Baker IDI Research Online
http://library.bakeridi.edu.au

This is the preprint version of the work. It is the manuscript that
was submitted to the journal and has not undergone peer review.

Mannan HR, Stevenson CE, Peeters A, McNeil JJ. A new set of risk equations for predicting long term risk of all-cause mortality using cardiovascular risk factors. Prev Med 2013;56(1):41-5.

http://hdl.handle.net/11187/1545

Copyright © Elsevier. This file is for personal use. Further distribution is not permitted.

A new set of risk equations for predicting long term risk of all-cause mortality using cardiovascular risk factors

**Haider R Mannan*

email: haider.mannan@monash.edu

Christopher E Stevenson

email: christopher.stevenson@deakin.edu.au

Anna Peeters

email: anna.peeters@bakeridi.edu.au

John J McNeil

email: john.mcneil@monash.edu

**Corresponding author:*

The first and last authors belong to: Department of Epidemiology & Preventive Medicine, Alfred Centre, Monash University, Melbourne, Victoria 3004, Australia. Co-author CE Stevenson belongs to Department of Health & Social Development, Deakin University, Burwood campus, Melbourne, Victoria 3125, Australia and co-author Anna Peeters belongs to Baker IDI, Melbourne, Victoria 3004, Australia.

Abstract:

Objectives: As population ages and treatment for cardiovascular disease improves the risk of all-cause mortality has become a more meaningful outcome. We develop all-cause mortality equations for predicting long term risk using cardiovascular risk factors.

Methods: The 24-year risk of all-cause mortality was evaluated using Cox model for participants aged 40-81 years at the 10th or 11th examination of the Framingham original cohort and the first examination of the offspring cohort-all of whom were free of major chronic diseases.

Results: The predictors of all-cause mortality were age, sex, systolic blood pressure, total cholesterol/HDL ratio and smoking status. Risk prediction improved significantly when intensity of smoking and time since quitting were included into smoking status. A reduced model based on non-laboratory risk factors also demonstrated good predictive performance. Our models had considerably higher discrimination than both European and Framingham cardiovascular risk scores which have already been used for predicting all-cause mortality.

Conclusions: All-cause mortality risk equations incorporating cardiovascular risk factors provide an improved tool to quantify risk and guide prevention of mortality. There are great potentials for prevention of the CVD epidemic and for increased longevity with health, through improved life-styles and consequent lower levels of blood pressure, cholesterol and smoking.

Key Words: All-cause mortality, predictive equation, detailed smoking measures, other modifiable risk factors, reduced equations.

Note: This study was conducted at *Department of Epidemiology & Preventive Medicine, Alfred Centre, Monash University, Melbourne, Victoria 3004, Australia during mid 2011*.

Introduction: Follow-up data from population studies have aided in developing a large number of risk equations in developed countries for major chronic diseases, based on health risk factors. However, there have only been a few studies (Kannel et al. 1986; Menotti et al. 2001; Janssen et al. 2005; Aktas et al. 2004) which have developed all-cause mortality risk equations based on these risk factors. But all-cause mortality is a more ultimate indicator of health than the occurrence of single diseases. Furthermore, predicting disease specific mortality is likely to be less accurate for older people as the allocation of causes of death becomes less accurate as age and the number of comorbidities increase.

Previous studies have found that risk equations can be developed for all-cause mortality based on cardiovascular disease (CVD) risk factors (Kannel et al. 1986; Menotti et al. 2001; Janssen et al. 2005; Aktas et al. 2004). This is because CVD is the leading cause of death worldwide (WHO) and many CVD risk factors increase the risk of death from other major causes like cancer and diabetes. However, these equations have limitations either in terms of age range of the population studied or use of CVD risk factors. Further, increasing life expectancy of the population in middle and high income countries suggests the need for an all-cause mortality equation with a longer prediction period which would allow prediction of mortality risk at older ages from risk behaviour in middle ages.

The aim of this paper is to develop new risk equations for all-cause mortality using detailed measures for major CVD risk factors which will allow the prediction of longer term (24 year) mortality risk for a population cohort initially aged 40-81. As a sub-aim, we will investigate the rate of reduction in risk following smoking cessation and when, or even whether, the risk for quitters reaches that of never smokers. To investigate the dose effects of smoking we will investigate how the excess risk associated with smoking increases with intensity and duration. Finally, we will develop a reduced equation based on non-laboratory risk factors as these are more readily available to primary care physicians than laboratory risk factors.

