Predictive modelling of SNP for myocardial infarction in SMuRFless patients

[1]Amra Elahi, [2]Elamathi Natarajan
[1]Student, [2]Professor
Department of Bioinformatics
Biotecnika info Labs, Bengaluru, India
[1]amraelahi@gmail.com, [2]elamathi@biotecnika.org

Abstract

The proportion of patients with myocardial infarction (MI) who have no standard modifiable risk factors such as smoking, hypertension, diabetes and dyslipidaemia has been increasing; this subgroup has been termed 'SMuRF-less MI'. Although these patients often receive guideline appropriate therapy there are poor short term outcomes; this indicates that current clinical risk stratification methods (i.e., traditional risk factors) are limited for these patients. Recent studies have demonstrated that the genetic predisposition to MI, specifically due to single nucleotide polymorphisms (SNPs), is a significant contributing factor among the SMuRF-less population. This review will explore the utility of SNPs as potential predictive markers for MI development and assess the effectiveness of polygenic risk score and machine learning algorithms for assessing genetic risk. The use of genome-wide approaches that aggregate numerous variants across various genes to better define the level of risk exhibited may identify those patients at high risk of developing MI who had been previously unidentified by conventional risk scoring approaches. However, current SNP-based prediction models show little incremental predictive value beyond existing methods and therefore require more careful variant selection, larger reference populations and more robust modelling techniques. The combination of genetic data with biomarker and clinical data would improve the predictive accuracy of MI risk scores and lead to a more personalised approach to prevention and treatment for patients with SMuRF-less MI.

Keywords: SMuRF-less myocardial infarction, single nucleotide polymorphisms (SNPs), polygenic risk score (PRS), genetic risk prediction, genome-wide association studies (GWAS), machine learning, cardiovascular risk stratification, precision medicine.

Introduction

Recent data indicates that a growing population of patients with myocardial infarction (MI) are not exhibiting any of the known modifiable cardiovascular risk factors (hypertension, diabetes, smoking, or dyslipidemia) referred to as SMuRF-less MI has been established [1]. Reported estimates suggest that a large, and in some cohorts, increasingly large percentage of individuals who sustain an MI will do so without any SMuRFs, indicating a serious limitation of the current system for the classification of cardiovascular risks [1]. Additionally, an increasing percentage of patients who have experienced an MI with SMuRF-less risk profiles exhibit a paradoxically higher rate of short-term mortality and poor outcome compared to those patients exhibiting traditional cardiovascular risk factors, indicating a significant role for other risk factor clusters, such as genetics, in the development of coronary artery disease in the SMuRF-less population. The inability of traditional clinical models to predict MI in patients with SMuRF-less profiles has resulted in the application of genomic data, including the use of single nucleotide polymorphisms (SNPs) in predictive modelling [1]. As a result of genomic epidemiology and continued advancements in machine learning, predictive models that synthesise SNPs into an integrative model, or polygenic risk scores, provide better identification of patients with an increased risk of MI than traditional cardiovascular risk factors alone [1]. The availability of these tools is particularly valuable for SMuRF-less populations where clinical indicators are absent or weak, and where there are no clear clinical signals that would indicate a patient's vulnerability to developing an MI. There is potential for early risk diagnosis and tailored intervention in SMuRF-less patients by combining SNP-based predictive models with clinical and demographic data [1].

Conventional methods of predicting risk for Myocardial Infarction (MI) are inadequate in regards to patients who have been diagnosed with an MI but do not have any of the usual Modifiable Risk Factors (MRF) (e.g., smoking, high cholesterol, hypertension) [2]. Traditional risk scores do not include those patients who do not meet the criteria outlined in clinical guidelines to determine whether they are at risk, and therefore do not provide any predictive value [2]. More people are beginning to recognize that genetic susceptibility, particularly through SNPs, contributes significantly to a person's overall risk of suffering from an MI, unrelated to other risk factors. Polygenic Risk Scores (PRS), along with other methods of analysis of SNPs, can assist with improving the risk stratification of patients with MRF-less ACS, and with identifying at-risk patients early in their disease course [3]. The vast majority of patients who experience an Acute Coronary Syndrome (ACS) do not have any of the usual MRFs, thus limiting the effectiveness of standard clinical scoring systems to identify at-risk individuals [3]. Clinical risk scores have been shown to be ineffective in identifying patients at risk for ACS because the majority of studies on population-level incidence of ACS show a high frequency of cases occurring in individuals without MRFs [3]. Genetic risk due to SNPs may also increase the likelihood of developing MI. Including SNP data into predictive models, along with the development of Polygenic Risk Scores, may help to better identify at-risk individuals within the SMuRF-less population [3].

