Interaction Between ECG and Genetic Markers of Coronary Artery Disease

Julia Ramírez[1], Stefan van Duijvenboden[1], William J Young[1,2], Andrew Tinker[1], Pier D Lambiase[3], Patricia B Munroe[1], Michele Orini[2,3]
[1] William Harvey Research Institute, Queen Mary University of London, London, United Kingdom
[2] Barts Heart Centre, St Bartholomew's Hospital, London, United Kingdom
[3] Institute of Cardiovascular Science, University College London, London, United Kingdom

Abstract

Coronary artery disease (CAD) is the main contributor to cardiovascular mortality in developed countries, making accurate diagnosis of utmost importance. We developed risk scores to assess CAD risk in a population without known cardiovascular disease by combining ECG and a genetic risk score (GRS) for CAD. We analysed data in 52,260 individuals in the UK Biobank study. ECG indices included heart rate, PR, QRS, QT and T-peak-to-T-end intervals, while we built the GRS from publicly available genome-wide association results for CAD that were derived in an independent population. In a training set (N = 39,195), the indices with the strongest CAD prognostic impact were the PR and QT intervals, and the GRS. When combined together into a Multivariate model, both the ECG markers and the GRS were independently associated with CAD. In an independent test set (N = 13,065), we then built three risk scores based on (1) ECG markers, (2) genetic data, and (3) a combination of ECG and genetic data, respectively. The hazard ratio (95% confidence interval) for CAD comparing high versus low-risk individuals was 6.5 (5.1 - 8.3), 8.4 (6.4 - 10.8) and 8.4 (6.5 - 10.8) for the three risk scores, respectively. In conclusion, the inclusion of genetic markers into risk scores with ECG markers independently contributes to CAD risk prediction in a large population of individuals without known cardiovascular disease.

1. Introduction

Cardiovascular mortality is the main cause of death in the general population [1], and it accounts for 31% of all deaths worldwide, with its estimated cost expected to be $1044 billion by 2030. Coronary artery disease (CAD) is the main contributor to cardiovascular mortality in developed countries, but prediction remains a critically important challenge.

Risk scores based on clinical and ECG variables have been proposed for risk stratifying individuals at risk of a cardiovascular event [2], and can be utilised as a first step to identify at risk subgroups. Genetic risk scores (GRS) have shown to be useful for examining the cumulative predictive ability of genetic variation on CAD [3], and they have been reported to be independent to conventional demographic risk factors [3]. However, whether they can independently contribute to refine risk prediction in combination with traditional electrocardiogram (ECG) markers is unknown.

The aim of this study was to test the predictive value of a model combining ECG and a GRS for CAD.

2. Materials and Methods

2.1. Study population

This work has been performed using data from N = 52,260 individuals without known CV disease from the UK Biobank [4]. The UK Biobank study was approved by the North West Multi-Centre Research Ethics Committee and all participants provided written informed consent.

Genotyping was performed by UK Biobank using the Applied Biosystems UK BiLEVE Axiom Array or the UK Biobank AxiomTM Array, and genetic variants were imputed centrally by UK Biobank using the 1000G phase III and UK10K Haplotype Reference panels. Information on UK Biobank array design and protocols is available on the UK Biobank website (www.ukbiobank.ac.uk).

The primary endpoint of this work was death with CAD as the main cause, or a diagnosis of CAD [5]. Data from health records and death certificates were sent to UK Biobank on a quarterly basis up to May 2020. Follow-up information was available from 2009 until 2020.

2.2. Methods

2.2.1. Clinical, ECG and genetic markers

Sex and age were obtained on the day of the exercise protocol (Table 1).

We pre-processed and analysed the raw ECG signals from the 52,260 individuals. Pre-processing of the ECG

	Overall population ($N = 52,260$)	CAD free ($N = 49,912$)	CAD ($N = 2,348$)	*P* value
Sex [men]	24,195 (46.3%)	22,541 (45.2%)	1,654 (70.4%)	**<0.0001**
Age [years]	59 (13)	58 (12)	63 (8)	**<0.0001**
Resting HR [bpm]	70.3 (15.1)	70.3 (15.0)	70.9 (16.9)	0.212
PR interval [ms]	130 (31)	129 (32)	133 (33)	**<0.0001**
QRS duration [ms]	68 (18)	68 (18)	69 (18)	0.130
QTc [ms]	396 (31)	396 (30)	400 (33)	**<0.0001**
Tpe interval [ms]	62 (12)	62 (12)	63 (13)	0.228
GRS [SD]	-0.012 (1.342)	-0.021 (1.341)	0.208 (1.333)	**<0.0001**

Data are presented as absolute frequencies and percentages and as median (interquartile range). HR = Heart rate, QTc = Corrected QT interval, Tpe = T-peak-to-T-end, GRS = genetic risk score.

