Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Yusuf S, Rangarajan S, Teo K, et al. Cardiovascular risk and events in 17 low-, middle-, and high-income countries. N Engl J Med 2014;371:818-27. DOI: 10.1056/NEJMoa1311890

Table of Contents

PURE Project Office Staff, National Coordinators, Investigators and Key Staff

Project office (Population Health Research Institute, Hamilton Health Sciences and McMaster University, Hamilton, Canada): S Yusuf* (Principal Investigator).

S Rangarajan (Project Manager); K K Teo, C K Chow, M O'Donnell, A Mente, D Leong, A Smyth, P Joseph, S Islam (Statistician), M Zhang (Statistician), W Hu (Statistician), G Wong (Statistician), N Manandhar (Statistician), M Dehghan (Nutritionist), M Royerr, J DeJesus, G Lewis, P Mackie, L Farago, I Kay, D Agapay, R Solano, S Ramacham, N Kandy, J Rimac, S Trottier, W ElSheikh, M Mustaha, T Tongana, N Aoucheva, J Swallow, E Ramezani, J Lindeman

Core Laboratories: M McQueen, K Hall, J Keys (Hamilton), X Wang (Beijing, China), J Keneth, A Devanath (Bangalore, India).

ARGENTINA: R Diaz*; A Orlandini, B Linetsky, S Toscanelli, G Casaccia, JM Maini Cuneo; **BANGLADESH:** O Rahman*, R Yusuf, AK Azad, KA Rabbani, HM Cherry, A Mannan, I Hassan, AT Talukdar, RB Tooheen, MU Khan, M Sintaha, T Choudhury, R Haque, S Parvin; **BRAZIL:** A Avezum*, GB Oliveira, CS Marcilio, AC Mattos; **CANADA:** K Teo***,** S Yusuf***,** J Dejesus, D Agapay, T Tongana, R Solano, I Kay, S Trottier, J Rimac, W Elsheikh, L Heldman, E Ramezani, G Dagenais, P Poirier, G Turbide, D Auger, A LeBlanc De Bluts, MC Proulx, M Cayer, N Bonneville, S Lear, D Gasevic, E Corber, V de Jong, I Vukmirovich, A Wielgosz, G Fodor, A Pipe, A Shane; **CHILE:** F Lanas*, P Seron, S Martinez, A Valdebenito, M Oliveros; **CHINA:** Li Wei*, Liu Lisheng*, Chen Chunming, Wang Xingyu, Zhao Wenhua, Zhang Hongye, JiaXuan, Hu Bo, Sun Yi, Bo Jian, Zhao Xiuwen, Chang Xiaohong, Chen Tao, Chen Hui, Chang Xiaohong, Deng Qing, Cheng Xiaoru, Deng Qing, He Xinye, Hu Bo, JiaXuan, Li Jian, Li Juan,Liu Xu, Ren Bing, Sun Yi, Wang Wei, Wang Yang, Yang Jun, Zhai Yi, Zhang Hongye, Zhao Xiuwen,Zhu Manlu, Lu Fanghong, Wu Jianfang, Li Yindong, Hou Yan, Zhang Liangqing, Guo Baoxia, Liao Xiaoyang, Zhang Shiying, BianRongwen, TianXiuzhen, Li Dong, Chen Di, Wu Jianguo, Xiao Yize, Liu Tianlu, Zhang Peng, Dong Changlin, Li Ning, Ma

Xiaolan, Yang Yuqing, Lei Rensheng, Fu Minfan, He Jing, Liu Yu, Xing Xiaojie, Zhou Qiang, ; **COLOMBIA:** P Lopez-Jaramillo*, PACamacho Lopez, R Garcia, LJA Jurado, D Gómez-Arbeláez, JF Arguello, R Dueñas, S Silva, LP Pradilla, F Ramirez, DI Molina, C Cure-Cure, M Perez, E Hernandez, E Arcos, S Fernandez, C Narvaez, J Paez, A Sotomayor, H Garcia, G Sanchez, T David, A Rico; **INDIA:** P Mony *, M Vaz*, A V Bharathi, S Swaminathan, K Shankar AV Kurpad, KG Jayachitra, N Kumar, HAL Hospital, V Mohan, M Deepa, K Parthiban, M Anitha, S Hemavathy, T Rahulashankiruthiyayan, D Anitha, K Sridevi, R Gupta, RB Panwar, I Mohan, P Rastogi, S Rastogi, R Bhargava, R Kumar, J S Thakur, B Patro, PVM Lakshmi, R Mahajan, P Chaudary, V Raman Kutty, K Vijayakumar, K Ajayan, G Rajasree, AR Renjini, A Deepu, B Sandhya, S Asha, HS Soumya; **IRAN:** R Kelishadi*, A Bahonar, N Mohammadifard, H Heidari; **MALAYSIA:** K Yusoff*, TST Ismail, KK Ng, A Devi, NM Nasir, MM Yasin, M Miskan, EA Rahman, MKM Arsad, F Ariffin, SA Razak, FA Majid, NA Bakar, MY Yacob, N Zainon, R Salleh, MKA Ramli, NA Halim, SR Norlizan, NM Ghazali, MN Arshad, R Razali, S Ali, HR Othman, CWJCW Hafar, A Pit, N Danuri, F Basir, SNA Zahari, H Abdullah, MA Arippin, NA Zakaria, I Noorhassim, MJ Hasni, MT Azmi, MI Zaleha, KY Hazdi, AR Rizam, W Sazman, A Azman; **OCCUPIED PALESTINIAN TERRITORY:** R Khatib*, U Khammash, R Giacaman, A Husseini, K Abu-Diab, M Warrad; **PAKISTAN:** R Iqbal*, A Afridi, R Khawaja, A Raza, K Kazmi; **POLAND:** W Zatonski*, A Szuba, K Zatonska, R Ilow[#], M Ferus, B Regulska-Ilow, D Różańska, M Wolyniec; **SOUTH AFRICA:** A Kruger*, H H Voster, A E Schutte, E Wentzel-Viljoen, FC Eloff, H de Ridder, H Moss, J Potgieter, AA Roux, M Watson, G de Wet, A Olckers, JC Jerling, M Pieters, T Hoekstra, T Puoane, E Igumbor, L Tsolekile, D Sanders, P Naidoo, N Steyn, N Peer, B Mayosi, B Rayner, V Lambert, N Levitt, T Kolbe-Alexander, L Ntyintyane, G Hughes, R Swart, J Fourie, M Muzigaba, S Xapa, N Gobile , K Ndayi, B Jwili, K Ndibaza, B Egbujie; **SWEDEN:** A Rosengren*, K Bengtsson Boström, U Lindblad, P Langkilde, A Gustavsson, M Andreasson, M Snällman, L Wirdemann, K Pettersson, E Moberg ; **TURKEY:** A Oguz*, AAK Akalin, KBT Calik, N Imeryuz, A Temizhan, E Alphan, E Gunes, H Sur, K Karsidag, S Gulec, Y Altuntas; **UNITED ARAB EMIRATES:** AM Yusufali*, W Almahmeed, H Swidan, EA Darwish, ARA Hashemi, N Al-Khaja, JM Muscat-Baron, SH Ahmed, TM Mamdouh, WM Darwish, MHS Abdelmotagali, SA Omer Awed, GA Movahedi, F Hussain, H Al Shaibani, RIM Gharabou, DF Youssef, AZS Nawati, ZAR Abu Salah, RFE

Abdalla, SM Al Shuwaihi, MA Al Omairi, OD Cadigal; R.S. Alejandrino; **ZIMBABWE:** J Chifamba*, L Gwaunza, G Terera, C Mahachi, P Mrambiwa, T Machiweni, R Mapanga.

