2019

Automated Identification of Unhealthy Drinking Using Routinely Collected Data: A Machine Learning Approach

Levi N. Bonnell  
*University of Vermont, lbonnell@uvm.edu*

Benjamin Littenberg  
*University of Vermont, Benjamin.Littenberg@uvm.edu*

Safwan R. Wshah  
*University of Vermont, Safwan.Wshah@uvm.edu*

Gail L. Rose  
*University of Vermont, Gail.Rose@uvm.edu*

Follow this and additional works at: [https://scholarworks.uvm.edu/comfac](https://scholarworks.uvm.edu/comfac)  
Part of the [Epidemiology Commons](https://scholarworks.uvm.edu/comfac), [Health Services Research Commons](https://scholarworks.uvm.edu/comfac), and the [Substance Abuse and Addiction Commons](https://scholarworks.uvm.edu/comfac)

Recommended Citation  
[https://scholarworks.uvm.edu/comfac/9](https://scholarworks.uvm.edu/comfac/9)

This Poster is brought to you for free and open access by the Larner College of Medicine at ScholarWorks @ UVM. It has been accepted for inclusion in Larner College of Medicine Faculty Publications by an authorized administrator of ScholarWorks @ UVM. For more information, please contact donna.omalley@uvm.edu.
Automated Identification of Unhealthy Drinking Using Routinely Collected Data: A Machine Learning Approach

Levi N. Bonnell MPH, Benjamin Littenberg MD, Safwan R. Wshah PhD, Gail L. Rose PhD, University of Vermont

INTRODUCTION

- Unhealthy drinking is highly prevalent and can lead to serious health and social consequences.
- Identifying unhealthy drinkers can be time-consuming for primary care providers.
- Unhealthy drinking is under-identified and under-treated.
- An automated tool for identification would allow attention to be focused on patients most likely to need care and therefore increase efficiency and effectiveness.

STUDY OBJECTIVE

To build a clinical prediction tool for unhealthy drinking based solely on routinely collected demographic and laboratory data.

METHODS

- Demographic and laboratory data on 13,833 adults seen at the University of Vermont Medical Center, 2011-2017.
- Predictors:
  - Demographic information: Smoking status, gender, age, race/ethnicity
  - 22 Lab values from routine clinical chemistry and hemograms.
- Reference test
  - Unhealthy drinking measured by AUDIT-C
- Definitions:
  - > 3 drinks per day or > 7 drinks per week for women
  - > 4 drinks per day or > 14 drinks per week for men
  - 22 lab values, 4 demographic variables.
  - Training set (68%), validation set (12%), test set (20%).
- Logistic regression, support vector machines (SVM), k-nearest neighbor, random forests were used to build clinical prediction models.
- Model with largest area under the curve (AUC) was selected.
- Support vector machine with polynomials of degree 3 produced the largest AUC.
- An operating point with greater specificity was prioritized.

RESULTS

- We chose an optimum operating point that produces:
  - Sensitivity 31.1%
  - Specificity 91.2%
  - Positive predictive value 50.4%
  - Negative predictive value 82.1%
  - Change in prevalence of unhealthy drinking from 22% to 50%
  - Reduced the target population by 86%
  - The most influential predictors were age, HDL cholesterol, white blood count, and hemoglobin.

CONCLUSION

The virtue of the clinical prediction rule is not that it is perfectly accurate but that it is fast, inexpensive, unobtrusive, and identifies subjects with a higher prevalence of unhealthy drinking than the normal population.

LIMITATIONS

- Primary care providers were not practicing universal screening.
- Prevalence of unhealthy drinking 60% among those with AUDIT-C.
- Sample may not be generalizable
- Sensitivity and specificity of AUDIT-C

ACKNOWLEDGEMENTS & CONTACT

Thanks to the many patients whose data we used and to the NIH for funding. SRRT 1R41AA025297-01A1
Gail Rose, PhD <gail.rose@uvmhealth.org>
Levi Bonnell, MPH <levi.bonnell@med.uvm.edu>
Given S-459
University of Vermont
89 Beaumont Ave, Burlington, VT, 05404