Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Medicine and Health Sciences
Department of Public Health and Nursing

Karianne Skjærstein

Polygenic score to improve prediction of coronary artery disease in women under 45 years

The HUNT Study

Master's thesis in Global Health
Supervisor: Brooke N. Wolford
May 2024

NTNU
Norwegian University of Science and Technology

Karianne Skjærstein

Polygenic score to improve prediction of coronary artery disease in women under 45 years

The HUNT Study

Master's thesis in Global Health
Supervisor: Brooke N. Wolford
May 2024

Norwegian University of Science and Technology
Faculty of Medicine and Health Sciences
Department of Public Health and Nursing

NTNU
Norwegian University of
Science and Technology

Abstract

Introduction: Cardiovascular diseases are an immense health burden across the world. One of the contributing diseases is coronary artery disease (CAD), with a mortality rate at 16% globally and 15% in Norway. Recent numbers suggest an increase in incidence and mortality among younger women. The use of a biomarker such as a polygenic score (PGS), which estimates disease risk based on the effect of multiple genetic variants associated with a disease, holds promise to improve early prediction and prevent disease.

Methods: This study uses a sample from The HUNT Study consisting of 65,923 participants with 53% females and 47% males, and a median age of 46 and 47 years respectively. With a median follow-up time of 21 years, the primary aim was to investigate associations between PGS, family history, established clinical risk factors, and CAD. A previously published PGS with over 1.7 million genetic variants was estimated in The HUNT Study participants. Logistic regression and Cox Proportional Hazard regression analyses stratified by age and sex were performed to investigate if the PGS can improve prediction of CAD in women of younger age.

Results: There were 1,706 prevalent events, and 7,139 incident events of CAD during follow-up. The PGS was significantly associated with CAD, but the effect as measured by both regression methods was stronger in individuals under 45 years as compared to over 45 years. The best model for all age and sex strata, as measured by Harrell's C-statistic, was PGS combined with family history and clinical risk factors. When estimating cumulative incidence based on PGS and family history, we see slightly better risk stratification in males than females.

Conclusion: The PGS can be a valuable predictor for CAD particularly among males and females under 45 years. More research is needed to understand additional risk factors that could improve prediction of CAD in young women.

Acknowledgements

I would like to thank my supervisor, Dr. Brooke N. Wolford for her guidance, for teaching me so much, and for sharing her knowledge and wisdom with me throughout this process of working on this thesis. It has been a privilege, and I am very grateful.

I want to thank all the individuals who have participated in The Hunt Study and by this contributing to invaluable understanding of human health.

For valuable contributions to this study and thesis, I would like to thank Dr. Ida Surakka for data processing and contributions to the study design, and Dr. Therese Haughdahl Nøst for providing the HUNT Cloud lab space.

Table of Contents

List of figures and tables

Abbreviations

CAD – coronary artery disease

CVD – cardiovascular disease

PGS – polygenic score, polygenic risk score

FH – Family history

GWAS – genome-wide association studies

SNP – single-nucleotide polymorphism

SCORE2 - The Systematic COronary Risk Evaluation model

HUNT –Trøndelag Health Study (Helseundersøkelsen i Trøndelag)

HDL-C – high-density lipoprotein cholesterol

SBP – systolic blood pressure

OR – odds ratio

HR – hazard ratio

1 Introduction

1.1 The global burden of coronary artery disease among women

Cardiovascular diseases (CVDs) are the leading cause of mortality and are an immense health burden across the world (1,2). CVDs encompass a range of non-communicable diseases, and the largest group is coronary artery disease (CAD), sometimes referred to as coronary heart disease or ischemic heart disease. This disease affects the heart muscle and the arteries around it and is caused by the narrowing or blockage of the coronary arteries (1,3). Atherosclerotic buildup in the arteries leads to ischemia or infarction, and treatment options can be revascularization, mediation, and management of modifiable risk factors. Primary prevention is identifying individuals at risk before they establish a disease and is an important public health responsibility (4). Whereas secondary prevention aims to prevent secondary events or manage already-established disease.

Worldwide, one-third of the mortality is due to CVDs, and CAD is responsible for 16% globally and 15% of deaths in Norway (5). CAD and stroke are the two leading causes of premature mortality from CVDs (6). A recent study spanning three decades using data from the Global Burden of Disease Study found that sex-specific disparities in mortality were greater in the younger age groups compared to all ages (6). From the year 1990 to 2019 deaths worldwide caused by premature CVD increased by 25%, with CAD being the largest contributor.

CAD is the leading cause of death globally in women and the prevalence of myocardial infarction in younger women is increasing (7,8). Given the significant mortality rates caused by premature CVD, there is a need for earlier detection and better strategies for prevention. An individual's sex and their gender influence health and disease outcomes through genetic, biological, hormonal, and social factors (9). Women's health has not gotten much attention, priority, and resources in the past. However, there is increasing consensus within healthcare and medical research that sex-differences are in need of more attention. Women and men do not have the same anatomy and physiology, and at the same time they live different lives, meet dissimilar exposures, and they experience their symptoms and disease progression differently than men (8,9). Number three of the 17 United Nations Sustainable Development Goals is:

"By 2030, reduce by one-third premature mortality from non-communicable diseases through prevention and treatment and promote mental health and well-being" (10).

To achieve good health and quality of life, for everyone, at all ages, it is crucial to build sustainable healthcare and address the burden of and reduce premature mortality from CVDs by measures like early prevention (10). At the same time, evidence continues to affirm the inequalities in healthcare delivered to women (11). The reasons for inequalities in health outcomes between women and men are many and complex. There are still knowledge gaps on issues such as women-specific risk factors and symptoms, the understanding of biological and genetic mechanisms, and in the diagnostic and treatment areas with prevention and screening being among the issues (12). Biological and genetic variations are increasingly recognized to both directly affect and modify disease risk.

Women worldwide deal with pregnancy and menopause throughout their lives, and this have been demonstrated to be associated with risk of CVDs (13). Three areas have been found to especially disadvantage women 1) personal, knowledge and behavior, 2) healthcare, screening and treatment; 3) risk factors, women-specific risk factors and other sex differences affecting disease outcomes. Even though the mortality caused by CVDs in women and the overall global CVD mortality has improved over the past 30 years, issues like the increasing age of the global population, accompanied by a rise in disability prevalence, pose significant implications and challenges to people, healthcare systems, and economies worldwide (2,8). Women have been and are underrepresented in biomedical research and there is a gap in prevention, treatment, and access to care, as well as detection of disease, and disease management for many diseases including CVDs (8). Historically, medical research has exhibited a bias towards men, leading to a lack of understanding of disease in women. Women have been underrepresented and even excluded from clinical trials (8,14). Both biological sex determined by genetics, and gender identity impact health across the life (15).

There are similarities between men and women, but also important biological and behavioral differences that influence the epidemiology and development of various diseases (16). Increasing evidence shows that a person's sex is one of the most important deciding factors when it comes to risk of disease and response to treatment (17). Women respond differently to medications due to body size, hormonal variations, and a difference in metabolic pathways in

the context of medication usage, for example. For overall CVD, women describe different symptoms and can therefore be delayed in receiving a diagnosis. The evolving recognition and awareness of sex differences in patient care is one of the greatest attributes of modern medicine, and it highlights the need to understand CVDs, particularly how they manifest and are managed in women (17). Although older individuals suffer more often from diseases of the cardiovascular system, it is still important to consider younger individuals for preventive measures because of the possible number of healthy years lost (18). One possible tool for timely risk stratification is genetics. In recent years there has been increasing interest in including genetic biomarkers in predicting diseases since both lifestyle, environment, and genetics influence disease risk (19,20). Genetic scores can quantify disease risk based on large numbers of genetic variants in the human genome and can within healthcare be used for many purposes within the clinical setting (21).

1.2 Cardiovascular Risk Factors

The Framingham Heart Study began in 1948 in Massachusetts, USA, and is an epidemiological cohort study that has contributed to a range of discoveries within heart health. While the previous clinical goals were to treat already manifested disease, the idea of prevention for individuals with an increased risk for heart disease became apparent (22). In the late 1960s, the concept of identifying individuals at risk through combining multiple risk factors into a clinical risk score to predict CVDs came about. Disease prevention or health promotion can be seen from both a population and the individual perspective. From a population perspective interventions such as vaccines, sugar taxes, smoking bans, and mandatory use of seat belts improve health on the population level, while a given individual may not necessarily notice a change (23). Conversely, there is the individual-level where the goal is to identify individuals at high risk for a certain health problem, which usually leads to intervention like medications or lifestyle advice.

The landscape of cardiovascular risk factors is a complex interplay with interactions between genetic, environmental, behavioral, and social determinants that influence an individual's health throughout their lifetime (24,25). These factors encompass a broad spectrum, from genetic predispositions, determining an individual's lifespan, overall health, and disease susceptibility, to socioeconomic and cultural influences, all contributing positively and negatively to health (26). The major risk factors for CAD can be categorized into non-

modifiable and modifiable. The largest modifiable risk factors for CAD, also referred to as environmental, behavioral or lifestyle are diet, high systolic blood pressure, and high cholesterol (27,28). Environmental risk factors range from personal choices regarding lifestyle and behaviors to living situation, income, and interpersonal relationships (29). It goes beyond only personal choices, these are factors also shaped and decided on by elements not controlled by the individual. Non-modifiable factors can be genetics, age, sex, ethnicity, and family history of heart disease (30).

Table 1. Risk factors associated with coronary artery disease *(8,27,30,31)*.

Modifiable or environmental	**Genetic**	**Women-specific**
Hypertension*	Sex	Socioeconomic factors.
Diabetes*	Ethnicity	Factors related to pregnancy
Cholesterol*	Age	and menopause (Gestational
Diet and physical activity	Family history of	diabetes and hypertension,
Smoking	heart disease	premature menopause)
Air pollution		Polycystic ovary syndrome
		Hormonal differences

*Can also be genetic or a combination of genetic and environmental exposure.

The majority of diseases are complex diseases, meaning they are caused by a range of genetic influences and environmental factors together (32). Cardiovascular health in women goes beyond only medical considerations, encompassing a wide array of factors including beliefs, behaviors, economic conditions, environmental aspects, political influences, and sociocultural dynamics (33). The genetic reasons behind CVDs are significant, yet they represent only a part of the risk landscape.

Women experience sex-specific risk factors, they include socioeconomics and factors regarding pregnancy and menopause (8,34). Examples include premature menopause, reproductive-related factors like gestational diabetes and hypertension and preeclampsia (35). Polycystic ovary syndrome (PCOS) is an intricate endocrine condition with a global prevalence estimated at 6 to 10% in women of reproductive age, among several other

comorbidities, this condition has additionally been linked to increased cardiovascular risk (36,37). Sex is a factor in and of itself. However, there are also sex-specific factors that can contribute to and influence CVDs. These include pregnancy and menopause, hormonal contraceptive use, autoimmune diseases, polycystic ovarian syndrome, and breast cancer are examples (38,39). The difference in CVD pathophysiology can lead to women to experience other symptoms than the classical symptoms and how they present, and also have different response to pharmacological interventions (39,40). A global study on sex-specific associations with cardiovascular risk factors found that 57% of CAD incidents among women may be attributable to only five modifiable risk factors, these being hypertension, high body mass index, smoking, cholesterol levels, and diabetes (41).

Social determinants of health are major systemic and economic forces such as gender, inequality, poverty, and psychological stressors (25,42). Socioeconomic status and sociocultural roles play a role as direct risk factors or as modifying factors, and they are influenced by gender and often disproportionately affect women (8). Psychosocial experiences like motherhood, the distribution of income and wealth which often fall with men affect women's health. Additionally, caregiving roles and differences in priorities. The lack of sex-specific knowledge on symptoms of CVDs can lead to women with chest pain to be misunderstood and diagnosed with anxiety. Gender as a separate concept as opposed to biological sex is important to consider because risk can be influenced by both biological sex and gender, which is how an individual presents themselves to the world, is conditioned to behave, and the constrained roles of society (15). Gender as a determinant of health entails the societal construct of gender and thereby roles and expectations of a given society, these are different in different times, places, and life stages of a person (15). The level of education can be used as an indicator of socioeconomic position, and it has been documented in Norway that there is a higher burden of modifiable risk factors for CAD among individuals with less education (43). Additionally, a higher incident rate of myocardial infarction among less educated have been observed in Norway, particularly among women aged 35 to 69 years (44).