*National Coordinator

[#] Deceased

Supplementary Methods: The Prospective Urban Rural Epidemiological Study (PURE Study) Design

The Prospective Urban Rural Epidemiological Study (PURE Study) enrolled 155,875 individuals between 35 and 70 years of age from 17 low, middle and high-income countries.[1,2] The study includes population samples from 628 communities from 17 countries from 5 continents representing a broad range of economic and social circumstances.[1,2] PURE includes countries in four income strata based on World Bank classification in 2006: four low-income countries (Bangladesh, India, Pakistan, and Zimbabwe), three lower middle-income countries (China, Colombia, and Iran), seven upper middle-income countries (Argentina, Brazil, Chile, Malaysia, Poland, South Africa, and Turkey), and three high-income countries (Canada, Sweden, and United Arab Emirates) (see Table S1). The study is coordinated by the Population Health Research Institute, Hamilton Health Sciences and McMaster University, Canada.

Supplementary Methods: PURE Study Participant Selection Methodology as Excerpted from Teo et al.[1]

Selection of Countries

The choice and number of countries selected in PURE reflects a balance between involving a large number of communities in countries at different economic levels, with substantial heterogeneity in social and economic circumstances and policies, and the feasibility of centers to successfully achieve long-term follow-up (see Table S2). Thus, PURE included sites in which investigators are committed to collecting good-quality data for a low-budget study over the planned 10-year follow-up period and did not aim for a strict proportionate sampling of the entire world.

Selection of Communities

Within each country, urban and rural communities were selected based on broad guidelines (see Table S2). A common definition for "community" that is applicable globally is difficult to establish.[3] In PURE, a community was defined as a group of people who have common characteristics and reside in a defined geographic area. A city or large town was not usually considered to be a single community, rather communities from low-, middle-, and high-income areas were selected from sections of the city and the community area defined according to a geographical measure (eg, a set of contiguous postal code areas or a group of streets or a village). The primary sampling unit for rural areas in many countries was the village. The reason for inclusion of both urban and rural communities is that for many countries, urban and rural

environments exhibit distinct characteristics in social and physical environment, and hence, by sampling both, we ensured considerable variation in societal factors across PURE communities.

The number of communities selected in each country varied, with the aim to recruit communities with substantial heterogeneity in social and economic circumstances balanced against the capacity of local investigators to maintain follow-up. In some countries (eg, India, China, Canada, and Colombia), communities from several states/provinces were included to capture regional diversity, in policy, socioeconomic status, culture, and physical environment. In other countries (eg, Iran, Poland, Sweden, and Zimbabwe), fewer communities were selected.

Selections of Households and Individuals

Within each community, sampling was designed to achieve a broadly representative sample of that community of adults aged between 35 and 70 years (see Table S2). The choice of sampling frame within each center was based on both "representativeness" and feasibility of long-term follow-up, following broad study guidelines. Once a community was identified, where possible, common and standardized approaches were applied to the enumeration of households, identification of individuals, recruitment procedures, and data collection.

The method of approaching households differed between regions. For example, in rural areas of India and China, a community announcement was made to the village through contact of a community leader, followed by in-person door-to-door visits of all households. In contrast in Canada, initial contact was by mail followed by telephone inviting members of the households to a central clinic. Households were eligible if at least 1 member of the household was between the

ages of 35 and 70 years and the household members intended to continue living in their current home for a further 4 years.

For each approach, at least 3 attempts at contact were made. All individuals within these households between 35 and 70 years providing written informed consent were enrolled. When an eligible household or eligible individual in a household refused to participate, demographics and self-reported data about CVD risk factors, education, and history of CVD, cancers and deaths in the households within the two previous years were recorded.

To ensure standardization and high data quality, we used a comprehensive operations manual, training workshops, DVDs, regular communication with study personnel and standardized report forms. We entered all data in a customized database programmed with range and consistency checks which was transmitted electronically to the Population Health Research Institute in Hamilton (Ontario, Canada) where further quality checks were implemented.

Supplementary Methods: Representativeness of the PURE Cohort

a) Are the countries included in the PURE cohort atypical?

We compared the countries participating in PURE with those participating in MONICA[4], the largest previous study of this nature conducted. We plotted national income (GDP/capita) against research output (number of publications recorded in SCOPUS between 1996 and 2010 per 100,000 population). These data are taken from a previous study undertaken by one of the authors (M. McKee) on global health research capacity to inform policy discussions within WHO[5]. These graphs (Figures S1 and S2) demonstrate that PURE has captured the full diversity of countries on these two dimensions (Figure S1), in marked contrast to MONICA, which was concentrated in high income countries with substantial research capacity (Figure S2).

b) Are PURE populations representative of the countries in which they are situated?

The PURE household population compared to national statistics had more women (sex ratio 95.1 men per 100 women vs 100.3) and was older (33.1 years vs 27.3), although age had a positive linear relationship between the two data sources (Pearson's r = 0.92). PURE was 59.3% urban compared to an average of 63.1% in participating countries. The distribution of education was less than 7% different for each category, although PURE households typically had higher levels of education. For example, 37.8% of PURE household members had completed secondary education compared to 31.3% in the national data. However, age-adjusted annual mortality rates showed positive correlation for men (r = 0.91) and women (r = 0.92) but were lower in PURE compared to national statistics (7.9 per 1000 vs 8.7 for men; 6.7 vs 8.1 for women) (Figure S3). These findings indicate that modest differences exist between the PURE household population

and national data for the indicators studied. These differences, however, are unlikely to have much influence on exposure-disease (or health systems assessments vs outcome) associations derived in PURE. In addition, mortality rates reported for the two years prior to enrolment in the PURE study were closely correlated with the mortality rates observed in the study participants during follow up (r=0.87 and 0.85 respectively) (Figure S4). Further, incidence estimates from PURE, adjusted or stratified according to sex and/or urban/rural location, will enable valid comparisons of the relative rates of various cardiovascular outcomes across countries.

Supplementary Methods: Collection of Demographics, Risk Factors and Outcome Events

We collected data at national, community, household, and individual levels with standardized questionnaires.[1] Questions about age, sex, education, smoking status, hypertension, diabetes, and obesity were identical to those in the INTERHEART and INTERSTROKE studies.[6,7] We obtained BP measurements in individuals and so hypertension was defined as those with a BP >140/>90 or those who were already on treatment. Fasting glucose was available in most individuals (76%) and so diabetes was defined as those who were reported as having diabetes and those with a fasting glucose >7.0 mmol/L. (Sensitivity analyses indicate a very high correlation between self report of diabetes alone versus self-report and fasting glucose >7.0 in the 110,000 people with both measures, and so self-report is a reasonable surrogate for the prevalence of diabetes) Total cholesterol was available in 122,640 individuals and a value of >5.2 mmol/L was considered to be elevated.

