

## Volume 16 - Issue 2, 2016 - Management Matters

### The Future of ICU Prediction Scores in the Era of 'Big Data'



**Prof Vitaly Herasevich, MD, PhD, FCCM**

\*\*\*\*\*@\*\*\*mayo.edu

Professor of Anesthesiology and Medicine - Department of Anesthesiology, Division of Critical Care, Mayo Clinic, Rochester, USA



**Prof. Mark Keegan, MD, MSc**

\*\*\*\*\*@\*\*\*mayo.edu

Professor of Anesthesiology - Department of Anesthesiology, Division of Critical Care, Mayo Clinic, Rochester, USA



**Prof Brian Pickering, MD, MSc**

\*\*\*\*\*@\*\*\*mayo.edu

Professor of Anesthesiology - Department of Anesthesiology, Division of Critical Care, Mayo Clinic, Rochester, USA

Fuelled by the growth in electronic medical record (EMR) adoption, the recent boom of Big Data analytics in healthcare affords unprecedented opportunities for developing new prediction scores for patients in the intensive care unit (ICU) and beyond. Based on a review of scientific literature reports, however, Big Data science has added no new scores to clinical practice in recent years. In this article, the authors will discuss barriers and potential solutions for the future development of meaningful clinical scores.

#### Do We Need Scores?

Clinical practice is concerned with resolving uncertainty around diagnosis, treatment options and prognosis. Clinical decisions are usually made by an individual such as a physician who has trained for many years in the science, art and practice of medicine. These individuals use their clinical experience to hone in on relevant cues from the patient history, physical examination, physiologic trends and laboratory and other diagnostic data. These cues often fall into some pattern the physician recognises, which forms the basis of a clinical diagnosis. The diagnosis informs treatment choices and the response to treatment determines prognosis.

The most useful prediction scores should combine medical signs, symptoms and other findings to reliably predict the probability of a specific disease or outcome (McGinn et al. 2000). Clinical scores became popular in the 1980s when computerisation of medical data made access and computation much easier than before. Electronic data could be processed more quickly, reliably and consistently than the equivalent manual process. Recently prediction models, clinical calculators and scores have attracted more attention as they become important parts of quality improvement initiatives and are used in administrative and financial management to standardise patient populations for comparison (Keegan et al. 2011).

We recently performed a systematic analysis of the peer-reviewed literature, and identified 176 validated clinical calculators used mostly in the specialties of critical care, emergency medicine and internal medicine (unpublished data). The most frequent outputs of clinical scores were outcome prediction, quantification of the severity of illness and estimation of disease progression as an aid to clinical decision-making. The most frequent computation technique utilised is branching logic with cutoff rules followed by regression analysis (unpublished data). The use of these techniques continues to grow in step with increases in computational power and availability of data.

#### Big Data Revolution

With the almost ubiquitous use of electronic medical records, clinical data have become accessible to analytics as never before. The more exotic potential applications of predictive analytics are almost within reach. The most important of those applications, in the authors' view, are those which will use large quantities of clinical data—collected and annotated automatically in data mining and machine learning algorithms—to build 'on-the-fly' predictive analytics tailored to the individual patient and to patient populations. The building blocks of these applications are well-established in industries which made an early switch from paper-based data storage and transfer to a business model in which electronic data processing is key (e.g. the banking and retail sectors). In these industries the tools are particularly focused on the prediction of future system and individual behaviours, on trajectory analysis, and on early detection of events which deviate from normal. Real-world examples include fraud detection in the banking industry and, in the retail sector, the prediction of what the customer might want to buy next. This often results in contact with the unsuspecting customer in the form of surprisingly relevant advertisements or phone calls. The success of Big Data analytics to deliver these results in other data-rich businesses, combined with the availability of clinical data in an electronic form, bodes well for the ability of machine-learning techniques to produce clinically important prediction models or scores. However, almost a decade into the Big Data revolution in healthcare we have not yet seen a new breakthrough described in the clinical scientific literature. Why?