Single nucleotide polymorphisms (SNPs) as predictive biomarkers

In this review article, the clinical relevance of both novel and traditional biomarkers for heart failure and acute myocardial infarction (AMI) will be highlighted, as well as the limitations associated with relying solely on conventional risk factors to diagnose these conditions [4]. With regard to SMuRF-less myocardial infarction, the study presents evidence that molecular and genetic analysis is necessary in combination with conventional clinical markers, as these markers do not fully explain the mechanisms involved in this condition [4]. This further supports the current project's rationale for developing SNP-based predictive models of the genetic predisposition to myocardial infarction, in order to provide an improved means of risk stratification and more accurately identifying individuals affected by AMI than current biomarker and risk factor approaches [4].

Predictive modelling approaches in genetic risk assessment

Using genome-level SNP data allows for a quantitative assessment of risk for complex diseases, since SNP data can integrate several loci's contributions to the overall risk; however, most of those loci will have minimal individual contributions to that overall risk [5]. The ability to accurately predict an individual's disease risk using SNPs based on the parameters outlined in the deterministic model is determined primarily by trait heritability, number of informative markers available to use with your model for training, and reference population size that was the basis of your model development [5]. The deterministic model provides a framework for establishing the limits of how accurately we can predict complex diseases and guiding the establishment of research studies and providing context for the use of genetic risk scores [5]. These principles are also applicable to multiple cardiovascular disease risk factors and myocardial infarction prediction where genome-wide studies can also identify susceptibility to Myocardial Infarction based on GWAS beyond traditional risk factor-based methods [5].

The occurrence of myocardial infarction in populations that do not have any of the standard modifiable cardiovascular risk factors (SMuRF) highlights the significance of nontraditional factors that determine susceptibility to disease, including genetic variability [6]. The use of aggregate SNPs to form a polygenic model allows researchers to utilize the total impacts of many common genetic variations that have minimal independent effects on an individual basis; however, this approach has revealed that the risk of a coronary event for an individual is correlated with the sum of the effects of specific SNPs and is independent of traditional clinical predictors [6]. Therefore, SNP-based prediction models will offer critical information about the presence of genetic factors in SMuRFless individuals that traditional risk score methods may not identify. In fact, because of the limited amount of predictive power shown to date, we will need to select additional genetic variants, expand our reference sample size, and utilize disease-specific modelling methods to identify the different levels of risk within this population [6].

The genetic basis for coronary artery disease (CAD) and myocardial infarction (MI) has many similarities at both a major level (through common gene variants) and at a minor level (through infrequent gene variants) [7]. Both of these disorders have been found to be highly inheritable based on studies of both families and single population samples over the years [7]. There are also numerous other loci found to be associated with coronary artery disease (CAD) and myocardial infarction (MI) within different ethnic/racial groups and populations. Cardiovascular disease (CVD) risk can be better predicted by combining genetic diversity with family history and environmental risk factors [7]. Researchers have sought to overcome the limitations of traditional statistical models by employing more advanced machine-learning techniques to estimate cardiovascular risk [8]. The use of evolutionary algorithms, such as genetic programming, enables the creation of automated predictive models that can identify and capture the complex and non-linear relationships between clinical variables throughout time-to-event data [8]. It has been established through various studies of comparison that predictive models created using these types of methods produce results that are comparable in quality to those generated by traditional regression techniques, as they offer a greater diversity of modelling constructs [8]. The development of these types of methods demonstrates how data-driven modelling frameworks may assist in enhancing cardiovascular risk stratification in cases where multiple aetiological factors combine to produce cardiovascular disease and where known risk factors do not completely explain the aetiology [8].

Current challenges in SNP-based prediction for SMuRF-less MI

The recent growth in how we are able to use genomics to predict the development of complex diseases has highlighted the limitations of the traditional way of using polygenic risk scores and how machine learning models may provide a better method of predicting the risk of these diseases [9]. Polygenic risk scores are composites formed by combining the combined effects of many different risk alleles to generate a single score indicating the overall amount of gene-based risk for developing a disease [9]. Because the calculations that derive polygenic risk scores are generally made utilizing the assumption that the effects of each allele in determining a person's chance of developing a disease will combine to produce an additive cumulative result for that individual, it is possible that many of the risk factors for a particular disease may have complex interactions that would not be apparent with a polygenic risk model [9]. Conversely, when we analyse the relationship between SNPs and the risk of developing disease using machine learning based methods we can build models that account for high dimensional non-linear relationships as well as being able to inform us about our risk for developing a disease as well as how likely it is that we will develop a disease based on SNPs, which is why machine learning models are increasingly being explored as options for precision medicine in complex genetic diseases [9]. Research has shown that multilocus genetic risk score methods that combine information from multiple loci with coronary risk alleles can help researchers identify individuals who are at risk for future cardiovascular events [10]. In cohort studies consisting of young individuals who survived their first myocardial infarction, higher combined genetic risk scores predicted higher rates of recurrence of CVD-related events [10]. The combination of genetic and clinical model information appears to add additional predictive power in identifying CVD patients that require close monitoring. Therefore, using multilocus genetic risk scores in conjunction with clinical data will likely improve the precision of secondary prevention [10].

