Visualizing patient-level risk factors from clinical risk prediction algorithms

Nate C. Apathy, BS[1,2]; Anna Roberts, MIS[2]; Ross Hayden, MS[2]; Christopher A. Harle, PhD, MS[1,2]; and Joshua R. Vest, PhD, MPH[1,2]
[1]Richard M. Fairbanks School of Public Health, Indiana University, Indianapolis, IN; [2]Regenstrief Institute, Indianapolis, IN

What might the attendee be able to do after being in your session?

Session attendees will understand why risk prediction models illustrating aggregate risk scores are unlikely to be actionable. We illustrate interactive visualizations of patient-level factors driving aggregate risk scores derived from prediction algorithms. By extending these examples to their own organizations, attendees will be better equipped to develop visualizations of actionable, patient-specific drivers of patient risk for use by clinicians and care managers.

Background and Problem

Algorithmic approaches to understanding patient risk are becoming increasingly popular in health care, with the goal of improving patient outcomes and reducing system costs.[1] Health systems and provider organizations often leverage data from electronic health records (EHRs), insurance claims, and social determinants of health to calculate patient risk scores for clinical outcomes as well as the needs for specific services.[2] Health systems have expanded their use of machine learning in predicting patient risk as tools have become more user-friendly. Despite their relative ease of implementation, these models are difficult to interpret. Clinicians often do not receive training on how to understand the inputs or outputs of prediction models, relegating these models to "black boxes." However, in health care it is critical to understand both the methods used to predict risk and the clinical utility of the information involved in those predictions. Only by understanding those factors can clinicians effectively judge the value of a given risk prediction. Nevertheless, most visualizations simply identify patients as "high risk" for a certain outcome, with no insight into why that particular patient is at risk.[3] For example, two patients may be at the same aggregate level of risk for developing diabetes, but the primary factor driving that risk may be genetics for one person and diet for another. We find few examples of risk visualization tools that enable dynamic visualization of intervenable or clinically relevant risk factors at the patient level. Therefore, we aimed to develop an interactive visualization tool to give clinicians and care managers insight into 1) population distribution of risk, 2) individual aggregate risk, and 3) the specific factors driving each patient's predicted risk.

Methods

We developed a prediction model and visualization application for examining aggregate risk scores and patient factors driving risk of congestive heart failure (CHF) among 66,800 diabetic patients. We used a random forest classification model with 10 percent holdout validation, and generated local importance factors for each patient observation in the analytic data set. Crucially, local importance factors capture the degree to which each variable in the prediction model contributed to the aggregate risk score, enabling visualization of patient-specific risk factors. Rather than emphasizing the risk prediction model in this abstract, we focus on the visualization application.

Results

Using the RShiny suite of tools, we developed a risk prediction visualization application that presented users with a population distribution of risk scores, with each point representing a single patient. The application simultaneously presents 1) a violin plot of risk distribution and 2) individual points for each patient, positioned vertically by the patient's predicted probability of developing CHF. The left panel is interactive; to explore patient-level factors driving aggregate risk, users double-click any point to generate a plot in the right panel. This plot dynamically selects the ten highest ranking variables contributing to that patient's aggregate risk score, ranked by absolute value of the local importance measure. Below, we show two patients with similar risk scores that have different primary factors driving their respective risk (**Figures 1 & 2**).