#### Challenges and Potential Solutions

**1. Scoring systems need to be generalisable.** Data models developed based on a specific clinical setting tend to work best in that setting. When applied outside of the development environment they often perform poorly. A universal scoring system requires clinical data from multiple healthcare settings. Notwithstanding the difficulties involved in the sharing of data among competing organisations, blended data can drag the

performance of the model down to the lowest common denominator. Specific tools will perform better than universal tools. With further development of healthcare data mining tools and machine learning, combined with the availability of domain experts, it is likely that highly calibrated tools will be generated 'on the fly' for very narrow problems and specific settings.

**2. Prediction models developed from electronically captured data do not always represent the whole picture of healthcare delivery.**

Observation of patients, discussion with clinical staff, patients and their families, and evaluation of the care delivery environment often provide essential information to decision-makers. Such information is typically poorly documented in the EMR. As new modalities of data capture (computer vision, radio-frequency identification (RFID) tags, accelerometers, social media, etc.) are integrated into the clinical record, Big Data analytic tools will have additional rich streams of data from which to generate predictions.

**3. Garbage in - garbage out.** Missing data, manually entered or delayed data and poorly validated data each play a role in decreasing the reliability of the final result. Often models work well on retrospective sets of data (collected and present as a single table), but when applied in real time suffer from delayed data presentation and cannot reliably predict an event earlier than a clinician at the bedside.

**4. Association is not causation.** Without biologic plausibility the 'black box' output from prediction models can be difficult to take seriously in the clinical setting where the stakes are very high. The incorporation of some basic hypothesis testing will add greatly to the credibility of 'black box' analytics.

**5. There are limits to the power of prediction.** As a clinician one needs to know that 'black box' predictions will not expose a patient to harm. With any test there are always false-positive and false-negative cases. Any diagnostic test can be described by its sensitivity and specificity. Prognostic scoring systems may be evaluated by assessment of discrimination and calibration. Big Data proponents have shied away from engaging with regulatory bodies, but must answer questions regarding test accuracy, safety, reliability and ease of implementation.

**6. Black box recommendations are poorly tolerated by clinicians.** Until artificial intelligence (AI) becomes standard in our lives, clinicians need to understand how the model produces a prediction. This is especially true if a model might be incorrect some of the time. Adjusting models to rule out true negative cases (for diagnostic models) and explanation of the steps behind survival predictions (for prognostic models) will enhance clinician acceptance. Engaging clinicians in model design and implementation are also useful intermediate steps on the path towards acceptance of AI recommendations.

**7. Actionable predictions will be highly valued by clinicians.** Predictions are most useful if they can be used by clinicians to correct course and prevent an adverse event or outcome for the patient. Clinicians do not highly value risk of dying or risk of deterioration predictions unless they come bundled with potential risk mitigation strategies or recommendations. Knowing that the elderly frail patient has a 45% chance of dying in the intensive care unit is not information that is actionable in most cases. When developing prediction algorithms, it is important that careful consideration be given to the associated expected course of action to reduce the risk of harm or suboptimal patient outcome.

**8. Presentation of final results.** Highly performing prediction models are useless if poorly implemented or presented (Ofoma et al. 2014). Careful consideration of the delivery mechanism of an alert will increase the likelihood that the output of a prediction model will be incorporated into clinical decision-making. Up-front consideration of "who, when, where and how" will greatly increase the potential impact of Big Data analytics predictions.

Although significant limitations exist in the current generation of EMRs, the move to a digital record opens up a world of possibilities for advanced analytics techniques. We have not yet seen significant breakthroughs at the bedside, but it is early days. As Big Data analytics experts from retail and finance incorporate learning from healthcare domain experts into their models, the complexity and human-centred nature of clinical care will become obvious. With these new insights and continuously advancing technology, clinicians and patients should expect to encounter ever more useful and meaningful tools supporting better health, better care and lower costs.

**Conflict of Interest**

Vitaly Herasevich declares that he has no conflict of interest. Mark T. Keegan declares that he has no conflict of interest. Brian W. Pickering declares that he has no conflict of interest.

**Abbreviations**

AI artificial intelligence

EMR electronic medical record

ICU intensive care unit

RFID radio-frequency identification

Published on : Mon, 30 May 2016