Table 1. Clinical, ECG and genetic variables in the study population. Significant differences between individuals with and without CAD are indicated in bold.

signals included low-pass filtering at 50 Hz to remove electric and muscle noise but still allow QRS detection [6]. Baseline wander was removed by further high-pass filtering the ECG signals at 0.5 Hz. We then signal-averaged the heartbeats to attenuate noise and artifacts. The onset, peak, and end timings of the waveforms were located using bespoke software [6,7]. We then automatically derived the following ECG markers: heart rate, PR interval, QRS duration, corrected QT interval (QTc), and resting T-peak-to-T-end interval (Tpe).

We then derived a GRS using 192 previously reported genome-wide significant variants for coronary artery disease (identified using an independent population) [8]. The GRS was calculated by summing the 192 risk alleles and weighting them by their respective effect sizes from the original genome-wide association study. The GRS was then standardised to have a mean of 0 and a standard deviation of 1 [9].

2.3. Statistical Analyses

We evaluated differences in the risk markers across cases (individuals with CAD) and controls (the rest) using Mann-Whitney test. Then, as a first step in the development of the risk models, we split the study population into training (N = 39,195, 75%) and test (N = 13,065, 25%) sets, keeping a balanced case-control ratio and similar distribution of variables in both sets.

2.3.1. Selection of risk markers

To find the optimal risk markers for the risk scores, in the training set we fit three different models for CAD: an ECG-based model (including ECG markers only), a GRS, and a combined model (combining the ECG markers and the GRS). All models included sex and age as well. For each model, we performed Univariate analyses using Cox regression to determine the relationship between each potential risk marker and the primary endpoint [2]. Then, we subsequently performed Multivariate Cox regression analyses, with risk markers significantly ($P < 0.05$) associated with outcomes in the Univariate analysis being selected and placed into backward stepwise elimination models, which eliminates markers with non-significant association with the endpoint.

2.3.2. Assessment of the risk scores

To calculate and assess each of the three risk scores (one score per model) for each individual in the test set, we used the beta coefficients estimated in the training set. The scores were calculated as the weighted sum of each of the significant risk markers in the Multivariate models from the training set, where the weights were the corresponding beta coefficients [2]. We then calculated the area under the receiver operating curve (AUC) by considering the sensitivity and specificity of the three risk scores (ECG, GRS and combined) in identifying individuals with CAD during the follow-up. Finally, for each risk score, we identified three CAD risk groups based on their distribution: low-risk (first and second quintiles), middle-risk (third and fourth quintiles) and high-risk (fifth quintile). Association between risk groups and the primary endpoint was evaluated using hazard ratios (HRs), calculated using the low-risk group as a reference.

3. Results

3.1. Characteristics of the study population

The detailed characteristics of the study population are shown in Table 1. During follow-up, there were 2,348 CAD events (4.5%). Individuals with CAD were more frequently men, were older, had a longer PR interval and QTc interval, and had a higher count of risk alleles for CAD (Table 1). Resting heart rate, QRS duration and the Tpe interval were not significantly different in the CAD risk group compared to the group free of CAD.

	ECG		GRS		Combined	
	Hazard ratio (95% CI)	*P* value	Hazard ratio (95% CI)	*P* value	Hazard ratio (95% CI)	*P* value
Sex [men]	**2.81** (2.53-3.12)	$<$ **0.0001**	**2.74** (2.47-3.04)	$<$ **0.0001**	**2.83** (2.55-3.14)	$<$ **0.0001**
Age [SD]	**1.74** (1.64-1.83)	$<$ **0.0001**	**1.79** (1.69-1.89)	$<$ **0.0001**	**1.74** (1.65-1.84)	$<$ **0.0001**
Resting HR [SD]	N.S.	N.S.	-	-	N.S.	N.S.
PR [SD]	**1.07** (1.02-1.12)	**0.005**	-	-	**1.06** (1.02-1.11)	**0.007**
QRS duration [SD]	N.S.	N.S.	-	-	N.S.	N.S.
QTc [SD]	**1.15** (1.11-1.19)	$<$ **0.001**	-	-	**1.15** (1.11-1.19)	$<$ **0.0001**
Tpe interval [SD]	N.S.	N.S.	-	-	N.S.	N.S.
GRS [SD]	-	-	**1.23** (1.18-1.29)	$<$ **0.001**	**1.23** (1.18-1.29)	$<$ **0.0001**

SD = standard deviation, HR = heart rate, QTc = corrected QT interval, Tpe = T-peak-to-T-end, GRS = genetic risk score.

