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. 2015;1-10.]

SEE EDITORIAL, P. 11.

INTRODUCTION

The US acute care system plays an increasingly important role in overall health care costs. Evidence suggests this trend will likely continue because of an aging population and increasing insurance enrollment under the Patient Protection and Affordable Care Act. 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.

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. 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. 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. 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. 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. Existing studies have explored how associations in big data can drive prediction-based, personalized improvements in care quality. Although this limitation for big data may be overcome in the future, caution is warranted because conflicting results between nonexperimental methods and randomized controlled trials are common.

In general, large observational data sets are not ideal for causal inference. However, they are often well suited for developing predictive models. 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. Similarly, Google uses its gigantic databases to predict stock market changes. 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. Other applications, such as machine-learning algorithms for predicting pneumonia mortality and presence of myocardial infarction, have also been developed.

Historically, predictive analytics have played a role in clinical medicine as simple heuristics for risk stratification and diagnostic rule-out. 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. 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 and the Pediatric Emergency Care Applied Research Network minor head injury CT rules for children, similar predictive tools exist in the ED for chest pain, pulmonary embolism, syncope, and potential cervical spine injury.

The process of deriving, validating, and implementing clinical decision rules like these has historically involved a maintenance of certain methodological standards, 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...
to automatically prompt for more information when specific orders or patterns of evaluation are detected. Past techniques have faced limited feasibility in overloaded and hectic emergency care settings. Historically, decision rules needed to be simple, with few variables, to ensure that providers could easily memorize and compute them. For example, the thrombolysis in myocardial infarction risk score uses 7 variables, restricted to yes or no answers, to estimate mortality or recurrent ischemia for patients with unstable angina and non–ST-elevation myocardial infarction. This framework for the prediction rule contains discontinuities that do not make sense clinically. In particular, a 66-year-old patient and 95-year-old patient are treated identically by the rule, which uses a cutoff of aged 65 years as one criterion.

Advances in technology will obviate the need to apply simple decision tools to patients with complex disease. Readily available resources, including electronic medical records and smartphones, can provide up-front computational power for new-wave predictive analytics. Even small changes in the flexibility of a decision rule can affect that rule’s performance. For example, the diagnosis of acute heart failure in undifferentiated patients with dyspnea, using models in which natriuretic peptides were treated as dichotomous variables with set threshold values, had lower sensitivity than those incorporating clinical factors and pretest probability. Prediction models can have tens or even hundreds of variables when an individual practitioner is not required to input each data element individually or perform complex computations. More generally, prediction tools could estimate risk pools according to patients’ entire clinical profiles. Even nonclinical factors, such as time of day, day of week, or season, may also meaningfully contribute to pretest probability and find use in emerging applications to improve ED operations. Prediction models built with the full range of variables available are likely to reveal surprising relationships that challenge conventional wisdom. Lists of current medications and allergies, for example, would provide many potential variables, although caution is warranted because inaccurate data entry may render them invalid for use in predictive analytics.

In addition to using more variables with more flexible criteria, new-wave predictive analytics could compute predictions from these criteria with more sophisticated techniques. Predictions could be derived that do not use standard linear or logistic regression models. For example, nonparametric classification and regression trees allow predictive models that require no assumptions about the mathematic relationship between predictors and outcome. Algorithms for the development and validation of decision rules based in complex systems and with incomplete information have already been developed. One technique, called counterfactual regret minimization, is designed specifically to function in situations with incomplete information in which outcomes can be known only in hindsight. More sophisticated analytic techniques, some of which are enumerated in Table 2, offer additional possibilities. Researchers have already found success with more flexible predictive analytics for acute coronary syndrome, in which risk estimation using attribute matching performed better than logistic regression. These techniques have been applied in a randomized trial in which clinicians estimated patients’ pretest probability for acute coronary syndrome, using such a model, and the results showed reduced resource use without increased rate of premature discharge.