Materials and Methods

Data

We used data from Framingham Heart Study (FHS) established in 1948 that followed a cohort of 5,209 adults from Framingham, Massachusetts, to examine the relationship between health risk factors and subsequent CVD. A further 5,124 people were added to the cohort in 1971 (the 'offspring cohort'). The details regarding the design, selection criteria and examination procedures of FHS have been elaborated elsewhere (D'Agostino et al. 2008; Wolf et al. 1991; Murabito et al. 1997; Kannel et al. 1999; Doll et al. 2004).

The outcome of the study was all-cause mortality. The study cohort consisted of people from examination 1 (1971-1975) of the offspring cohort and from examinations 10 and 11 (1968-1971) of the original cohort for whom HDL was measured for the first time. For the original cohort, in most cases (81.3%), HDL was measured for the first time at examination 11, while for some cohort members, it was examination 10. Follow-up was performed through the 22nd examination cycle, a span of approximately 24 years. For the offspring cohort, risk factor measurements were from the first examination cycle (1971-1975), whereas follow-up was performed through the sixth examination cycle, approximately 24 years later. Participants were considered eligible if at the baseline they were aged 40-81 years, were free of CVD, cancer and chronic kidney disease and had complete information on covariates. After exclusions, the study included 3,243 persons (1473 women and 1770 men).

Selection of risk factors

For assessing the overall risk associated with hypertension we considered both systolic and diastolic blood pressures (SBP and DBP) because both sometimes correlate with major chronic diseases (D‘Agostino et al 2008; Levy and Kannel 1988). In case of cholesterol we considered total and HDL cholesterol or the total cholesterol/HDL ratio and LDL because total cholesterol only provides a part of the picture (Levy and Kannel 1988). The other risk factors examined were age, sex, cigarette smoking, diabetes status, triglycerides and body mass index. To incorporate current intensity and duration into smoking status, a five and four category smoking status variable were created with categories never smokers, quitters and current smokers with ≤15, >15 to ≤30 and more than 30 cigarettes being smoked and categories never smokers, quitters and current smokers with ≤30 years and >30 years durations. To include time since quitting into smoking status, the quitters were split into years since quitting ≤5 years and >5 years respectively. All these categorizations yielded best risk prediction.

The Risk Prediction Models

First, we fitted Model1 by including a current/past/never smoker smoking variable and age, SBP and total cholesterol/HDL ratio in logarithmic scale as predictors. We also fitted three more models containing the same risk factors but with different measures of smoking:

Model 2: incorporates intensity of smoking among current smokers and time since quitting among past smokers in the smoking status variable

Model 3: incorporates duration of smoking among current smokers and time since quitting among past smokers in the smoking status variable

Model 4: incorporates both intensity and duration of smoking.

Simpler Risk Prediction Models Using Non-laboratory Predictors Routinely Ascertained in Primary Care

We developed a simplified model based on the one that gave the best predictive performance of the four models described above. This model used only simple office-based predictors that are routinely obtained in primary care and do not require laboratory testing. These variables included age, sex, systolic blood pressure and current smoking.

Measurement of risk factors

The examination blood pressure was calculated as the average of 2 physician-obtained measures. Standardized enzymatic methods were used to determine total and HDL cholesterol levels. In the original Framingham cohort, diabetes was diagnosed if a casual whole blood glucose measurement was ≥150 mg/dl or the individual was being treated with insulin or oral hypoglycemics. In the offspring Framingham cohort, a more recent definition, requiring treatment or a fasting glucose level of ≥140 mg/dl from plasma measurement, was used. The various measures of cigarette smoking were obtained by self-report. Details are provided in Appendix.

Development and assessment of predictive models

We used Cox's proportional-hazards model (Cox 1972) to relate risk factors to risk of all-cause mortality during a 24 year follow up. All continuous covariates were logarithmically transformed to improve model discrimination and calibration and to minimize the influence of outliers. For a nominal covariate the proportional hazards was tested by examining whether there is a non-significant interaction between that covariate and log of survival time (Hosmer and Lemeshow 1999) and for a continuous covariate plotting Schoenfeld residuals against log(survival time). The formula for estimating absolute risk of all-cause mortality is given in the Appendix.

The predictors of all-cause mortality were age, sex, SBP, total cholesterol/HDL ratio and smoking status. The addition of quadratic terms for age and total cholesterol/HDL ratio produced significantly improved model fit.

