk prediction [49]. In several studies, genome predictive models that consider all accessible genetic variants were shown to identify more efficiently individuals at high risk of complex diseases [50,51]. More recently, very large GRSs have been constructed using more than 1 million SNPs. However, because large GRSs included many SNPs below the genome-wide significance threshold for association with CVD, many SNPs might not contribute to the explanatory power of GRSs [15]. Our review shows a recent trend in constructing GRS using a large number of SNPs, but there has been no noticeable trend for predictive ability. These inconsistent results prevent prediction of the number of SNPs required for accurate and robust GRSs for CVD. Although there is a number of challenges in this regard [52], due to the genetic structure of CVD, much larger sample sizes will be required to detect a sufficiently large number of variants to make meaningful contributions to risk prediction models and to construct useful predictive risk scores [53]. Selecting the best set of truly susceptibility SNPs to increase the impact of GRS on clinical decision-making is likely to be stable only after GWAS has reached huge sample sizes containing hundreds of thousands of individuals [54].

The choice of phenotypes for deriving SNPs to be used for GRSs should be considered. GRSs can be constructed using SNPs that are clinically disease-associated [55]. In the current review, studies were included in which GRSs consisting of only CVD-associated SNPs, as well as intermediate risk factor-associated SNPs, or a combination of the two types of SNPs were used. Although GRSs consisting of only CVD-associated SNPs as well as CVD plus intermediate risk factor-associated SNPs showed improvement in discrimination over conventional risk scores, GRSs consisting of only CVD-associated SNPs were the best predictor of CVD. These results indicate that intermediate risk factor-associated SNPs do not improve the prediction of CVD. Although intermediate traits might not be useful for predicting individual future risk of CVD, studies in which the risk variants linked to underlying causal genes are evaluated could identify new therapeutic targets for preventing disease [49]. Early discrimination of dyslipidemia patients with or without CVD can lead to timely treatment with lipid-lowering drugs and consequently lower the risk of CVD to the level equivalent to that of the general public [56]. Therefore, further studies are needed to identify SNPs

strongly associated with symptoms, because SNPs associated with intermediate risk factors might be useful for explaining variations in the subclinical phenotype [57].

A family history is not always a risk factor, but can be easily identified [58]. According to the American Cholesterol Guidelines, family history of CVD is a relative indicator in the evaluation for primary preventive support to reinforce statin recommendations [59]. In several studies, family history was included in the base model to identify genetic associations. However, inclusion of family history did not lead to a difference in predicted values. Phenotypes are the result of both genetic and environmental interactions [60]. Significant risk factors due to the shared nature of family genes should be elucidated in future genetic studies.

Furthermore, studies in which CVD was predicted using GRSs mainly included subjects of Caucasian or European ancestry and only three studies included Asians. The predictive capacity and diagnostic accuracy of findings in GWAS show biases when tested in non-European cohorts using GRSs derived from European-based GWAS. Until recently, 80% of the subjects involved in genetic studies were of European ancestry, 14% were Asians, and 6% were others [61]. The involvement of individuals from diverse ethnicities in medical genomics is needed to evaluate the link between disease and related genetic variants for various populations for use in generalized risk prediction models [62].

The current review was limited to participants without CVD or intermediate risk disease in a population-based cohort because hyperselection of patients with these diseases can overestimate the effect size and prediction value [63]. Although meta-analysis is the best evidence-based method to confirm the clinical utility of predictive models, such research has not been possible due to the methodological heterogeneity of GRSs between studies. Studies published after the literature search in 2020 were not included, and studies for which inclusion criteria were not identified through the abstract and title were likely excluded, which is another limitation. Despite these limitations, current development of CVD prediction models using the GRS and comprehensive evaluation based on models were systematically reviewed to emphasize the use of genetic information in predictive models for CVD traits. The findings would be helpful for future investigations and clinically useful if considered in the appropriate context.