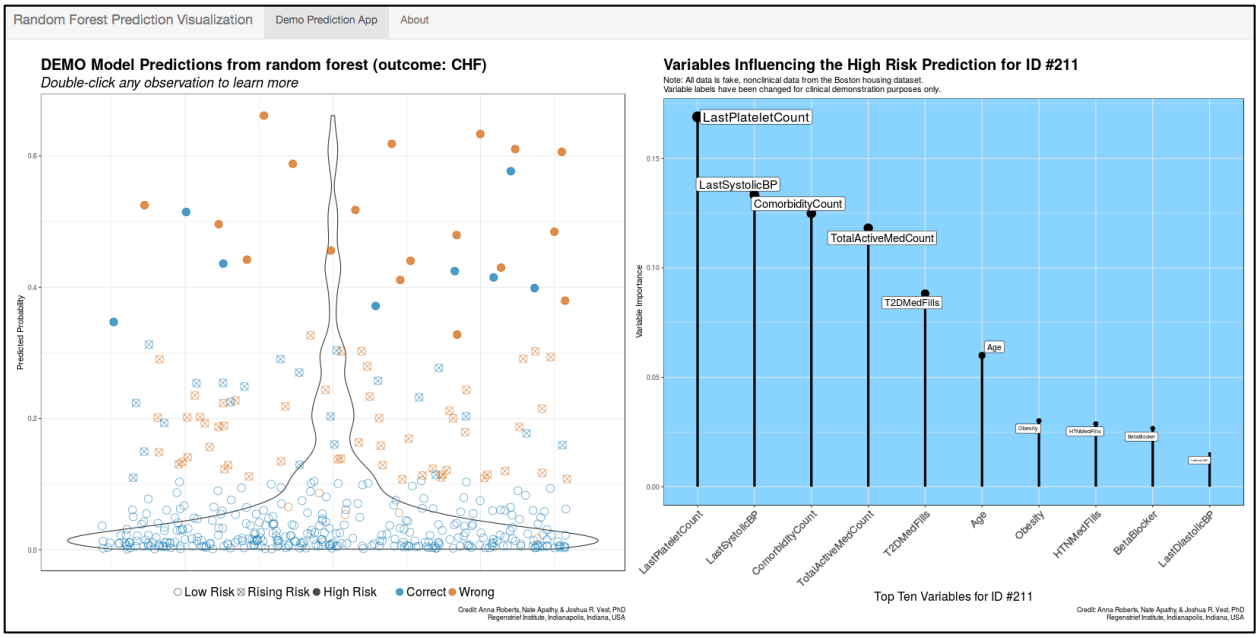
Discussion

We developed a tool that concurrently presented population risk distribution, risk scores for individual patients, and specific factors driving each patient's risk. Tools like these that contextualize patient risk are likely to be more useful than aggregate risk scores in determining the highest-yield clinical interventions for specific patients.

Conclusion

When developing patient risk prediction models to inform patient care decisions, informaticians should consider carefully the actionable information providers and care managers can use. Furthermore, informaticians should leverage interactive data visualization tools to illustrate not only aggregate risk but the underlying clinical and social factors driving risk scores for each patient, to better inform care and improve population health outcomes.

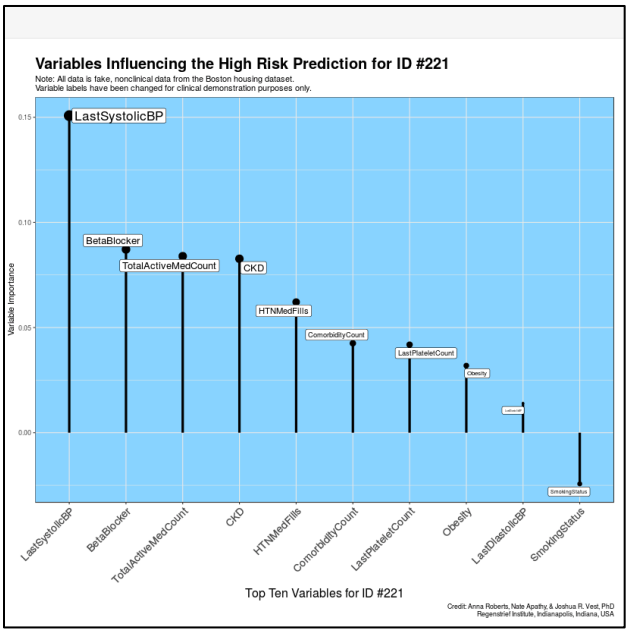
Figure 1. Demo Model of Risk Prediction Tool and Patient-Level Risk Factors, Patient #211



Note: For patient #211, the primary driver of CHF risk is the value of their last platelet count.

Figure 2 (below, right)**.** Demo Model of Risk Prediction Tool and Patient-Level Risk Factors, Patient #221

Note: For patient #221, the primary driver of CHF risk is their last systolic blood pressure reading.



Attendee's take-away tool

Attendees are welcome to view our publicly available application online at http://bit.ly/risk_pred. The tool presents a demonstration of patient-level risk factors driving overall risk scores using a random forest prediction algorithm. When developing risk prediction models and tools for informing clinicians and care managers, this demonstration application can be a useful reference to illustrate a method to present data regarding population risk, individual risk predictions, and patient-level risk factors.

References

1. Cutler DM. Practical Steps to Make Health Care Better. JAMA. 2019 May 28;321(20):1962-3.
2. Vest JR, Harris LE, Haut DP, Halverson PK, Menachemi N. Indianapolis provider's use of wraparound services associated with reduced hospitalizations and emergency department visits. Health Affairs. 2018 Oct 1;37(10):1555-61.
3. Harle CA, Neill DB, Padman R. Information visualization for chronic disease risk assessment. IEEE Intelligent Systems. 2012 Nov 29;27(6):81-5.