After model estimation, decision rules require prospective validation: models are estimated in one study and their predictive value is verified in a second, usually with an entirely separate patient population. New-wave predictive analytics could be prospectively validated with data from a separate cohort within the electronic medical record database from which the model was derived. This process would require care because single-site homogeneity could compromise the scientific rigor of the process. Simultaneously, electronic medical record vendors are already capable of compiling data from many regional health systems. Information sharing across hospitals and health systems could enable cross-validation with largely similar but distinct populations. Consequently, EDs and providers will no longer be restricted to, for example, a head CT rule developed from a population much different from their own, and sites will be able to derive more precise prediction models for given conditions. Overall, simplicity in prediction tools can be valuable, but it is no longer a necessity.

<table>
<thead>
<tr>
<th>Technique</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayesian networks</td>
<td>Relates a set of random variables by conditional dependencies</td>
</tr>
<tr>
<td>Discriminant analysis</td>
<td>Predicts which objects fall into different classes</td>
</tr>
<tr>
<td>Decision tree learning</td>
<td>Maps observations about an item to some quantitative conclusion about that item</td>
</tr>
<tr>
<td>Matching</td>
<td>Compares differences between groups of observations that otherwise share common features</td>
</tr>
<tr>
<td>Markov and Monte Carlo simulations</td>
<td>Simulates multidimensional analyses with stochastic methods to develop approximation models</td>
</tr>
</tbody>
</table>
OPPORTUNITIES IN EMERGENCY CARE

Nowhere is the need to explore the value of new-wave predictive analytics more evident than in high-flow, acute care settings such as the ED. The Figure displays one idealized model for diagnostic uncertainty and the utility of prediction during a clinical encounter, beginning with a hypothetical presentation to the ED. When a patient first arrives, a series of decisions must be made, each committing further resources and subjecting the patient to increased risk of adverse events, incidental findings, costs, and potentially avoidable diagnostic testing and empiric treatments. Stronger predictions, using data gathered from the ED, allow clinicians to avoid marginal tests, procedures, and admissions. The utility of prediction depends both on accuracy and time to the clinical encounter. At initial presentation, the utility of prediction is low because insufficient patient data exist to drive accuracy. Triage and initial evaluation gather data to increase the accuracy of prediction. Later, as resource-committing diagnostic tests and therapeutic interventions are completed, the utility of prediction decreases again. However, there is likely substantial variation in the significance and time course of uncertainty for ED patients, including the presence of discontinuities in which a single finding substantially increases or decreases uncertainty.

The ED has already been identified as a primary target area for cost containment, with the most significant opportunities in clinical conditions associated with high degrees of uncertainty. One challenge for cost containment in the ED is the need for tools and evidence that can support clinicians as they work to make more efficient and cost-saving disposition decisions. Developing alternatives to admission for relatively low-risk patients, in the form of hospital-type home care and improved discharge planning, will hinge on quality risk-stratification tools. Predictions that could help fuel protocol-based management of intermediate-risk patients to achieve optimal hospital admissions rates are especially needed.

Aside from the relatively narrow task of diagnosis and prognosis for individual patients, predictive analytics have the potential to improve the operational flexibility and throughput quality of ED services. In the context of

![Figure](image-url)
predicting demand for ED services \(^{52}\) and problems related to crowding. \(^{2,53,54}\) Emergency care operations research has enjoyed increasing attention. Some models have focused on crowding with the aim of helping administrators predict the effects of changes to ED operations. \(^{55}\) Models have also been developed to predict at triage patients’ likelihood of inpatient admission, which could be used to improve real-time bed management. \(^{56}\) Early work has not demonstrated an operations benefit for community EDs that have invested in electronic medical records. \(^{57}\) Better information technology infrastructure, using tools from predictive analytics, may be necessary to realize gains from that investment. Overall, many aspects of emergency care might benefit from new-wave predictive analytics.