In most of the LIC and MIC there was no central system of death or event registration. We therefore; 1) obtained information on prior medical illness and medically certified cause of death where available, 2) captured best available information from reliable sources in those instances where medical information was not available in order to be able to arrive at a probable diagnosis or cause of death. Event documentation was based on information from household interviews and medical records, death certificates and other sources. We also used Verbal Autopsies to ascertain cause of death in addition to medical records which were reviewed by a health professional. This approach has been used in several studies conducted in LIC and MIC.[8,9]

To ensure a standard approach and accuracy for classification of events across all countries and over time, the first 100 CVD events (deaths, MI, strokes, heart failure or cancers) for China and India, and 50 cases for other countries were adjudicated both locally and also by the adjudication chair, and if necessary further training was provided. Thereafter, every year, 50 cases for China and India and 25 cases for each of the remaining countries were adjudicated as above.

Supplementary Methods: Event Definitions

FATAL EVENTS

Death due to cardiovascular events

Sudden unexpected cardiovascular death

Death that occurred suddenly and unexpectedly without evidence of other cause of death (examples: witnessed collapse, persons resuscitated from cardiac arrest who later died) or persons seen alive less than 12 hours prior to discovery of death (example persons found dead in his/her bed).

Non-sudden unexpected cardiovascular death.

Death that occurred unexpectedly without evidence of other cause of death in persons seen alive more than 12 hours but less than 24 hours.

Fatal myocardial infarction (one of the following)

- Autopsy demonstrating fresh myocardial infarction and/or recent coronary occlusion, or
- ECG showing new and definite sign of MI (Minnesota code 1.11), or
- Symptoms typical or atypical or inadequately described but attributed to cardiac origin lasting more than 10 minutes and cardiac enzymes at least twice above the upper limit of normal or troponin at least at the lower level of necrosis, or

- ECG with new ischemic changes (new ST depression or T wave inversion ≥ 2 mm) and cardiac enzymes at least twice above the upper limit of normal or troponin at least at the lower level of necrosis.

Fatal stroke

Diagnosis of stroke by a physician based on sudden neurological deficit of vascular origin with or without neuroimaging studies (CT scan/MR scan/angiography/Doppler) lasting 24 hours and more, occurring within 30 days of signs or symptoms of stroke or autopsy evidence of a recent stroke. *If death occurred within 24 hours of onset of stroke signs, this will be considered a definite death due to stroke.*

Congestive heart failure

Death due to heart failure in absence of myocardial infarction or other causes was attributed to fatal heart failure.

The diagnosis of congestive heart failure required 2 of the 3 following criteria:

- Signs (rales, increased jugular venous pressure or ankle edema) or symptoms (nocturnal paroxysmal dyspnea, dyspnea at rest of ankle edema) of congestive heart failure,
- Radiological signs of pulmonary congestion,
- Treatment of heart failure with diuretics

Other cardiovascular death *(other causes having been excluded)*

Arterial rupture of aneurysm, Pulmonary embolism, Arrhythmic death (A-V block, sustained ventricular tachycardia in absence of other causes), Death after invasive cardiovascular intervention, Congenital heart disease, Heart valve disease (including rheumatic heart disease, Endocarditis, Myocarditis, Tamponade.

Non-fatal cardiovascular events

Non-periprocedural myocardial infarction

- ECG showing new and definite sigh of MI (Minnesota code 1.11), or
- Symptoms typical or atypical or inadequately described but attributed to cardiac origin lasting more than 10 minutes and cardiac enzymes at least twice above the upper limit of normal (ULN) or troponin at least at the lower level of necrosis, or
- ECG with new ischemic changes (new ST depression or T wave inversion $\geq$ 2 mm) and cardiac enzymes at least twice above the upper limit of normal or troponin at least at the lower level of necrosis.

Periprocedural myocardial infarction

ECG showing new and definite sign of MI (Minnesota code 1.11), OR cardiac marker values:

- Percutaneous coronary intervention, CKMB should be ≥ 3X ULN of troponin ≥ 5 X above lower level of necrosis,
- Coronary surgery cardiac markers CKMB should be ≥ 10X ULN or troponin ≥ 10 X above lower limit of necrosis.

Stroke

Diagnosis of stroke by a physician based on sudden neurological deficit of vascular origin with or without neuroimaging studies (CT scan/MRI scan/angiography/Doppler) lasting 24 hours and more.

Congestive heart failure

The diagnosis of congestive heart failure requires 2 of the 3 following criteria:

- Signs (rales, increased jugular venous pressure or ankle edema) of symptoms (nocturnal paroxysmal dyspnea, dyspnea at rest of ankle edema) of congestive heart failure,
- Radiological signs of pulmonary congestion,
- Treatment of heart failure with diuretics

Non-Major CVD

Includes all hospitalizations for CVD (other than those due to myocardial infarction, stroke, or heart failure) or related investigations or procedures.

Supplementary Methods: The INTERHEART Risk Score (IHRS) (McGorrian et al.)[10]

We chose the INTERHEART Risk Score for use in the PURE study because it has been developed and initially validated in an international case-control study of MI (28,000 people) and then in an external dataset of an international prospective cohort study, called EpiDREAM involving 18,000 people. Briefly, in the INTERHEART study a brief parsimonious risk score was developed and validated.[10] The short risk score included (apolipoproteins, smoking, second hand smoking, history of HTN, and DM). In addition some secondary models were developed to enable use for individuals with cholesterol measurements instead of apolipoproteins (a cholesterol risk score), and for those individuals where no blood measure of lipids was taken (non-lab based risk score) (Table S3). The "non-lab" based model contains a variable for family history, plus variables that can be obtained from history or simple physical examination, but no lipid variable.

This non-lab was created and tested to determine its potential for use in resource poor settings. Thus it is most appropriate to be used in the PURE study. The ROC C statistics for the "full modifiable risk", the "cholesterol" score, and the "non-laboratory" based scores are similar across geographic regions (Table S4). Non-laboratory-based scores have an added advantage that they do not require lab testing facilities in order to estimate risk, making them ideal for the patient's first visit in primary care settings, for community worker use, and for use in resource-poor settings. The US NHANES prospective cohort showed that a non-lab based risk score (age, SBP, smoking, DM, HTN, and BMI) was equally predictive of first CVD as compared to a lab–based risk score which included LDL and HDL cholesterol instead of BMI[11].

Note that in the PURE subset which had cholesterol values available, the correlation between the cholesterol-based IHRS and non-laboratory IHRS is 0.77.

Development of the IHRS

The INTERHEART risk score was derived from the INTERHEART case-control study which recruited cases of first acute MI admitted to coronary care or equivalent units, and at least one sex and age matched (within 5 years) control, who had no history of heart disease. Participants were recruited from 252 centers, in 52 countries worldwide.

Stage 1: Assessment of each proposed risk factor:

After data cleaning and exclusion of subjects with missing risk factor data, the INTERHEART dataset was split randomly into a 2/3s derivation set and a 1/3 test set. Splitting of the data was performed in a paired manner, to maintain the matched nature of the data, and was stratified by sex and geographic area.

Using the 2/3s derivation set, each of the 9 modifiable INTERHEART risk factors was examined in a simple logistic regression model, adjusting for sex, age and geographic region, and with acute MI as the dependent or outcome variable. The variables relating to each risk factor were chosen from the questionnaire data based both on the variables used in previous INTERHEART analyses and publications, and on the expected clinical utility of the variable. Categorical variables were examined using design variables. Proposed variables were retained if they achieved the set criterion of alpha ≤ 0.05 for statistical significance in the simple model.

