Surgical risk is a here and now proposition; you get the outcome of your surgical management right away; the acuity of the timing makes it difficult to claim an adverse result was solely due to the disease or the patient. So you might forgive surgeons, for being a bit obsessed with identifying better ways to determine the risk for the patient sitting across from us, population aggregates are of little comfort. There are any number of algorithmic calculators to quantify a surgeon’s qualitative judgment, but most involve elective care when you have a few more options. For emergency surgery, time is not on anyone’s side, it means the surgeon may not have all the information they would like, and the patient is busy dying with limited options.

A study in the Annals of Surgery applies a novel form of machine learning to the task. The American College of Surgeons (disclaimer I am a member and big fan) has one of the most complete prospectively collected, rigorously monitored databases, the National Surgical Quality Improvement Program (NSQIP). The College using their statisticians have developed a risk calculator using the dataset, but did it “old school.” Their risk calculation is based on the premise that each variable and the database collects 150 of them on each patient, act independently of one another to accumulate a risk. And while their statistical analysis is, you’ll forgive the expression, “bleeding edge,” the underlying assumption of the independence of the risk factors and how they accumulate is not, risk is not a $1 + 1 = 2$ approach.

In this instance, the researcher’s assumption was that risk factors interact with one another, that $1 + 1$ can equal three or $1/2$, it all depends on which elements are present. It is a different way of getting at an individual’s risk from population data, and it is made possible by the increasing power of computation that allows a machine to construct a more robust decision tree. Perhaps, that is the
best way to think of the change in approach. The old-school regressions, for all their sophistication, yielded a stick figure tree; using a different form of optimization, the new approach creates a very thick bush.

Using this different analysis resulted in an improvement in the accuracy of their predictions compared to the other available risk calculators.

It increased the predictability of a greater than 50% risk of death from 89 to 91% and the prediction of morbidity (a complication of surgery) from 80 to 84%. Those numbers alone should show you how well the old-school linear model was, and how hard it is to go that last mile from 89% into the mid to upper 90th percentiles. When asked to predict specific morbidities, such as a heart attack or infection, all the models had difficulty. Even this model was only accurate 68% of the time in predicting superficial surgical infections. Every complication had its own accuracy.

To the researchers’ credit, their model is evidence-based and can be rerun as more evidence is acquired. They demonstrated improved accuracy and developed an easy user interface that a surgeon can refer to through their phone. The questions asked by the algorithm vary with what prediction the surgeon desires, mortality, morbidity or a specific complication and those questions provide the surgeon with the traces of the algorithm’s “thought-process,” eliminating the “black box” quality of many of these systems. Finally, the underlying data points are readily found in most electronic health records, making it a practical tool.

The authors mention several limitations, most importantly, the quality of the data underlying the model. But NSQIP is perhaps the tightest, cleanest dataset we could hope form, all variables have agreed-upon definitions, and the nurses doing the data input, are not only trained on the system but undergo intra and inter-examiner testing before submitting data to the registry. That too is a measure of how difficult it is to provide meaningful data so that machine learning can get right.

The final limitation is that correction of an underlying variable may not only be “too little, too late,” but have unintended consequences. I would argue that this is just the tip of the limitation. As the researchers state, the calculator “can equip surgeons with personalized and highly accurate risk estimates that will allow them to counsel emergency surgery patients and families before surgery.” (Italics added) In my experience, a patient or family unprepared for a life and death surgical emergency wants everything done, irrespective of the risks. From their perspective, the chance of losing a loved one is 100% if they do not go forward, so even a 5 or 10% chance of saving them is a better deal. And rarely do unprepared families feel comfortable with an active decision to do nothing, no one wants to "kill" grandpa. The time to discuss what to do is not in the Emergency Department, it is in the office, long before it becomes a reality, that is why end-of-life discussions are so crucial to compassionate care.