Genetic variation, identifiable from birth, plays a pivotal role in an individual's susceptibility to CAD (20). The genetic risks a person is born with can be considered "The First Risk Factor", laying a lifelong foundation for disease susceptibility (45,46). Family history also serves as a valuable indicator, that captures both environmental and genetic risks. However, it is essential to recognize the dynamic interplay and interactions between genetic

predispositions and environmental factors, emphasizing that genetics do not solely dictate one's health destiny. Evidence suggests that the absence of traditional risk factors at age 50 is associated with a lower lifetime risk for CVDs (47). This finding supports the strategy of focusing on early intervention and prevention, particularly in younger populations, to address the escalating challenge effectively. Furthermore, the limitations of risk factors as prognostic tools warrant discussion. While they provide valuable insights, their predictive accuracy is not absolute. Understanding these limitations is essential for developing more effective strategies for CVD prevention and management (48).

Historically, medical research and understanding of diseases have predominantly centered around the male physiology, leading to the misconception that men and women are biologically identical (49,50). It is now recognized that biological differences in health and disease are not just specialized areas but are crucial for a comprehensive understanding of medical and health research and there exist major sex and gender differences in the prevalence and burden of CVDs and this is increasingly guiding preventive work (51). The genetic divergence between male and female begins at conception with the assignment of XX or XY chromosomes, making an individual's sex a significant modifier of health and disease due to genetic, epigenetic, and hormonal differences (49).

The inclusion of sex as a biological variable in healthcare and research will aid in improving sex-specific care and developing beneficial treatments. In health research, including participants' sex and age as variables is vital for accurate representation (50). The categorization of premature mortality from CVD varies across the world due to differences in life expectancy. Age is, after all, the biggest risk factor for most non-communicable diseases. In Norway, deaths caused by non-communicable diseases before 75 years of age are considered premature and these individuals are a target group for prevention (52). The evolving demographics characterized by higher life expectancy in both high-, middle- and low-income countries is a testament to successful public health interventions, but also presents new challenges. People are living longer, but not necessarily healthier lives. Life expectancy in Norway is high, and one of the biggest challenges onwards is the disproportional age distribution in the population and more individuals living with chronic disease and disability at high ages (53). Incorporating a genetic risk score can enhance early prediction and prevention, thereby reducing morbidity and mortality (54).

1.3 Polygenic Scores and coronary artery disease

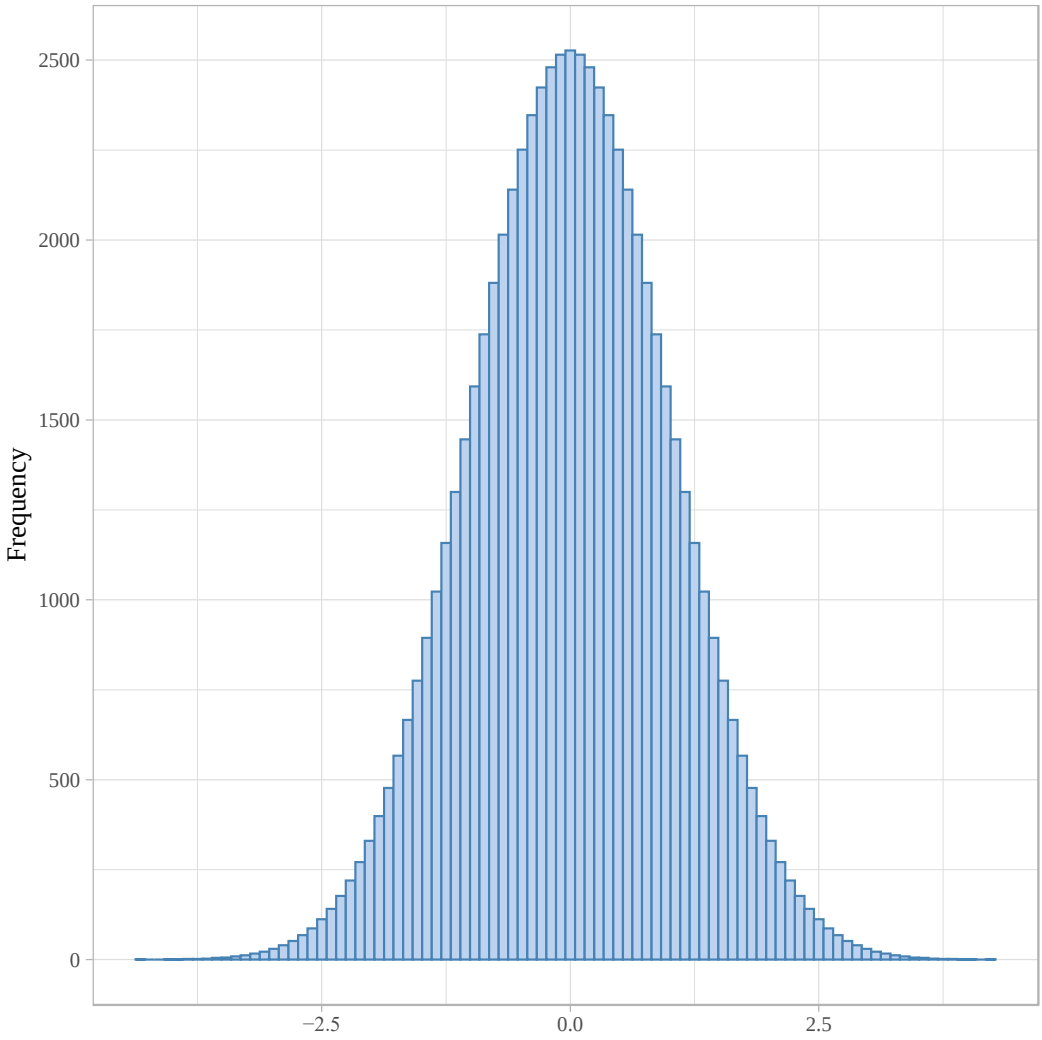
The relationship between genetic predispositions and environmental factors is intricate. While genetics may indicate a higher likelihood of developing CAD, individual behavior and environmental conditions play a significant role in actual disease manifestation and progression. This dynamic interaction highlights the importance of considering both genetic and non-genetic factors. Monogenic and polygenic are central descriptive measures of disease within biomedicine and CAD has a considerable heritable and polygenic architecture (55). This can be estimated in an individual using a polygenic score (PGS), also called a genetic risk score (GRS) or polygenic risk score (PRS) in the literature. The origins of the PGS first started with the completion of The Human Genome Project which was finished in 2003 (56). The project successfully mapped the entire human genome. From this, genome-wide association studies (GWAS) laid the foundation for developing and calculating PGS. The first GWAS was completed in 2005 and compared allele frequencies at genome-wide variants for age-related macular degeneration (57). Variants refer to variations within the genome and the most common type presently used is called single nucleotide polymorphism (SNP) (30). The presence or absence of specific variants connected to diseases, conditions, and other traits tells the tale of lifetime genetic risk without considering other factors like socioeconomic status and environment. GWAS studies have systematically found associations between genetic variants and disease (58). These associations between variants and diseases are measured with an estimated effect size.

A PGS is a mathematical, single-value estimate calculated from an individual's genetic variants and their association with the given disease, this is based on the current knowledge of the genetic composition of the disease (59). Most diseases are polygenic, with numerous variants across the genes responsible, not monogenic, like familial hypercholesterolemia, where only one gene is accountable (60). The first study testing a genetic score was published in 2010, this was a genetic score based on 13 SNPs associated with CAD (61).

PGS offer a measure of disease susceptibility, particularly for complex, polygenic diseases like CVDs (62). Since most diseases are at least partially heritable, finding individuals with high risk at an earlier stage of disease development will be important in the years to come because of the disease burden to societies (62). One of the key benefits of PGS is its ability to predict complex diseases over an individual's entire life course (30). Within a population, a

PGS is normally distributed where most individuals are at the mean, with some people being on the tails of the distribution, representing a low and high risk for a particular disease (59). This allows for the visualization and comparison of an individual's PGS against a reference population. Figure 1 provides a visualization of the distribution of the CAD PGS used in this thesis study sample (n = 65,923). To date there are over 60 scores for CAD published in the PGS Catalog (63), an open access host of published scores. Some of these scores are currently in clinical trials to establish efficacy in clinical care. Genetic profiling and the systematic cataloging of polygenic scores are increasingly used in research and eventually in clinical settings (30).

Figure 1. The distribution of polygenic scores for CAD in the study sample.



Studies have shown the value of PGS in guiding patient care, predicting disease onset, and screening for conditions like cancers and other non-communicable diseases, but their clinical utility is in the beginning stages and further research and validation of current scores is

needed (64). An optimal strategy to implement PGS is in primary prevention and the possible clinical advantages are active areas of inquiry. One study estimated a PGS for CAD in a cohort with over 300,000 individuals found that the most predictive power belonged to individuals younger than 50 years and concluded this be used to improve early identification of those who will benefit from medication (65). Evidence demonstrates that most PGS for CVDs have similar or better predictive ability compared to traditional clinical risk factors alone (66).

A positive family history is a risk factor for many non-communicable diseases, including CAD, and this can be used in preventive strategies (67). Having a close relative, parent, or sibling experience myocardial infarction before the age of 60, can suggest elevated risk (52). Many epidemiological studies include self-reported questionnaires on family history status on various diseases because it is a key variable that captures both genetic and environmental exposure to disease risk and how they might modify each other. Studies have given insights into the value of including both genetics and family history of CAD in prediction (68,69). Family history can capture not only genetics but also the social and environmental aspects of risk, for instance, the exposure to negative socioeconomic life circumstances and cultural behaviors, as well as shared traditional risk factors for CVDs.

A large twin study from Sweden with a follow-up time of 26 years looked at the role of genetics and the degree of hereditability of CAD in different ages and both in twin pairs of women and men (70). Results demonstrated that the disease is highly connected to genetics and is the result of heritability for both sexes. Current evidence states the estimated heritability between 40% to 60% (55). Genetic heritability in medicine can be compared across traits to understand the influence of genetic versus environmental factors (71). Questions on how well a PGS can be integrated alongside established clinical risk factors and who will benefit most are important to consider (72). In overworked healthcare systems worldwide, the cost-effectiveness equation is essential. To summarize, evidence on the comparison of associations and contributions of FH and PGS to disease prediction from a large cohort study demonstrates that the effects of these two are not necessarily interchangeable measures, but FH and PGS are independent, complement each other, and can be used together in risk assessment (73). It has been suggested that the inclusion of family history increase the accuracy of a PGS, especially in populations other than the one the PGS was derived from, usually of European ancestry (74). Inclusion of a genetic score to

traditional risk factors and a positive family history of CVDs has previously shown a higher predictive ability than traditional risk factors alone (75).

1.4 Prediction and Screening

Advances within healthcare and public health work targeting CVD risk factors have made the disease burden reduced (2,76). This has been accomplished through public health efforts and primary prevention with, for example, risk assessment tools, which aim to identify individuals with high-risk profiles and thereby provide risk-reducing measures (77). Screening is a part of primary prevention and is used to identify asymptomatic or subclincally symptomatic individuals and thereby intervene before it is too late (78). For example, the screening of women for breast and cervical cancer has had a major impact on health, by successfully decreasing the cancer incidence and mortality. This is of both importance to individuals and also of public health importance. But screening can also impose harm in the form of psychological stress in the waiting process, overdiagnosis and overtreatment of individuals who might never have noticed the disease on their own, and unfortunate false negative and false positive tests (78,79). The good outcomes must outweigh the negative consequences. Precision medicine describes treatment or prevention that is individually tailored to a person's disease process or symptoms and can be based on genetics (80). Biomarkers like PGS can add valuable insights into an individual's disease risk compared to clinical markers alone, especially in the case of a positive FH of CVDs. The use of large population-based cohort surveys, biobanks, and electronic health records is revolutionizing the ability to investigate traits and diseases on a large scale (81). Cohort studies including genetic information, like the United Kingdom Biobank and The HUNT Study make research on PGS possible (82,83).

Representation of global diversity within genetics has major deficits, and most of the GWAS to date are derived from European ancestry (45). The current state of genomic population-based biobanks is that they are highly skewed towards participants of European ancestry and therefore are not very diverse. This limits the global understanding of heritable diseases (84). As of 2021, 86% of all studies including genetic information, have been done on populations of European descent (85). This amplifies already large health disparities around the world, and greater diversity is therefore needed (86). There are several challenges and, especially ethical challenges tied to the use of genetics, and genetic risk scores in general, there are ethical, social and legal considerations to be made (87). This can be tied to cost, psychological

impact, the understanding of risk estimates for both patients and healthcare professionals, and how to prioritize the patients who will benefit the most (45). Receiving the message of being a genetically high-risk individual may create a sense of fatalism and belief that genetic predisposition cannot be intervened with, even though it can be by targeting modifiable risk factors or by pharmaceutical interventions (80). Genetics in the context of risk prediction for diseases offers possibilities to discover, potentially earlier in life, the individuals who can benefit from prediction and early intervention, disease surveillance, and individualized treatment plans for common, complex diseases like CVDs. As an illustration, a woman with a high PGS for breast cancer may get earlier or more frequent mammograms for surveillance, while a woman with a high PGS for CVD may be prescribed treatment of statins. Including sex-specific risk factors is essential when trying to identify and enable early detection of women with a high-risk profile (31). Optimal risk assessment of women remains a challenge despite the multiple predictive screening and risk assessment tools available.