Stage 2: Creation of risk factor definitions to be used in the INTERHEART modifiable risk score

The following risk factor definitions were selected. Apolipoprotein B:A1 ratio (included as a continuous variable in the final model, and as quartiles for the final score) was selected as the measure of lipid status of choice. A secondary model (the "cholesterol" score) replaced apolipoproteins with low density lipoprotein (LDL) and high density lipoprotein (HDL) measurements, for use in regions where apolipoprotein testing is not readily available. Current smokers were defined as individuals who had smoked any tobacco in the last 12 months, including those people who had quit within that time. Current smoking was categorized into number of cigarettes or beedies smoked per day. Former smokers were those persons who had quit 12 or more months prior to the interview. Second hand smoke was defined as exposure for one or more hours per week, versus less or no exposure. Hypertension and diabetes were both defined by self-report. Waist-hip ratio was chosen as the optimal index of abdominal obesity (included as a continuous variable in the final model and as quartiles for the final score).

Psychosocial factors included details of work/home stress, depression, perceived locus of control, incidence of adverse life events, and financial stress. Diet-related variables chosen were those which had previously been identified as having the strongest association with case status. Consumption was measured in frequency of eating the foodstuff in question, and the diet variables were ultimately included as dichotomous variables. Physical exercise was defined as the level of regular physical activity during leisure time. Alcohol consumption was categorized into persons who never drink alcoholic drinks, those who drink rarely or less than once a month,

and those who drink on one or more occasion a month. Family history was defined as a history of MI in either parent, at any age. This variable was only used in the "non-laboratory" based score.

Stage 3: Calculation of the risk equation

Methods similar to those of Sullivan, Massaro and D'Agostino[12] from the Framingham Heart study were used to develop the INTERHEART Modifiable Risk Score. The risk factor variables were examined in multivariable unconditional logistic regression models, with MI case status as the dependent variable. Variables were added to the multivariable model in a forward stepwise method, with variable entry chosen manually by the investigators as per the relative importance and effect size attributed to each risk factor from the INTERHEART study. The criterion for statistical significance was set at alpha=0.05. The effect of interaction terms was assessed, with terms relating to risk factor*age/sex included in the model building. Model fit was checked at each step using the area under the receiver operator characteristic (ROC) curve c statistic, and the integrated discrimination index (IDI) was used to measure improvement in model discrimination as variables were added. The final model was determined as that which provided the best fit in terms of the ROC and IDI, with due regard to the aim of model parsimony.

A base or reference category was assigned for each variable. The other categories were then valued on how far they were from the base category and this value was weighted by multiplying it by the beta coefficient from the multiple regression equation. The points were derived by multiplying by an empirical constant, and rounding to the nearest integer, for ease of use of the risk score and this method was applied to the creation of the secondary models including the

"non-laboratory" based score and the "cholesterol" score. The "non-laboratory" based model contains a variable for family history, and no lipid variable. This was created and tested to determine its potential for use in resource poor settings. The "cholesterol" model contains variables for LDL and HDL cholesterol, but not for the apolipoproteins. The ROC c statistics for the "full modifiable risk", the "cholesterol" score and the "non-laboratory" based scores are consistent across geographic regions. Non-laboratory-based scores have an added advantage that they do not require lab testing facilities in order to estimate risk, making them ideal for the patient's first visit in primary care settings, for community worker use, and for use in resource-poor settings. Gaziano et al have also shown in the US NHANES prospective cohort that a non-lab based risk score (age, SBP, smoking, DM, HTN, and BMI) was equally predictive of first CVD as compared to a lab–based risk score which included LDL and HDL cholesterol instead of BMI.[11]

INTERHEART used a matched design for age (± 5 years) and sex, with the result that the true effects of age and sex cannot be accurately estimated in this analysis. Nevertheless, we wished to include an age and sex-related variable, to maintain optimal face validity of the score. Furthermore, because matching was not achieved in all recruited subjects, an effect of both age and sex was evident on logistic regression analysis. Therefore, unconditional logistic regression was used to optimize use of the data, and all models were adjusted by age and sex. A single age and sex variable was created which classified men and women into younger vs older groups (men younger than 55 and women younger than 65, or else older) based on epidemiological evidence of this age gap of risk from both INTERHEART and the Framingham study. Analyses were performed using Intercooled Stata 9 (StataCorp, Texas, USA).

Validation within the INTERHEART Study

Stage 4: Internal Validation

Validation of the model was performed in the 1/3 test set. Validity of the model and score were assessed using measures of calibration and discrimination. Calibration is the agreement between the expected probabilities of disease and the actual event rate seen in the test set, and refers to the extent that bias influences the model. Calibration is often assessed by the Hosmer Lemeshow test. However, the results of the Hosmer-Lemeshow test can vary by the statistical software used, and the test is over-sensitive to small deviations in fit as the sample size increases. To address these problems, the deciles of risk were compared separately, and displayed as a calibration plot of the observed versus expected events. Discrimination, or refinement, is a measure of a model's ability to rank subjects correctly in terms of risk. Model discrimination (i.e., the model's ability to rank persons appropriately, from low to high risk) was assessed using the c statistic from ROC testing, and further estimates of model discrimination were made on subgroups of the population. Measures of global fit were also examined, including the Akaike information criteria (AIC), the Bayesian information criteria (BIC), and the Brier score. The Brier score quantifies the overall accuracy of predictions, and ranges from 0 (perfect accuracy) to 0.25 (worthless). Competing models were examined and compared using these measures, and also by the integrated discrimination improvement (IDI) which is a measure of improvement in average sensitivity and "one minus specificity".

External Validation

Stage 5: External validation in a large cohort study of 19,000 subjects

The derived score was externally assessed in an independent study population. EpiDREAM is an international prospective cohort study which includes follow up data on 18,990 participants who were screened for eligibility for the DREAM (Diabetes REduction Assessment with ramipril and rosiglitazone Medication) clinical trial.[13] Subjects were recruited from 21 countries and 191 centers, from North America, South America, Europe, Australia, and Asia. All participants completed a questionnaire with information collected on medical history, physical activity, and diet. Furthermore, all participants underwent a 75 gram oral glucose tolerance test (OGTT), and physical measurements including weight, height, waist and hip circumference were taken using a standardized protocol. 6,800 subjects (35.8%) had impaired fasting glucose or impaired glucose tolerance (5,269 of whom were randomized into DREAM), 2,563 subjects (13.5%) had type 2 diabetes, and 9,627 subjects (50.7%) were normoglycemic. Definitions of modifiable risk factors in EpiDREAM were matched to the INTERHEART definitions (Table S3). The INTERHEART Modifiable Risk Score (IHMRS) was evaluated against incident MI (n=95), and incident CHD (including MI, new angina, and revascularization) (n=289). The models were adjusted for trial status and region.

Internal validation studies were performed for the three tested modifiable risk scores: the "short" score, the "full" score and the "non laboratory-based" score in the development of the various IHRS. These studies (Table S4) demonstrated that the non-lab IHRS performs just as well as the full IHRS, making the former a very useful and efficient approach in resource poor settings.