Genetic association studies of large adult populations have identified multiple single nucleotide polymorphisms (SNPs) on a population basis which are significantly associated with an increased risk for future heart attacks [11]. In examining prospective cohorts of adults aged 60 years and older, several SNPs that were previously associated with cardiovascular disease had nominal associations with new heart attacks. In addition, several of the loci associated with heart attacks (such as those related to lipoprotein metabolism) had consistently demonstrated elevated risk for heart attacks independent of traditional clinical risk factors [11]. The data indicates that some genetic variants can increase an individual's susceptibility to a heart attack, further supporting the need to incorporate SNP information into genetic profiling of complex cardiovascular disease processes [11]. Studies have looked into using genetics-based risk scores that combine the information of many individual single nucleotide polymorphisms (SNPs) to improve the prediction of risk of developing early coronary artery disease in addition to the typical clinical risk factors [12]. In earlier attempts at developing these risk scores, the methods involved the use of several selected candidate genes. More recently, many of the new methods are using genome-wide SNP data to create comprehensive biological

risk scores for genes [12]. These polygenic-risk-type measures reflect the very complicated multi-factorial nature of the genetic architecture of coronary disease [12]. These polygenic-risk types of scores can help to identify individuals that may not show any conventional signs of being at risk, and thus demonstrate the future ability of genetic profiles to improve the ability to classify patients according to their coronary-disease risk [12].

Conclusion

The contributing factors for myocardial infarction vary greatly [13]. There are many individual factors (such as genetics) as well as the combined effects of many different types of factors (both common and rare variants). These factors act through pathways including, but not limited to, lipid metabolism, inflammation, vascular function, thrombosis and plaque stability [13]. There have been a number of studies that have examined myocardial infarction risk at the genomic level and have identified several risk loci as well as SNP-based models of predictability that combine multiple genetic factors into a single polygenic risk score [13]. The use of genetic risk scores provides a powerful tool to predict myocardial infarction, especially for SMuRFless myocardial infarction, where there is a limited number of traditional risk factors. Though it is well established that genetic risk scores have a direct correlation with myocardial infarction, the additional clinical predictive value above and beyond that provided by traditional clinical models of risk assessment is modest [13]. Finally, integrating genetic data with clinical and biomarker data should ultimately improve how we personalize our risk stratification and preventive strategies for myocardial infarction [13].

Patients who have ST-elevation myocardial infarction (STEMI) but do not have any of the usual modifiable risk factors such as hypertension; diabetes; hyperlipidemia or smoking are a unique and important clinical group [14]. There is evidence that patients without SMuRFs, including younger adults, have a much poorer short-term prognosis than males and females with conventional backgrounds; have much higher early mortality and have significantly more early complications (including cardiac arrest, need for pressor support, either inotropic and/or mechanical support; admission to ITU) [14]. The poorer outcomes will continue to exist even though the treatments provided to these patients are equivalent to the guideline-directed and revascularization rates in patients with conventional risk factor profiles. In addition, differences in the size of infarcts and the presence of non-traditional risk factors could be responsible for the adverse prognosis in patients who do not have SMuRFs [14]. These results show the need for further research to increase our understanding of the non-modifiable and emergent risk factors, including the recent advances in genetics; atypical disease processes and non-measurable biological pathways, and the need for development of specific strategies to improve early detection, prevention, and the accuracy of treatment of myocardial infarctions in patients who do not have SMuRFs [14].

