Exploring the Potential of Predictive Analytics and Big Data in Emergency Care

Alexander T. Janke, BS*; Daniel L. Overbeek, MD; Keith E. Kocher, MD, MPH; Phillip D. Levy, MD, MPH

*Corresponding Author. E-mail: atjanke@gmail.com, Twitter: @ATJeconmed.

Clinical research often focuses on resource-intensive causal inference, whereas the potential of predictive analytics with constantly increasing big data sources remains largely unexplored. Basic prediction, divorced from causal inference, is much easier with big data. Emergency care may benefit from this simpler application of big data. Historically, predictive analytics have played an important role in emergency care as simple heuristics for risk stratification. These tools generally follow a standard approach: parsimonious criteria, easy computability, and independent validation with distinct populations. Simplicity in a prediction tool is valuable, but technological advances make it no longer a necessity. Emergency care could benefit from clinical predictions built using data science tools with abundant potential input variables available in electronic medical records. Patients' risks could be stratified more precisely with large pools of data and lower resource requirements for comparing each clinical encounter to those that came before it, benefiting clinical decisionmaking and health systems operations. The largest value of predictive analytics comes early in the clinical encounter, in which diagnostic and prognostic uncertainty are high and resource-committing decisions need to be made. We propose an agenda for widening the application of predictive analytics in emergency care. Throughout, we express cautious optimism because there are myriad challenges related to database infrastructure, practitioner uptake, and patient acceptance. The quality of routinely compiled clinical data will remain an important limitation. Complementing big data sources with prospective data may be necessary if predictive analytics are to achieve their full potential to improve care quality in the emergency department. [Ann Emerg Med. 2016;67:227-236.]

A **podcast** for this article is available at www.annemergmed.com.

0196-0644/$-see front matter

http://dx.doi.org/10.1016/j.annemergmed.2015.06.024

SEE EDITORIAL, P. 237.

INTRODUCTION

The US acute care system plays an increasingly important role in overall health care costs.[1-3] Evidence suggests this trend will likely continue because of an aging population[4] and increasing insurance enrollment under the Patient Protection and Affordable Care Act.[5-7] Unlike in many other health care settings, diagnostic and prognostic uncertainty permeate acute care and may be key contributors to health care expenditures. Serious conditions, such as stroke and myocardial infarction, frequently manifest with nonspecific symptoms. Resource-intensive testing is often needed to enable timely diagnosis and intervention. Although relatively few patients who present with such nonspecific symptoms ultimately receive a diagnosis of a serious condition, the expectation of a low or zero miss rate for serious conditions prompts liberal use of expensive diagnostic testing.[8,9]

With ongoing payment reform efforts aimed at improving the value of health care services delivered in the United States, clinicians and departments are under pressure to efficiently use limited resources. Simple heuristics for clinical decisionmaking have been developed with the goal of curbing health care costs without sacrificing quality, but to date results have been variable.[10-17] In the ideal case, decision tools based on predictive analytics can provide actionable information without excessive resource expenditure, helping to limit otherwise wasteful spending that may stem from diagnostic and prognostic uncertainty.[18] Although big data have received significant attention at the health care administrative and policy level, clinical research often focuses on resource-intensive causal inference with primary data. The potential for development of effective, clinically applicable predictive analytics with big data, in the service both of improved clinician decisionmaking and health systems operations, remains largely unexplored. This article provides an introduction to predictive analytics and big data, an exploration of opportunities in emergency medicine at the clinical and emergency department (ED) operations levels, and an overview of challenges, limitations, and potential disadvantages of information technology–driven predictive analytics in emergency care.