Figure S1: Countries Participating in PURE by National Income and Medical Research Capacity

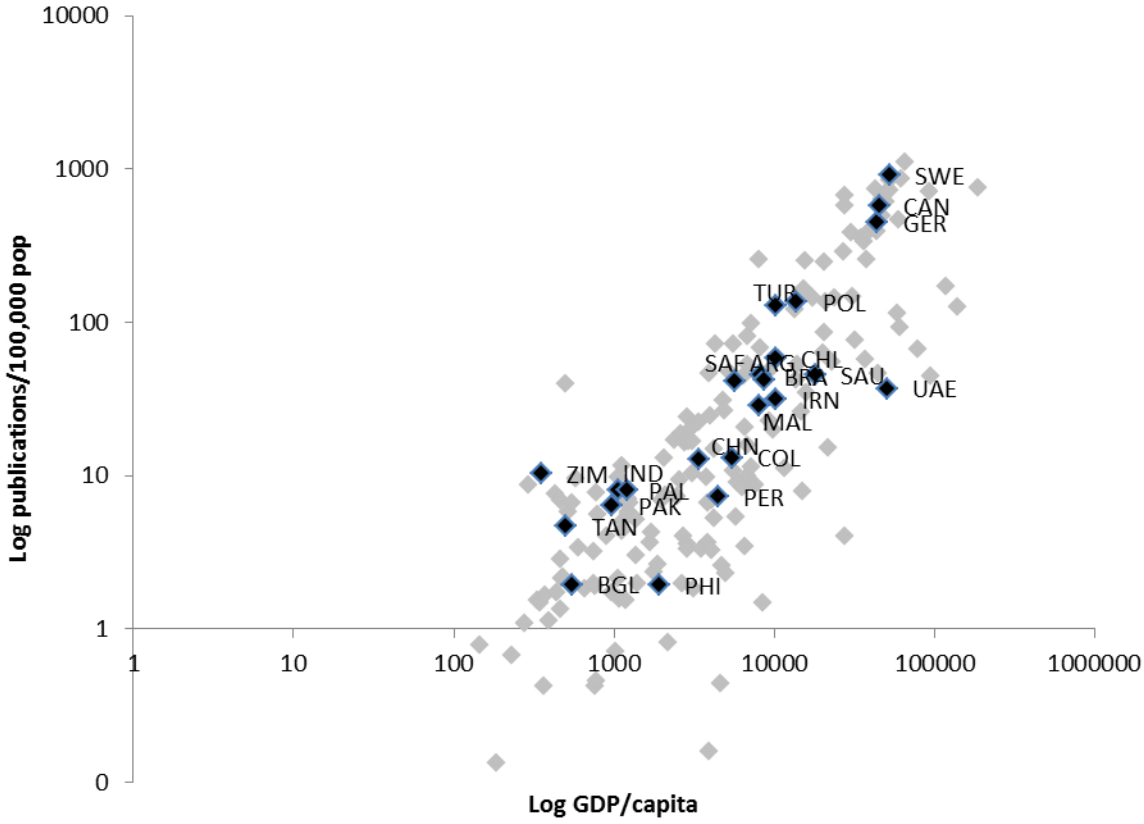
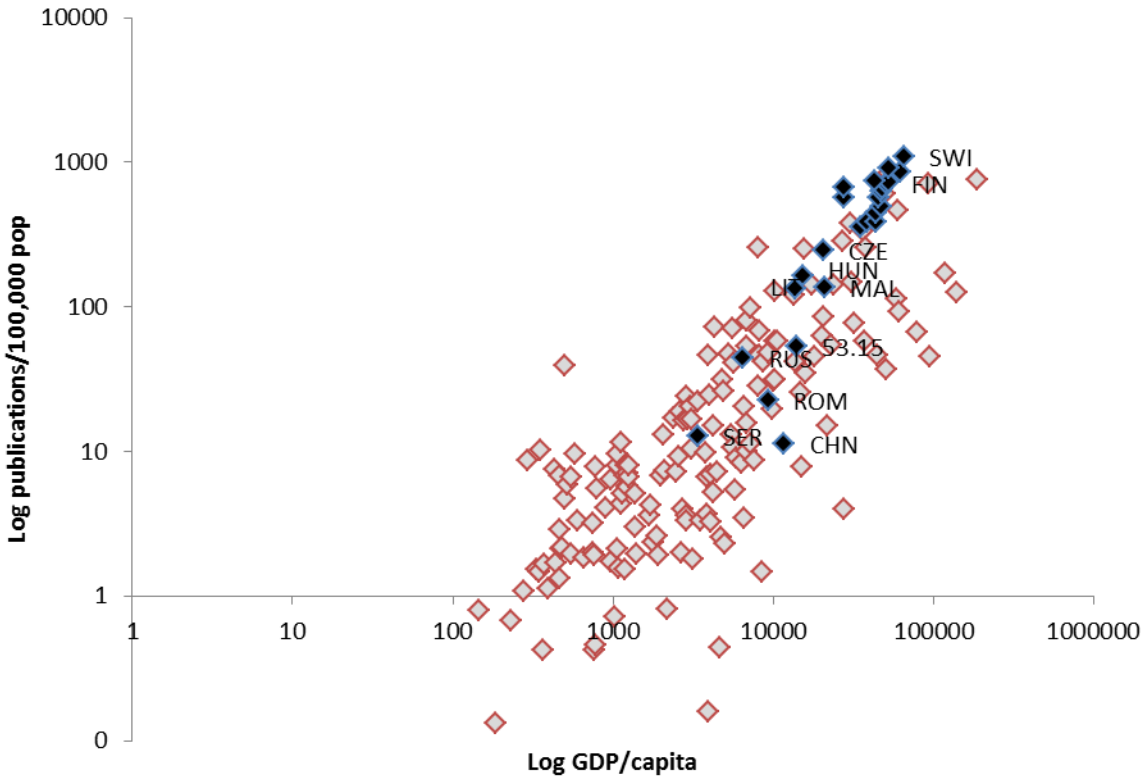


Figure S2: Countries Participating in MONICA by National Income and Medical Research Capacity



For clarity, only a few of the richest and most productive countries in the MONICA study are labeled in this graph. However, all of the middle-income or lower-income countries, and those with lower levels of research productivity, are labeled.

Figure S3: Mortality in PURE Households Reported in the Previous Two Years Versus National Mortality Data[5]