Contemporary clinical risk scores to estimate individual risk of CVDs used in healthcare today include SCORE2, the Framingham Heart Risk, and NORRISK2 (88,89). They have similarities but include some different measures known to be a risk factors and are made to be used in different populations around the world and in different ages. Currently, Norway uses NORRISK2, a sex-specific calculator that includes factors like age, family history, and cholesterol to calculate a 10-year risk of myocardial infarction or stroke (89).

The Systematic COronary Risk Evaluation (SCORE2) model provides a framework for well-established risk factors and a 10-year prediction model specific to CVD within European populations and was available from August 2021 (90,91). It is made to be used in healthy individuals aged 40-69 without previous CVD or diabetes to predict fatal or non-fatal events of CVD. The HUNT Study was part of the external validation of the score model alongside 24 other cohorts. A similar model exists for “older persons (OP)”, the SCORE2-OP, ages 70 and above (92). These risk prediction tools rely heavily on already-established factors associated with heart disease, like cholesterol, which is typically measured in middle age and beyond. Meanwhile, a PGS can be implemented from birth and used to prevent risk factor development (66). All of these predictive models have their own limitations, and most are not designed for younger individuals under 40 years of age. Evidence suggests that these scores tend to underestimate risk in younger individuals, specifically if an individual belongs to a family of heightened risk (18,88). Clinical risk score validity is also an issue for women and

minority populations due to biases in the derivation sample. Therefore, developing ancestry-specific, sex-specific, and age-specific disease risk thresholds should improve primary prevention (8).

2 Study aims

2.1 Rationale

This study uses a polygenic score (PGS) calculated for coronary artery disease (CAD) and includes associated risk factors to investigate the disease in young women. This work explores the impact of genetics, sex, and age on the risk of CAD, and contributes to the current understanding of their interplay in women. It aims to address gaps in research and highlight the importance of sex-specific approaches in cardiovascular healthcare, and medical and epidemiological research.

The role of sex in the overall risk profile and presentation of CAD is important and needs further investigation to better understand how the disease presents itself, progresses, and causes death among all populations worldwide. The integration of genetics offers a promising avenue for advancing our understanding and management of CAD among women today. This approach aligns with the evolving paradigm of precision medicine, emphasizing the need for personalized healthcare strategies based on individual genetic profiles and demographic characteristics.

Given the mortality due to the disease and current challenges within healthcare systems, it is important to reduce the disease burden. In addition, with the aging population and shift in age demographics, the need for better healthcare solutions is increasingly important.

2.2 Aims

The main aim is to investigate the contributions of a PGS for the primary prevention of CAD, with a particular focus on women under 45 years of age. This thesis aims to enhance the understanding of PGS as a better predictor of CAD in women.

This will be done through two specific aims:

1) Investigate associations between a polygenic score, established clinical risk factors and coronary artery disease in a cohort with a median follow-up of 21 years.
2) Stratify by age and sex to investigate disease risk in women of younger age.

3 Methods

This chapter describes the study methods. This study aims to investigate the use of a polygenic score (PGS) to primary prevention of coronary artery disease (CAD) within a sub-population of The HUNT Study.

3.1 The HUNT Study

The Trøndelag Health Study (The HUNT Study) is a population-based cohort health survey from the former Nord Trøndelag County in Norway (83,93). Adult residents were invited to participate in HUNT1 (1984- 86), HUNT2 (1995-97), HUNT3 (2006-08), and HUNT4 (2017-19). The study covers a range of different health-related topics measured in interviews, questionnaires, clinical examinations, and biological samples. To this date, the study holds information on more than 229,000 individuals with follow-up over nearly 40 years.

The HUNT Biobank consists of DNA, RNA, proteins, and other biological samples from participants (83). Genetic material extracted from blood samples has been sequenced in 2,201 participants and genotyped in 88,000 participants. Further imputation resulted in 33 million variants with estimated allele dosage. The HUNT Study encompasses a range of health measurements and is special because of the extensive follow-up time. Additionally, Norway's universal healthcare enables the connection to national health registries for its population which is valuable for epidemiological research purposes.

3.2 The study sample

This study uses data from HUNT2 and HUNT3 with information from surveys, electronic health records, and a PGS calculated for CAD using genetic data derived from blood samples. The participants included are all with available imputed allele dosages. Participants with poor genotyping and imputation metrics, and those with missing self-reported family history of CVDs were excluded. 65,923 participants met the criteria, and characteristics of the study sample include the proportion 52,8% females and 47,2% males, with the median age of 46, and 47 years respectively. The specific age threshold of 45 years old is based on the median age in the study sample and the need for improved prediction in younger populations. The participant's sex in this study is derived from genetics, the XX and XY chromosomes found

in genetic testing. While the gendered term “women” is used in this thesis, biological sex is used in the research.

3.3 Coronary artery disease as outcome

The outcome variable is defined as the first event of fatal or nonfatal ischemic heart disease or myocardial infarction documented in The Norwegian Patient Registry (Norsk Pasientregister) or The National Cause of Death Registry (Dødsårsaksregisteret). The complete list of ICD (International Classification of Diseases and Related Health Problems) codes that the definition is based upon is found in appendix 1.

3.4 Covariates

Analyses are stratified by sex and age. Baseline age is when participants entered The HUNT Study. Information on family history of CVD was collected in self-reported questionnaires. The variable was constructed to represent the most complete presence or absence of first-degree family history CVD. It is made up of information from questionnaires across HUNT1, HUNT2 and HUNT3. There were some differences across HUNT surveys, so this is the best approximation available. The specific questions asked are found in appendix 1. Briefly, the participants were asked if parents, siblings, or offspring had myocardial infarction or angina pectoris. In the most recent questionaries, it was specified as occurring before 60 years old.

The clinical covariates are the same as those in the clinical SCORE2 prediction model (91). These are well-established risk factors for cardiovascular disease (CVD), and are total- and HDL cholesterol, smoking status, and systolic blood pressure. Clinical measurements were taken at the study entry. Smoking status was collected in self-reported questionnaires. While family history of CVD is an established clinical risk factor, we differentiate in this study between the clinical factors from SCORE2 and family history.

Table 2. Description of study covariates.

	Description	**Measurement**
Sex	Male or female	Biological samples and genetics

Age	Continuous	Year of birth
Polygenic score	Single value estimate	Biological samples and mathematical calculations
Family history of CAD	Positive, negative, or unknown	Self-reported in questionnaires
Ever smoked	Yes or no	Self-reported in questionnaires
Total- and HDL cholesterol	Continuous	Biological samples
Systolic blood pressure	Continuous	Clinical measurements
Follow-up time	Continuous	Time from participants entry into HUNT2 or HUNT3

3.5 Calculation of the polygenic score

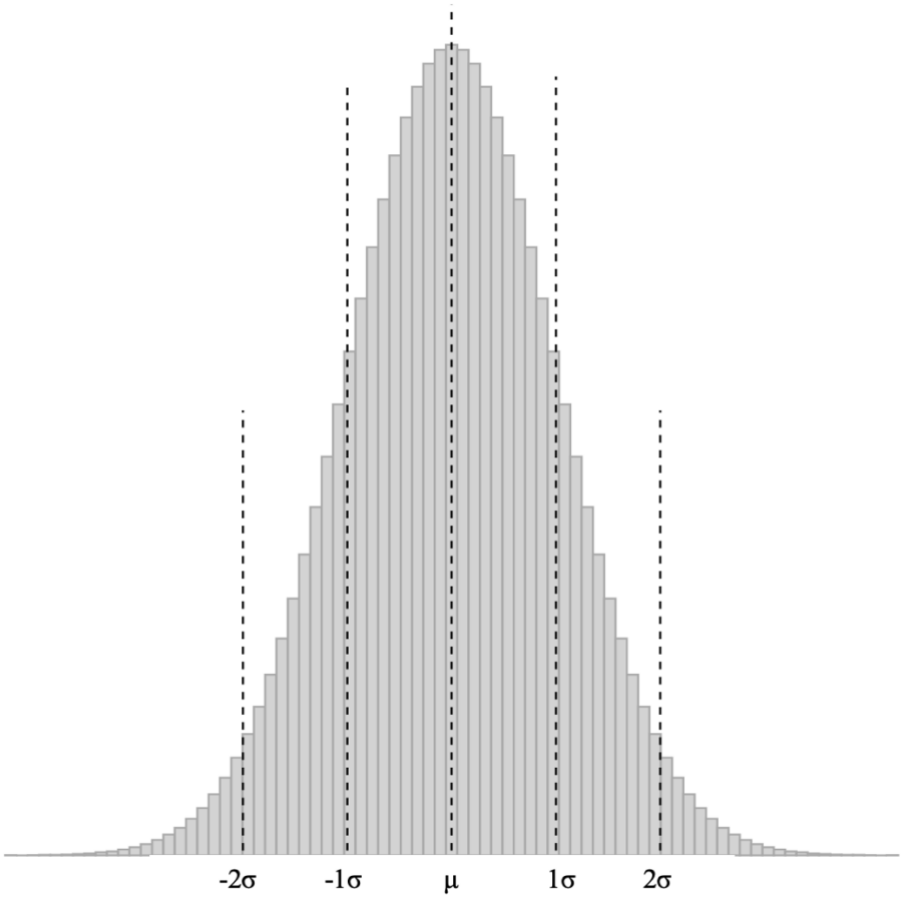
The published polygenic score from the PGS Catalog has been calculated for The HUNT Study participants using the tool pgs_calc (63,94). The polygenic score used in this study is named metaGRS_CAD (PGS Catalog ID: PGS000018) and was developed to include over 1,7 million single nucleotide polymorphisms associated with CAD (55,95). The score was published in 2018 and when tested alone and accompanying clinical risk factors for CAD, it demonstrated promise as a complementary risk prediction tool. By blood samples acquired from study participants and further genotyped, the score gives an estimate of lifetime risk for CAD. Importantly, the GWAS underlying the PGS and the tuning of PGS parameters was independent from The HUNT Study sample so results are not subject to overfitting. The following is the equation for calculation:

$$PGS_i = \sum_{j=0}^{M} \hat{B}_j \times D_{ij}$$

Where M is selected markers, $\hat{B}_j$ is the estimated effect size from GWAS, and D_{ij} is the dosage probability at a given marker for a given individual across i individuals in the cohort.

When included in statistical analyses, the polygenic score has been standardized by inverse normalization to fit the standard normal distribution curve (59). The benefit of inverse normalizing is the interpretability of results and using scores across studies for better comparability. Since PGS is continuous and normalized, interpretations of effect estimates need to be interpreted as a one-unit change in the predictor (PGS) as one standard deviation (SD). In analyses, one standard deviation in PGS corresponds with decreasing or increasing odds ratio (OR) or hazard ratio (HR).

Figure 2. The mean and standard deviations in the study sample polygenic scores.



3.6 Statistical Analyses

The prevalence of CAD was estimated by combining prevalent and incident CAD events. The PGS distribution was into 10 equal groups and then estimated the prevalence of CAD within each group. 95% confidence intervals around these prevalence estimates were constructed using the standard error defined as:

$$SE(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Two regression methods were applied to investigate the association between PGS and CAD. First, multivariable logistic regression aimed at investigating associations between prevalent and incident CAD and PGS. Second, Cox Proportional Hazard regression models to investigate CAD incidence over time in association with PGS. The relationships between CAD and PGS will be assessed in analyses by their effect size presented in ORs and HRs, their 95% confidence intervals (CIs), and p-values. The significance level of the p-value is set to >0.05. Analyses are stratified by sex and age, and in groups of women and men, under and over 45 years of age.

Table 3. The statistical analyses.

Analysis	**Sample**	**Model**
Logistic regression	Women < 45 and > 45	PGS
	Men < 45 and > 45	PGS + FH
		PGS + FH + Clinical risk factors
Cox proportional	Women < 45 and > 45	PGS
hazard regression	Men < 45 and > 45	PGS + FH
		PGS + FH + Clinical risk factors

PGS = Polygenic score. FH = Family history of CAD. Clinical = Total and HDL-cholesterol, systolic blood pressure and smoking status.