Table 2. Multivariate Cox Regression analyses results for the three risk models, ECG, GRS and combined in the training set. The logarithm of the hazard ratios (beta coefficients) were used to weight each risk marker in the calculation of each score in the test set. Statistically significant values are marked in bold.

.

3.2. Selection of risk markers

All the risk markers that were significant after the Mann-Whitney test remained significant in the Univariate model. Resting heart rate, QRS duration and the Tpe interval were not significantly associated with CAD, and therefore were not taken forward into the Multivariate analyses. Table 2 shows the HRs and P-values for the variables that had a significant association with CAD after including them together in each corresponding Multivariate model. As Table 2 shows, both the PR and the QTc intervals remained significantly associated with CAD in the ECG model. Regarding the combined model, the PR and the QT interval remained significantly associated with CAD after the inclusion of the GRS.

3.3. CAD risk scores

Figure 1 shows the AUC of the three scores, ECG (black), GRS (Gen, red) and combined (ECGGen, green) in the test set when classifying CAD risk. As shown in Figure 1, the GRS provided a more accurate classification of CAD than the ECG score, and this was slightly increased after combining both together. However, the improvements were not statistically significant.

Figure 2 shows the HRs of the middle (blue) and high (red) risk groups for each score, relative to the low risk group. As shown in the Figure, the HR of the high risk group for the GRS was higher than the HR for the ECG score, remaining stable when combining ECG markers and the GRS (Figure 2). However, this improvement was not statistically significant. Finally, the HRs for the middle-risk groups were similar across the three scores.

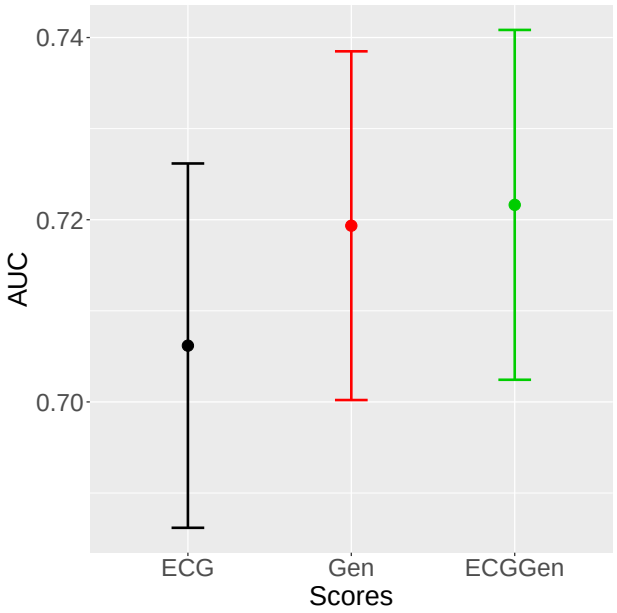

Figure 1. Area under the receiver operating curve (AUC) for the three scores, ECG, GRS (Gen) and combined (ECGGen)

4. Discussion and Conclusions

The main result of this study is that, in a middle-aged population without known cardiovascular disease, the combination of ECG and genetic markers shows a trend towards improving CAD risk prediction, as compared to the individual use of ECG or genetic variables. This indicates a potential new strategy to identify individuals in the general population specifically at risk of suffering CAD.

We found that the ECG variables included in the ECG and in the combined score were the PR interval and the QT interval, supporting previous studies that reported these ECG indices as CAD risk factors [10, 11]. Our results confirm that CAD risk manifests as abnormalities in ventricular conduction (PR prolongation), and in ventricular repolarization (QT prolongation). However, the individuals included in the study could potentially have subclinical disease. Therefore, the prolongations in PR and QTc intervals could be reflecting an underlying CAD substrate.

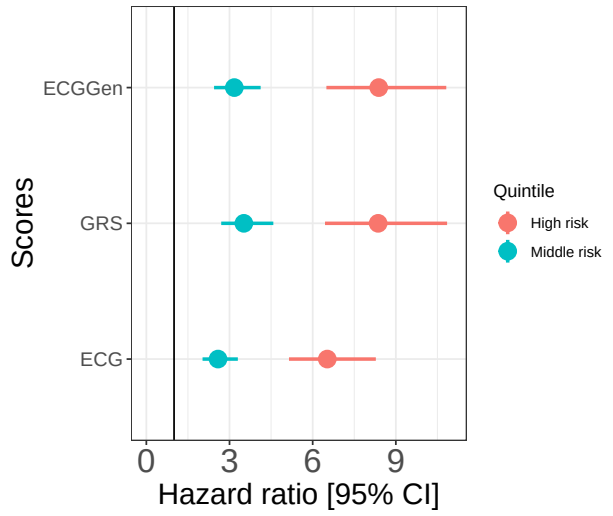

Figure 2. Hazard Ratios of the middle and high risk groups with respect to the low risk group defined for each of the three risk scores, ECG, GRS (Gen) and combined (ECGGen)

We also validated the association of the GRS with CAD in previous studies [3, 8], confirming CAD has an important genetic predisposition.