The discriminative and calibration abilities of a model were assessed by Harrell's C (Harrell et al. 1996; D'Agostino and Nam 2004) and Hosmer-Lemeshow statistic and its modification (Harrell Jr et al. 1996) while internal validation was assessed by a bootstrapping method. We found negligible overoptimism in all C-statistic estimates with the maximum being 0.0003. Finally, the C-statistic was corrected by subtracting the degree of overoptimism from the original C-statistic. We used a nonparametric test of difference in two correlated C-statistics to assess the improvement in discrimination between two nested models (Antolini et al. 2004). One limitation of this test is that for models having reasonably good discrimination, even significant "independent" association of the new covariate with the outcome may not result in a meaningfully larger C-statistic (Pepe et al. 2004; Greenland and O'Malley 2005; Ware 2006). To overcome this we used category-free Net Reclassification Improvement using 95% bootstrap percentile interval (Pencina et al. 2011) and Integrated Discrimination Improvement (Pencina et al. 2008). We computed conditional likelihood ratio tests to evaluate the global fit of models and AIC to rank their predictive ability adjusting for model complexity.

Results

Table 1 shows the risk factor characteristics in our sample at the baseline examination. There are 44.9% males, 2.1% diabetics, 28.2% quitters and 37.2% current smokers. The average values for age, SBP and total cholesterol/HDL ratio are 52.9, 132.2 and 4.7 respectively.

Table 2a shows the likelihood ratio statistics, Hosmer Lemeshow and its modification and AIC for these models. All models have both good discrimination and calibration.

Table 2b presents the conditional likelihood ratio statistics, NRI and IDI comparing these models, along with differences in their C-statistics and AIC for each model. While model 1 has the best calibration, model 2 is the preferred model as it has the best discrimination which is significantly higher than other models. The AIC values demonstrate that model 2 (see Table 2c for parameters estimates) has the best overall fit and parsimony. The risk associated with quitting declined with time since quitting and almost reached that of never smokers after longer than 5 years since quitting.

Simpler risk prediction equation using non-laboratory risk factors

The reduced non-laboratory risk factor-based equation (Model 5; see Table 2d for parameter estimates) performed reasonably well in terms of both discrimination and calibration (Table 2a). So, this would be the preferred risk equation when quick assessment of all-cause mortality risk of an individual is needed. However, the corresponding full model (Model 2) predicts significantly better (Table 2b).

Sensitivity Analyses

A three-category categorization of smoking duration done to examine whether this improved risk prediction revealed no major improvement. Our risk equations were largely insensitive to exclusion of deaths both in the first 5 and 10 years of follow-up with no noticeable change in the significance level of any of the predictors. A sub-analysis excluding persons aged ≥75 years to assess possible differences in risk factors in this older group and its potentially large influence in the algorithm determination did not reveal any noticeable change in the significance level of any of the predictors.

Point Score Algorithm

To provide a tool easy to use in clinical practice, score sheets for prediction of absolute 24-year risk of all-cause mortality were developed from Model 2. Appendix shows average risk values, by age, sex, systolic blood pressure, total cholesterol/HDL ratio and smoking status for the Framingham population. For example, the risk is 99.29% for a 57 year old man who is currently smoking more than 30 cigarettes per day, has SBP 167.0 mmHg and total cholesterol/HDL ratio 9.9. This person may require to be referred for an aggressive management of risk factors, but for less extreme situations, the main question would be to settle the threshold to define high risk individuals. If high cardiovascular risk has been assessed as a 24-year probability of fatal CVD event as 10%, and given that CVD mortality represents about 40% of all-cause mortality in Framingham people aged 40-81, we propose to consider that subjects with a 24-year risk of all-cause mortality of 25%, according to our algorithm, are at high risk (i.e. slightly below 30%, to take into account that our equation may underestimate probabilities of death).

Discussion

This study demonstrated that the major CVD risk factors can successfully predict long-term risk of all-cause mortality. The prediction algorithm has good internal validation and is a simple tool for daily clinical practice, requiring solely a clinical examination and a blood test to estimate 24-year absolute risk of mortality for a given subject. A reduced non-laboratory risk factor-based equation was also provided.

Risk prediction improved significantly when smoking intensity and time since quitting were included into smoking status. Some smoking related studies have highlighted that for lung cancer, smoking duration is more important than intensity in causing death (Flanders et al. 2003), however, for CVD which is the leading cause of death in our study cohort, this is unknown. We further analyzed our sample and found that smoking duration did not predict CVD mortality but predicted lung cancer mortality independent of intensity (data not given). Diabetes was not found as a predictor probably because of it having low prevalence in our cohort.

