Development of a Cardiovascular Risk Score for Use in Low- and Middle-Income Countries

Stella Aslibekyan, Hannia Campos, Eric B. Loucks, Crystal D. Linkletter, Jose M. Ordovas, and Ana Baylin

Abstract

Summary measures of cardiovascular risk have long been used in public health, but few include nutritional predictors despite extensive evidence linking diet and heart disease. Study objectives were to develop and validate a novel risk score in a case-control study of myocardial infarction (MI) conducted in Costa Rica during 1994–2004. After restricting the data set to healthy participants (n = 1678), conditional logistic regression analyses modeled associations of lifestyle factors (unhealthy diet, decreased physical activity, smoking, waist:hip ratio, low or high alcohol intake, and low socioeconomic status) with risk for MI. Using the estimated coefficients as weights for each component, a regression model was fit to assess score performance. The score was subsequently validated in participants with a history of chronic disease. Higher risk score values were associated with a significantly increased risk of MI [OR = 2.71 (95% CI = 2.26–3.26)]. The findings were replicated in a model (n = 1392) that included the best covariate measures available in the study [OR = 2.71 (95% CI = 2.26–3.26)]. Performance of the score in different subsets of the study population showed c-statistics ranging from 0.63 to 0.67. The new score presents a quantitative summary of modifiable cardiovascular risk factors in the study population. J. Nutr. 141: 1375–1380, 2011.

Introduction

Summary measures of cardiovascular risk such as the Framingham score have long been used in public health research and practice (1–4). These risk scores play an important role in screening programs, identifying susceptible individuals before the onset of clinical symptoms and facilitating primary prevention (5). Cardiovascular risk scores are also used in epidemiologic research, namely as measures of exposure, stratification variables, or measures of potential confounders (6,7).

The predictive ability of existing cardiovascular risk scores varies greatly between populations and is particularly decreased in ethnic minority communities (8). Such variability merits further investigation, because most cardiovascular risk measures in use were derived in high-income Western cohorts (9,10) and evidence of their generalizability to the context of low- or middle-income populations worldwide, or even to some ethnically and economically diverse communities of high-income countries, is limited (11).

Additionally, no commonly used cardiovascular risk score incorporates nutritional predictors despite overwhelming evidence, including studies from developing countries, linking diet to the risk of heart disease (12). Dietary risk factors are modifiable and less costly to measure than many of the intermediate risk factors (e.g. blood lipid levels), which makes them an attractive option for use in resource-poor settings. As more populations begin the nutritional and epidemiologic transitions associated with modernization, new methods of risk stratification may be necessary to address the increasing global burden of cardiovascular disease.

The purpose of our study is to derive and validate a novel cardiovascular risk score comprised of predictors such as diet, physical activity, smoking, alcohol consumption, waist:hip ratio, and socioeconomic status in a population-based, case-control study of myocardial infarction (MI) in Costa Rican adults. Although an argument can be made that socioeconomic status is not truly a modifiable risk factor, it is included into the score as an important upstream determinant of cardiovascular risk that could be affected by economic development and policy changes.

Methods

Study population. The population of the Costa Rica Study included 4547 Hispanics who resided in the Central Valley of Costa Rica between 1994 and 2004 (13–16). Cases of first nonfatal acute MI were ascertained by 2 independent cardiologists in the participating hospitals.
and deemed eligible if they met the WHO criteria (17), survived hospitalization, were under 75 years of age on the day of their first MI, and able to answer the questionnaire. Eligible cases (n = 2273) were matched by 5-year age group, sex, and area of residence to population controls (n = 2274), identified randomly using data from the National Census and Statistics Bureau of Costa Rica. Women comprised 27% of all study participants (1209 total, 605 controls, and 604 cases). After the cases were discharged from the hospital, all cases and controls received home visits, during which trained study workers collected lifestyle and medical history data, anthropometric measurements, and biological specimens. Information on diet, physical activity, smoking, alcohol intake, socioeconomic status, and medical history was collected using questionnaires (18–20). Dietary exposures were ascertained both via FFQ and biological markers, specifically adipose tissue concentrations of selected fatty acids (21,22). To avoid reverse causation and recall bias, data on exposures among cases were recorded as close to the time of MI as possible. Participation was 98% for cases and 88% for controls. All participants provided written informed consent. The study has been approved by the Human Subjects Committee of the Harvard School of Public Health and the University of Costa Rica.