DATABASE DEVELOPMENT AND MANAGEMENT

As depicted in Table 3, implementation of new-wave predictive analytics would require efforts at several levels of engagement. Database infrastructure could be built within or across regional hospital systems, modeling similar approaches for comparative effectiveness research. \(^{58,59}\) Payers, many of whom are already operating in payment models designed to reward high-value care, may be interested in these infrastructure investments. The Centers for Medicare & Medicaid Services and other government entities may also be important investors. The Centers for Medicare & Medicaid Services has already targeted something similar for the outpatient setting with its Transforming Clinical Practices Initiative, a program that seeks to improve health care by creating a self-learning, nationwide network of providers. \(^{60}\) Prediction database infrastructure could also be a major component of the potential cost savings of health information exchange, \(^{61,62}\) a recent focus of federal policy. \(^{63,64}\) As part of the 2009 Health Information Technology for Economic and Clinical Health Act, the federal government has invested more than \$500 million in one-time grants to states for health information exchange. \(^{65}\) Strategic investment in predictive analytics may be one mechanism to achieve otherwise tenuous financial sustainability \(^{66}\) of investments such as these.

These initiatives would need to rely on expertise in massive data management and coordination with health systems leadership and electronic medical record vendors. Existing projects, such as the National Patient-Centered Clinical Research Network, which aims to “improve the nation’s capacity to conduct comparative effectiveness research efficiently by creating a large, highly representative network for conducting clinical outcomes research,” \(^{59}\) provide precedent for investment in coordinated clinical data exchange across health systems. Other projects, such as the Healthcare Cost and Utilization Project, demonstrate that routinely collected data, already built into electronic medical records, can be aggregated from many sources to provide useful insights related to, for example, patterns of admission for heart failure across different regions and hospitals \(^{67,68}\) and ED use after hospital discharge. \(^{69}\) In addition, the National Institutes of Health launched the Big Data to Knowledge initiative in 2012 with the aim of “enhancing the utility of biomedical big data,” \(^{70}\) with, among others, the task of building a working group for President Obama’s Precision Medicine Initiative. \(^{71}\)

Investments in predictions infrastructure will be most successful if they build on existing health information technology. Electronic medical records have been designed largely to collect and organize clinical information for efficient clinical operations but can also be used for data collection and analysis. Data mining of all routinely collected information within electronic medical records could be combined with prospective registries, operating within individual departments, to enable even greater clarity in potential predictor and outcome variables and to facilitate validation of developing clinical tools. Prospective registries could improve on incomplete, nonstandardized, and error-filled documentation within EDs \(^{72}\) and difficulty in measuring adherence to newly implemented decision tools. \(^{73}\) In the most ambitious case, technologies such as IBM’s Watson, \(^{40}\) capable of interpreting complex data types such as free text and imaging to benefit cancer diagnoses, \(^{74}\) may be built into this developing system. Coordination with electronic medical records will allow databases to be updated in real time, with a wide variety of information about the patient encounter. At the same time,
increasing consolidation of physician practices and hospital systems\textsuperscript{75-77} may make it easier to connect information gathered in the ED, for example, to outcome variables.

Data management would need to comply with health information privacy regulations, such as the Health Information Portability and Accountability Act. Researchers could access disseminated data sets, with patient identifiers removed, for prototyping new models. A formal procedure could be developed for more resource-intensive variables to be added to the prospective registries maintained by particular EDs. Registries may be able to address unique regional, local, and departmental needs. After a trial period of narrow data collection, new data elements could be evaluated in a process of prospective validation. Predictive analytics applications prototyped in low-risk settings could be evaluated within the database infrastructure. Observed heterogeneity in patient and provider decisionmaking could be used to inform the direction of ongoing modifications to data collection procedures.

