FIC Federazione Italiana di Cardiologia
Italian Federation of Cardiology

Beyond the cardiovascular risk charts: the new way of hybrid profiles

Francesco Romeo[a], Giuseppe Novelli[a], Maurizio Ferrari[c] and Maurizio Talamo[b]

J Cardiovasc Med 2016, **17**:000–000

[a]University of Rome Tor Vergata, [b]INUIT Tor Vergata Foundation, Rome and [c]Vita Salute University of Milan, Italy

Correspondence to Francesco Romeo, MD, Policlinico Tor Vergata - Viale Oxford, 81 - 00133 Roma, Italy
E-mail: romeocerabino@tin.it

Received 23 June 2015 **Revised** 7 April 2016
Accepted 22 April 2016

Cardiovascular disease (CVD) is a major contributor to morbidity and mortality in the United States and Europe. In Italy, CVD is the leading cause of death in adults and accounts for 44% of total deaths.[1–3] Identifying patients at high risk of CVD is a key priority of primary prevention and the first step toward reducing CVD risk by targeting modifiable risk factors.[4] Well assessed studies identify principle factors that determine the risk of CVD. Univariable statistical analysis confirms the value of any single risk factor (e.g. odd ratios in the cardiovascular risk charts of Italian CUORE project).[5] Accordingly, the guidelines for CVD prevention have traditionally focused on single factor assessment rather than on the overall risk based on a combination of risk factors (http://www.World-heart-federation.Org/cardiovascularhealth/cardiovascular-disease-risk-factors/).

A risk factor may be defined as a characteristic of a person that is associated with an increased risk of developing a specific disease such as atherosclerotic CVD. To be clinically relevant, it should be accepted as causal,[2] modifiable, and a defined benefit should result from such modification. Most risk-estimation systems in CVD include age, sex, smoking, blood lipids, and blood pressure (BP) as their core variables (Table 1).[6] In this context, age is a measure of exposure time and not a risk factor as such. However, large prospective studies[4,7] have clearly shown that CVD is a multifactorial disease and that risk factors do not act in isolation but rather in conjunction with each other. It is, therefore, important to consider all risk factors, including the genetic ones, when assessing a patient global CVD risk. This way of identifying high-risk individuals recognizes not only CVD multifaceted nature and the multiplicative effect of risk factors but also the importance of assessing and managing patients beyond mere treatment of risk factors.[4]

Over the last decades, numerous epidemiological studies[2,4,8] have focused their attention on global absolute risk for identifying high-risk individuals. Several risk-prediction methods have been developed to identify individuals at high risk of CVD (or of small variants of this disease). Among those risk-prediction methods, the Framingham Heart Study[9] and the PROCAM study[10] are two well-known examples of predictive equations derived from large prospective studies. Additional risk charts and scores have also been developed from other widely known studies such as the systematic coronary risk evaluation project,[11] the United Kingdom prospective diabetes study,[12] and the Italian CUORE project.[13] All these risk charts and scores are useful tools for evaluating an individual global coronary heart disease (CHD) risk in clinical practice. The CUORE study is a large prospective cohort follow-up study, including cohorts from the northwest, northeast, centre, and south of Italy. The aim of the CUORE project[13] was to develop a 10 years coronary risk-predictive equation specific for the Italian population. Among the sample of 6865 men, aged from 35 to 69 years and CHD-free at baseline, 312 first fatal and nonfatal major coronary events occurred in the 9.1 years median follow-up. Women were excluded from the analysis because of the small number of events and the shorter follow-up period. The follow-up period is currently being extended and more reliable estimates are available also for women. The CUORE predictive equation includes well-known CHD risk factors such as age, total cholesterol, SBP, cigarette smoking, high density lipoprotein (HDL) cholesterol, type 2 diabetes, hypertension medications, and family history of CHD. In the CUORE equation, BMI was not an independent risk factor and did not improve CHD prediction. This could be explained by the fact that other obesity-related risk factors, such as type 2 diabetes, elevated total cholesterol, high BP, and low HDL cholesterol, were already considered in the model. Triglyceride levels were also not included in the model, because their inclusion eliminated the protective effect of HDL cholesterol. Once HDL cholesterol was included, adding triglycerides to the model did not improve CHD prediction. A comparison between these models shows that there are divergences in the exact problem they front, in the definition of the explicative variables of their models, in the use of these models in different regions of the world.[4] Also results are different in subgroups of populations.[4] Many countries have developed their own model to be used; all these