Descriptive statistics. To assess lifestyle habits in the Costa Rica Study population, the study used WHO guidelines on nutrients and physical activity (23,24). Participants were considered compliant with a specific guideline if their self-reported values fell within the recommended range. Frequency tables and a histogram were constructed to ascertain the proportion of the study sample that reported adherence to WHO guidelines on nutrient intake and healthy lifestyle, as well as fell above the national poverty line threshold. Participants were excluded from this analysis if they were missing information on any of the covariates, yielding a sample size of 4091.

The distribution of risk factors by case-control status was compared using McNemar’s test. Cases were rematched to controls on age, sex, and area of residence to preserve the study design, resulting in a sample size of 3968.

Measurement of the cardiovascular risk score components. Two versions of the cardiovascular risk score were developed, 1 based on WHO nutrient intake and physical activity recommendations and poverty line standards outlined above (score 1) and 1 incorporating biomarker measures and more refined socioeconomic status measures available in the Costa Rica Study (score 2).

Cardiovascular risk score components were selected based on a prior analysis of modifiable MI risk factors in our study population as well as international guidelines for healthy lifestyle (14,23). The selected risk score components showed expected associations with the risk of MI in our study population (Supplemental Table 1). Score 1 was derived to ensure a simple, low-cost risk estimation algorithm that could be adapted to a variety of populations, whereas score 2 was designed to include the most reliable measures of risk factors available for the Costa Rica Study population (Supplemental Table 2). The healthy diet score used in score 2 was derived as a composite measure of total dietary intake of saturated fats, cholesterol, polyunsaturated fats, fiber, folate, and adipose tissue 𝛼-linolenic acid (ALA) and total trans fats. Adipose tissue ALA was chosen for inclusion in score 2 due to its importance as a cardioprotective factor in the study population, characterized by low ALA intake (25). Although adipose tissue was used in this population as the most reliable measure of long-term intake, ALA intake as estimated by FFQ could also constitute a valid measure. Physical activity information was collected using a questionnaire described in more detail in previous publications (18). Briefly, participants reported the average frequency and time spent on several occupational and leisure time activities during the last year. Energy expenditure for each activity was calculated as the product of frequency, time, and intensity measured in METS (metabolic equivalents, defined as the energy expenditure for sitting quietly or approximately 1 kg·kg body$^{-1}$·h$^{-1}$) (18). The physical activity questionnaire was validated by its ability to predict fitness level measured by the Harvard Step test, plasma lipids, and obesity, in our previous studies in Puriscal, Costa Rica (19,20). Smoking and alcohol intake were measured using questionnaires. Anthropometric measurements, including waisthip ratio, were collected the morning after an overnight fast by trained fieldworkers while subjects wore light clothing and no shoes. Measurements were performed in duplicate, with the average used in the analysis. Finally, income was measured by showing participants index cards with ranges of income (in US dollars/mo) and asking them to select the appropriate index card.

Derivation of the cardiovascular risk score. The derivation data sets were developed using the complete case method, i.e. participants with missing values for any of the covariates were excluded from the analysis. Additionally, participants with a self-reported history of diabetes, hypertension, or current use of medication for chronic conditions were excluded from the derivation data set to avoid reverse causation (14). After restriction, remaining cases were rematched to controls on age, sex, and area of residence to preserve the study design. To derive score 1, 456 of the original 4547 participants were excluded due to missing covariate values, 2167 were excluded due to history of chronic disease, and 246 were lost to the rematching process, yielding the final sample size of 1678 (839 cases and 839 controls). For the analyses involving score 2, 1114 of the original 4547 participants were excluded due to missing covariate values, 1840 were excluded due to history of chronic disease, and 201 were lost to the rematching process, yielding the final sample size of 1392 (696 cases and 696 controls) to derive score 2.