CHALLENGES, LIMITATIONS, AND POTENTIAL DISADVANTAGES

Data Quality and Model Derivation

Databases capable of delivering on the potential of predictive analytics will contain a staggering wealth of information, likely with similarly large variation in the quality and quantity of individual data elements. Examples of relatively successful applications of predictive analytics in emergency medicine have been built with prospective data. For example, the aforementioned Canadian CT Head Rule was developed after careful documentation of specific physical examination findings relevant to the prediction of interest. Data regularly collected within electronic medical records are not of the same quality. A naive application of the same analytic tools used to derive the Canadian CT Head Rule, applied to electronic medical record data, would not be appropriate. Extra scrutiny will be required in light of the limited quality of individual data elements. In addition, the sheer number of different variables available for model estimation will provide extraordinary flexibility. Caution is warranted, though, because flexibility with input variables could increase the risk of model overfit. Individual models may enjoy extraordinarily high goodness of fit, but short of validation with new data not used to estimate the original model, this is not a measure of success. Even after the predictive value of a model is validated, the clinical value of a predictive analytics tools must still be justified. This calls attention to an important general limitation of new-wave predictive analytics: although a passive health information technology infrastructure may be capable of deriving and validating predictive models without the need for experimental evaluation, it cannot by itself evaluate the effects of implementing those predictive models in clinical practice. That last step requires the same level of clinical research rigor necessary in any investigation seeking to establish cause and effect.

Clinical Uptake

Uptake for the average department and provider may be challenging. Use of clinical decision tools has historically been low,\textsuperscript{78,79} whereas early evidence has failed to demonstrate the value of electronic medical record data in ED operations.\textsuperscript{57} Clinicians have cited performance reproducibility, cumbersome implementation, and exclusion of and inferiority to clinician gestalt as reasons for this.\textsuperscript{80-82} Some barriers may be more easily addressed with new-wave predictive tools. Automated message alerts within electronic medical records and data feedback about diagnostic and prognostic information, when carefully used, have the potential to enhance implementation and uptake. Tools such as these, integrated into computerized provider order entry systems, have already been implemented with consensus standards for CT scans in mild traumatic brain injury\textsuperscript{83,84} and have been shown to reduce use and improve yield of the scans.\textsuperscript{85,86} In some cases, however, implementation of these tools may actually increase use, especially where the precise decisionmaking value of the rule is not well defined. Overall, in systematic review, only 13\% of trials of computerized decision support systems found outcome benefits.\textsuperscript{87} Given the existing burden of documentation for clinicians and imperfect record of predictive analytics, it will be incumbent on proponents of these new applications of information technology to justify their usefulness.

Although information technology may improve on some past limitations of predictive analytics, certain aspects of predictive analytics are not made easier by simply computerizing the process. In clinical applications of predictive analytics, the decisionmaking value of new tools must be explicit to clinicians. For example, prediction tools might be used to rule out the need for diagnostic testing, but this does not necessarily imply that failure to rule out necessitates performing the testing. Trade-offs among sensitivity, specificity, positive predictive value, and negative predictive value must be well understood or the benefits of predictive analytics may not be realized.

Increasing integration of these tools into daily practices may have important implications for standard of care, the emergency medicine community’s consensus of what reasonable and prudent emergency physicians would do in
given clinical circumstances. Does a given provider consistently order expensive or invasive diagnostic testing out of proportion to patients’ risk? Is that provider ordering significantly more or fewer tests? Clinicians in emergency care might enjoy an even clearer understanding of their medical decisionmaking, both with respect to the quantitative risks their patients face and to the practices of their peers. This may facilitate clinicians’ achieving, for example, miss rates that are consistent with standard of care while forcing the emergency medicine community to contend more directly with the costs of certainty. However, the widespread implementation of predictive analytics certainly could obscure standard of care, especially given heterogeneous uptake. Carefully implemented, new-wave predictive analytics may avoid the pitfalls of the one-size-fits-all approach that has plagued past applications.