Reference

[1] Savonitto, S., Giampaoli, S., & De Luca, G. (2025b). Standard acute myocardial infarction without standard modifiable risk factors. *European Journal of Internal Medicine*, *141*, 106465. https://doi.org/10.1016/j.ejim.2025.106465

[2] Anderson, J. L., Knight, S., May, H. T., Le, V. T., Almajed, J., Bair, T. L., Knowlton, K. U., & Muhlestein, J. B. (2022b). Cardiovascular Outcomes of ST-Elevation Myocardial Infarction (STEMI) Patients without Standard Modifiable Risk Factors (SMuRF-Less): The Intermountain Healthcare Experience. *Journal of Clinical Medicine*, *12*(1), 75. https://doi.org/10.3390/jcm12010075

[3] Ball, J., Dinh, D. T., Brennan, A., Ajani, A., Clark, D. J., Freeman, M., Oqueli, E., Hiew, C., Nanayakkara, S., Walton, A., Shaw, J. A., Chan, W., Reid, C. M., & Stub, D. (2024). Prevalence and outcomes of patients with SMuRF-less acute coronary syndrome undergoing percutaneous coronary intervention. *Open Heart*, *11*(1), e002733. https://doi.org/10.1136/openhrt-2024-002733

[4] Wang, X., Zhang, F., Zhang, C., Zheng, L., & Yang, J. (2020). The biomarkers for acute myocardial infarction and heart failure. *BioMed Research International*, *2020*(1), 2018035. https://doi.org/10.1155/2020/2018035

[5] Daetwyler, H. D., Villanueva, B., & Woolliams, J. A. (2008). Accuracy of predicting the genetic risk of disease using a Genome-Wide approach. *PLoS ONE*, *3*(10), e3395. https://doi.org/10.1371/journal.pone.0003395

[6] Morrison, A. C., Bare, L. A., Chambless, L. E., Ellis, S. G., Malloy, M., Kane, J. P., Pankow, J. S., Devlin, J. J., Willerson, J. T., & Boerwinkle, E. (2007). Prediction of Coronary Heart Disease Risk using a Genetic Risk Score: The Atherosclerosis Risk in Communities Study. *American Journal of Epidemiology*, *166*(1), 28–35. https://doi.org/10.1093/aje/kwm060

[7] Dai, X. (2016). Genetics of coronary artery disease and myocardial infarction. *World Journal of Cardiology*, *8*(1), 1. https://doi.org/10.4330/wjc.v8.i1.1

[8] Bannister, C. A., Halcox, J. P., Currie, C. J., Preece, A., & Spasić, I. (2018). A genetic programming approach to development of clinical prediction models: A case study in symptomatic cardiovascular disease. *PLoS ONE*, *13*(9), e0202685. https://doi.org/10.1371/journal.pone.0202685

[9] Ho, D. S. W., Schierding, W., Wake, M., Saffery, R., & O'Sullivan, J. (2019). Machine Learning SNP based prediction for precision medicine. *Frontiers in Genetics*, *10*, 267. https://doi.org/10.3389/fgene.2019.00267

[10] Rincón, L. M., Sanmartín, M., Alonso, G. L., Rodríguez, J. A., Muriel, A., Casas, E., Navarro, M., Carbonell, A., Lázaro, C., Fernández, S., González, P., Rodríguez, M., Jiménez-Mena, M., Fernández-Golfín, C., Esteban, A., García-Bermejo, M. L., & Zamorano, J. L. (2019). A genetic risk score predicts recurrent events after myocardial infarction in young adults. *Revista Española De Cardiología (English Edition)*, *73*(8), 623–631. https://doi.org/10.1016/j.rec.2019.08.006

[11] Shiffman, D., O'Meara, E. S., Bare, L. A., Rowland, C. M., Louie, J. Z., Arellano, A. R., Lumley, T., Rice, K., Iakoubova, O., Luke, M. M., Young, B. A., Malloy, M. J., Kane, J. P., Ellis, S. G., Tracy, R. P., Devlin, J. J., & Psaty, B. M. (2007). Association of gene variants with incident myocardial infarction in the cardiovascular health study. *Arteriosclerosis Thrombosis and Vascular Biology*, *28*(1), 173–179. https://doi.org/10.1161/atvbaha.107.153981

[12] Gupta, R. (2023). Genetics-based risk scores for prediction of premature coronary artery disease. *Indian Heart Journal*, *75*(5), 327–334. https://doi.org/10.1016/j.ihj.2023.08.003

[13] Narula, N., Rapezzi, C., Tavazzi, L., & Arbustini, E. (2012). "My Parents Died of Myocardial Infarction: Is that My Destiny?" *Medical Clinics of North America*, *96*(1), 67–86. https://doi.org/10.1016/j.mcna.2011.11.001

[14] Kelly, C., Lan, N. S., Phan, J., Hng, C., Matthews, A., Rankin, J. M., Schultz, C. J., Hillis, G. S., Reid, C. M., Dwivedi, G., Figtree, G. A., & Ihdayhid, A. R. (2023b). Characteristics and outcomes of young patients with ST-Elevation myocardial infarction without standard modifiable risk factors. *The American Journal of Cardiology*, *202*, 81–89. https://doi.org/10.1016/j.amjcard.2023.06.045