Participants were categorized according to each risk score component as follows. All dietary variables included in score 1 (trans fats, polyunsaturated fats, saturated fats, cholesterol, fiber, and folate) were included as categorical variables according to the quintile of intake. For each component of the healthy diet measure used in score 2 (polyunsaturated fats, saturated fats, cholesterol fiber, folate, and adipose tissue trans fats and LA), the participants were assigned scores from 0 to 4 corresponding to the quintile of intake, with 4 representing the lowest risk quintile (Supplemental Table 1). Quintiles of the healthy diet score were based on all participants. The assigned quintile values were then summed to produce the dietary score for each participant (26). Therefore, the resulting dietary score ranged from 0, which indicated the lowest possible adherence to dietary guidelines, to 28, which represented the highest possible adherence to dietary guidelines. For score 1, being physically active was defined as expending >10% of daily energy in the performance of moderate- and high-intensity physical activities (at least 4 times the basal metabolism rate) (27). For score 2, physical activity was included as a continuous variable, defined as total METS expended over a 24-h period. For both score 1 and score 2, smoking was defined as a dichotomous variable (currently smoking vs. not) and alcohol intake was measured in grams per day and defined as a categorical variable with the following cutoffs: 0 (not drinkers), 0.1–5.0, 5.1–10, over 10. For both scores, participants were classified as healthy if their waisthip ratio value lay below the cutoff of ≤0.90 for men and ≤0.85 for women as per WHO guidelines (28). For score 1, socioeconomic status was classified as low if a participant’s self-reported annual income fell below the threshold of twice the national poverty line for the year of recruitment into the study (29,30). For score 2, we used the socioeconomic status index (continuous variable), described in previous publications from our group as a comprehensive measure of education, occupation, income, and household possessions (31).

For each version of the score, conditional logistic regression models were fit with MI as the outcome and cardiovascular risk score components as predictors, while matching on age, sex, and area of residence to control for potential confounding by these demographic factors. The obtained regression coefficients for each score component were then multiplied by the values of cardiovascular risk score components and summed across components to produce the final value of the cardiovascular risk score. Thus, the final cardiovascular risk score value represents a weighted sum of individual risk score components (32). Two regression models, each adjusted for age, sex, and area of residence, were fit to assess the discriminatory ability of score 1 and score 2: 1 with each score as a continuous variable and 1 with indicator variables corresponding to quintiles of each score’s distribution. A test for linear trend was performed on categorical models, using the median value of each quintile as a continuous predictor. Sensitivity analyses were conducted using different subsets of components of the cardiovascular risk scores. A conditional logistic regression model was fit including foods as categorical variables (using type of oil used in the household as proxy for fat intake and quintiles of intake of fruits,
vegetables, and beans as proxies for fiber and folate) in an effort to make cardiovascular risk assessment more accessible in resource-limited settings. In another sensitivity analysis, a conditional logistic regression model was fit replacing waist:hip ratio with waist circumference, which was dichotomized using Latino-specific cutoffs of 89 cm for men and 84 cm for women (33). Because the models were not nested, we compared their fit using the Akaike Information Criterion (AIC) (34).

