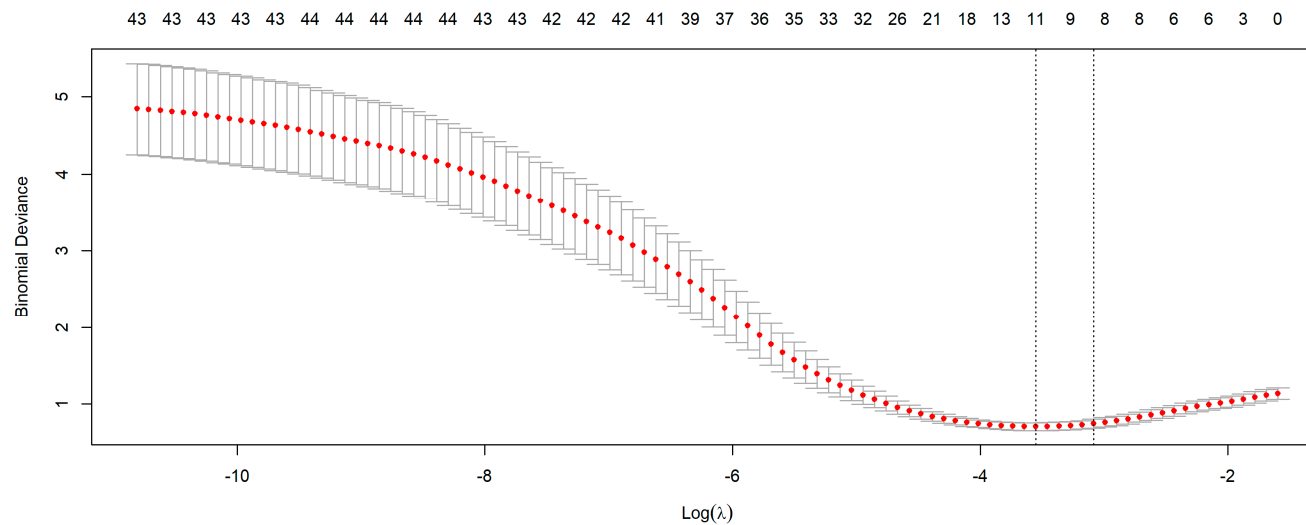
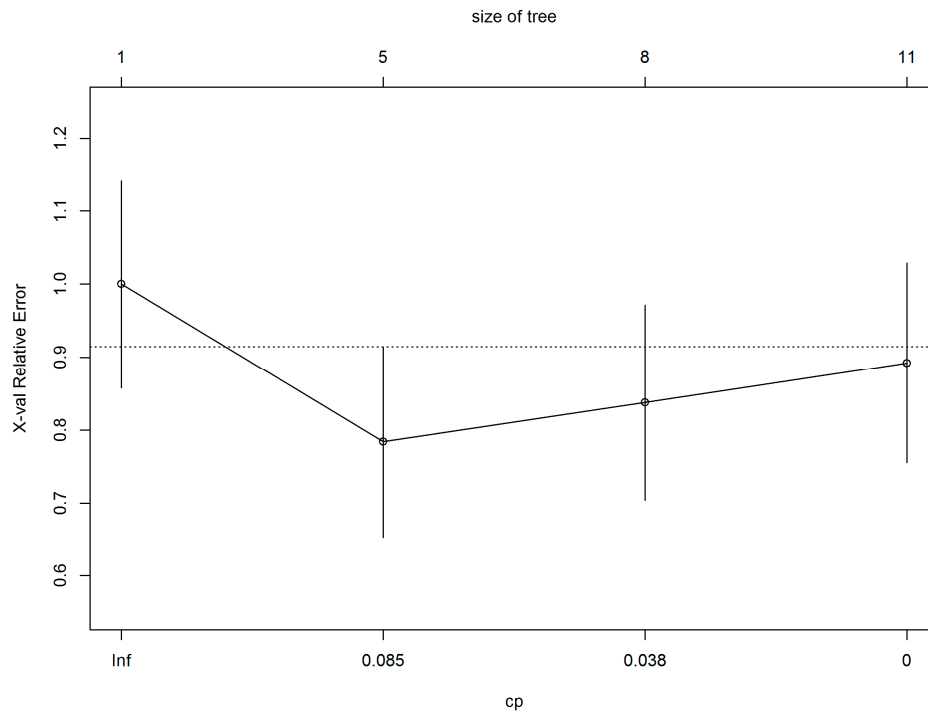


**Figure S1.** Plot of the 10-fold cross-validation curve of the mean errors (binomial deviance) made by different multivariable LASSO models (red dotted line) estimated at varying penalty parameters ( $\lambda$ ), along with the upper and lower standard deviation around the estimates (error bars).



Two special values along the  $\lambda$  sequence (log-values) are indicated by the vertical dotted lines. The leftmost  $\lambda$  value provides the minimum mean cross-validated error, while the rightmost  $\lambda$  value provides the most regularized model such that the cross-validated error is within one standard error of the minimum. The axis above indicates the number of nonzero coefficients detected at the current  $\lambda$ .

**Figure S2.** Plot of the 10-fold cross-validated mean relative error made by the tree at varying values of the complexity parameter (cp), along with the upper and lower standard error around the estimates (error bars).



The complexity parameter is used to control the size of the decision tree and to select the optimal tree size. When  $cp=0$  a tree with its maximum depth is built (i.e. large tree). The axis above indicates the number of terminal nodes of the tree at the current cp. The minimum cross-validation error was reached at  $cp=0.085$  and this returned a pruned final tree with 5 terminal nodes. The horizontal dotted line is drawn at 1 standard error above the minimum of the curve (i.e. at  $cp=0.085$ ).