Provider and Patient Perspectives
The implementation of predictive analytics will be most successful if each step of the process is informed by clinician and patient perspectives. Applications of predictive analytics must remain flexible in the context of nonclinical factors that influence clinician decisionmaking. This is a fundamental limitation of information technology–driven predictive analytics. Insofar as patients’ values vary, generalizing approaches to clinical decisionmaking cannot be optimal. Especially in the context of relevant nonclinical factors, clinician risk aversion, and diverse patient values, predictive analytics cannot serve as an alternative but must instead complement clinician discretion and expertise, lest it deserve the label “cookbook medicine.” Similarly, department-level applications of predictive analytics will be most successful if they address the unique needs of individual EDs.

Emergency care providers may be concerned about additional medicolegal liability associated with less resource-intensive approaches to care. Although more uniform, protocol-driven approaches to risk management can be implemented in the context of institutional or enterprise liability, medicolegal issues will need to be addressed if implementation is to be successful. Predictive analytics could be used as a tool in malpractice litigation to justify caricatures of complex clinical situations, fueling lawsuits against physicians. However, predictive analytics may provide a mechanism for clinicians to better clarify their decisionmaking in difficult situations, such as those subject to quality assurance review and legal allegation of negligence. Clinicians could compare decisionmaking and outcomes among patients with similar presentations, which may be especially relevant for low-risk scenarios resulting in bad outcomes.

Patient acceptance of these prediction tools will also be key to their long-term success. Predictive analytic tools could be built from anonymized patient data without researchers’ obtaining individual informed consent from patients. This would represent an important ethical challenge because financial or personal motivations of researchers or administrators may not be congruent with patients’ rights. In disseminating patients’ health information, supplementing shared decisionmaking, and informing day-to-day operations of health care services, risks to patients must be justified. Patient-facing applications of predictive analytics must have benefits relevant to patients. For example, consider a patient presenting with possible myocardial infarction whose initial evaluation result is negative. A patient-facing prediction tool might be presented as follows: “When we look at patients like you who have presented here at our ED and the other EDs in our area with similar symptoms and test results, approximately 1 in 50 needed to be treated for a heart attack within 14 days.” Use of such an approach has already been shown to help decrease stress test use in ED patients, and a follow-up multicenter trial is currently under way (NCT01969240).

CONCLUSION
Obstacles remain in applying the principles of predictive analytics to emergency care. Data collection and management requires substantial resources and commitment by providers and hospital systems. The quality of data collected may be the most important rate-limiting factor in applications of big data because even the most sophisticated statistical techniques cannot overcome the consequences of error-filled data. In addition, privacy regulations add a layer of complexity to the task. In the status quo, health care providers lack the tools and training to effectively and responsibly use big data for predictive analytics. These challenges can be overcome, and the extraordinary potential of new-wave predictive analytics is too great to ignore. Nowhere is this potential more salient than in emergency and acute care settings, where better predictions could make health care decisionmaking more cost-effective and, at the same time, more responsive to patients’ needs.
Cardiovascular Research Institute, Wayne State University, Detroit, MI (Levy).

**Funding and support:** By Annals policy, all authors are required to disclose any and all commercial, financial, and other relationships in any way related to the subject of this article as per ICMJE conflict of interest guidelines (see www.icmje.org). The authors have stated that no such relationships exist.

**Publication dates:** Received for publication February 13, 2015. Revisions received May 26, 2015; and June 6, 2015. Accepted for publication June 22, 2015.

**REFERENCES**


38. Moe GW, Howlett J, Januzzi JL, et al. N-terminal pro-B-type natriuretic peptide testing improves the management of patients with suspected acute heart failure: primary results of the Canadian Prospective
Janke et al
Exploring the Potential of Predictive Analytics and Big Data in Emergency Care


47. Cryer L, Shannon SB, Van Amsterdam M, et al. Costs for “hospital at home” patients were 19 percent lower, with equal or better outcomes compared to similar inpatients. Health Aff (Millwood). 2012;31:1237-1243.


79. Runyon MS, Richman PB, Kline JA. Emergency medicine practitioner knowledge and use of decision rules for the evaluation of patients with