Validation of the cardiovascular risk score. The validation data set comprised all study participants excluded from the derivation data set, i.e., participants with a self-reported history of hypertension, diabetes, and/or hypercholesterolemia. The models for both score 1 and score 2, derived in healthy participants, were used to predict the probability of MI in the validation data set and receiver operating characteristic (ROC) curves were constructed from all combinations of sensitivity and (1–specificity) characterizing score 1 and score 2, respectively. Areas under the ROC curve (c-statistic) were used to assess the models’ predictive ability in the validation data set and compared with similar measures of performance in currently used cardiovascular risk stratification models (8,10). For a comprehensive validation of the score models, the validation and derivation data sets were then switched, with scores 1 and 2 now derived in the subpopulation with a history of chronic disease and validated in the healthy subpopulation using the ROC method described above. All statistical analyses were conducted using SAS version 9.2.

Results

Descriptive statistics. The majority of participants (total n = 4091) complied with approximately one-half of all lifestyle recommendations (Fig. 1).

Compared with controls (n = 1984 after restriction to participants with complete information on covariates and rematching), cases (n = 1984) were less likely to follow the recommendations regarding intake of saturated fats, polyunsaturated fats, trans fats, and cholesterol, as well as smoking, but were more likely to follow guidelines on fiber and folate (Table 1). Additionally, cases were more likely to exhibit higher waist:hip ratios and to be below the income threshold of twice the national poverty line. Physical activity and alcohol consumption patterns did not significantly differ by case/control status. Overall, compliance was lowest for saturated fat intake (23% for cases, 30% for controls) significantly differ by case/control status. Overall, compliance was lowest for saturated fat intake (23% for cases, 30% for controls) and moderate alcohol consumption (16% for both). The mean age of the study population was 58 y (range: 18–86 y).

Score derivation. For most factors, a gradient of risk was observed between the categories of the variable (Supplemental Table 1). Of all risk score components, current smoking was associated with the biggest increase in the risk of MI [multivariate-adjusted OR = 3.17 (95% CI = 2.43–4.13)] (n = 1386).

Higher values of cardiovascular risk scores (n = 1678 for score 1 and n = 1392 for score 2) were associated with an increase in the risk of MI in both the continuous and categorical models [score 1: OR for the continuous variable = 2.72 (95% CI = 2.28–3.24; 2-sided P < 0.0001), for the categorical model, 2-sided P-trend < 0.0001. Score 2: OR for the continuous variable = 2.71 (95% CI = 2.26–3.26, 2-sided P < 0.0001), for the categorical model, 2-sided P-trend < 0.0001] (Table 2). The estimated OR represent the exponentiated slope of the corresponding regression line. The fit of the continuous models was preferred to the categorical ones due to their lower AIC values. The score was robust to the substitution of type of oil for fat intake and quintile of intake of fruits, vegetables, and beans for fiber and folate (OR for the continuous variable = 2.72 (95% CI = 2.28–3.24)). Finally, replacing waist:hip ratio with waist circumference slightly increased the estimates for both scores [OR for score 1 (continuous) = 2.73 (95% CI = 2.27–3.27); OR for score 2 (continuous) = 2.74 (95% CI = 2.25–3.33)].

Score validation. ROC curves were constructed to evaluate the performance of both scores in the validation data set (Fig. 2A,B). The area under the curve was estimated at 0.63 and 0.64 for score 1 and score 2, respectively. When the derivation and validation data sets were switched, the c-statistics for the corresponding ROC curves (not shown) were estimated at 0.65 for score 1 and 0.67 for score 2. In the sensitivity analysis replacing nutrients in score 1 with foods, c-statistics were estimated at 0.60 and 0.65 (ROC curves not shown).

Discussion

We developed and validated a new cardiovascular risk score in a population from Costa Rica, a middle-income country undergoing nutritional and epidemiologic transition. An increase of 1 unit in the score value was associated with a >2-fold increase in the risk of first nonfatal MI in the study population. The results were robust to inclusion of participants with known risk factors for cardiovascular disease, and the discrimination ability of the score as evaluated by the area under the ROC curve was typically somewhat lower than that of other established risk measures. For example, the c-statistic for the Framingham risk score in diverse ethnic subgroups ranges from 0.63 to 0.79 in men and 0.66 to 0.83 in women (8). It is also important to note that the reported c-statistic estimates are conservative, because the predictive quality of our models may have been adversely affected by differences in characteristics (specifically, chronic disease history) between the derivation and validation data sets. Additionally, some of the components of the Framingham score (e.g. blood lipids) are considerably more difficult and expensive to measure than lifestyle variables. Other proposed cardiovascular risk measures, such as novel biomarkers, are even more costly to implement and offer only modest improvements in predictive ability beyond the Framingham risk score (35). In a limited-resource setting, our proposed score provides a feasible and effective alternative to currently available risk measures.

