ANNALS OF FAMILY MEDICINE

Online Supplementary Material

Van Voorhees BW, Paunesku D, Gollan J, et al. Predicting future risk of depressive episode in adolescents: the Chicago Adolescent Depression Risk Assessment (CADRA). *Ann Fam Med*. 2008;6(6):503-511.

http://www.annfammed.org/cgi/content/full/6/6/503/DC1

Supplemental Appendix 1. Chicago Adolescent Depression Risk Prediction Assessment (CADRA)

Chicago Adolescent Depression Risk Assessment

	Question *Response option [response value]*		
1	How old are you?	[	years]
2	How much do you weigh?	[	pounds]
3	How tall are you?	[	inches]
4	How close do you feel to your father? *Not at all [1] Very little [2] Somewhat [3] Quite a bit [4] Very much [5]*		
5	How close do you feel to your mother? *Not at all [1] Very little [2] Somewhat [3] Quite a bit [4] Very much [5]*		
6	Compared with other people your age, how intelligent are you? *Moderately below average [1]* *Slightly below average [2]* *About average [3]* *Slightly above average [4]* *Moderately above average [5]* *Extremely above average [6]*		
7	Over the past 12 months, how often were you moody? *Never [0] Rarely [1] Occasionally [2] Often [3] Every day [4]*		
8	Rate your agreement with this statement: "I feel socially accepted." *Strongly agree [1] Agree [2] Neither agree or disagree [3] Disagree [4] Strongly Disagree [5]*		
Over the past week, how often have you experienced the following:			
9	You felt sad. *Never or rarely [0] Sometimes [1] A lot of the time [2] Most or all of the time [3]*		
10	You were bothered by things that usually don't bother you. *Never or rarely [0] Sometimes [1] A lot of the time [2] Most or all of the time [3]*		
11	You felt that you could not shake off the blues, even with help from your family and your friends. *Never or rarely [0] Sometimes [1] A lot of the time [2] Most or all of the time [3]*		
12	You felt depressed. *Never or rarely [0] Sometimes [1] A lot of the time [2] Most or all of the time [3]*		
13	You felt fearful. *Never or rarely [0] Sometimes [1] A lot of the time [2] Most or all of the time [3]*		
14	You felt that you were too tired to do things. *Never or rarely [0] Sometimes [1] A lot of the time [2] Most or all of the time [3]*		

15	You felt hopeful about the future. *Never or rarely [3] Sometimes [2] A lot of the time [1] Most or all of the time [0]*
16	You didn't feel like eating or your appetite was poor. *Never or rarely [0] Sometimes [1] A lot of the time [2] Most or all of the time [3]*
17	It was hard to get started doing things. *Never or rarely [0] Sometimes [1] A lot of the time [2] Most or all of the time [3]*
18	You had trouble keeping your mind on what you were doing. *Never or rarely [0] Sometimes [1] A lot of the time [2] Most or all of the time [3]*
19	People were unfriendly to you. *Never or rarely [0] Sometimes [1] A lot of the time [2] Most or all of the time [3]*
20	You enjoyed life. *Never or rarely [3] Sometimes [2] A lot of the time [1] Most or all of the time [0]*
21	You talked less than usual. *Never or rarely [0] Sometimes [1] A lot of the time [2] Most or all of the time [3]*
22	You thought your life had been a failure. *Never or rarely [0] Sometimes [1] A lot of the time [2] Most or all of the time [3]*

Note: To score this assessment, download the Excel scoring file at: http://catchit-public.bsd.uchicago.edu.

.

Supplemental Appendix 2. Statistical Methods

Boosting Regression

Boosting is a flexible regression model that allows selection of covariates without specifying the functional relationship (nonlinear or interaction) to the outcome. This method has also shown satisfactory results predicting relatively rare outcomes involving complex interaction between independent variables. Boosting involves iteratively fitting models to the combined residuals generated by all the previous models fitted to the data. At each iteration, the difference between current predictions and the observed values are reduced, leading ultimately to the best fit model.[1,2] This method has shown superiority over other methods for the development of regression models in complex data sets.[1,3]

Development of the Classification and Regression Trees Modes

We followed several sequential steps to arrive at our final classification and regression trees (CART) model. First, we randomly divided our analysis sample into a training sample (60%) and a test sample (40%). Using all of our covariates, we fitted a variety of boosted models to determine the optimal level of interaction between the independent variables and the outcome in the training sample. We compared the performance among these alternative models using the pseudo-R^2 criteria based on predictions in the test sample to select the model with the optimal level of interactions. Next, we used the predictions from this optimal model in the training sample to rank each covariate based on its influence statistics.[1] The influence statistic summarizes the relative importance of a variable, including both its main effects and all of its interactions.

Selection of Final Mode

In the final step, we selected the 20 most influential covariates to refine our prediction model. We evaluated models with 20 covariates and those with fewer or more covariates to determine whether more parsimonious or larger models had greater predictive power. In each case, the 20-covariate model had superior predictive power. Next, analogous to our previous approach, we determined the optimal interaction level when only these 20 variables are present. Finally, we generated predictions based on these 20 variables with the corresponding optimal level of interaction. We present the receiver operating characteristic curve, the area under the curve, and the sensitivity and specificity of these predictions in the test sample.

Conversion Into Prediction Index and Validation in the Test Sample and Full Sample

We constructed a simplified index that could be readily computed using a spreadsheet or a handheld device (technical details in the appendix). For each level of each of the 20 covariates, we obtain the predicted probability of developing depression based on our final boosted regression model. We then form a weighted average of these predicted probabilities after weighting each of the probability estimates with the influence-statistic of the corresponding covariate that it represents. A simple index is then formed; a 5-level risk stratification model was constructed on this weighted probability estimate. Our predictions and the development of the index were based on the training sample. We validated the index using the test sample.

References

1. Friedman J, Hastie T, Tibshirani R. Additive logistic regression: a statistical view of boosting. *Ann Statistics.* 2000;28(2):337-407
2. Bauer E, Kohavi R. An empirical comparison of voting classification algorithms: bagging, boosting, and variants. *Mach Learn.* 1999;36(1-2):105-139.
3. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Statistics.* 2001;29(5):1189-1232.