Use of Classification Trees for Prediction of Violence

Charles Broderick^{1*}, Benjamin Rose¹, Marie C. Schur¹, Katherine Warburton¹

1. California Department of State Hospitals *Contact author: charles.broderick@dsh.ca.gov

Keywords: classification trees, prediction, survival analysis, violence risk assessment

Prediction of future violence, or dangerousness, in mentally ill offenders who have committed previous violent acts is a serious endeavor with both public safety and individual rights issues. The seminal work on violence risk assessment was the MacArthur study [1], which used an early commercial implementation of classification trees [2]. While the MacArthur study was groundbreaking, the findings are not always applicable to all settings. With such a sensitive topic as prediction of future violence risk, it is incumbent upon professionals and organizations to act as responsibly as possible, and ensure that prediction findings are valid and accurate for their setting. The use of *R* [3], along with specialized packages such as **party** [4] and **caret** [5] has enabled our organization to conduct analyses within our system, using our own patient data, to ensure applicability to our setting.

Typically, in any given year, about 30% of our patients have one or more episodes of physically violent behavior. Starting with archived data, results of our pilot studies with R and the party package have produced classification tree models in a patient group (n=1277) based on only five demographic variables available pre-admission with adequate predictive ability (AUC = 0.67), when evaluating prediction models derived from a training set (70%) applied to a separate test set (30%). Use of these tools has enabled us to go beyond simple "yes/no" categorizations, i.e., we can predict groups of patients at higher risk of multiple aggressive/violent incidents, and by using conditional inference trees with a survival function [6] we can also evaluate patients for time to first violent act. In this manner, our hospitals can potentially better allocate resources in an attempt to prevent violent acts among higherrisk groups before violence even occurs. With the success of initial pilot models, our focus now is turning to identifying clinical variables that can be added to our model, to enhance prediction as well as to make the model more relevant to our clinicians and treatment teams working with these patients. In summary, access to free, open source tools such as R, with packages such as party and caret, has enabled our organization to undertake analyses aimed at predicting violent behavior before the first displays of violence occur. Our goal is to direct enhanced treatments towards individuals identified as high risk, to mitigate risk, before any such violent behaviors even occur.

References

- [1] John Monahan, Henry J. Steadman, Eric Silver, .et.al., (2001). Rethinking Risk Assessment. New York: Oxford University Press.
- [2] SPSS, Inc. (1993). SPSS for Windows CHAID (Release 6.0). Chicago: SPSS.
- [3] R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.
- [4] Torsten Hothorn, Kurt Hornik and Achim Zeileis (2006). Unbiased Recursive Partitioning: A Conditional Inference Framework. Journal of Computational and Graphical Statistics, 15(3), 651--674.
- [5] Max Kuhn. Contributions from Jed Wing, Steve Weston, Andre Williams, Chris Keefer, Allan Engelhardt, Tony Cooper, Zachary Mayer and the R Core Team (2014). caret: Classification and Regression Training. R package version 6.0-24. http://CRAN.R-project.org/package=caret
- [6] Brian S. Everitt and Torsten Hothorn (2013). HSAUR: A Handbook of Statistical Analyses Using R. R package version 1.3-3. http://CRAN.R-project.org/package=HSAUR