Statistical methods have been chosen because they are appropriate for investigating binary outcomes such as disease presence and absence. They are statistical methods commonly used in epidemiology to assess the strength of associations between disease and risk factors in population-based observational studies. All analyses will be adjusted for age and the genotyping batch, the DNA samples were prepared and genotyped in 5 batches so adjustment for the genotyping batch is done to control for technical variation. Failure to account for this in analysis can lead to bias and decreased internal validity (82).

Logistic regression will be used to investigate significant associations between CAD and the PGS, and covariates. The models are done with the binomial link function and additive explanatory variables. The outcome variable is all prevalent and incident CAD events (n = 8845) among participants. Sex- and age-stratified models are used to investigate differences and compare exposure to covariates. ORs are reported for PGS adjusted for baseline covariates (age and genotyping batch), adjusted for family history, and adjusted for the clinical risk factors (Table 3). Logistic regression models were assessed using Nagelkerkes R^2 which is a coefficient of determination and quantifies how much the PGS and covariates can explain the variance in the outcome (96,97).

Cox regression, a type of survival analysis methodology, were made to investigate associations between time to dependent on explanatory variables (98, p. 100-108). Survival analysis is a widely used statistical method in epidemiological research where the goal is to investigate questions where there is an event of disease, death, or recovery, and the time until that event happens, days, months, or years (98, p. 4-5). The advantage is the inclusion of time as a variable to investigate how hazardous given exposures are to experience an outcome. Participants in the study sample have been followed up for a median time of 20.8 years. At the end of the follow-up, the outcome has been coded as 1 for an incident event and 0 for no incidence of CAD. Follow-up time or time-on-study was the time scale for Cox regression. In contrast with the logistic regression analyses, prevalent outcome events from before the study entry are excluded. Cumulative incidence figures were made to visualize CAD incidence over the follow-up time for both sexes and stratified by positive or negative family history together with the PGS tertiles representing low, medium, high genetic risk of CAD. For Cox proportional hazard regression, the concordance index, specifically Harrel`s C-statistic, is used to evaluate predictive strength (99,100). The Schoenfeld residuals was used to assess the proportional hazard assumption (98, p. 181-183).

Logistic regression and Cox regression analyses were performed in the statistical software program RStudio version 4.3.1. Statistical analyses have been conducted on the HUNT Cloud. The main R packages used for statistical analysis are the tidyverse (101) and survival (102). ggplot2 (103) and gtsummary (104) for tables and figures.

3.7 Ethical Considerations

The participants in The HUNT Study have given written informed consent to data storage and the linkage to other health registries. Personal information is pseudonymized before data is delivered for research. Internal routines for the safe storing and use of health data at the Norwegian University of Science and Technology (NTNU) and St. Olav´s Hospital are followed. The General Data Protection Regulation (GDPR) regulations are met on the HUNT Cloud. This thesis has approval from the Regional Committee for Medical Research Ethics and the HUNT Research Center (REK number 2016/885, HUNT student project number 2023/35293, and HUNT research project number 2016/20667 with PID: 106887).

4 Results

This chapter presents descriptive information about the study sample and results from statistical analyses.

4.1 Descriptive information on sample

The descriptive characteristics of the sample are presented in Table 4. Categorical variables are presented in numbers (n) and percentages (%), and continuous variables are presented in the median and interquartile range.

Table 4. Characteristics of the study sample.

Characteristic	Overall, n = 65,923[1]	Female, n = 34,806[1]	Male, n = 31,117[1]
Coronary artery disease events	7,139 (11%)	2,729 (7.8%)	4,410 (14%)
Prevalent coronary artery disease events	1,706 (2.6%)	435 (1.2%)	1,271 (4.1%)
Age (years)	47 (35, 60)	46 (34, 60)	47 (35, 61)
Age distribution			
Under 45	30,772 (47%)	16,794 (48%)	13,978 (45%)
Over 45	35,151 (53%)	18,012 (52%)	17,139 (55%)
Family history of cardiovascular disease	23,453 (36%)	12,948 (37%)	10,505 (34%)
Ever smoked	37,335 (57%)	18,355 (53%)	18,980 (61%)
Systolic blood pressure (mmHg)	131 (120, 146)	127 (116, 144)	135 (125, 147)
Total cholesterol (mmol/L)	5.60 (4.80, 6.50)	5.60 (4.80, 6.60)	5.60 (4.90, 6.40)
HDL-cholesterol (mmol/L)	1.30 (1.10, 1.60)	1.40 (1.20, 1.70)	1.20 (1.00, 1.40)
Follow-up time for incident coronary artery disease (years)	20.8 (10.5, 21.6)	20.9 (10.7, 21.6)	20.6 (10.2, 21.5)

[1] n (%); Median (IQR)

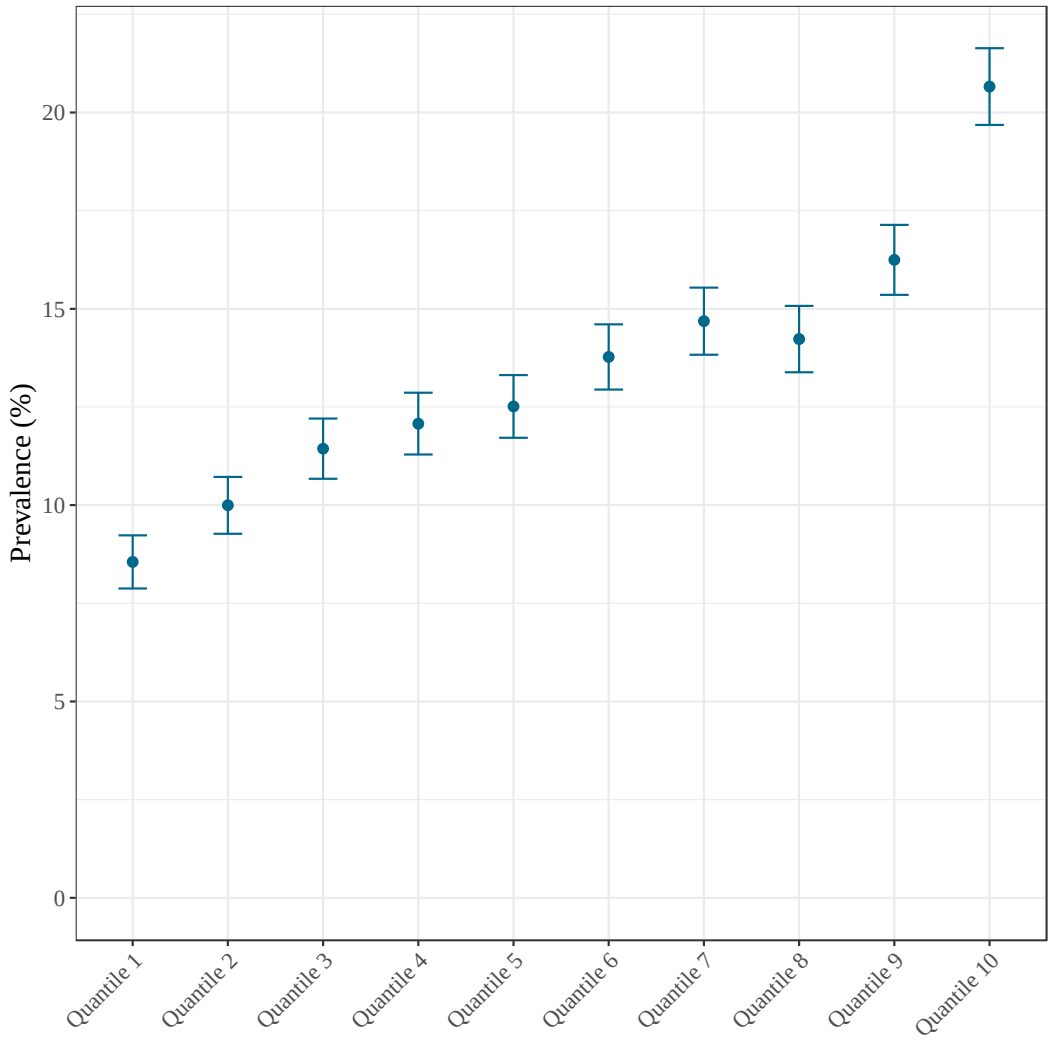
The study sample encompasses a population of 65,923 participants, divided nearly equally between 34,806 (52.8%) females and 31,117 (47.2%) males. The incidence of CAD within the sample is observed as 11%, with a higher prevalence among males (14%) than females (7.8%). Prevalent cases of CAD (observed before enrollment into The HUNT Study) are observed in 2.6% of the sample, with males exhibiting a greater prevalence (4.1%) than females (1.2%). The median overall follow-up time for CAD is 20.8 years, with negligible differences between sexes. The median age of the participants is 47 years, with no major sex disparity. Approximately 47% of the sample is below 45 years of age, while 53% are older. A substantial proportion, 36%, reports a family history of CVD. Smoking history is documented for the majority, with 57% having smoked at some point, this behavior is more prevalent among males (61%) than females (53%).

The median systolic blood pressure across the sample is approximately 131 mmHg, showing minimal sex variation. Total cholesterol levels average at 5.60 mmol/L, with no significant difference between sexes. The median HDL-cholesterol levels vary slightly by sex, with females at 1.40 mmol/L and males at 1.20 mmol/L.

4.2 Coronary artery disease prevalence and the polygenic score

To assess the value of the PGS all study participants were split into 10 quantiles and the prevalence of CAD across the study time was calculated within each quantile (Figure 4). Quantile 1 represent a low PGS and quantile 10 a high PGS. Quantile 1 has a CAD prevalence under 10% while quantile 10 has a prevalence over 20%. The overall trend suggests a substantially increased occurrence of CAD among participants within each decile of PGS.

Figure 3. Polygenic scores and the prevalence of CAD (95% CI).



4.3 Logistic regression models

Logistic regression models with PGS alone and adjusted for covariates and other explanatory covariates were made to investigate the association between CAD and PGS. The covariates are family history (FH) and clinical CAD risk factors from SCORE2 which are total- and HDL cholesterol, smoking status, and systolic blood pressure (91). Both prevalent and incident CAD are included in the analyses. Participants are only measured once, and all events are first-time events.

Table 5. Odds ratios for prevalent and incident CAD per 1 SD in the PGS.

Sample (Sample size)	**Model**	**PGS Odds Ratios per SD**	**P-value**	**Nagelkerke´s R^2**

		(95% CI)		
Women under 45 (16794)	PGS	1.71 (1.50, 1.95)	7.01e-16	0.088
	PGS + FH	1.64 (1.44, 1.87)	1.31e-13	0.104
	PGS + FH + Clinical	1.54 (1.35, 1.77)	2.30e-10	0.149
Women over 45 (18012)	PGS	1.38 (1.32, 1.44)	4.78e-51	0.135
	PGS + FH	1.34 (1.29, 1.40)	1.58e-42	0.149
	PGS + FH + Clinical	1.31 (1.25, 1.37)	1.23e-34	0.178
Men under 45 (13978)	PGS	1.75 (1.60, 1.91)	2.52e-36	0.134
	PGS + FH	1.67 (1.53, 1.83)	2.00e-30	0.152
	PGS + FH + Clinical	1.60 (1.46, 1.75)	1.53e-24	0.196
Men over 45 (17139)	PGS	1.52 (1.47, 1.58)	9.70e-117	0.116
	PGS + FH	1.48 (1.43, 1.53)	1.52e-99	0.137
	PGS + FH + Clinical	1.46 (1.41, 1.51)	2.26e-91	0.160

PGS = Polygenic score. FH = Family history of CAD. Clinical = Total and HDL-cholesterol, systolic blood pressure and smoking status.

In women under the age of 45, the PGS OR was 1.71 (95% CI: 1.50, 1.95), which slightly decreased to 1.64 (95% CI: 1.44, 1.87) after adjustment for FH, and to 1.54 (95% CI: 1.35, 1.77) with both FH and clinical factors considered. Women over 45 exhibited a lower PGS OR of 1.38 (95% CI: 1.32, 1.44), which similarly decreased with adjustments to 1.34 (95% CI: 1.29, 1.40) for FH and 1.31 (95% CI: 1.25, 1.37) for FH and clinical factors.

In men under 45, the PGS OR per SD was 1.75 (95% CI: 1.60, 1.91), decreasing to 1.67 (95% CI: 1.53, 1.83) after FH adjustment and to 1.60 (95% CI: 1.46, 1.75) with FH and clinical factors. In the group of men over 45, the PGS only was 1.52 (95% CI: 1.47, 1.58), which was attenuated to 1.48 (95% CI: 1.43, 1.53) after FH adjustment and reduced slightly to 1.46 (95%

CI: 1.41, 1.51) upon including both FH and clinical factors. The PGS was statistically significant (p-value < 0.05) associated with CAD in all models.