DOI:10.2459/JCM.0000000000000405

Table 1 Risk factors for coronary heart disease in the Amsterdam Growth and Health Study

	Biologic risk factors			
	TC (mmol/l)	HDL (mmol/l)	SBP (mmHg)	DBP (mmHg)
13–16 years	≥5.2	≤1.1	≥126	≥82
21–27 years	≥6.2	≤0.9	≥140	≥90

	Lifestyle risk factors			
	Fat intake (% of energy)	Carbohydrated intake (% of energy)	Cholesterol intake (mg/mJ)	Daily physical activity
13–16 years	≥35	≤50	≥33	≤P25
21–27 years	≥35	≤50	≥33	≤P25

Reproduced from [6]. HDL, high density lipoprotein cholesterol; P25, 25th percentile; TC, total cholesterol.

models have a common core but every model preserves its peculiar specifics. The existence and the persistence over the time of so many risk-estimation systems highlight the need to better recognize hidden sets of contributing factors that are not included in current models, taking into account also the interaction with geographical specific factors.

Risk-prediction systems

The main feature of a good risk-prediction system is to associate one or more precise decisions to each of its outputs. Given a patient (and a set of measurements), there are many potential outputs (or risk levels) and, for each of them, there is a different associated action. Necessary conditions to obtain this are: the sharpness of results and the calibration of them, that is, the appropriateness of results calculated for the input patient. What is sharpness? For better understanding it, we need to face the model working without information, and we call it the crazy doctor model. How does work our crazy doctor model? Let's consider a risk system that decides for a patient simply tossing a regular coin. The output face of the coin will decide if the patient will have a CVD in next 10 years or not. This crazy risk-estimation system intercepts patient future for 50%. Now let's imagine having a sample of 1000 patients of our crazy doctor. Overall, 500 of them will not have any CVD in the next following 10 years; the other 500 patients will be affected by some CVD in the next 10 years. Our crazy doctor will predict the risk for each patient tossing a coin and the expected result will be that his diagnoses will be correct for 500 patients. For the remaining 500 patients the crazy doctor with his coin will get it wrong: for healthy patients he will predict that they will be ill and for patients that will have CVD he will predict that they will be healthy. Even changing proportions between healthy and ill patients the result will be the same: one half of the sample will have a correct risk prediction and the other half of the sample will have an incorrect risk prediction. The evidence coming from this scenario is that a good risk-prediction model can give relevant results only if it is able to give correct prediction better than the random model of our crazy doctor. This evidence is relevant if we consider the impact of giving medicaments, sometimes even with important collateral effects, to an unhealthy person, or, vice versa, not foreseeing a therapy to a risky patient. There are also other important aspects that must be considered: we involve the 500 risky patients of our crazy doctor in a clinical trial. The surprising result is that, at the end of the trial, the healthy patients of the sample will think that they are healthy because of the pharmacological treatment effect during the trial. This evaluation can drive to incorrect conclusions on the effects of some pharmacological trials. This extreme example

Table 2 Sample data grouped by age class and sex

Age	49–55	56–60	61–65	66–70	71–80	Total
Hypertension/familiar hypertension						
Male	11/10	15/14	14/13	13/11	3/3	56/51
Female	5/5	11/8	5/4	5/5	1/1	27/23
Col/HDL (average values)						
Male	212/42	164/36	200/41	198/46	148/43	188/41
Female	208/55	199/50	209/53	214/51	190/52	205/51
Smoke						
Male	10	10	8	6	1	35
Female	1	4	1	0	0	6
Diabetes						
Male	6	11	6	8	0	31
Female	2	6	3	4	0	15
Case/control						
Male	13/6	14/10	16/1	12/9	4/0	59/26
Female	2/3	0/15	3/2	1/5	1/0	7/25

HDL, high density lipoprotein cholesterol.

Table 3 Patient risks obtained using CUORE

Cuore	Min	Max	Avg	Std
Case	1.9	44.4	19.06	9.72
Control	1.2	37.1	10.79	9.61
Total	1.2	44.4	13.2	9.90

Min, minimum; Max, maximum; Avg, average; Std, standard deviation.

allows us to understand, on one hand, why models on predictive risk are so difficult to design and to develop and require a rigorous scientific method. On the other hand, we understand that a model strongly diverging from a random classification allows us to assess the risk of a single patient and, at the same time, allows us to select the right samples of patients for clinical trials. Furthermore, our crazy doctor with his incorrect diagnosis has a very bad economic impact on health costs, because he gives medical therapies to healthy patients and vice versa in ill patients, increasing related costs.