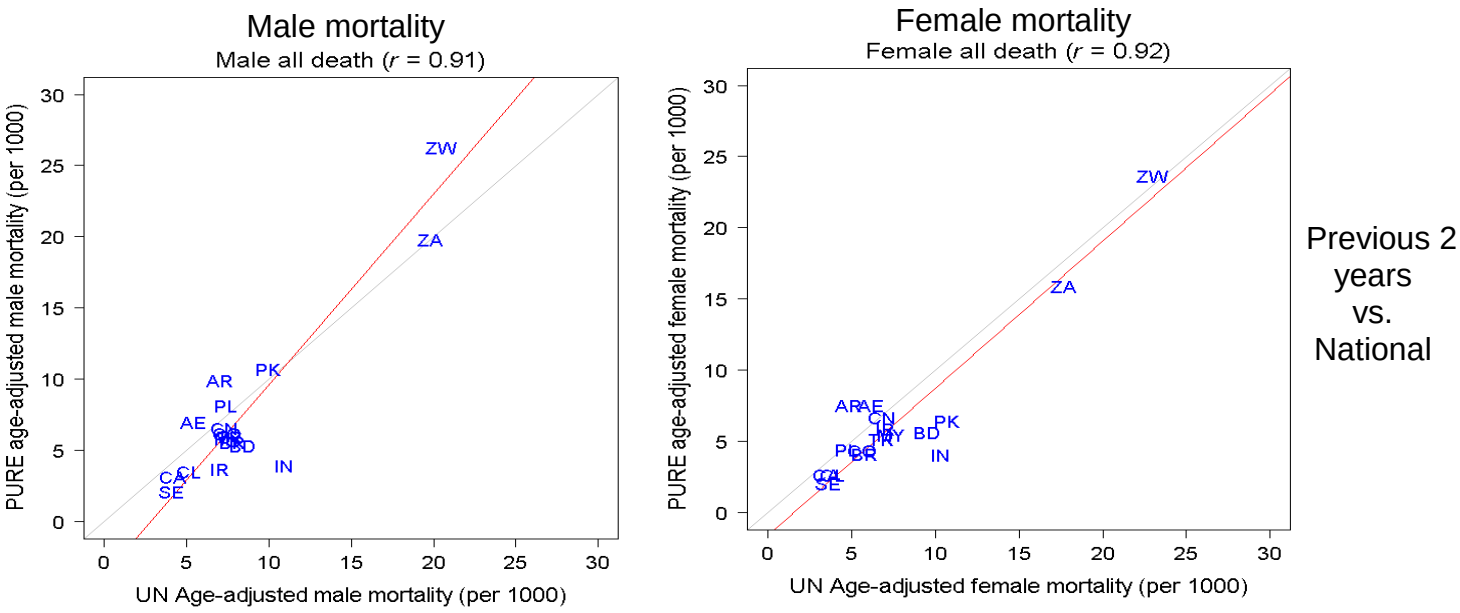
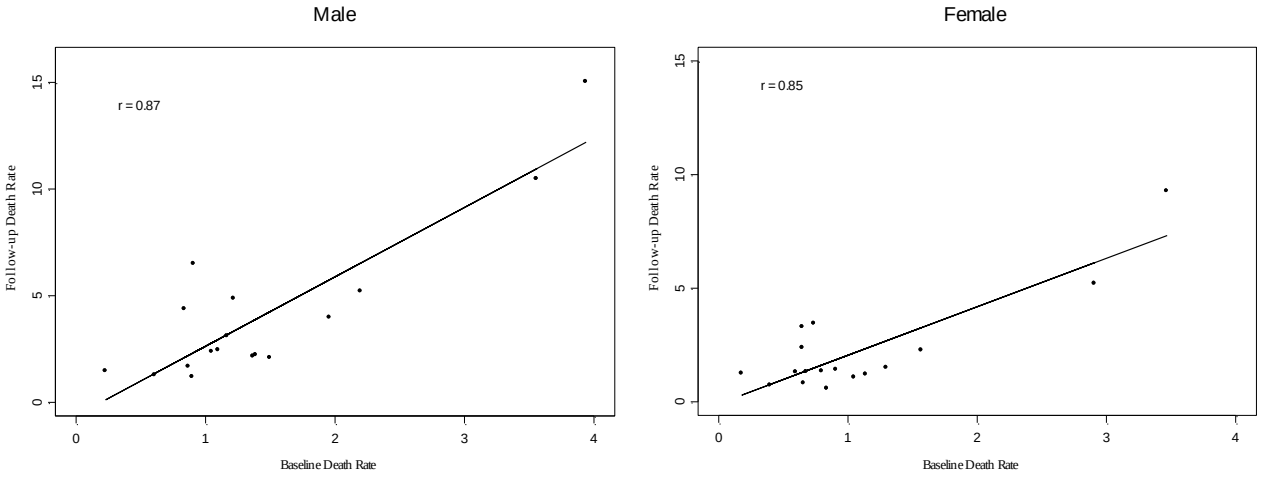


Figure S4: Mortality in the Household in the Previous Two Years (Baseline) vs Deaths During Follow-Up in PURE*



Previous 2 years vs. Follow-up

* Unpublished data.

Table S1: Number of PURE Participants by Country

Country	**N**
India	29228
China	47620
South Africa	4697
Colombia	7524
UAE	1499
Zimbabwe	1263
Brazil	6081
Sweden	4153
Chile	3570
Iran	6013
Canada	10463
Argentina	7540
Poland	2036
Malaysia	15826
Bangladesh	2936
Turkey	4234
Pakistan	1741
All	**156,424**

Table S2: Guidelines for Selection of Countries, Communities, Households, and Individuals Recruited to PURE

Countries
1. High-income countries, middle-income countries, and low-income countries, with the bulk of the recruitment from low- and middle-income regions.
2. Committed local investigators with experience in recruiting for population studies.
Communities
1. Select both urban and rural communities. Use the national definition of the country to determine urban and rural communities.
2. Select rural communities that are isolated (distance of >50 km or lack easy access to commuter transportation) from urban centers. However, consider ability to process bloods samples, eg, villages in rural developing countries should be within 45-min drive of an appropriate facility.
3. Define community to a geographical area, eg, using postal codes, catchment area of health service/clinics, census tracts, areas bordered by specific streets or natural borders such as a river bank.
4. Consider feasibility for long-term follow-up, eg, for urban communities, choose sites that have a stable population such as residential colonies related to specific work sites in developing countries. In rural areas, choose villages that have a stable population. Villages at greater distance from urban centers are less susceptible to large migration to urban centers.
5. Enlist a community organization to facilitate contact with the community, eg, in urban areas, large employers (government and private), insurance companies, clubs, religious organizations, clinic or hospital service regions. In rural areas, local authorities such as priests or community elders, hospital or clinic, village leader, or local politician.
Individual
1. Broadly representative sampling of adults 35 to 70 years within each community unit.
2. Consider feasibility for long-term follow-up when formulating community sampling

framework, eg, small percentage random samples of large communities may be more difficult to follow-up because they are dispersed by distance. In rural areas of developing countries that are not connected by telephone, it may be better to sample entire community (ie, door-to-door systematic sampling).
3. The method of approach of households/individuals may differ between sites. In MIC and HIC, mail, followed up by phone contact may be the practical first means of contact. In LIC, direct household contact through household visits may be the most appropriate means of first contact.
4. Once recruited, all individuals are invited to a study clinic to complete standardized questionnaires and have a standardized set of measurements.

Table S3: The "Non-Laboratory" Based INTERHEART Modifiable Risk Score*

Risk factor	Question	Points for the answer		Points for each section
Age	Are you a man 55 years or older OR woman 65 years or older?		2	**Points:**
	OR Are you a man younger than 55 years or woman younger than 65 years		0	
Smoking. Pick the description which matches you best:	I never smoked		0	**Points:**
	OR I am a former smoker (last smoked more than 12 months ago)		2	
	OR I am a current smoker or I smoked regularly in the last 12 months, and I smoke…	1-5 cigarettes per day	2	
		6-10 cigarettes per day	4	
		11-15 cigarettes per day	6	
		16-20 cigarettes per day	7	
		More than 20 cigarettes per day	11	
Second hand smoke	Over the past 12 months, what has been your typical exposure to other people's tobacco smoke?	Less than 1 hour or exposure per week or no exposure	0	**Points:**
		OR One or more hours of secondhand smoke exposure per week	2	
Diabetes	Do you have diabetes mellitus?	Yes	6	**Points:**
		No or unsure	0	
High Blood Pressure	Do you have high blood pressure	Yes	5	**Points:**
		No or unsure	0	
Family history	Have either or both of your biological parents had a heart attack?	Yes	4	**Points:**
		No or unsure	0	
Waist to hip ratio	Pick one only:	Quartile 1: Less than 0.873	0	**Points:**
		Quartile 2 &3: 0.873 - 0.963	2	
		Quartile 4: greater than or =0.964	4	
Psychosocial factors	How often have you felt work or home life stress in the last year? Pick one only	Never or some periods	0	**Points:**
		OR Several periods of stress or permanent stress	3	

