**Using Random Forest Method to Select Most Valued Items on a Healthy Weight [Obesity Risk] Survey for Young Children**  
  
Christopher Aden1, Christiana Drake1, Mical Shilts3 and Marilyn S Townsend2. 1Statistics, University of California, Davis, CA, United States; 2Nutrition, University of California, Davis, CA, United States and 3Family and Consumer Sciences, California State University, Sacramento, CA, United States.   
  
Random forest is an alternative statistical method to generalized linear models. The objective was to use and compare the two statistical methods that identify the most predictive questions for various outcomes including predicting overweight status at a later time point. Head Start and WIC parents and their young children provided data: parent self-administered 43-item Healthy Kids (HK) at 4 time points, nine 24-hour diet + activity recall logs, 3 blood samples and child anthropometrics collected at 3 time points. An importance plot determined the relative importance of each question by computing the average error when the particular question is omitted from a tree. Of the 43 behavioral items tested, 16 are selected for the final model. The chosen model has a similar classification rate as the stepwise regression procedure in looking at overweight status; it also provides the most important variables in this prediction task, along with weights. The advantage of random forest is that it selected models reasonably supported by nutritional knowledge, and avoided problems faced by stepwise regression in fitting large models. Random forest has potential for use in nutritional studies.  
  
Supported by National Research Initiative #2009-55215-05019 & Agriculture and Food Research Initiative #2010-85215-20658 from USDA National Institute of Food and Agriculture, Human Nutrition & Obesity 93330.