Keeping clinical and genetic data together

What is evident, with this risk concept in mind, is that clinical predictive multivariate models leave gaps in the ability to stratify cardiovascular risk. Many studies individuate in genetic profile the main source to improve this imprecision in estimation.[14,15] The aim of this study is to test a new hybrid mathematical model for reclassification of cardiovascular risk using both clinical and genetic data. The hybrid mathematical model has been implemented using a new software called Race. To test the hybrid mathematical model we analyzed 117 Italian patients, divided into controls and CVD trial (cases). They were categorized according to nine features of demographic, social, and clinical nature. For each of them, we had at our disposal also the genetic data of 11 polymorphisms, selected among the set of higher-risk polymorphism for CVD. Our analysis was conducted in four main steps.

In the first step, starting from a sample of 117 patients with their social and clinical profiles, we used the CUORE software (Istituto Superiore di Sanità, Italy), available on the Internet (http://www.cuore.iss.it/sopra/calc-rischio.asp), to obtain a risk level for each of them and then we used this obtained risk level to classify them into case and control (Tables 2 and 3). Once patients were assigned to the case or control categories, statistical analysis was used to evaluate the goodness of the classification. With this approach 66 patients were correctly classified as case or control and 51 were classified incorrectly.

Table 4 Patient risks obtained using only their genetic data

Genetic profile	Min	Max	Avg	Std
Case	11.8	13.53	12.75	0.37
Control	11.59	13.65	12.73	0.43
Total	11.59	13.65	12.74	0.40

Min, minimum; Max, maximum; Avg, average; Std, standard deviation.

Table 5 Patient risks obtained using integration of clinical and genetic data

RACE	Min	Max	Avg	Std
Case	0	9	4.5	2.08
Control	0	8	2.92	2.13
Total	0	9	3.81	2.24

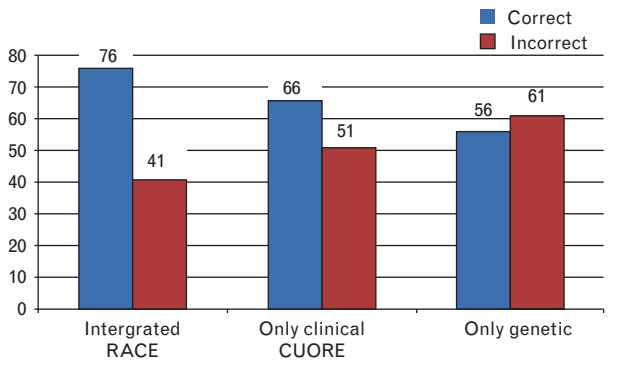
Min, minimum; Max, maximum; Avg, average; Std, standard deviation.

In the second step, trying to improve the results for the classification tool of step 1, based only on the clinical data, we used a statistical technique, based on entropy analysis,[14] to evaluate all theoretical classifications of sample patients based on the available data considering only the social and clinical parameters. The goal of this step was to explore the existence of a classification model based only on social and clinical data. This did not give conclusive results or significant improvements in comparison to that used at step 1.

In the third step, we classified patients using only their genetic profiles obtaining a modest result, as shown in Table 4. With this approach 56 patients were correctly classified as case or control and 61 were incorrectly classified. So we tried to find evidence for the need of considering genetic data together with the clinical data and to discover the nature of their mathematical correlation. We found that the mean value of the risk genetic profile in the subpopulation of the CVD patients is significantly higher than the mean value of CVD in the subpopulation of high-risk genetic profile patients. In causal analysis, this asymmetry signals the presence of an incomplete causal graph in which the set of genetic variables is only a subset of all the causal variables (presence of hidden variables). Furthermore, like for the relationships behind the clinical risk evaluation procedures, the asymmetry is an aggregate result that has no direct use in single patient profiling. Previous results showed that integration is highly required, but, at the same time, it is not an easy task, for many reasons. First, the nature of integration is not known[16,17] and it is hard to discover. From a statistical point of view this means difficulties in model specification, such as limitations in the regression model. Also in the Bayesian approach the question is open; in this case, the set of causal variables in the specification of a causal graph is not clear. Second, there are multicollinearity problems related to the nature of genetic polymorphism in the interaction with clinical factors. Third, there may be computational problems related to the elevated number of causal variables.