5. Conclusions

Based on the results obtained in this review, GRSs were a significant predictor of CVD, and the predictive ability was improved but modest compared with traditional models. However, the methodological heterogeneity was too high to use the model as a guideline. The slight improvement and methodological heterogeneity of the predictive model limit the generalization of GRS as predictor and the application of GRS predictive models in clinical prac-

tice. Therefore, further research is needed to develop a standardized protocol of GRS constructs and to validate the findings in various cohorts from diverse populations.

Author contributions

HY and EYL designed the research. All authors contributed towards analyzing the data (HY, EYL, NIN), and HY wrote this manuscript. All authors read, made changes and approved the final version of the manuscript.

Ethics approval and consent to participate

Not applicable.

Acknowledgment

Thanks to everyone who contributed to this study.

Funding

This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIT) (No. 2020R1A2C1014449).

Conflict of interest

The authors declare no conflict of interest.

References

- [1] World Health Organization. World Health Statistics 2021: monitoring health for the SDGs, sustainable development goals. Geneva: World Health Organization. 2021; 28-31. Available at: <https://www.who.int/data/gho/publications/world-health-statistics> (Accessed: 8 July 2021).
- [2] De La Vega FM, Bustamante CD. Polygenic risk scores: a biased prediction? *Genome Medicine*. 2018; 10: 100.
- [3] Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation*. 1998; 97: 1837-1847.
- [4] Goff DC, Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, Gibbons R, *et al.* 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. *Journal of the American College of Cardiology*. 2014; 63: 2935-2959.
- [5] Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, Brindle P. Performance of the QRISK cardiovascular risk prediction algorithm in an independent UK sample of patients from general practice: a validation study. *Heart*. 2008; 94: 34-39.
- [6] Siontis GCM, Tzoulaki I, Siontis KC, Ioannidis JPA. Comparisons of established risk prediction models for cardiovascular disease: systematic review. *British Medical Journal*. 2012; 344: 14.
- [7] Leiter LA, Fitchett DH, Gilbert RE, Gupta M, Mancini GJ, McFarlane PA, *et al.* Cardiometabolic risk in Canada: a detailed analysis and position paper by the cardiometabolic risk working group. *Canadian Journal of Cardiology*. 2011; 27: e1-e33.
- [8] Lloyd-Jones DM, Wilson PW, Larson MG, Beiser A, Leip EP, D'Agostino RB, *et al.* Framingham risk score and prediction of lifetime risk for coronary heart disease. *The American Journal of Cardiology*. 2004; 94: 20-24.
- [9] Abraham G, Havulinna AS, Bhalala OG, Byars SG, De Livera AM, Yetukuri L, *et al.* Genomic prediction of coronary heart disease. *European Heart Journal*. 2016; 37: 3267.
- [10] Marenberg ME, Risch N, Berkman LF, Floderus B, de Faire U. Genetic susceptibility to death from coronary heart disease in a study of twins. *The New England Journal of Medicine*. 1994; 330: 1041-1046.
- [11] Lambert SA, Abraham G, Inouye M. Towards clinical utility of polygenic risk scores. *Human Molecular Genetics*. 2019; 28: R133-R142.
- [12] Musunuru K, Kathiresan S. Genetics of Common, Complex Coronary Artery Disease. *Cell*. 2019; 177: 132-145.
- [13] Morrison AC, Bare LA, Chambless LE, Ellis SG, Malloy M, Kane JP, *et al.* Prediction of coronary heart disease risk using a genetic risk score: the Atherosclerosis Risk in Communities Study. *American Journal of Epidemiology*. 2007; 166: 28-35.
