The 7 Habits of Effective Predictive Model Implementations: Lessons from the Clinical Trenches

Karandeep Singh, MD, MMSc[1], Sean Meyer, MBA[1], Michael Burns, MD, PhD[1], Jeremy B. Sussman, MD, MSc[1], Akbar K. Waljee, MD, MSc[1], and Brahmajee K. Nallamothu, MD, MPH[1]

[1] Michigan Integrated Center for Healthcare Analytics and Medical Prediction, University of Michigan, Ann Arbor, MI

Background. Implementing predictive models within a health system is fraught with challenges that go well beyond model discrimination and calibration. While scientific literature and popular media often highlight successful deployments, these are far outweighed by the failure to both deploy models in clinical environments and to use these models to drive clinical decision-making. Implementing predictive models effectively requires clinicians and computer scientists to overcome numerous sociotechnical barriers. We highlight key lessons learned from our collective successes and failures in implementing predictive models in a tertiary care health system.

The 7 Habits of Effective Predictive Model Implementations.

1. *Select a "treatable" problem where physicians agree on the right course of action.* Many clinical diagnoses and outcomes can be accurately predicted for which no effective or agreed-upon treatment exists. In the face of conflicting scientific evidence, for example, physicians disagree on several facets of sepsis treatment, including how to assess adequate fluid resuscitation and whether to administer corticosteroids. Even though predicting the onset of sepsis has been a focus area in the clinical and machine learning community due to the availability and high quality of MIMIC-III data, it is unlikely that physicians would agree on the right course of action even if a perfect model existed.
2. *Make sure the health system has sufficient health IT infrastructure and resources.* Tree-based ensemble models and neural networks are common in the clinical literature, yet many electronic health record (EHR) systems have no mechanism to deploy such models in their IT infrastructure. When such infrastructure is unavailable, it must be budgeted for, prioritized, and purchased by health system leadership. Even if such infrastructure has been implemented, health IT staff must spend time initially and after each EHR version update connecting predictive models to data elements in the EHR and validating the accuracy of the connection, which may not always be possible in light of competing priorities.
3. *Anticipate the need and challenges of "local" validation.* Models being considered for implementation in a health system have often been developed and validated in other health systems. Although external validation is of supreme interest to model developers as a measure of generalizability, health systems primarily care about whether models will work well locally in their environments and in their patients.
4. *Deliver the model output in the right format to the right people.* Model output can be presented to clinical decision-makers in several ways using the EHR. The most common ways are through interruptive pop-up alerts (termed clinical decision support) and population health management approaches (sometimes referred to as panel management or registry-based approaches). An example of a population health management approach would be to sort outpatients in order of highest to lowest risk of emergency department visits and then offer complex care management services to those with the highest risk. Whereas pop-ups interrupt clinician workflow, a population health management approach may allow non-urgent predictions to be deferred and acted upon asynchronously. The right way to deliver the model's output depends on the clinical context, urgency, and the individuals who will act on the information.
5. *Balance the workload budget when possible.* Interventions linked to predictive models often focus additional resources on patients identified at highest risk. Unfortunately, "doing more" for patients at highest risk results in a net increase in the workload that is not usually offset by a reduction in other work-related responsibilities. Whenever possible, the plan to implement predictive models should consider a way to "do less" for very low-risk patients to partially offset the new workload. When this is not possible, removing existing work or hiring additional staff should be strongly considered.
6. *Start with a low-fidelity pilot.* Blind spots in model implementation are often revealed at the moment the model is "turned on." Starting with a "shadow pilot" (model running in the background reporting hypothetical results) followed by a (paper-based) low-fidelity pilot minimizes start-up time and may teach lessons crucial for eventual success.
7. *Deploy the model, monitor its use, and adapt.* Initially, an early warning system for maternal hemorrhage implemented at Michigan Medicine unexpectedly generated an alert every 10 minutes. Having a governance structure and technical capability to make rapid changes to reduce alert frequency was key to its eventual high acceptability among clinicians.