Our new risk score is exclusively comprised of lifestyle risk factors, namely diet, physical activity, smoking, alcohol consumption, socioeconomic status, and waist:hip ratio. The modifiable nature of this score has several important implications. First, it empowers clinical and public health practice, because it clearly illustrates how the risk of coronary heart disease in the study population can be reduced by addressing each of the score...
components. Second, it presents a convenient summary estimate of cardiovascular risk due to lifestyle factors, facilitating future epidemiologic research on heart disease in Costa Rica. For example, the new score could be used as a risk stratification variable, a measure of lifestyle confounders, or a convenient summary of environmental factors in studies investigating gene-environment interactions. Third, it enables risk estimation for the entire lifestyle pattern, from diet and health behaviors to socioeconomic status. Fourth, the choice of total physical activity rather than solely recreational activity provides a more reliable estimate of caloric expenditure in the setting of middle- or low-income countries, where recreational physical activity is less common (36). Finally, the score can be easily adapted to incorporate risk factors important to other populations or to include FFQ measures of dietary variables when biomarker data are not available. Depending on specific study designs and objectives, new adaptations of our score can also include variables like age and sex, which are not modifiable but highly predictive of cardiovascular risk across populations.

The limitations of the study include the use of the complete case method to construct the validation and derivation data sets. Participants with missing values for any of the covariates were dropped from the analysis, which requires the assumption that the data are missing at random. Although there is no strong evidence to suggest otherwise, this assumption has not been tested in this study. Another limitation are the possible inaccuracies in self-report of nutritional intake, because all dietary covariates could not be ascertained using biomarker data. In addition to nondifferential misclassification associated with the use of an FFQ, under-reporting of caloric intake and adverse health habits has been well documented in a variety of populations and is especially prevalent among high-risk participants, potentially resulting in a risk estimate that is biased toward the null (37). Additionally, our definition of the outcome as nonfatal MI did not include other adverse coronary heart disease events such as fatal MI or stroke, limiting the scope of our risk score. Further, because the blood lipid measures among cases in the Costa Rica Study were taken post-MI, we could not directly compare the performance of our score with that of the Framingham risk score in our study population. Moreover, the estimates provided by the 2 versions of the score are also not directly comparable, because the analyses were based on slightly different study populations due to availability of covariate information. Finally, because this was a case-control study with participants matched on age and sex, these findings apply primarily to the age groups investigated in this study (ages 18–86 y). To be incorporated for use in the general population, these findings should be replicated and researchers should consider providing age- and sex-specific cardiovascular risk score algorithms.

Although the burden of cardiovascular disease in Costa Rica is substantially lower than the average for the Americas (188 deaths/100,000 population vs. 281 deaths/100,000 population, age-standardized) (38), compliance with international healthy lifestyle guidelines in our study population is similar to or worse than other middle-income countries. A recent study of modifi-

---

### TABLE 1

Frequency of compliance with international guidelines for healthy lifestyle and income standards in the Costa Rica Study population by case-control status