- [14] Roberts R, Chang CC, Hadley T. Genetic Risk Stratification: A Paradigm Shift in Prevention of Coronary Artery Disease. *JACC: Basic to Translational Science*. 2021; 6: 287-304.
- [15] Mosley JD, Gupta DK, Tan J, Yao J, Wells QS, Shaffer CM, *et al.* Predictive Accuracy of a Polygenic Risk Score Compared with a Clinical Risk Score for Incident Coronary Heart Disease. *The Journal of the American Medical Association*. 2020; 323: 627.
- [16] Elliott J, Bodinier B, Bond TA, Chadeau-Hyam M, Evangelou E, Moons KGM, *et al.* Predictive Accuracy of a Polygenic Risk Score-Enhanced Prediction Model vs a Clinical Risk Score for Coronary Artery Disease. *The Journal of the American Medical Association*. 2020; 323: 636.
- [17] Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, *et al.* The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Journal of Clinical Epidemiology*. 2009; 62: e1-e34.
- [18] Allen MT, Patterson SM. Hemoconcentration and stress: a review of physiological mechanisms and relevance for cardiovascular disease risk. *Biological Psychology*. 1995; 41: 1-27.
- [19] Trinder M, Uddin MM, Finneran P, Aragam KG, Natarajan P. Clinical Utility of Lipoprotein(a) and LPA Genetic Risk Score in Risk Prediction of Incident Atherosclerotic Cardiovascular Disease. *JAMA Cardiology*. 2021; 6: 287.
- [20] Kim SY, Seo HJ, Lee YJ, Park JE. Study Design Algorithm for Medical Literature of Intervention (DAMI) and Risk of Bias for Non-randomized studies (RoBANS) ver 2.0. *Health Insurance Review & Assessment Service*. 2013: 65-103.
- [21] Kathiresan S, Melander O, Anevski D, Guiducci C, Burt NP, Roos C, *et al.* Polymorphisms associated with cholesterol and risk of cardiovascular events. *New England Journal of Medicine*. 2008; 358: 1240-1249.
- [22] Paynter NP, Chasman DI, Paré G, Buring JE, Cook NR, Miletich JP, *et al.* Association between a literature-based genetic risk score and cardiovascular events in women. *The Journal of the American Medical Association*. 2010; 303: 631-637.
- [23] Ripatti S, Tikkanen E, Orho-Melander M, Havulinna AS, Silander K, Sharma A, *et al.* A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. *Lancet*. 2010; 376: 1393-1400.
- [24] Vaarhorst AAM, Lu Y, Heijmans BT, Dollé MET, Böhringer S, Putter H, *et al.* Literature-Based Genetic Risk Scores for Coronary Heart Disease: the Cardiovascular Registry Maastricht (CAREMA) prospective cohort study. *Circulation: Cardiovascular Genetics*. 2012; 5: 202-209.
- [25] Brautbar A, Pompeii LA, Dehghan A, Ngwa JS, Nambi V, Virani SS, *et al.* A genetic risk score based on direct associations with coronary heart disease improves coronary heart disease risk prediction in the Atherosclerosis Risk in Communities (ARIC), but not in the Rotterdam and Framingham Offspring, Studies. *Atherosclerosis*. 2012; 223: 421-426.
- [26] Lluís-Ganella C, Subirana I, Lucas G, Tomás M, Muñoz D, Sentí M, *et al.* Assessment of the value of a genetic risk score in improving the estimation of coronary risk. *Atherosclerosis*. 2012; 222: 456-463.
- [27] Thanassoulis G, Peloso GM, Pencina MJ, Hoffmann U, Fox CS, Cupples LA, *et al.* A genetic risk score is associated with incident cardiovascular disease and coronary artery calcium: the Framingham Heart Study. *Circulation: Cardiovascular Genetics*. 2012; 5: 113-121.
- [28] Hughes MF, Saarela O, Stritzke J, Kee F, Silander K, Klopp N, *et al.* Genetic markers enhance coronary risk prediction in men: the MORGAM prospective cohorts. *PLoS ONE*. 2012; 7: e40922.
- [29] Ganna A, Magnusson PKE, Pedersen NL, de Faire U, Reilly M, Arnlöv J, *et al.* Multilocus genetic risk scores for coronary heart

disease prediction. *Arteriosclerosis, Thrombosis, and Vascular Biology*. 2013; 33: 2267–2272.