PREDICTIVE ANALYTICS AND BIG DATA

Big data, broadly construed here as the full range of available information in electronic medical records and other health care–related databases, has extraordinary potential to improve patient care.[19-23] Many large EDs experience more than 100,000 patient visits per year, yielding a vast amount of real-world observational data. However, accessing this potential resource involves significant logistic challenges. Big data come with limitations, including problems with accuracy, interpretation, missing values, and data handling methodologies.[24] Additionally, big data are drawn from observational sources, limiting their use for causal inference. In clinical research, the criterion standard for causality is the randomized controlled trial. Causal inference techniques relying on nonexperimental methods are subject to confounding. Although quasi-experimental methods using naturally occurring randomness as a means for causal inference hold promise, they have not yet found wide application in health care.[25] Existing studies have explored how associations in big data can drive prediction-based, personalized improvements in care quality.[22] Although this limitation for big data may be overcome in the future,[23] caution is warranted because conflicting results between nonexperimental methods and randomized controlled trials are common.[26]

In general, large observational data sets are not ideal for causal inference. However, they are often well suited for developing predictive models.[27] Basic prediction, divorced from causal inference, has found wide application in a variety of fields. For example, Nate Silver, an American statistician whose predictive analytics work has moved from baseball to US politics, maintains an incredible record of elections predictions at FiveThirtyEight.[28] Similarly, Google uses its gigantic databases to predict stock market changes.[29] Big data tools are on the horizon in medicine as well. IBM's Watson, contributing to cancer diagnoses at Sloan-Kettering, is perhaps the best known popular example in medicine.[30] Other applications, such as machine-learning algorithms for predicting pneumonia mortality[31] and presence of myocardial infarction,[32] have also been developed.

Historically, predictive analytics have played a role in clinical medicine as simple heuristics for risk stratification and diagnostic rule-out.[33] Clinical heuristics based on prediction models have enjoyed substantial attention in EDs. One canonic example is the Canadian CT Head Rule for patients presenting with minor head injury.[10] Derived with carefully collected prospective data, the rule makes use of information readily available at presentation and attempts to identify whether patients have sufficiently high risk of positive head computed tomography (CT) results to warrant performing a scan. In this case, we need not know that, for example, intracranial hemorrhage causes vomiting, only that empirically the 2 events are linked and that there is a plausible mechanistic underpinning that supports an apparent association. For prediction, models need to have high goodness of fit but need not provide unbiased estimates of causation. This distinction between simple prediction and the relatively more difficult task of causal inference lays the framework for the potential and the limitations of prediction models. Prediction models cannot by themselves inform clinicians about the effects of clinical interventions. In addition to the Canadian CT Head Rule and other rules for head trauma, such as the New Orleans minor head trauma rule[11] and the Pediatric Emergency Care Applied Research Network minor head injury CT rules for children,[12] similar predictive tools exist in the ED for chest pain,[13] pulmonary embolism,[14] syncope,[15] and potential cervical spine injury.[16]

The process of deriving, validating, and implementing clinical decision rules like these has historically involved a maintenance of certain methodological standards,[34] such as parsimonious criteria, easy computability, and independent validation with distinct patient populations. Technology now exists to build predictive models without any preference for this standard approach. Data routinely collected within electronic medical records, potentially supplemented with disease-specific information prospectively derived with registries, could be used to define both predictor and outcome variables for new model estimation. Table 1 depicts how past limitations to predictive analytics may be overcome in the future. Rather than relying on parsimonious criteria, new-wave predictive analytics could take advantage of the huge number of variables already being collected in electronic medical records, along with the ability

Table 1. Past limitations mapped to future opportunities in predictive analytics for emergency care.

Limitation	Opportunities to Overcome
Parsimonious criteria	EMRs could provide a huge number of potential variables for deriving predictions. Data entry forms and model estimation could be designed to accept more flexible criteria.
Easy computability	With enhanced computational power available to clinicians, predictive models could use sophisticated techniques, such as machine learning, to compute risk estimates.
Independent validation	Routine data collection within EMRs, strategic prospective registries, and health information exchanges could facilitate model development and cross-validation among relatively similar populations across hospital systems.

EMR, Electronic medical record.