<table>
<thead>
<tr>
<th>Healthy diet</th>
<th>Cases, n = 1984</th>
<th>Controls, n = 1984</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saturated fats &lt;10% of total energy</td>
<td>450 (23)*</td>
<td>593 (30)</td>
</tr>
<tr>
<td>Polyunsaturated fats = 6–10% of total energy</td>
<td>1044 (53)*</td>
<td>1114 (56)</td>
</tr>
<tr>
<td>Trans fats &lt;1% of total energy</td>
<td>679 (34)*</td>
<td>744 (38)</td>
</tr>
<tr>
<td>Cholesterol &lt;300 mg/d</td>
<td>1014 (51)*</td>
<td>1252 (63)</td>
</tr>
<tr>
<td>Fiber &gt;25 g/d</td>
<td>951 (48)*</td>
<td>862 (41)</td>
</tr>
<tr>
<td>Folate &gt;400 μg/d</td>
<td>944 (56)*</td>
<td>818 (51)</td>
</tr>
<tr>
<td>Physically active</td>
<td>702 (35)</td>
<td>751 (38)</td>
</tr>
<tr>
<td>Waist:hip ratio ≤0.90 for men, ≤0.85 for women</td>
<td>202 (10)*</td>
<td>327 (16)</td>
</tr>
<tr>
<td>Currently nonsmoking</td>
<td>1182 (60)*</td>
<td>1557 (78)</td>
</tr>
<tr>
<td>Alcohol intake = 10–50 g/d</td>
<td>323 (16)</td>
<td>327 (16)</td>
</tr>
<tr>
<td>Income &gt; twice the national poverty line</td>
<td>1734 (87)*</td>
<td>1798 (91)</td>
</tr>
</tbody>
</table>

* Expended >10% of daily energy intake in the performance of moderate- and high-intensity physical activities (at least 4 times the basal metabolism rate) (29). * P < 0.05 (McNemar’s test).

### TABLE 2

OR estimates of nonfatal MI associated with the value of the cardiovascular risk score in healthy participants of the Costa Rica Study^1

<table>
<thead>
<tr>
<th>Quintile 1</th>
<th>Quintile 2</th>
<th>Quintile 3</th>
<th>Quintile 4</th>
<th>Quintile 5</th>
<th>AIC</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score 1</td>
<td>335</td>
<td>336</td>
<td>336</td>
<td>336</td>
<td>335</td>
<td></td>
</tr>
<tr>
<td>Score 2</td>
<td>278</td>
<td>279</td>
<td>278</td>
<td>279</td>
<td>278</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>1984</td>
<td>1984</td>
<td>1984</td>
<td>1984</td>
<td>1984</td>
<td></td>
</tr>
</tbody>
</table>

Model^1

<table>
<thead>
<tr>
<th>Score 1, categorical</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score 1, referent</td>
<td>1.00</td>
</tr>
<tr>
<td>Score 1, non-referent</td>
<td>1.25 (0.90, 1.74)</td>
</tr>
<tr>
<td>Score 2, categorical</td>
<td>1.00</td>
</tr>
<tr>
<td>Score 2, non-referent</td>
<td>1.78 (1.21, 2.62)</td>
</tr>
</tbody>
</table>

^1 All models were adjusted for age, sex, and residence (by matching).
A novel cardiovascular risk score should entail a comparison of this method with established scores such as the Framingham or SCORE equations, as well as applications of the score to other populations, especially in developing countries.

**Acknowledgments**

S.A. conducted the data analysis and wrote the manuscript; H.C. contributed to the interpretation of the data and proofread and edited the manuscript; E.B.L. helped devise the study’s analytic strategy and proofread and edited the manuscript; C.L. provided assistance with the statistical methods used in the analysis; J.M. O. proofread and edited the manuscript; and A.B. supervised the data analysis and main aspects of data interpretation, proofread, and edited the manuscript. All authors read and approved the final manuscript.

**Literature Cited**


17. Tunstall-Pedoe H, Kuulasmaa K, Amouyel P, Arveiler D, Rajakangas AM, Pajak A. Myocardial infarction and coronary deaths in the World...


30. Panorama social [Social panorama]. Santiago (Chile): Comisión Económica de las Naciones Unidas para América Latina (CEPAL); 1999.


