

Artificial Intelligence and Surgical Decision-making

Tyler J. Loftus, MD; Patrick J. Tighe, MD, MS; Amanda C. Filiberto, MD; Philip A. Efron, MD; Scott C. Brakenridge, MD; Alicia M. Mohr, MD; Parisa Rashidi, PhD; Gilbert R. Upchurch Jr, MD; Azra Bihorac, MD, MS

IMPORTANCE Surgeons make complex, high-stakes decisions under time constraints and uncertainty, with significant effect on patient outcomes. This review describes the weaknesses of traditional clinical decision-support systems and proposes that artificial intelligence should be used to augment surgical decision-making.

OBSERVATIONS Surgical decision-making is dominated by hypothetical-deductive reasoning, individual judgment, and heuristics. These factors can lead to bias, error, and preventable harm. Traditional predictive analytics and clinical decision-support systems are intended to augment surgical decision-making, but their clinical utility is compromised by time-consuming manual data management and suboptimal accuracy. These challenges can be overcome by automated artificial intelligence models fed by livestreaming electronic health record data with mobile device outputs. This approach would require data standardization, advances in model interpretability, careful implementation and monitoring, attention to ethical challenges involving algorithm bias and accountability for errors, and preservation of bedside assessment and human intuition in the decision-making process.

CONCLUSIONS AND RELEVANCE Integration of artificial intelligence with surgical decision-making has the potential to transform care by augmenting the decision to operate, informed consent process, identification and mitigation of modifiable risk factors, decisions regarding postoperative management, and shared decisions regarding resource use.

JAMA Surg. 2020;155(2):148-158. doi:10.1001/jamasurg.2019.4917
Published online December 11, 2019.

- +** [Supplemental content](#)
- +** [CME Quiz at jamanetwork.com/learning](#)
and [CME Questions page 180](#)

Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Author: Azra Bihorac, MD, MS, Precision and Intelligent Systems in Medicine, Division of Nephrology, Hypertension, and Renal Transplantation, Department of Medicine, University of Florida Health, PO Box 100224, Gainesville, FL 32610-0224 (abihorac@ufl.edu).

Surgeons make complex, high-stakes decisions when offering an operation, addressing modifiable risk factors, managing complications and optimizing resource use, and conducting an operation. Diagnostic and judgment errors are the second most common cause of preventable harm incurred by surgical patients.¹ Surgeons report that lapses in judgment are the most common cause of their major errors.² Surgical decision-making is dominated by hypothetical deductive reasoning and individual judgment, which are highly variable and ill-suited to remedy these errors. Traditional clinical decision support tools, such as the National Surgical Quality Improvement Program (NSQIP) Surgical Risk Calculator, can reduce variability and mitigate risks, but their clinical adoption is hindered by suboptimal accuracy and time-consuming manual data acquisition and entry requirements.³⁻⁸

Although decision-making is one of the most difficult and important tasks that surgeons perform, there is a relative paucity of research investigating surgical decision-making and strategies to improve it. The objectives of this review are to describe challenges in surgical decision-making, review traditional clinical decision-support systems and their weaknesses, and propose that artificial intelligence models fed with live-streaming electronic health record data (EHR) would obviate these weaknesses and should be integrated with bedside assessment and human intuition to augment surgical decision-making.

Methods

PubMed and Cochrane Library databases were searched from their inception to February 2019 (eFigure in the [Supplement](#)). Articles were screened by reviewing their abstracts for the following criteria: (1) published in English, (2) published in a peer-reviewed journal, and (3) primary literature or a review article. Articles were selected for inclusion by manually reviewing abstracts and full texts for these criteria: (1) topical relevance, (2) methodologic strength, and (3) novel or meritorious contribution to existing literature. Articles of interest cited by articles identified in the initial search were reviewed using the same criteria. Forty-nine articles were included and assimilated into relevant categories (Table 1).¹⁻⁴⁹

Observations

The quality of surgical decision-making is influenced by patient values and emotions, patient-surgeon interactions, decision-making volume and complexity, time constraints, uncertainty, hypothetical deductive reasoning, and individual judgment. There are effective and ineffective methods for dealing with each of these factors, which lead to positive and negative outcomes, respectively (Figure 1).