- [30] Havulinna AS, Kettunen J, Ukkola O, Osmond C, Eriksson JG, Kesäniemi YA, *et al*. A blood pressure genetic risk score is a significant predictor of incident cardiovascular events in 32,669 individuals. *Hypertension*. 2013; 61: 987–994.
- [31] Tikkanen E, Havulinna AS, Palotie A, Salomaa V, Ripatti S. Genetic risk prediction and a 2-stage risk screening strategy for coronary heart disease. *Arteriosclerosis, Thrombosis, and Vascular Biology*. 2013; 33: 2261–2266.
- [32] Shah S, Casas JP, Gaunt TR, Cooper J, Drenos F, Zabaneh D, *et al*. Influence of common genetic variation on blood lipid levels, cardiovascular risk, and coronary events in two British prospective cohort studies. *European Heart Journal*. 2013; 34: 972–981.
- [33] Goldstein BA, Knowles JW, Salfati E, Ioannidis JPA, Assimes TL. Simple, standardized incorporation of genetic risk into non-genetic risk prediction tools for complex traits: coronary heart disease as an example. *Frontiers in Genetics*. 2014; 5: 254.
- [34] Beaney KE, Cooper JA, Ullah Shahid S, Ahmed W, Qamar R, Drenos F, *et al*. Clinical Utility of a Coronary Heart Disease Risk Prediction Gene Score in UK Healthy Middle Aged Men and in the Pakistani Population. *PLoS ONE*. 2015; 10: e0130754.
- [35] Krarup NT, Borglykke A, Allin KH, Sandholt CH, Justesen JM, Andersson EA, *et al*. A genetic risk score of 45 coronary artery disease risk variants associates with increased risk of myocardial infarction in 6041 Danish individuals. *Atherosclerosis*. 2015; 240: 305–310.
- [36] de Vries PS, Kavousi M, Ligthart S, Uitterlinden AG, Hofman A, Franco OH, *et al*. Incremental predictive value of 152 single nucleotide polymorphisms in the 10-year risk prediction of incident coronary heart disease: the Rotterdam Study. *International Journal of Epidemiology*. 2015; 44: 682–688.
- [37] Lu X, Huang J, Wang L, Chen S, Yang X, Li J, *et al*. Genetic predisposition to higher blood pressure increases risk of incident hypertension and cardiovascular diseases in Chinese. *Hypertension*. 2015; 66: 786–792.
- [38] Iribarren C, Lu M, Jorgenson E, Martínez M, Lluís-Ganella C, Subirana I, *et al*. Clinical Utility of Multimarker Genetic Risk Scores for Prediction of Incident Coronary Heart Disease: a Cohort Study among over 51 000 Individuals of European Ancestry. *Circulation: Cardiovascular Genetics*. 2016; 9: 531–540.
- [39] Tada H, Melander O, Louie JZ, Catanese JJ, Rowland CM, Devlin JJ, *et al*. Risk prediction by genetic risk scores for coronary heart disease is independent of self-reported family history. *European Heart Journal*. 2016; 37: 561–567.
- [40] Antiochos P, Marques-Vidal P, McDaid A, Waeber G, Vollenweider P. Association between parental history and genetic risk scores for coronary heart disease prediction: the population-based CoLaus study. *Atherosclerosis*. 2016; 244: 59–65.
- [41] Beaney KE, Cooper JA, Drenos F, Humphries SE. Assessment of the clinical utility of adding common single nucleotide polymorphism genetic scores to classical risk factor algorithms in coronary heart disease risk prediction in UK men. *Clinical Chemistry and Laboratory Medicine*. 2017; 55: 1605–1613.
- [42] Chang X, Salim A, Dorajoo R, Han Y, Khor C, van Dam RM, *et al*. Utility of genetic and non-genetic risk factors in predicting coronary heart disease in Singaporean Chinese. *European Journal of Preventive Cardiology*. 2017; 24: 153–160.
- [43] Iribarren C, Lu M, Jorgenson E, Martínez M, Lluís-Ganella C, Subirana I, *et al*. Weighted Multi-marker Genetic Risk Scores for Incident Coronary Heart Disease among Individuals of African, Latino and East-Asian Ancestry. *Scientific Reports*. 2018; 8: 6853.
