Development of a model to predict the likelihood of a genetic variant causing familial hypercholesterolaemia

Rachel O'Leary[1,2], Samuel G. Urwin[2,3], Clare Lendrem[2,3], Ahai Luvai[4], R. Dermot G. Neely[2], A. Joy Allen[2]

[1]Northern Medical Physics and Clinical Engineering, The Newcastle-upon-Tyne Hospitals NHS Foundation Trust, UK.
[2]The National Institute for Health Research Newcastle In Vitro Diagnostics Co-operative, Newcastle-upon-Tyne, UK.
[3]Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, UK.
[4]Laboratory Medicine, The Newcastle-upon-Tyne Hospitals NHS Foundation Trust, UK.

Background: Familial hypercholesterolaemia (FH) is a common, life-threatening genetic condition associated with long-term elevation of cholesterol levels in the blood. A diagnosis of FH can be confirmed by genetic testing; however, it is expensive, and can often be mistargeted due to limitations of the scoring systems used to refer patients.
Aim: To develop a model using clinical data to improve the targeting of genetic testing by predicting the likelihood of a patient having a variant causing FH.
Methods: Data were obtained from 243 patients referred for genetic testing on suspicion of having FH. Forward stepwise logistic regression was performed, with variant status (binary) as the dependent variable, and age, sex, individual components of the Dutch Lipid Clinic Network (DLCN) criteria, total cholesterol (TC), high-density lipoprotein cholesterol (HDL-C), triglycerides, and low-density lipoprotein cholesterol (LDL-C) as independent variables. Variables were added to the model until their inclusion was not significant ($p>0.05$), and the Bayesian information criterion (BIC) increased. Backward stepwise logistic regression was performed to verify the results and ensure consistency. Receiver operating characteristic (ROC) curve analysis and cross validation (CV) were performed.
Results: Data for 170 patients remained after exclusion of missing data and outliers. The optimal model contained the variables age, LDL-C, and triglycerides. The regression equation for this model was: Probability of FH causing variant = (0.74768 x LDL-C) – (0.06656 x age) – (1.26284 x triglycerides) – 0.06555. The area under the ROC curve (AUROC) for the model was 0.82, with an R^2 of 0.25 and test error rate following CV of 0.25.
Conclusions: The model displayed promising results, and shows potential for improving the targeting of genetic testing in patients suspected of having FH. Interestingly, no individual components of the DLCN criteria were retained in the optimal model. Further work is required to develop and validate the model.

Keywords
Familial hypercholesterolaemia, genetic testing, variant, modelling