The integration of ECG and genetic markers into a combined score could also be potentially used to study the evolution and the mechanisms of CAD risk (genotype-phenotype interactions) [12]. In addition, recent publications on the genetics of ECG markers of cardiovascular risk have reported GRSs for these ECG risk markers [5, 9, 13, 14]. However, the role that these GRSs play when integrated into combined risk scores with independent clinical and non-invasive cardiac markers remains to be elucidated.

Our results confirm the hypothesis that a combination of non-invasive markers capturing complementary information about the cardiovascular condition can potentially improve CAD risk prediction. Future studies with more power and including additional complementary risk factors are needed.

Acknowledgements

This research has been conducted using the UK Biobank Resource (application 8256) and is supported by the MRC grant MR/N025083/1. JR acknowledges support from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007/2013) under REA grant agreement 608765. PDL is supported by UCL/UCLH Biomedicine NIHR. WJY is supported by an MRC grant MR/R017468/1. We also wish to acknowledge the NIHR Cardiovascular Biomedical Research Unit at Barts and QMUL, UK.

References

[1] N. Townsend et. al., "Cardiovascular disease in Europe: epidemiological update 2016," *European Heart Journal*, vol. 37, p. 3232–3245, Nov 2016.

[2] J. Ramírez et. al., "Sudden cardiac death and pump failure death prediction in chronic heart failure by combining ECG and clinical markers in an integrated risk model," *PLoS ONE*, vol. 12, no. 10, p. e0186152, 2017.

[3] M. Inouye et. al., "Genomic risk prediction of coronary artery disease in 480,000 adults - implications for primary prevention.," *JACC*, vol. 72, no. 16, pp. 1883–1893, 2018.

[4] C. Bycroft et. al., "Genome-wide genetic data on 500,000 UK Biobank participants," *bioRxiv*.

[5] J. Ramírez et. al., "Cardiovascular predictive value and genetic basis of ventricular repolarization dynamics," *Circ Arr Electrophysiol*, vol. 12, p. e007549, 2019.

[6] M. Orini et al., "Long-term intra-individual reproducibility of heart rate dynamics during exercise and recovery in the UK Biobank cohort," *Plos One*, vol. 12, p. e0183732, 2017.

[7] M. Orini et al., "A time-varying nonparametric methodology for assessing changes in QT variability unrelated to heart rate variability," *IEEE Trans Biomed Eng*, vol. 65, pp. 1443–1451, 2018.

[8] M. Paquette et. al., "Polygenic risks core predicts prevalence of cardiovascular disease in patients with familial hypercholeterolemia.," *J Clin Lipidol*, vol. 11, no. 3, pp. P725–732, 2017.

[9] J. Ramírez et. al., "Thirty loci identified for heart rate response to exercise and recovery implicate autonomic nervous system," *Nat Comms*, p. In Press, 2018.

[10] Y. Chan et. al, "PR interval prolongation in coronary patients or risk equivalent: excess risk of ischemic stroke and vascular pathophysiological insights," *BMC Cardiovasc Disord*, vol. 17, p. 233, 2017.

[11] Y. Zhang et. al., "Electrocardiographic QT interval and mortality: a meta-analysis," *Epidemiology*, vol. 22, no. 5, pp. 660–670, 2011.

[12] R. Elosua, "Road to unravel gene-environment interactions on cardiovascular complex diseases," *Circ Gen Prec Med*, vol. 11, no. 1, 2018.

[13] J. Ramírez et. al., "Common genetic variants modulate the electocardiographic Tpeak-to-T-end interval," *Am J Human Gen*, vol. 106, pp. 764–778, 2020.

[14] S. van Duijvenboden et. al., "Genetic basis and prognostic value of exercise QT dynamics," *Circ Gen Prec Med*, vol. 13, no. 4, 2020.

Address for correspondence:

Julia Ramírez, Charterhouse Square,
London EC1M 6BQ, United Kingdom
j.ramirez@qmul.ac.uk