	During the past 12 months, was there ever a time when you felt sad, blue, or depressed for two weeks or more in a row?	Yes	3	**Points:**
		No	0	
Dietary factors. Pick one answer for each food group mentioned	Do you eat salty food or snacks one or more times a day	Yes	1	**Points:**
		No	0	
	Do you eat deep fried foods or snacks or fast foods 3 or more times a week?	Yes	1	**Points:**
		No	0	
	Do you eat fruit one or more times daily?	Yes	0	**Points:**
		No	1	
	Do you eat vegetables one or more times daily?	Yes	0	**Points:**
		No	1	
	Do you eat meat and/ or poultry 2 or more times daily?	Yes	2	**Points:**
		No	0	
Physical activity	How active are you during your leisure time?	I am mainly sedentary or perform mild exercise (requiring minimal effort)	2	**Points:**
		OR I perform moderate or strenuous physical activity in my leisure time	0	

* In this table, the categories of the risk factors are presented in the first column, and the specific questions to be asked in the middle columns. Only one answer is chosen for every question, and inserted into the "points" column. All questions must be answered for the most accurate risk score estimate. The IHRS for an individual is calculated by adding the points from each of the above items.

Table S4: Validation of the Three Versions of the INTERHEART Risk Score

	"Short" INTERHEART modifiable risk score	**"Full" INTERHEART modifiable risk score**	**"Cholesterol" risk score**	**"Non laboratory-based" INTERHEART modifiable risk score**
	Validation studies within the 2/3 derivation set			
Odds increase of MI for a 1-point increase in score (95% CI)	15.2% (14.4%, 16.1%)	14.3% (13.5%, 15.1%)	12.9% (12.2%, 13.7%)	14.4% (13.9%, 15.3%)
ROC c statistic (95% CI)	0.708 (0.700, 0.716)	0.721 (0.712, 0.729)	0.693 (0.683, 0.702)	0.710 (0.701, 0.719)
Brier score	0.213	0.209	0.217	0.212
	Validation studies within the 1/3 test set			
Odds increase of MI for a 1-point increase in score (95% CI)	15.8% (14.7%, 17.0%)	14.0% (13.0%, 15.1%)	13.0% (12.0%, 14.3%)	14.2% (13.1%, 15.3%)
ROC c statistic (95% CI)	0.713 (0.701, 0.726)	0.712 (0.708, 0.732)	0.694 (0.681, 0.707)	0.710 (0.697, 0.722)
Brier score	0.212	0.210	0.217	0.213

Table S5: Key Characteristics of Eligible vs Enrolled Individuals by Country Income Level

	Overall		HIC		MIC		LIC	
	Eligible	**Enrolled**	**Eligible**	**Enrolled**	**Eligible**	**Enrolled**	**Eligible**	**Enrolled**
N	200905	155713	20334	17116	136525	114207	44046	24390
Mean age (years)	50.20	50.67	52.3	52.6	50.5	50.8	48.3	48.7
% Females	52.9	56.0	52.9	53.9	53.1	56.2	52.1	56.7
% Current Smokers	22.6	21.2	15.0	14.2	23.6	22.0	23.3	22.5
% Low education	41.6	42.1	11.4	11.3	43.7	44.6	52.5	51.8
% Hypertension	13.7	15.2	17.4	18.1	13.8	14.8	11.7	15.2
% Diabetes	5.4	5.6	7.2	7.3	4.2	4.5	8.1	9.8
% Stroke	1.2	1.4	1.2	1.2	1.0	1.0	2.1	3.1
% CHD	3.6	4.1	4.5	4.7	3.7	3.9	3.1	4.3
% Cancer	1.3	1.2	3.7	3.9	0.5	0.5	2.7	2.6

Due to a mismatch in the identification numbers between the Adult and Family census, we were not able to identify the status of 711 (0.45%) individuals who were considered to be eligible.

Table S6: Comparison of All Participants Enrolled and All Participants with Follow-Up by IHRS and Educational Level

	IHRS(mean and SD)		Higher Education (%)	
	All enrolled (156,424, 100%)	With Follow Up (n=152,606, 97.56%)	All enrolled (156,424, 100%)	With Follow Up (n=152,606, 97.56%)
HIC	12.89(6.2)	12.89(6.2)	59.48	59.48
MIC	10.47(5.8)	10.53(5.8)	14.67	14.79
LIC	8.28(5.1)	8.29(5.1)	13.48	13.34
Urban	10.73(5.9)	10.78(5.9)	30.66	30.86
Rural	9.67(5.6)	9.71(5.6)	6.19	6.26
Overall	10.23(5.8)	10.27(5.8)	19.0	19.19

Higher Education defined as Trade or College or University Education.

Of those who had follow-up (N=152,606), in 143 (0.1%) age, sex or date of the questionnaire was missing. Age and Sex adjusted analyses do not include these individuals. Therefore age and sex adjusted analyses are based on 152,463 individual.

Table S7: Standardized Event Rates Per 1000 Person-Years of Follow-Up for Death and CVD

*Major CVD = Fatal CVD + MI + Stroke + Heart Failure. MI, strokes and heart failure include fatal and nonfatal events
‡Total CVD = MI + Stroke + Heart Failure + other hospitalized

		Death		Myocardial Infarction		Stroke		Heart Failure		Major CVD*		Hospitalization for other CVD		Total CVD‡	
Country Economic Levels	**Total Number Followed**	**No.**	**Rate**	**No.**	**Rate**	**No.**	**Rate**	**No.**	**Rate**	**No.**	**Rate**	**No.**	**Rate**	**No.**	**Rate**
Total	152463	3900	6.15	1736	2.73	1317	2.07	414	0.65	3483	5.50	1163	1.83	4646	7.36
HIC	16114	216	2.43	172	1.92	129	1.49	53	0.59	351	3.99	317	3.69	668	7.76
MIC	101921	2052	5.59	816	2.21	881	2.37	277	0.75	1984	5.38	645	1.72	2629	7.14
LIC	34428	1632	9.23	748	4.13	307	1.72	84	0.48	1148	6.43	201	1.10	1349	7.55
P for trend			<.0001		<.0001		0.82		0.06		<.0001		<.0001		0.58

Table S8: Ten Most Frequent Causes of Hospitalization for Non-Major CVD Per 1000 Participants

	Overall N=152,463		HIC N=16,114		MIC N=101,921		LIC N=34,428	
Medical Term	N	Per 1000	N	Per 1000	N	Per 1000	N	Per 1000
Angina	474	3.11	123	7.63	281	2.76	70	2.03
Hypertension	246	1.61	28	1.74	152	1.49	66	1.92
Atrial fibrillation or flutter	105	0.69	74	4.59	31	0.30	0	0.00
Chronic ischemic heart disease	55	0.36	5	0.31	29	0.28	21	0.61
Thromboplebitis	54	0.35	20	1.24	33	0.32	1	0.03
Other cardiac arrhythmias	39	0.26	9	0.56	30	0.29	0	0.00
Pulmonary embolism	31	0.20	26	1.61	5	0.05	0	0.00
Hypotension	27	0.18	4	0.25	9	0.09	14	0.41
Varicose veins of lower extremities	21	0.14	4	0.25	12	0.12	5	0.15
Other peripheral vascular diseases	15	0.10	4	0.25	9	0.09	2	0.06
Other causes	135	0.89	40	2.48	62	0.61	29	0.84
Total No.	1202	7.88	337	20.91	653	6.41	208	6.04

