ESM Methods 2. Additional statistical methods

1) Logistic regression: both forward and backward stepwise approaches were used to enter in the explanatory variables described in the main methods. How useful the individual predictors were in classifying MODY was determined by the beta coefficients, and their corresponding z-statistic and p value. The overall best model was determined by the Deviance statistic[1]. The models were assessed using both logit and probit link functions. Goodness of fit statistics (Hosmer-Lemeshow[2] and the dispersion parameter[1]) were used to assess model fit. Lowess plots were used to examine the shape of the predictors and determine whether a linear fit was appropriate[3], and plots of fraction of positive cases against average of fitted values[4] were used to further assess model fit across all probabilities.

2) Discriminant analysis: The same explanatory variables were added into these models, and their significance was determined by the F statistic[5]. Plots of the discriminant coordinate (crimcoords) were used to aid interpretation and allowed inspection of the discriminative ability of the model.

3) Classification trees: Dendograms were produced to demonstrate the split criteria. Overly elaborate classification trees were simplified using pruning[6].

Analysis was carried out using the R statistical computing package. Specific libraries used included 'ROCR' for ROC analysis, 'MASS' for discriminant analysis, 'rpart' for classification trees, and the 'boot' library for cross-validation.

References:

[1]   Collett D (2002) Modelling binary data. Chapman & Hall/CRC, Boca Raton, FL
[2]   Hosmer DW, Lemeshow S (2000) Applied logistic regression. Wiley, New York ; Chichester
[3]   Diggle P, Liang K-Y, Zeger SL (1994) Analysis of longitudinal data. Clarendon, Oxford
[4]   Gillmann G, Minder CE (2009) On Graphically Checking Goodness-of-fit of Binary Logistic Regression Models. Methods Inf Med 48: 306-310
[5]   Everitt B, Dunn G (2001) Applied multivariate data analysis. Arnold, London
[6]   Breiman L (1984) Classification and regression trees. Wadsworth, Belmont, Calif.