The hypothesis that cardiovascular risk factors are also major risk factors for all-cause mortality, a previous Framingham cohort (Hoes et al. 1999) pointed out that age, sex, smoking and SBP are long-term predictors of all-cause mortality in middle aged subjects. These risk factors except for SBP also increase the risk of cancer mortality which is the second most common cause of death in our study cohort.

The advantages of our equations over existing equations are the inclusion of a wider age range and more detailed measures for smoking and cholesterol. The use of total cholesterol/HDL ratio over total cholesterol and HDL is advantageous in populations where the latter variables are highly correlated. We found dose and temporal effects of smoking intensity and cessation. The C-statistic of our risk models ranged from 0.8443 to 0.8470 which is higher than the C-statistic of 0.57 for the Framingham Risk Score (D‘Agostino Sr et al. 2001) based on age, sex, smoking status, total cholesterol, HDL cholesterol, SBP and diabetes or the C-statistic of 0.73 for the European SCORE (Conroy et al. 2003) based on age, sex, total cholesterol, SBP and smoking status, for predicting all-cause mortality (Aktas et al. 2004). Risk equations developed from Framingham for CVD have been found to recalibrate well to other similar populations (Doll et al. 2004, NCEP 2001; Anderson et al. 1991; Kannel et al. 1992). However, the Framingham sample is mostly Caucasian and so the generalizability of our equations in other ethnic groups is uncertain.

Previously developed all-cause mortality risk equations based on CVD risk factors include the MRFIT equation (Kannel et al. 1986) restricted to males with a maximum age of 57 years examining total cholesterol, DBP and smoking intensity as risk factors. However, SBP should not be ignored as it is a more reliable measure of blood pressure than DBP. A similar study (Menotti et al., 2001) conducting 25 year prediction of all-cause mortality was restricted to males aged 40-59 years, excluded HDL cholesterol and included prevalence or intensity of smoking. The cohort sizes and number of events for these studies were considerably smaller than ours. The discriminatory ability of our models cannot be compared with any of these models because these did not report C-statistic.

Conclusion

Algorithms to estimate probability of all-cause mortality are of major importance because these help to screen those people who will benefit from prevention of major chronic diseases and subsequently have higher longevity, and consequently guide practitioners towards a proper management of risk factors. These risk algorithms therefore are key for developing cost-effective disease prevention as they assist in screening only those people who are at high risk of mortality rather than screening the entire population. Thus, assessing all-cause mortality enables a more global approach of health prevention and a better identification of priorities for health policies. The tool we developed will be complementary to cardiovascular algorithms we published previously (Mannan et al. 2010; Mannan et al. 2011).

Acknowledgements

The Framingham Heart Study - Offspring (FHS-O) is conducted and supported by the NHLBI in collaboration with the FHS-O Study Investigators. This Manuscript was prepared using a limited access dataset obtained from the NHLBI and does not necessarily reflect the opinions or views of the FHSO or the NHLBI. This research was supported by an NHMRC health services research grant (no. 465130) and a Vichealth Senior Fellowship. We also thank A/Prof Michael Pencina of Boston University, USA to provide with SAS Macros for estimating category-free NRI and its associated bootstrap confidence intervals.

Contributorship

HRM was involved in all stages of this research as the principal author. CES, AP and JJM read the draft of the paper and provided useful suggestions. JJM is the Principal Investigator of the grant which enabled this research to be carried out. All authors read and approved the final manuscript.

Data Sharing Statement

There is no data to share.

Competing interests

The authors declare that they have no competing interests.

Submission declaration

The work described has not been published previously and is not under consideration for publication elsewhere and its publication is approved by all authors. If accepted, it will not be published elsewhere including electronically in the same form, in English or in any other language, without the written consent of the copyright-holder.

References

Aktas MK, Ozduran V, Pothier CE, Lang R, Lauer MS, 2004. Global Risk Scores and Exercise Testing for Predicting All-Cause Mortality in a Preventive Medicine Program. JAMA 292(12): 1462-1468.

Anderson KM, Odell PM, Wilson PW and Kannel WB, 1991. Cardiovascular disease risk profiles. Am Heart J 121:293-8.

Antolini L, Nam BH, Agostino RB, 2004. Inference on correlated discrimination measures in survival analysis: a nonparametric approach. Commun Stat Theory Methods 33:2117-35.