HIC=High Income Countries, MIC=Middle Income Countries and LIC=Low Income Countries

Table S9: Numbers of Deaths by Condition and Case Fatality Rates Per 100 Participants

	MI			Stroke			Heart Failure			Major CVD			All CVD		
Country Economic Levels	**No. Events**	**No. Fatal**	**Yearly Fatality Rate**	**No. Events**	**No. Fatal**	**Yearly Fatality Rate**	**No. Events**	**No. Fatal**	**Yearly Fatality Rate**	**No. Events**	**No. Fatal**	**Yearly Fatality Rate**	**No. Events**	**No. Fatal**	**Yearly Fatality Rate**
	N	**N**	**%**	**N**	**N**	**%**	**N**	**N**	**%**	**N**	**N**	**%**	**N**	**N**	**%**
Total	1736	846	16.53	1317	395	11.34	414	144	13.42	3483	1499	15.60	4646	1499	11.90
HIC	172	24	5.12	129	12	3.99	53	9	6.19	351	56	6.50	668	56	3.28
MIC	816	366	18.37	881	242	10.84	277	80	12.06	1984	762	15.86	2629	762	12.35
LIC	748	456	17.17	307	141	14.82	84	55	20.79	1148	681	17.28	1349	681	14.43
P for trend			0.20			<0.001			<0.001			0.01			<0.001

Table S10: Age and Gender Standardized Event Rates per 1000 Person-Years by Country Economic Category and by Urban Versus Rural Location

		Major CVD		Other CVD		Fatal CVD		All CVD		Death	
	Nos. Followed	N	Rate	N	Rate	N	Rate	N	Rate	N	Rate
Location											
Urban	79945	1609	4.83	712	2.12	575	1.71	2321	6.99	1498	4.48
Rural	72518	1874	6.25	451	1.50	924	3.09	2325	7.79	2402	8.01
P for difference			<0.001		<0.001		<0.001		<0.001		<0.001
Income Region and Location											
HIC – Urban	11583	260	4.14	237	3.80	43	0.67	497	8.03	148	2.35
HIC - Rural	4531	91	3.63	80	3.46	13	0.48	171	7.17	68	2.64
P for difference			0.27		0.32		0.37		0.13		0.20
MIC-Urban	52401	900	4.77	358	1.87	283	1.50	1258	6.67	808	4.28
MIC-Rural	49520	1084	6.09	287	1.59	479	2.72	1371	7.70	1244	7.01
P for difference			<0.001		0.05		<0.001		<0.001		<0.001
LIC-Urban	15961	449	5.55	117	1.37	249	3.16	566	6.95	542	6.81
LIC-Rural	18467	699	7.21	84	0.86	432	4.49	783	8.08	1090	11.30
P for difference			<0.001		<0.001		<0.001		0.002		<0.001

Major CVD = Fatal CVD + MI + Stroke + Heart Failure
Other CVD = Hospitalizations for CVD reasons other than MI, stroke or heart failure
Fatal CVD = Fatal MI + Fatal Stroke + Fatal Heart Failure + other fatal CVD
All CVD = Fatal CVD + hospitalizations for CVD
Based on 152,144 individuals

Table S11: Case Fatality Rates for Major CVD

	Urban			Rural			P value
	No. Events	No. Fatal	Yearly Fatality Rate	No. Events	No. Fatal	Yearly Fatality Rate	

HIC	260	43	6.72	91	13	5.86	0.66
MIC	900	283	13.41	1084	479	17.78	<0.001
LIC	449	249	16.58	699	432	17.71	0.85
Overall	1609	575	13.52	1874	924	17.25	<0.001

References

1. Teo K, Chow CK, Vaz M, Rangarajan S, Yusuf S, Group PI-W. The Prospective Urban Rural Epidemiology (PURE) study: examining the impact of societal influences on chronic noncommunicable diseases in low-, middle-, and high-income countries. *Am Heart J* 2009;158(1):1-7 e1.

2. Yusuf S, Islam S, Chow CK, Rangarajan S, Dagenais G, Diaz R, et al. Use of secondary prevention drugs for cardiovascular disease in the community in high-income, middle-income, and low-income countries (the PURE Study): a prospective epidemiological survey. *Lancet* 2011;378(9798):1231-43.

3. MacQueen KM, McLellan E, Metzger DS, et al. What is community? An evidence-based definition for participatory public health. *Am J Public Health* 2001;91:1929-38.

4. Tunstall-Pedoe H, Vanuzzo D, Hobbs M, Mahonen M, Cepaitis Z, Kuulasmaa K, et al. Estimation of contribution of changes in coronary care to improving survival, event rates, and coronary heart disease mortality across the WHO MONICA Project populations. *Lancet* 2000;355(9205):688-700.

5. Corsi DJ, Subramanian SV, Chow CK, McKee M, Chifamba J, Dagenais G, et al. Prospective Urban Rural Epidemiology (PURE) study: Baseline characteristics of the household sample and comparative analyses with national data in 17 countries. *Am Heart J* 2013;166(4):636-46 e4.

6. Yusuf S, Hawken S, Ounpuu S, Dans T, Avezum A, Lanas F, et al. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *Lancet* 2004;364(9438):937-52.

7. O'Donnell MJ, Xavier D, Liu L, Zhang H, Chin SL, Rao-Melacini P, et al. Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study. *Lancet* 2010;376(9735):112-23.

8. Gajalakshmi V, Peto R. Verbal autopsy of 80,000 adult deaths in Tamilnadu, South India. *BMC Public Health* 2004;4:47.

9. Jha P, Gajalakshmi V, Gupta PC, Kumar R, Mony P, Dhingra N, et al. Prospective study of one million deaths in India: rationale, design, and validation results. *PLoS Med* 2006;3(2):e18.

10. McGorrian C, Yusuf S, Islam S, Jung H, Rangarajan S, Avezum A, et al. Estimating modifiable coronary heart disease risk in multiple regions of the world: the INTERHEART Modifiable Risk Score. *Eur Heart J* 2011;32(5):581-9.

11. Gaziano TA, Young CR, Fitzmaurice G, Atwood S, Gaziano JM. Laboratory-based versus non-laboratory-based method for assessment of cardiovascular disease risk: the NHANES I Follow-up Study cohort. *Lancet* 2008;371(9616):923-31.

12. Sullivan LM, Massaro JM, D'Agostino RB Sr. Presentation of multivariate data for clinical use: The Framingham Study risk score functions. *Stat Med* 2004;23(10):1631-60.

13. Bosch J, Yusuf S, Gerstein HC, Pogue J, Sheridan P, Dagenais G, Diaz R, Avezum A, Lanas F, Probstfield J, Fodor G, Holman RR. Effect of ramipril on the incidence of diabetes. *N Engl J Med* 2006; 355(15): 1551-62.