In the fourth step, given the weaknesses of the models previously used, we introduce and apply our model that combines genetic profile and clinical categories to obtain a new risk classification. Our model is a new approach for combining genetic and clinical data in a single risk profile using an extensional approach substituting the analytics one, introducing a new class of generators that have the

Fig. 1



Comparison between clinical, genetic, and integrated risk assessment.

same behavior of the complex analytics function calculating the risk. It outputs a risk level with a temporal risk profile, given a patient with a set of clinical and genetic data (Table 5). With our approach, 76 patients were correctly classified as case or control and 41 were incorrectly classified, showing better accuracy, compared with CUORE classification (66 correct vs. 51 incorrect classification) (Fig. 1).

In conclusion, our model is able to correctly classify patients with an increased performance compared with currently available risk-prediction models.

Acknowledgements

The authors gratefully acknowledge Fondazione Roma for its support to this research [Project: Development of a integrated protocol based on environmental and genetic/epigenetic data for the risk prediction of acute myocardial infarction (AMI) in patients with coronary atherosclerosis. Sn.NCDS-2013–00000333].

There are no conflicts of interest.

References

1 WHO. World health statistics 2009. Geneva: World Health Organization. 2009e.
2 Grundy SM, Pasternak R, Greenland P, *et al.* AHA/ACC scientific statement: assessment of cardiovascular risk by use of multiple-risk-factor assessment equations: a statement for healthcare professionals from the American Heart Association and the American College of Cardiology. *J Am Coll Cardiol* 1999; **34**:1348–1359.
3 Palmieri L, Rielli R, Dematte L, *et al.* Cardiovascular risk assessment: preliminary results. *G Ital Cardiol (Rome)* 2010; **11**:154–161.
4 Cooney MT, Dudina AL, Graham IM. Value and limitations of existing scores for the assessment of cardiovascular risk: a review for clinicians. *J Am Coll Cardiol* 2009; **54**:1209–1227.
5 Donfrancesco C, Palmieri L, Vanuzzo D, *et al.* Italian cardiovascular mortality charts of the CUORE project: are they comparable with the SCORE charts? *G Ital Cardiol (Rome)* 2010; **11**:148–153.
6 Twisk JW, Kemper HC, van Mechelen W, Post GB. Tracking of risk factors for coronary heart disease over a 14-year period: a comparison between lifestyle and biologic risk factors with data from the Amsterdam growth and health study. *Am J Epidemiol* 1997; **145**:888–898.
7 Donfrancesco C, Palmieri L, Cooney MT, *et al.* Italian cardiovascular mortality charts of the CUORE project: are they comparable with the SCORE charts? *Eur J Cardiovasc Prev Rehabil* 2010; **17**:403–409.
8 WHO. Prevention of cardiovascular disease: guidelines for assessment and management of cardiovascular risk. Geneva, Switzerland: World Health Organization, 2007.
9 D'Agostino RB Sr, Vasan RS, Pencina MJ, *et al.* General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation* 2008; **117**:743–753.
10 Assmann G, Cullen P, Schulte H. Simple scoring scheme for calculating the risk of acute coronary events based on the 10-year follow-up of the prospective cardiovascular Munster (PROCAM) study. *Circulation* 2002; **105**:310–315.
11 Conroy RM, Pyorala K, Fitzgerald AP, *et al.* Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *Eur Heart J* 2003; **24**:987–1003.
12 King P, Peacock I, Donnelly R. The UK prospective diabetes study (UKPDS): clinical and therapeutic implications for type 2 diabetes. *Br J Clin Pharmacol* 1999; **48**:643–648.
13 Palmieri L, Donfrancesco C, Lo Noce C, *et al.* Il progetto cuore: 15 anni di attivita' per la prevenzione e la riduzione del rischio cardiovascolare. *Not Ist Super Sanita* 2013; **26**:3–8.
14 Chen L, Li C, Miller S, Schenkel F. Multipopulation genomic prediction using a multitask Bayesian learning model. *BMC Genet* 2014; **15**:53.
15 Radha V, Mohan V. Predictive value of genomics in the screening of type 2 diabetes: limitations and current status. *Adv Genomics Genet* 2014; **4**:45–57.
16 Schnabel RB, Baccarelli A, Lin H, *et al.* Next steps in cardiovascular disease genomic research: sequencing, epigenetics, and transcriptomics. *Clin Chem* 2012; **58**:113–126.
17 Seo D, Ginsburg GS, Goldschmidt-Clermont PJ. Gene expression analysis of cardiovascular diseases: novel insights into biology and clinical applications. *J Am Coll Cardiol* 2006; **48**:227–235.