
**Abstract**

BACKGROUND: Logistic regression and, more recently, Cox regression have been the predominant methods for identifying risk factors and developing risk estimation equations for coronary heart disease (CHD). Software for the regression tree method is now available for binary and survival outcomes and thus offers an alternative methodology. This paper compares these four methods for identifying significant risk factors from among a set of candidate factors and for estimating the risk of death from CHD using baseline and mortality follow-up data on 1,701 men participating in the Busselton Health Study. The candidate risk factors were age, body mass index, systolic and diastolic blood pressure, treatment for hypertension, cholesterol and smoking.

METHODS: Logistic regression, Cox proportional hazards regression, binary regression tree, and survival regression tree analyses have been applied to data obtained from the same cohort of men for CHD death risk estimation and prediction. The four methods are compared in terms of the variables selected, goodness-of-fit of models, similarity of cross-validated estimated risks for individuals, and ability to discriminate between those who died from CHD and those who did not die from CHD during the follow-up period, including the comparison of Receiver Operating Characteristic (ROC) curves.

RESULTS: Although age and a blood pressure variable were selected by all four methods, body mass index was also selected by the regression tree methods and smoking was also selected by Cox regression. There was good, but not excellent, agreement between methods in estimates of risk for individuals, the areas under the ROC curves were 0.66 for the binary tree, 0.72 for logistic regression, 0.71 for the survival tree method and 0.78 for Cox regression. The average differences in estimated risk between those who died from CHD and those who did not die from CHD during the follow-up period were 0.051 for logistic regression, 0.070 for the binary tree method, 0.073 for the survival tree method and 0.088 for Cox regression.

CONCLUSION: For a moderately sized cohort typical of many applications of these methods in the literature, the two methods which used the survival outcome performed better than the methods using a binary outcome. Despite selecting some different variables and showing moderate differences in risk estimates for individuals, the two binary approaches were similar in performance. Cox regression appeared to be superior to the survival tree method, but further larger studies of completely separate samples for model development and evaluation of prediction performance are required to confirm this finding.