4.4 Cox proportional hazard regression

Cox regression analyses were used to estimate the HRs of PGS as an explanatory variable for CAD. The analyses include PGS, with subsequent analyses adjusting for family history (FH) and a combination of family history and clinical CAD risk factors, including total and HDL cholesterol, smoking status, and systolic blood pressure. In analyses, results are HRs per SD in the PGS for participants.

Table 6. Hazard ratios for incident CAD per 1 SD in the PGS.

Sample Sample size (number of CAD events)	**Model**	**PGS Hazard Ratio per SD (95% CI)**	**P-value**	**Harrel`s C-statistic (95% CI)**
Women under 45 16788 (229)	PGS	1.76 (1.54, 2.00)	1.62e-17	0.75 (0.72, 0.78)
	PGS + FH	1.70 (1.49, 1.93)	2.12e-15	0.77 (0.74, 0.79)
	PGS + FH + Clinical	1.59 (1.40, 1.82)	4.87e-12	0.80 (0.78, 0.83)
Women over 45 17583 (2500)	PGS	1.32 (1.27, 1.37)	1.24e-42	0.75 (0.74, 0.76)
	PGS + FH	1.29 (1.24, 1.34)	2.47e-36	0.76 (0.75, 0.78)
	PGS + FH + Clinical	1.26 (1.21, 1.31)	2.45e-29	0.77 (0.77, 0.78)
Men under 45 13956 (537)	PGS	1.70 (1.56, 1.85)	1.01e-34	0.76 (0.74, 0.78)
	PGS + FH	1.63 (1.50, 1.78)	1.00e-29	0.77 (0.75, 0.79)
	PGS + FH + Clinical	1.56 (1.43, 1.70)	4.04e-24	0.81 (0.79, 0.83)
Men over 45 15890 (3873)	PGS	1.39 (1.34, 1.43)	4.83e-87	0.70 (0.69, 0.71)
	PGS + FH	1.36 (1.31, 1.40)	2.77e-76	0.70 (0.70,0.71)
	PGS + FH + Clinical	1.33 (1.29, 1.38)	1.35e-67	0.72 (0.71, 0.73)

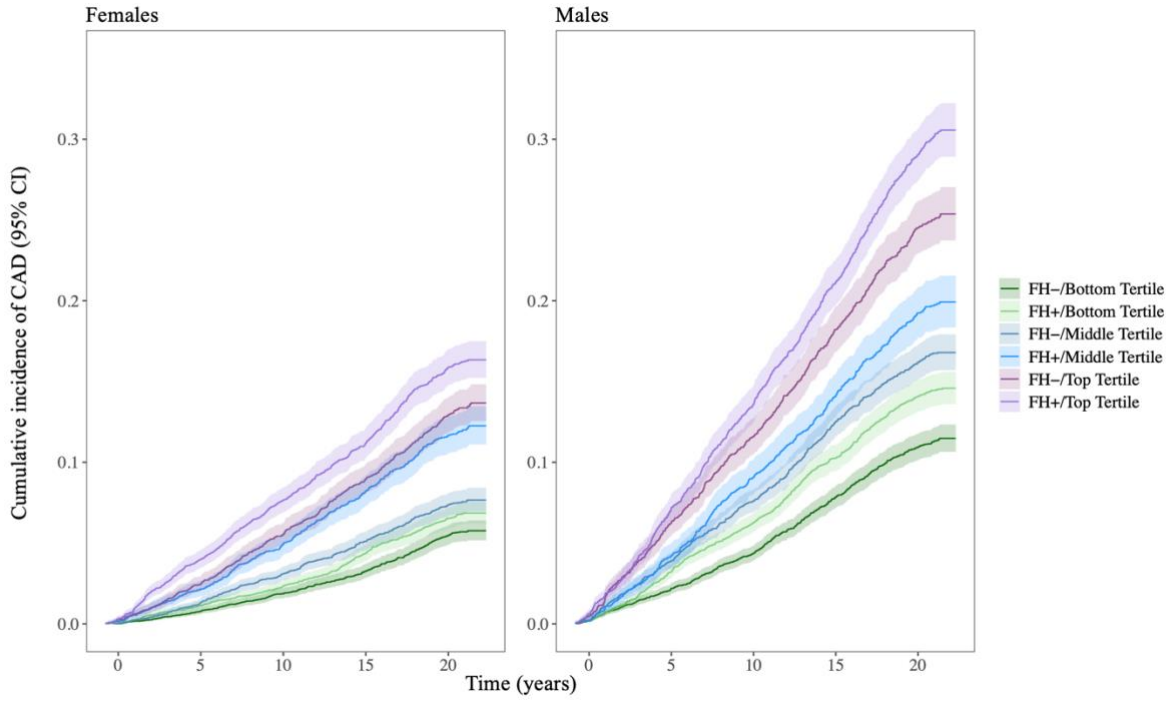
PGS = Polygenic score. FH = Family history of CAD. Clinical = Total and HDL-cholesterol, systolic blood pressure and smoking status.

In the unadjusted model, women under 45 had a PGS HR per SD of 1.76 (95% CI: 1.54, 2.00). Upon adjustment for family history, the HR decreased to 1.70 (95% CI: 1.49, 1.93), and further reduced to 1.59 (95% CI: 1.40, 1.83) when both family history and clinical risk factors were considered. For women over 45, the initial PGS HR was 1.32 (95% CI: 1.27, 1.37). This was modestly attenuated to 1.29 (95% CI: 1.24, 1.34) after adjusting for family history and further attenuated to 1.26 (95% CI: 1.21, 1.31) with the inclusion of clinical factors.

Among men under 45, the PGS HR was 1.70 (95% CI: 1.56, 1.85), which was minimally reduced to 1.63 (95% CI: 1.50, 1.78) after accounting for family history and further to 1.56 (95% CI: 1.43, 1.70) with the addition of clinical factors. In the group of men over 45, HR was initially 1.39 (95% CI: 1.34, 1.43) and showed a slight decrease after adjustments to 1.36 (95% CI: 1.31, 1.40) with family history and to 1.33 (95% CI: 1.29, 1.38) with both family history and clinical factors included. The PGS was statistically significantly (p-value < 0.05) associated with CAD in all models. Schoenfeld residuals were used to test the proportional hazard assumption. Within age and sex strata, only the women under 45 years met the proportional hazards assumptions with a non-significant global p-value. However, this test is known to perform conservatively in large sample sizes. Visual inspection of the Schoenfeld residual figures suggests the violation of assumptions to be small.

Furthermore, the cumulative incidence of CAD was estimated over study time for males and females, and stratified by positive or negative family history together with the PGS tertile representing low, medium, high genetic risk of CAD (Figure 4).

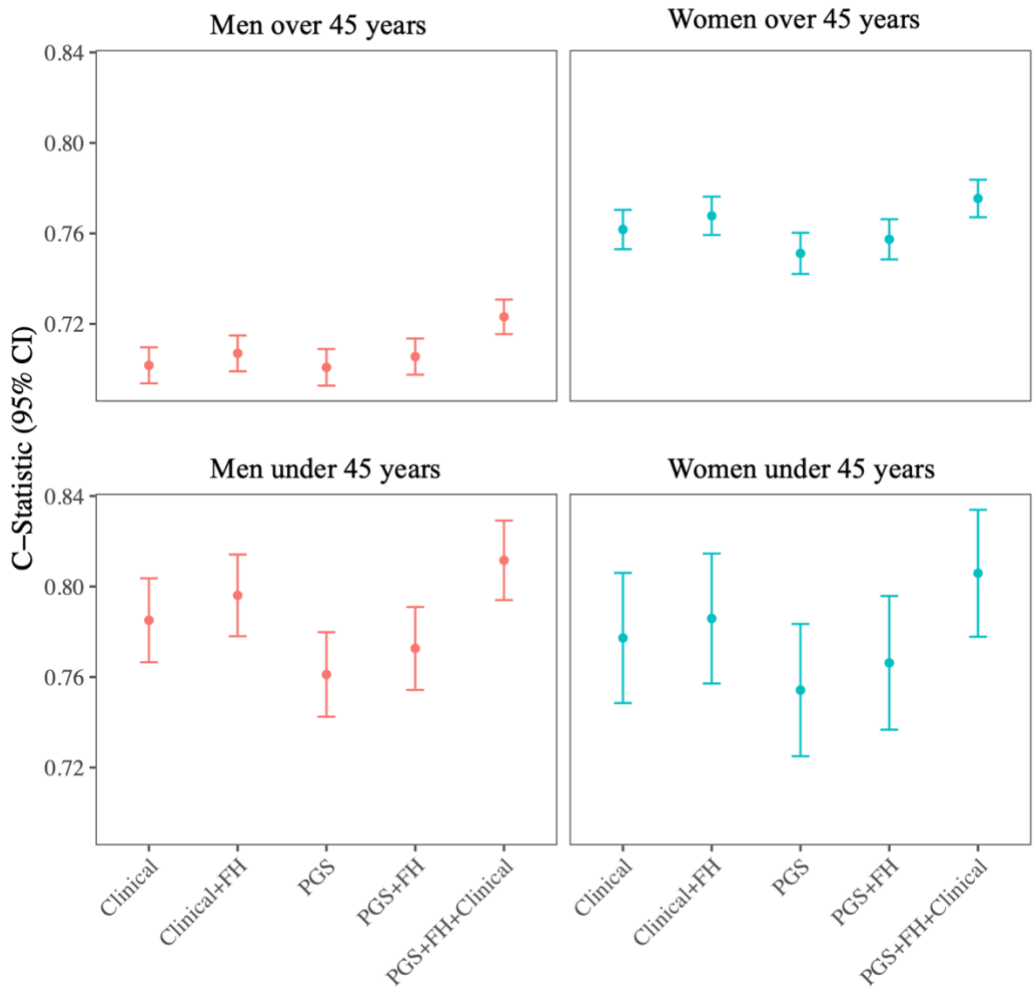
Figure 4. Cumulative incidence of CAD in the study sample (95% CI).



FH+ = Positive family history, FH- = Negative family history of CVD.

Cox regression models were assessed using the Harrel's C-Statistic (Figure 5). In addition to the three models in each sex and age group, the C-Statistic was calculated for models containing only clinical covariates, and clinical and family history of CAD for comparison within each group. The assessment is aimed at comparisons within each group, it is not to be compared across sex and age strata. Precise numbers and their 95% CI is in appendix 3.

Figure 5. Harrel`s C-Statistic from Cox Proportional Hazard regression and their 95% CI.



PGS = Polygenic score. FH = Family history of CAD. Clinical = Total and HDL-cholesterol, systolic blood pressure and smoking status.

5 Discussion

This study aimed to investigate the use of a polygenic score (PGS) to primary prevention of coronary artery disease (CAD) with a particular focus on women under 45 years of age. This chapter will present the findings and discuss their value and implications. Additionally, address the strengths and limitations of the methodology and comment upon future research.

5.1 Main findings

This study it was observed that the PGS for CAD had the highest effect among men and women under 45 years of age, indicating an opportunity for early incorporation of a disease-specific genetic score in primary prevention of CAD. The predicative performance as measured by Harrell's C-statistic and Nagelkerke's R^2 was greatest when added to already established clinical risk factors for CAD. There was not a large difference in PGS effect between men and women, although the family history and PGS together provide stratification into independent risk groups when considering cumulative incidence of disease.

5.1.1 Age-dependent genetic effects

PGS estimate the combined effect of risk associated with multiple genetic variants (30). As one of the main findings, this study found that for individuals under the age 45, the ORs for the association between the PGS for CAD and CAD was significantly higher, compared to the individuals over age 45, for both men and women. This finding can have many explanations. One theory is that the accumulated environmental and lifestyle factors over the lifetime have a larger effect on risk in older individuals than the genetics (105).

It is documented that genetic risk scores can be most useful in younger individuals (106). An individual's genetics do not change in a lifetime, so this observed difference across age groups could imply that genetics vary on their amount of influence on disease in different life stages. Previous research on a range of diseases and traits supports similar findings on PGS and the interaction with age, suggesting that the risk might be more influenced by modifiable environmental or behavioral factors because estimated effect decrease with age. The age-dependent effect of the PGS could also be the result of survivorship bias, in which individuals with high a genetic risk of disease experience earlier onset or more severe disease and do not survive to be well represented in the oldest age groups. This could be part of the reason since

this study has an age split at 45, it could be especially obvious and highlighted in the results. In a multi-biobank study, investigations were done on the effect of PGS and 13 diseases, including CAD. Results showed that the decreasing effect of PGS was approximately linear with age in all diseases, but for CAD the age-specific effect was present in men but not women, In fact, the association between PGS and CAD decreased with age in men but not in women (107).