- [44] Inouye M, Abraham G, Nelson CP, Wood AM, Sweeting MJ, Dudbridge F, *et al*. Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults. *Journal of the American College of Cardiology*. 2018; 72: 1883–1893.
- [45] Liu R, Cheng J, Muzlera C, Robinson JF, Ban MR, Hegele RA. Clinical Utility and Practical Considerations of a Coronary Artery Disease Genetic Risk Score. *CJC Open*. 2019; 1: 69–75.
- [46] Elliott J, Bodinier B, Bond TA, Chadeau-Hyam M, Evangelou E, Dehghan A, *et al*. Predictive Accuracy of a Polygenic Risk Score-Enhanced Prediction Model vs a Clinical Risk Score for Coronary Artery Disease. *Journal of the American Medical Association*. 2020; 323: 636.
- [47] Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Research*. 2007; 17: 1520–1528.
- [48] Chatterjee N, Shi J, Garcia-Closas M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nature Reviews Genetics*. 2016; 17: 392–406.
- [49] Erdmann J, Kessler T, Munoz Venegas L, Schunkert H. A decade of genome-wide association studies for coronary artery disease: the challenges ahead. *Cardiovascular Research*. 2018; 114: 1241–1257.
- [50] Abraham G, Tye-Din JA, Bhalala OG, Kowalczyk A, Zobel J, Inouye M. Accurate and robust genomic prediction of celiac disease using statistical learning. *PLoS Genetics*. 2014; 10: e1004137.
- [51] Goldstein BA, Yang L, Salfati E, Assimes TL. Contemporary Considerations for Constructing a Genetic Risk Score: an Empirical Approach. *Genetic Epidemiology*. 2015; 39: 439–445.
- [52] Börnigen D, Karakas M, Zeller T. The challenges of genetic risk scores for the prediction of coronary heart disease. *Clinical Chemistry and Laboratory Medicine*. 2017; 55: 1450–1452.
- [53] Dudbridge F. Power and predictive accuracy of polygenic risk scores. *PLoS Genetics*. 2013; 9: e1003348.
- [54] Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. *Nature Reviews Genetics*. 2018; 19: 581–590.
- [55] Paynter NP, Ridker PM, Chasman DI. Are Genetic Tests for Atherosclerosis Ready for Routine Clinical Use? *Circulation Research*. 2016; 118: 607–619.
- [56] Louter L, Defesche J, Roeters van Lennep J. Cascade screening for familial hypercholesterolemia: Practical consequences. *Atherosclerosis Supplements*. 2017; 30: 77–85.
- [57] Smith JA, Ware EB, Middha P, Beacher L, Kardina SLR. Current Applications of Genetic Risk Scores to Cardiovascular Outcomes and Subclinical Phenotypes. *Current Epidemiology Reports*. 2015; 2: 180–190.
- [58] Hippisley-Cox J, Coupland C, Robson J, Brindle P. Derivation, validation, and evaluation of a new QRISK model to estimate lifetime risk of cardiovascular disease: cohort study using QRResearch database. *British Medical Journal*. 2011; 341: c6624.
- [59] Stone NJ, Grundy SM. The 2018 AHA/ACC/Multi-Society Cholesterol guidelines: Looking at past, present and future. *Progress in Cardiovascular Diseases*. 2019; 62: 375–383.
- [60] Banerjee A. A review of family history of cardiovascular disease: risk factor and research tool. *International Journal of Clinical Practice*. 2012; 66: 536–543.
- [61] Popejoy AB, Fullerton SM. Genomics is failing on diversity. *Nature*. 2016; 538: 161–164.
- [62] Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, *et al*. Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *American Journal of Human Genetics*. 2017; 100: 635–649.
- [63] Janssens ACJ, van Duijn CM. Genome-based prediction of common diseases: methodological considerations for future research. *Genome Medicine*. 2009; 1: 20.

Temporary page!

L^AT_EX was unable to guess the total number of pages correctly. As there was some unprocessed data that should have been added to the final page this extra page has been added to receive it.

If you rerun the document (without altering it) this surplus page will go away, because L^AT_EX now knows how many pages to expect for this document.