Conroy RM, Pyo¨ra¨la¨ K, Fitzgerald AP, Sans S, Menottid A, De Backere G, De Bacquere D, Ducimetièref P, Jousilahtig P et al, 2003. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. Eur Heart J 24:987–1003.

Cox DR, 1972. Regression models and life tables. J Royal Stat Soc 34(series B):187-220.

D'Agostino RB Sr, Grundy S, Sullivan LM and Wilson P, 2001. Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. JAMA 286:180-187.

D'Agostino RB Sr.,Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, Kannel WB, 2008. General Cardiovascular Risk Profile for Use in Primary Care. Circulation 117:743-53.

D'Agostino R, Nam BH, 2004. Evaluation of the performance of survival analysis models: discrimination and calibration measures. In: Handbook of Statistics. Elsevier, Amsterdam:1-25.

Doll R, Peto R, Boreham J and Sutherland I, 2004. Mortality in relation to smoking: 50 year's observations on male British doctors. BMJ 328:1519.

Executive summary of the Third Report of The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III), 2001. JAMA 285:2486-97.

Flanders WD, Lally CA, Zhu BP, Henley SJ, Thun MJ, 2003. Lung cancer mortality in relation to age, duration of smoking, and daily cigarette consumption: results from Cancer Prevention Study II. Cancer Research 63(19):6556-62.

Greenland P, O'Malley PG, 2005. When is a new prediction marker useful? A consideration of lipoprotein-associated phospholipase A2 and C-reactive protein for stroke risk. Arch Intern Med 165:2454-6.

Harrell FE Jr, Lee KL, Mark DB, 1996. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 15:361-87.

Hoes AW, Grobbee DE, Valkenburg HA, Lubsen J and Hofman A, 1999. Cardiovascular risk and all-cause mortality: A 12 year follow-up study in the Netherlands. Eur J Epidemiol 9:285-92.
Hosmer DW Jr, Lemeshow S, 1999. Applied Survival Analysis: Regression Modeling of Time to Event Data. Wiley, New York.
Janssen I, Katzmarzyk PT, Church TS, Blair SN, 2005. The Cooper Clinic Mortality Risk Index Clinical Score Sheet for Men. Am J Prev Med 29(3):194–203.
Kannel WB, D'Agostino RB, Silbershatz H, Belanger AJ, Wilson PW, Levy D, 1999.
Profile for estimating risk of heart failure. Arch Intern Med 159:1197- 204.
Kannel WB, Neaton JD, Wentworth D, Thomas HE, Stamler J, Hulley SB, Kjelsberg MO, 1986. Overall and coronary heart disease mortality rates in relation to major risk factors in 325,348 men screened for the MRFIT. Am Heart J 112(4):825-36.
Levy D, Kannel WB, 1988. Cardiovascular risks: New insights from Framingham. Am Heart J 116(1),Part 2:266-272.
Mannan H, Stevenson C, Peeters A and McNeil JJ, 2010. Framingham risk prediction equations for CVD incidence using detailed measures for smoking. Heart International 5(2):49-57; doi:10.4081/hi.2010.e11.
Mannan HR, Stevenson CE, Peeters A and McNeil JJ, 2011. Age at quitting smoking as a predictor of risk of cardiovascular disease incidence independent of smoking status, time since quitting and pack-years. BMC Research Notes 4:39.
Menotti A, Blackburn H, Kromhout D, Nissinen A, Adachi H, Lanti M, 2001. Cardiovascular risk factors as determinants of 25-year all-cause mortality in the seven countries study. European J Epidemiol 17:337-346.
Murabito JM, D'Agostino RB, Silbershatz H and Willson PWF, 1997. Intermittent claudication: a risk profile from the Framingham Heart Study. Circulation 96:44-9.
Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr and Vasan RS, 2008. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med 27: 157-72.

Pencina MJ, D'Agostino Sr RB, Steyerberg EW, 2011. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. Stat Med 30:11–21.
Pepe MS, Janes H, Longton G, Leisenring W and Newcomb P, 2004. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. Am J Epidemiol 159:882-90.
Ware JH, 2006. The limitations of risk factors as prognostic tools. New Eng J Med 355:2615-7.
WHO: The top 10 causes of death. URL: http://www.who.int/mediacentre/factsheets/fs310/en/index.html.
Wolf PA, D'Agostino RB, Belanger AJ and Kannel WB, 1991. Probability of stroke: a risk profile from the Framingham study. Stroke 22:312-8.