This study finds a stronger difference in genetic effects across age than between the sexes, suggesting that a genetic score can be useful for both sexes of younger age. Previous research confirms PGS can identify younger individuals (<55) at risk of CAD before clinical risk scores alone can, which underlines an opportunity for the use of PGS either alone or added to already established clinical risk scores, and in result improve primary prevention for this age group (64). The effect sizes of PGS estimated from Cox regression are consistent with those estimated from logistic regression models with a similarly increased PGS effect in individuals under 45 years of age versus over when considered time to event. This is in line with previous research showing the earlier disease age of onset is associated with increased genetic risk (64). This has been seen in monogenic diseases where one mutation causes disease that is often more severe and has an earlier onset than a polygenic form of given disease. As an example, individuals with a monogenic mutation causing familial hypercholesterolemia, have the highest standardized incidence rate for acute myocardial infarction from ages 25 to 39 (108). This can be compared to CAD where individuals in the top 5% of the PGS have an odds of CAD equivalent to that of someone carrying the familial hypercholesterolemia gene (62). It might imply that individuals with early-onset CAD may be more genetically predisposed than those experiencing CAD later in life.

5.1.2 PGS can augment existing clinical risk scores

The ORs and HRs for PGS are attenuated when clinical covariates are added to the model. This is because the clinical covariates explain some of the phenotypic variability, thereby decreasing the PGS effect estimate in both regression methods. This is observed across all age and sex strata. However, the PGS and all covariates are still positively associated with CAD and are statistically significant with p-values below 0.05. Findings from statistical analyses highlight the importance of considering PGS, family history, and clinical factors in the risk assessment for CAD. Indeed, the best model fit is observed when PGS is added alongside

clinical risk factors and family history. The reduction in OR and HR after incorporating family history and clinical factors into the regression models confirms that CAD is a complex condition influenced by multiple factors, and interventions targeting these factors could potentially reduce the burden of CAD. However, even though an individual has a high PGS or genetic predisposition to disease, evidence still shows that adherence to a healthy lifestyle and a focus on avoiding clinical risk factors can significantly reduce the risk of CAD (20). Furthermore, family history can play an important role of identifying individuals with genetic and environmental risk factors in a clinical setting where genotyping is not available for estimating a PGS. Information on family history has the opportunity to combine both genetic and environmental risk factors, as persons from the same family not only share DNA but are usually living in similar places and in similar manner in terms of shared behaviors and social and cultural beliefs (73). But, for prevention among younger individuals, reliance on family history can become challenging if their close relatives are also of young age (109).

One of the key benefits of a genetic score is that it is not time-dependent and can therefore be measured from birth, as opposed to other non-genetic risk factors (87). To decrease the burden of complex diseases like CAD, PGS could be used to further characterize risk in specific populations, such as the young or for those with a specific family history of premature heart disease (32). It is important to note that the ORs and HRs estimated in this study, cannot determine causality, but they do indicate the strength and direction of associations. Based on the results from this study and previous studies, PGS can be valuable and an additional measure for identifying individuals who may benefit from preventive measures such as lifestyle modifications or pharmaceutical intervention (e.g. cholesterol medications) and who are not detected by clinical risk models alone (72). CVDs continue to be the leading cause of death globally for men and women. Our research shows that PGS can improve prediction, particularly in individuals under 45 years of age and should be further explored to estimate CAD risk in Norway and globally. PGS is not a non-modifiable factor and cannot be modified like blood pressure and cholesterol. However, because an individual's DNA is present from birth, a PGS is therefore a tool available early in life to identify high-risk individuals with timely risk prediction. This approach need not be exclusively for CAD but is also possible for other complex diseases (32). For example, the recent BOADICEA risk prediction model for breast cancer incorporates a PGS of 313 SNPs (110). The risk tool incorporates a PGS for breast cancer, cancer family history, and both environmental, reproductive, and hormonal risk factors.

The contribution of genes versus environment and lifestyle exposures contribute to disease is an ongoing field of research and is important to note. Based on the results, PGS can contribute to preventive strategies, especially for younger individuals and in both men and women. Complex diseases like CAD are the result of both genetic and environmental factors, which supports the idea of including both in risk prediction (60). The clinical risk factors included in these analyses are only a handful of known risk factors for CVD and based on the currently used clinical risk score in Europe, SCORE2 (91). There are most likely many other unobserved and unmeasured determining factors which would further optimize risk prediction. Family history and PGS together provide stratification into independent risk groups when considering cumulative incidence of disease.

5.1.3 Predictive performance

Logistic regression and Cox regression models were tested for their model fit and predictive performance. Harrell's C-statistic and Nagelkerke's R^2 was greatest when added to already established clinical risk factors for CAD. The discrimination of Cox regression models was measured by C-statistics ranging from 0.70 to 0.80. These are overall similar to previous metrics found and summarized in a recent scoping review of PGS and CAD (66). The review also found that adding a PGS to clinical risk scores improved predictive performance. In all four groups the C-statistic increased when adding the other explanatory covariates to PGS. The largest increases of predictive performance in this study were among the men and women under 45 years, which confirms that PGS can be especially valuable in younger individuals. However, assessment beyond only C-statistic and in other settings and populations should be performed in the future since results can be influenced by a range of factors (111). The performance of clinical risk models usually gets weaker when applied to or and externally validated in other settings or populations than it was derived from (112).

5.2 Value and implications

This study aimed to investigate the use of a genetic score in primary prevention of CAD. While results are generally similar between men and women, it is important to report sex-specific results for a disease such as CAD with greatly documented sex-differences (113). Further investigations into complex disease and sex-specific risk prediction (54). Previous

research suggests the clinical utility of PGS to be best in collaboration with other non-genetic risk factors, only in individuals with very high scores is the score alone clinically reasonable, and most studies conclude that genetic scores should only be supplemental to traditional risk factors (66). This study shows that PGS can be a useful marker of CAD disease risk in a Norwegian population. However, further studies need to evaluate the clinical applicability. Small improvement of risk prediction on individual level could be valuable and increase health benefits if used at population level (114).

Genetics have the opportunity to improve disease prevention, but the complexity of health and healthcare must not be forgotten. Responsible use involves that risk is outweighed by benefit, and the benefits should not be accessible to only certain individuals, and behind this is a lot of ethical considerations. This study utilizes genetic data and there are several ethical considerations must be made because of this and for the implications that may come from it. An important part of the Human Genome Project was establishing ELSI (Ethical, Legal and Social implications) research program at the same time (115). Genetic research can be subject to a range of problems, such as misuse, exploitation, and discrimination within healthcare (86,87). As examples, insurance companies could use the results from this study to deny health insurance, or it can also be used to discriminate against people with high-risk. Additionally, PGS for traits such as IQ and height and diseases such as CVD are already used for embryo selection in vitro fertilization (116). Therefore, this study should be viewed in the context of its limitations and should not be used to support the idea that anyone with a high PGS is destined to have a given disease. A healthy lifestyle (no smoking, normal body mass index, health diet and physical activity in everyday life) have been found to substantially decrease risk of CAD, even though being in the high-risk tail (20).

5.3 Strengths and limitations

One of the strengths of this study is the large sample which allows for statistically powerful and precise effect estimates representative of the sample population. The HUNT Study has had a historically high participation rate, which decrease concern for selection bias (93). Another strength of the study, and studies in general using electronic health records, there is a low chance of misclassification bias of the outcome since connecting to medical registries to attain ICD-codes of CAD events (93). But ICD codes can still have misclassification particularly if they are used incorrectly. Nevertheless, combining information-rich medical

records with PGS is promising for the future of precision medicine approaches for complex diseases (81). Statistical analyses were done by stratifying on sex and age which is valuable because CAD prevalence, outcomes, and risk factors are known to differ across these groups (117,118). Analyses were adjusted for age, which can decrease possible age-related confounding of results. The HUNT Study recruited adult individuals over 18 years of age and this study has a study population median of 47 years, so there is not much age bias in terms of participant ascertainment.

5.4 Limitations

This study has several limitations. The possibility of bias and confounding are common challenges within population-based observational studies.

The polygenic risk score for CAD is not one universal score, and as of May 2024, the PGS Catalog contains nearly 60 scores for the trait CAD (63). The specific score used in this thesis was aimed at screening in primary prevention (55). The PGS is also estimated based on the combined effect sizes of males and females. The sex-specific genetic effects we anticipate might be more apparent if a male PGS and female PGS were used instead. The score used in this thesis was primarily developed from genome-wide association studies in European ancestry participants. Therefore, the generalization of this score to global populations is limited (86). There are over 20 million genetic variants associated with CAD, and this is continuously growing due to research and advancements in the field of human genetics (119). Clinical implementation of a PGS will need to be dynamic with updated risk estimates based on updated PGS. This research and conclusions are based on a one-time risk stratification by PGS. The study sample consists of an ethnically homogenous cohort of European descent and therefore lacks global transferability (93). To ensure future accessibility and not further health disparities, the inclusion of other populations is crucial (86). Even though this study might have high internal validity, generalizability to other populations can become problematic. Global populations contain genetic variation not seen in Norway or represented in the PGS. Results should be interpreted in line with the study sample and their characteristics.

It is anticipated to have measurement error in blood pressure and cholesterol covariates. The family history covariate is vulnerable to information bias, specifically recall bias, since it is self-reported through questionaries (118). This can lead to both under- and overestimation of

the results. Clinical measurements of blood pressure and cholesterol are time-dependent and might change over time. While we used the earliest available measurement and controlled for the age at time of measurement, this is a potential confounding factor. The study sample has not been connected to medication registries. This could mean that blood pressure and cholesterol values could be from individuals treated with blood pressure or cholesterol medications. The SCORE2 model and other clinical risk models typically require adjustment of the blood pressure and cholesterol measurements if use of medications is known (91). This study sample only has information on medications from 2004 while sample enrollment began in 1995. Due to incomplete medication information, we did not use these to adjust our quantitative measurements. Lastly, we did not consider diabetes as an explanatory covariate for CAD or an exclusion factor. As an example, the SCORE2 was specifically designed for individuals without prior history of CAD or diabetes, but this study did not exclude individuals with prevalent diabetes so this could lead to an underestimation of their CAD risk. Evidence has suggested that as a risk factor, both smoking and diabetes can negatively affect cardiovascular health more in women than men (31), so including diabetes in the future would be useful. Lastly, because the proportional hazards assumption of the Cox regression models was not met in all groups, additional tests for optimal model calibration should be done in the future.

5.5 Further research

To our knowledge, this is the first study using a Norwegian sample utilizing a genetic risk score to investigate CAD primarily in women. Future research should involve narrower age groups, 10- or 5-year age bins, to investigate the specific age for optimal PGS implementation. This will increase knowledge of when prediction or possible stratified screening can be used for best results. Further cost-benefit analysis is needed to estimate the reduction in events and health care expenditures saved by investing in genotyping for certain risk groups. Finally, more research is needed to understand more sex-specific risk factors for CVD that can be used to further refine the clinical risk scores.

6 Conclusion

This study highlights the potential that a PGS can have in improving prediction and prevent CAD in individuals under the age of 45. The results reveal that incorporating a PGS alongside traditional risk factors, can be an effective measure for early identification of risk and give possibility for early intervention. The results indicate that it can be used as a supplement in existing clinical risk models, and especially in younger individuals as genetic predispositions exist before the manifestation of lifestyle and environmental risk factors. The age-dependent genetic effects highlighted in this study suggest that the effect have a strong influence in both sexes under 45 years. Further research is needed for PGS in clinical settings to reduce the major burden of this disease.

6.1.1 References

1. Organization WH. Cardiovascular diseases (CVDs) [Internet]. 2021 [cited 2024 May 13]. Available from: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

2. Vos T, Lim SS, Abbafati C, Abbas KM, Abbasi M, Abbasifard M, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. 396(10258):1204–22.

3. Libby P, Theroux P. Pathophysiology of Coronary Artery Disease. Circulation. 2005;111(25):3481–8.

4. Hobbs FDR. Cardiovascular disease: different strategies for primary and secondary prevention? Heart. 2004;90(10):1217.

5. Network GB of DC. Global Burden of Disease Study 2019 (GBD 2019). Seattle, United States: Institute for Health Metrics and Evaluation (IHME). 2020;

6. Lababidi H, Salerno PRVO, Wass SY, Hasani NS, Bourges-Sevenier B, Al-Kindi S. The Global Burden of premature cardiovascular disease, 1990–2019. Int J Cardiol Cardiovasc Risk Prev. 2023;19:200212.

7. Arora S, Stouffer GA (Rick), Kucharska-Newton A, Qamar A, Vaduganathan M, Pandey A, et al. Twenty Year Trends and Sex Differences in Young Adults Hospitalized with Acute Myocardial Infarction: The ARIC Community Surveillance Study. Circulation. 139(8):1047–56.

8. Vogel B, Acevedo M, Appelman Y, Merz CNB, Chieffo A, Figtree GA, et al. The Lancet women and cardiovascular disease Commission: reducing the global burden by 2030. Lancet. 2021;397(10292):2385–438.

9. NOU 2023: 5. Den store forskjellen. Om kvinners helse og betydningen av kjønn for helse [The big difference. On women´s health in Norway and why sex and gender matter]. Oslo: Helse- og omsorgsdepartementet. Norwegian. [Internet]. [cited 2024 Apr 1]. Available from: https://www.regjeringen.no/no/dokumenter/nou-2023-5/id2964854/

10. Nations U. The Sustainable Development Goals. Goal 3: Ensure healthy lives and promote well-being for all at all ages [Internet]. [cited 2024 Apr 1]. Available from: https://sdgs.un.org/goals/goal3#targets_and_indicators

11. Gerdts E, Regitz-Zagrosek V. Sex differences in cardiometabolic disorders. Nat Med. 2019;25(11):1657–66.

12. Harvey RE, Coffman KE, Miller VM. Women-Specific Factors to Consider in Risk, Diagnosis and Treatment of Cardiovascular Disease. Women's Heal. 2015;11(2):239–57.

13. Woodward M. Cardiovascular Disease and the Female Disadvantage. Int J Environ Res Public Heal. 2019;16(7):1165.

14. Melloni C, Berger JS, Wang TY, Gunes F, Stebbins A, Pieper KS, et al. Representation of women in randomized clinical trials of cardiovascular disease prevention. Circ Cardiovasc Qual outcomes. 2010;3(2):135–42.

15. Phillips SP. Defining and measuring gender: A social determinant of health whose time has come. Int J Equity Heal. 2005;4(1):11.

16. Regitz-Zagrosek V. Sex and gender differences in health. Science & Society Series on Sex and Science. EMBO Rep. 2012;13(7):596–603.

17. Legato MJ, Johnson PA, Manson JE. Consideration of Sex Differences in Medicine to Improve Health Care and Patient Outcomes. JAMA. 2016;316(18):1865.

18. Crooijmans J, Singh S, Naqshband M, Bruikman CS, Pinto-Sietsma SJ. Premature atherosclerosis: An analysis over 39 years in the Netherlands. Implications for young individuals in high-risk families. Atherosclerosis. 2023;384:117267.

19. Villiers CB de, Kroese M, Moorthie S. Understanding polygenic models, their development and the potential application of polygenic scores in healthcare. J Méd Genet. 2020;57(11):725–32.

20. V. KA, A. EC, Isabel D, Pradeep N, G. BA, R. CN, et al. Genetic Risk, Adherence to a Healthy Lifestyle, and Coronary Disease. N Engl J Med. 2016;375(24):2349–58.

21. Dudbridge F, Pashayan N, Yang J. Predictive accuracy of combined genetic and environmental risk scores. Genet Epidemiology. 2018;42(1):4–19.

22. Mahmood SS, Levy D, Vasan RS, Wang TJ. The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. Lancet (Lond, Engl). 2013;383(9921):999–1008.

23. ROSE G. Sick Individuals and Sick Populations. Int J Epidemiology. 1985;14(1):32–8.

24. Yusuf S, Joseph P, Rangarajan S, Islam S, Mente A, Hystad P, et al. Modifiable risk factors, cardiovascular disease, and mortality in 155 722 individuals from 21 high-income, middle-income, and low-income countries (PURE): a prospective cohort study. Lancet. 2020;395(10226):795–808.

25. Marmot M, Friel S, Bell R, Houweling TAJ, Taylor S, Health C on SD of. Closing the gap in a generation: health equity through action on the social determinants of health. Lancet (Lond, Engl). 2008;372(9650):1661–9.

26. Harris KM, McDade TW. The Biosocial Approach to Human Development, Behavior, and Health Across the Life Course. RSF: Russell Sage Found J Soc Sci. 2018;4(4):2–26.

27. Dai H, Much AA, Maor E, Asher E, Younis A, Xu Y, et al. Global, regional, and national burden of ischaemic heart disease and its attributable risk factors, 1990–2017: results from the Global Burden of Disease Study 2017. Eur Hear J - Qual Care Clin Outcomes. 2020;8(1):50–60.

28. Roth GA, Mensah GA, Johnson CO, Addolorato G, Ammirati E, Baddour LM, et al. Global Burden of Cardiovascular Diseases and Risk Factors, 1990–2019 Update From the GBD 2019 Study. J Am Coll Cardiol. 2020;76(25):2982–3021.

29. Riggs DW, Yeager RA, Bhatnagar A. Defining the Human Envirome. Circ Res. 2018;122(9):1259–75.

30. Khera AV, Kathiresan S. Genetics of coronary artery disease: discovery, biology and clinical translation. Nat Rev Genet. 2017;18(6):331–44.

31. Appelman Y, Rijn BB van, Haaf ME ten, Boersma E, Peters SAE. Sex differences in cardiovascular risk factors and disease prevention. Atherosclerosis. 2015;241(1):211–8.

32. Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk of complex disease. Curr Opin Genet Dev. 2008;18(3):257–63.

33. Wenger NK. Cardiovascular Disease in Women: Understanding the Journey. Clin Chem. 2020;67(1):24–9.

34. Young L, Cho L. Unique cardiovascular risk factors in women. Heart. 2019;105(21):1656.

35. Wu P, Haththotuwa R, Kwok CS, Babu A, Kotronias RA, Rushton C, et al. Preeclampsia and Future Cardiovascular Health: A Systematic Review and Meta-Analysis. Circ Cardiovasc Qual outcomes. 2016;10(2).

36. Teede H, Deeks A, Moran L. Polycystic ovary syndrome: a complex condition with psychological, reproductive and metabolic manifestations that impacts on health across the lifespan. BMC Med. 2010;8(1):41.

37. Goodarzi MO, Dumesic DA, Chazenbalk G, Azziz R. Polycystic ovary syndrome: etiology, pathogenesis and diagnosis. Nat Rev Endocrinol. 2011;7(4):219–31.

38. Agarwala A, Michos ED, Samad Z, Ballantyne CM, Virani SS. The Use of Sex-Specific Factors in the Assessment of Women's Cardiovascular Risk. Circulation. 2020;141(7):592–9.

39. Saeed A, Kampangkaew J, Nambi V. Prevention of Cardiovascular Disease in Women. Methodist DeBakey Cardiovasc J. 2017;13(4):185–92.

40. Pravda NS, Karny-Rahkovich O, Shiyovich A, Pravda MS, Rapeport N, Vaknin-Assa H, et al. Coronary Artery Disease in Women: A Comprehensive Appraisal. J Clin Med. 2021;10(20):4664.

41. Consortium TGCR, Magnussen C, Ojeda FM, Leong DP, Alegre-Diaz J, Amouyel P, et al. Global Effect of Modifiable Risk Factors on Cardiovascular Disease and Mortality. N Engl J Med. 2023;389(14):1273–85.

42. Braveman P, Gottlieb L. The Social Determinants of Health: It's Time to Consider the Causes of the Causes. Public Heal Rep. 2014;129(1_suppl2):19–31.

43. Ernstsen L, Strand BH, Nilsen SM, Espnes GA, Krokstad S. Trends in absolute and relative educational inequalities in four modifiable ischaemic heart disease risk factors: repeated cross-sectional surveys from the Nord-Trøndelag Health Study (HUNT) 1984–2008. BMC Public Heal. 2012;12(1):266.

44. Igland J, Vollset SE, Nygård OK, Sulo G, Ebbing M, Tell GS. Educational Inequalities in Acute Myocardial Infarction Incidence in Norway: A Nationwide Cohort Study. PLoS ONE. 2014;9(9):e106898.

45. Klarin D, Natarajan P. Clinical utility of polygenic risk scores for coronary artery disease. Nat Rev Cardiol. 2022;19(5):291–301.

46. Natarajan P. Polygenic Risk Scoring for Coronary Heart Disease The First Risk Factor ∗. J Am Coll Cardiol. 2018;72(16):1894–7.

47. Lloyd-Jones DM, Leip EP, Larson MG, D'Agostino RB, Beiser A, Wilson PWF, et al. Prediction of Lifetime Risk for Cardiovascular Disease by Risk Factor Burden at 50 Years of Age. Circulation. 2006;113(6):791–8.

48. Ware JH. The limitations of risk factors as prognostic tools. N Engl J Med. 2006;355(25):2615–7.

49. Mauvais-Jarvis F, Merz NB, Barnes PJ, Brinton RD, Carrero JJ, DeMeo DL, et al. Sex and gender: modifiers of health, disease, and medicine. Lancet (Lond, Engl). 2020;396(10250):565–82.

50. Hawkes S, Haseen F, Aounallah-Skhiri H. Measurement and meaning: reporting sex in health research. Lancet. 2019;393(10171):497–9.

51. Mosca L, Barrett-Connor E, Wenger NK. Sex/gender differences in cardiovascular disease prevention: what a difference a decade makes. Circulation. 2011;124(19):2145–54.

52. Ariansen IKH, Olsen K, Selmer RM. Folkehelserapporten. Hjerte- og karsykdommer i Norge. [Internet]. Folkehelseinstituttet; 2021 Nov [cited 2024 Feb 25]. Available from: https://www.fhi.no/he/folkehelserapporten/ikke-smittsomme/Hjerte-kar/?term=

53. Knudsen AK, Sulo G, Klitkou ST, Vollset SE. Disease burden in Norway in 2050: A forecasting analysis based on the Global Burden of Disease (GBD) Study. Eur J Public Heal. 2023;33(Supplement_2):ckad160.649.

54. Surakka I, Wolford BN, Ritchie SC, Hornsby WE, Sutton NR, Gabrielsen ME, et al. Sex-Specific Survival Bias and Interaction Modeling in Coronary Artery Disease Risk Prediction. Circ: Genom Precis Med. 2023;16(1):e003542.

55. 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. J Am Coll Cardiol. 2018;72(16):1883–93.

56. Collins FS, Green ED, Guttmacher AE, Guyer MS, Institute UNHGR. A vision for the future of genomics research. Nature. 2003;422(6934):835–47.

57. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, et al. Complement Factor H Polymorphism in Age-Related Macular Degeneration. Science. 2005;308(5720):385–9.

58. Uffelmann E, Huang QQ, Munung NS, Vries J de, Okada Y, Martin AR, et al. Genome-wide association studies. Nat Rev Methods Prim. 2021;1(1):59.

59. Choi SW, Mak TSH, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. Nat Protoc. 2020;15(9):2759–72.

60. Muse ED, Chen SF, Torkamani A. Monogenic and Polygenic Models of Coronary Artery Disease. Curr Cardiol Rep. 2021;23(8):107.

61. 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(9750):1393–400.

62. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet. 2018;50(9):1219–24.

63. The Polygenic Score (PGS) Catalog [Internet]. [cited 2024 Apr 14]. Available from: https://www.pgscatalog.org

64. Mars N, Koskela JT, Ripatti P, Kiiskinen TTJ, Havulinna AS, Lindbohm JV, et al. Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers. Nat Med. 2019;26(4):549–57.

65. Marston NA, Pirruccello JP, Melloni GEM, Koyama S, Kamanu FK, Weng LC, et al. Predictive Utility of a Coronary Artery Disease Polygenic Risk Score in Primary Prevention. JAMA Cardiol. 2023;8(2):130–7.

66. Phulka JS, Ashraf M, Bajwa BK, Pare G, Laksman Z. Current State and Future of Polygenic Risk Scores in Cardiometabolic Disease: A Scoping Review. Circ: Genom Precis Med. 2023;16(3):286–313.

67. Valdez R, Yoon PW, Qureshi N, Green RF, Khoury MJ. Family History in Public Health Practice: A Genomic Tool for Disease Prevention and Health Promotion*. Annu Rev Public Heal. 2010;31(1):69–87.

68. 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.

69. Woodward M, Brindle P, Tunstall-Pedoe H, estimation S group on risk. Adding social deprivation and family history to cardiovascular risk assessment: the ASSIGN score from the Scottish Heart Health Extended Cohort (SHHEC). Heart. 2007;93(2):172.

70. Marenberg ME, Risch N, Berkman LF, Floderus B, Faire U de. Genetic susceptibility to death from coronary heart disease in a study of twins. N Engl J Med. 1994;330(15):1041–6.

71. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era — concepts and misconceptions. Nat Rev Genet. 2008;9(4):255–66.

72. Aragam KG, Dobbyn A, Judy R, Chaffin M, Chaudhary K, Hindy G, et al. Limitations of Contemporary Guidelines for Managing Patients at High Genetic Risk of Coronary Artery Disease. J Am Coll Cardiol. 2020;75(22):2769–80.

73. Mars N, Lindbohm JV, Parolo P della B, Widén E, Kaprio J, Palotie A, et al. Systematic comparison of family history and polygenic risk across 24 common diseases. Am J Hum Genet. 2022;109(12):2152–62.

74. Hujoel MLA, Loh PR, Neale BM, Price AL. Incorporating family history of disease improves polygenic risk scores in diverse populations. Cell Genom. 2022;2(7):100152.

75. Abraham G, Havulinna AS, Bhalala OG, Byars SG, Livera AMD, Yetukuri L, et al. Genomic prediction of coronary heart disease. Eur Hear J. 2016;37(43):3267–78.

76. Lopez AD, Adair T. Is the long-term decline in cardiovascular-disease mortality in high-income countries over? Evidence from national vital statistics. Int J Epidemiology. 2019;48(6):1815–23.

77. Hussain, Rifai A and A, Khalid M and, Virani U and, Salim S. Cardiovascular Risk Assessment in Primary Prevention. ["Shapiro, D."] Michael, editors. 2022;3–19. Available from: https://doi.org/10.1007/978-3-030-98824-1_1

78. Europe WRO for. Screening programmes: a short guide. Increase effectiveness, maximize benefits and minimize harm. Copenhagen; 2020.

79. Brodersen J, Siersma VD. Long-Term Psychosocial Consequences of False-Positive Screening Mammography. Ann Fam Med. 2013;11(2):106–15.

80. Ramaswami R, Bayer R, Galea S. Precision Medicine from a Public Health Perspective. Annu Rev Public Heal. 2017;39(1):153–68.

81. Wolford BN, Willer CJ, Surakka I. Electronic health records: the next wave of complex disease genetics. 2018;27(R1):R14–21. Available from: https://doi.org/10.1093/hmg/ddy081

82. Collister JA, Liu X, Clifton L. Calculating Polygenic Risk Scores (PRS) in UK Biobank: A Practical Guide for Epidemiologists. Front Genet. 2022;13:818574.

83. Brumpton BM, Graham S, Surakka I, Skogholt AH, Løset M, Fritsche LG, et al. The HUNT study: A population-based cohort for genetic research. Cell Genom. 2022;2(10):100193.

84. Carress H, Lawson DJ, Elhaik E. Population genetic considerations for using biobanks as international resources in the pandemic era and beyond. BMC Genom. 2021;22(1):351.

85. Fatumo S, Chikowore T, Choudhury A, Ayub M, Martin AR, Kuchenbaecker K. A roadmap to increase diversity in genomic studies. Nat Med. 2022;28(2):243–50.

86. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet. 2018;51(4):584–91.

87. Alliance PRSTF of the ICD, Adeyemo A, Balaconis MK, Darnes DR, Fatumo S, Moreno PG, et al. Responsible use of polygenic risk scores in the clinic: potential benefits, risks and gaps. Nat Med. 2021;27(11):1876–84.

88. Rossello X, Dorresteijn JA, Janssen A, Lambrinou E, Scherrenberg M, Bonnefoy-Cudraz E, et al. Risk prediction tools in cardiovascular disease prevention: A report from the ESC Prevention of CVD Programme led by the European Association of Preventive Cardiology (EAPC) in collaboration with the Acute Cardiovascular Care Association (ACCA) and the Association of Cardiovascular Nursing and Allied Professions (ACNAP). Eur J Prev Cardiol. 2020;26(14):1534–44.

89. Selmer R, Igland J, Ariansen I, Tverdal A, Njølstad I, Furu K, et al. NORRISK 2: A Norwegian risk model for acute cerebral stroke and myocardial infarction. European Journal of Preventive Cardiology [Internet]. 2017 May 1;24(7):773–82. Available from: https://doi.org/10.1177/2047487317693949

90. Visseren FLJ, Mach F, Smulders YM, Carballo D, Koskinas KC, Bäck M, et al. 2021 ESC Guidelines on cardiovascular disease prevention in clinical practice. Eur Hear J. 2021;42(34):3227–337.

91. collaboration S working group and EC risk, Hageman S, Pennells L, Ojeda F, Kaptoge S, Kuulasmaa K, et al. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. Eur Hear J. 2021;42(25):ehab309-.

92. collaboration SO working group and EC risk, Vries TI de, Cooney MT, Selmer RM, Hageman SHJ, Pennells LA, et al. SCORE2-OP risk prediction algorithms: estimating incident cardiovascular event risk in older persons in four geographical risk regions. Eur Hear J. 2021;42(25):2455–67.

93. Åsvold BO, Langhammer A, Rehn TA, Kjelvik G, Grøntvedt TV, Sørgjerd EP, et al. Cohort Profile Update: The HUNT Study, Norway. Int J Epidemiology. 2022;52(1):e80–91.

94. Lambert SA, Gil L, Jupp S, Ritchie SC, Xu Y, Buniello A, et al. The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. Nat Genet. 2021;53(4):420–5.

95. Catalog P. Polygenic Score (PGS) ID: PGS000018. [Internet]. [cited 2024 Apr 14]. Available from: https://www.pgscatalog.org/score/PGS000018/

96. Hughes G, Choudhury RA, McRoberts N. Summary Measures of Predictive Power Associated with Logistic Regression Models of Disease Risk. Phytopathology®. 2019;109(5):712–5.

97. NAGELKERKE NJD. A note on a general definition of the coefficient of determination. Biometrika. 1991;78(3):691–2.

98. Kleinbaum DG, Klein M. Survival Analysis, A Self-Learning Text. 2012;

99. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. 1996;15(4):361–87.

100. Caetano SJ, Sonpavde G, Pond GR. C-statistic: A brief explanation of its construction, interpretation and limitations. Eur J Cancer. 2018;90:130–2.

101. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the Tidyverse. J Open Source Softw. 2019;4(43):1686.

102. Therneau TM. A Package for Survival Analysis in R [Internet]. 2020. Available from: https://CRAN.R-project.org/package=survival.

103. Wickham H. ggplot2, Elegant Graphics for Data Analysis. 2016;

104. Sjoberg D D, Whiting K, Curry M, Lavery J A, Larmarange J. Reproducible Summary Tables with the gtsummary Package. R J. 2021;13(1):570.

105. Jiang X, Holmes C, McVean G. The impact of age on genetic risk for common diseases. PLoS Genet. 2021;17(8):e1009723.

106. 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. Eur Hear J. 2016;37(6):561–7.

107. Jermy B, Läll K, Wolford B, Wang Y, Zguro K, Cheng Y, et al. A unified framework for estimating country-specific cumulative incidence for 18 diseases stratified by polygenic risk. medRxiv. 2023;2023.06.12.23291186.

108. Mundal LJ, Igland J, Veierød MB, Holven KB, Ose L, Selmer RM, et al. Impact of age on excess risk of coronary heart disease in patients with familial hypercholesterolaemia. Heart. 2018;104(19):1600–7.

109. Wolford BN, Surakka I, Graham SE, Nielsen JB, Zhou W, Gabrielsen ME, et al. Utility of family history in disease prediction in the era of polygenic scores. medRxiv. 2021;2021.06.25.21259158.

110. Mavaddat N, Ficorella L, Carver T, Lee A, Cunningham AP, Lush M, et al. Incorporating alternative Polygenic Risk Scores into the BOADICEA breast cancer risk prediction model. Cancer Epidemiology, Biomark Prev. 2023;32(3):422–7.

111. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. Circulation. 2007;115(7):928–35.

112. Moons KGM, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: II. External validation, model updating, and impact assessment. Heart. 2012;98(9):691.

113. Woodward M. Rationale and tutorial for analysing and reporting sex differences in cardiovascular associations. Heart. 2019;105(22):1701–8.

114. Sun L, Pennells L, Kaptoge S, Nelson CP, Ritchie SC, Abraham G, et al. Polygenic risk scores in cardiovascular risk prediction: A cohort study and modelling analyses. PLoS Med. 2021;18(1):e1003498.

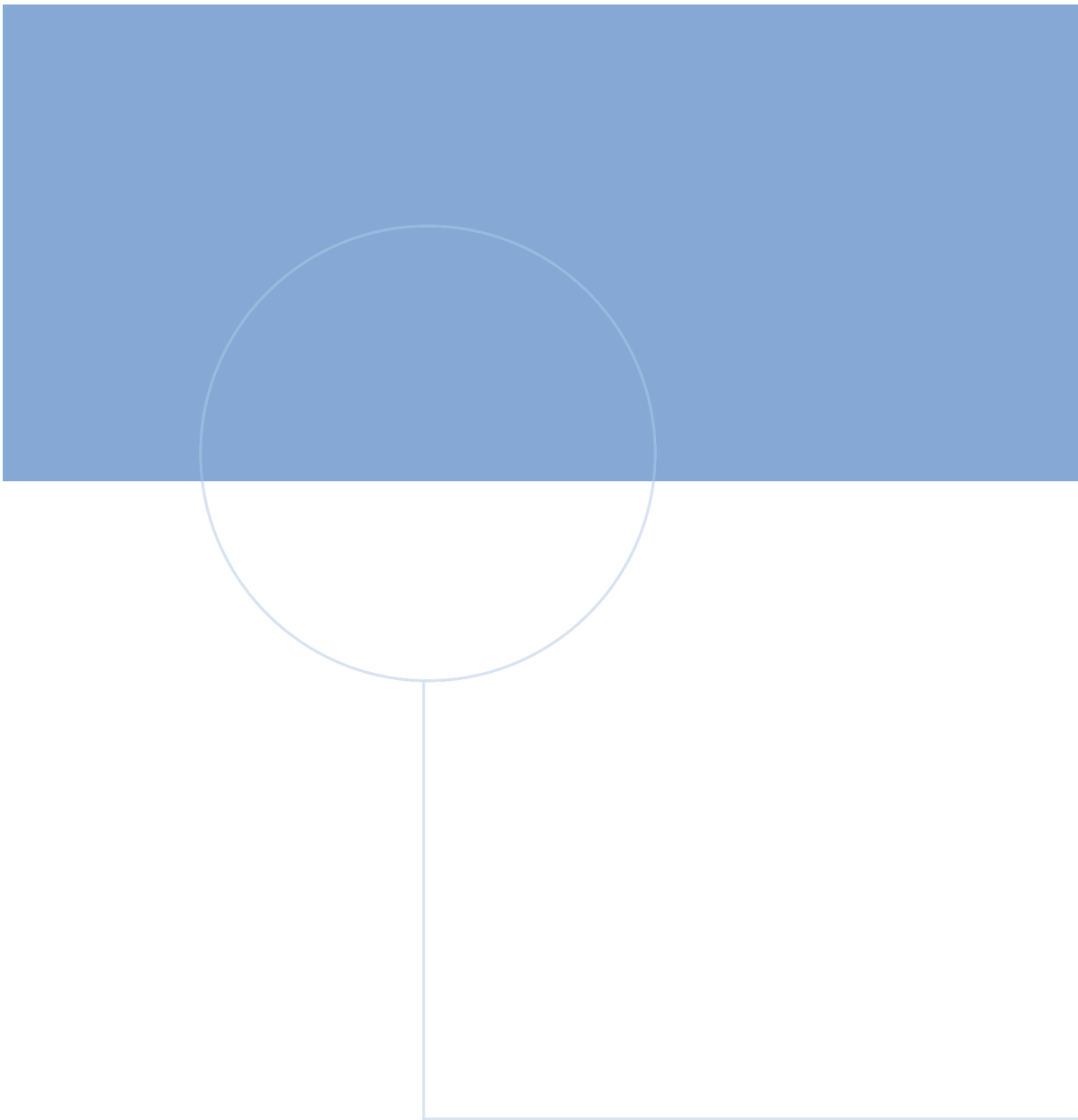
115. Dolan DD, Lee SSJ, Cho MK. Three decades of ethical, legal, and social implications research: Looking back to chart a path forward. Cell Genom. 2022;2(7):100150.

116. Grebe TA, Khushf G, Greally JM, Turley P, Foyouzi N, Rabin-Havt S, et al. Clinical utility of polygenic risk scores for embryo selection: A points to consider statement of the American College of Medical Genetics and Genomics (ACMG). Genet Med. 2024;26(4):101052.

117. Rohrer JM. Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data. Adv Methods Pr Psychol Sci. 2017;1(1):27–42.

118. Grimes DA, Schulz KF. Bias and causal associations in observational research. Lancet (Lond, Engl). 2002;359(9302):248–52.

119. Aragam KG, Jiang T, Goel A, Kanoni S, Wolford BN, Atri DS, et al. Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants. Nat Genet. 2022;54(12):1803–15.



NTNU

Norwegian University of
Science and Technology