Table 1. Summary of Included Studies

Source	Study Design	Population	Sample Size	Major Findings Pertinent to This Scoping Review	Sources of Funding; Conflicts of Interest
Adhikari et al ⁹	Retrospective	Patients undergoing inpatient surgery	2911	A machine learning algorithm accurately predicted postoperative acute kidney injury using preoperative and intraoperative data	NIGMS, University of Florida CTSI; NCATS; I Heerman Anesthesia Foundation; SCCM Vision Grant
Artis et al ¹⁰	Observational	Trainee ICU presentations	157	Potentially important data were omitted from 157 of 157 presentations; missing an average 42% of all data elements	AHRQ
Bagnall et al ¹¹	Retrospective	Patients who had colorectal surgery	1380	Six traditional risk models used to predict postoperative morbidity and mortality had weak accuracy with AUC, 0.46-0.61	St Mark's Hospital Foundation
Bechara et al ¹²	Observational	Healthy volunteers and participants with prefrontal cortex damage	16	Participants began to decide advantageously before they could consciously explain what they were doing or why they were doing it	National Institute of Neurological Diseases and Stroke
Bertrand et al ¹³	Prospective	ICU patients and attending surgeons	419	Clinicians thought 45% of all patients had decision-making capacity; a minimal status examination found that 17% had decision-making capacity	Pfizer, Fisher & Paykel; Pfizer; Alexion; Gilead; Jazz Pharma; Baxter; Astellas
Bertsimas et al ¹⁴	Retrospective	Emergency surgery patients	382 960	A app-based machine learning model accurately predicted mortality and 18 postoperative complications (AUC, 0.92)	None reported
Bihorac et al ¹⁵	Retrospective	Patients undergoing major surgery	51 457	A machine learning algorithm using automated EHR data predicted 8 postoperative complications (AUC, 0.82-0.94) and predicted mortality at 1, 3, 6, 12, and 24 mo (AUC, 0.77-0.83)	NIGMS; NCATS
Blumenthal-Barby et al ¹⁶	Review	Articles about heuristics in medical decision-making	213	Among studies investigating bias and heuristics among medical personnel, 80% identified evidence of bias and heuristics	Greenwall Foundation; Pfizer
Brennan et al ¹⁷	Prospective	Physicians	20	A machine learning algorithm was significantly more accurate than physicians in predicting postoperative complications (AUC, 0.73-0.85 vs AUC, 0.47-0.69)	NIGMS, University of Florida CTSI; NCATS
Che et al ¹⁸	Retrospective	Pediatric ICU patients	398	A gradient boosting trees method allowed for quantification of the relative importance of deep model inputs in determining model outputs	NSF; Coulter Translational Research Program
Chen-Ying et al ¹⁹	Retrospective	Clinic patients	840 487	A deep model predicted 5-y stroke occurrence with greater sensitivity (0.85 vs 0.82) and specificity (0.87 vs 0.86) than logistic regression	None reported
Christie et al ²⁰	Retrospective	Trauma patients	28 212	A machine learning ensemble accurately predicted mortality among trauma patients in the United States, South Africa, and Cameroon with AUC \geq 0.90 in all settings	NIH
Clark et al ⁴	Retrospective	Surgical patients	885 502	The ACS Surgical Risk Calculator accurately predicted mortality (AUC, 0.94) and morbidity (AUC, 0.83)	None reported
Cohen et al ⁶	Retrospective	Studies assessing the ACS Surgical Risk Calculator	3	External validation studies assessing ACS Surgical Risk Calculator performance may have been compromised by small sample size, case-mix heterogeneity, and use of data from a small number of institutions	None reported
Delahanty et al ²¹	Retrospective	ICU patients	237 173	A machine learning algorithm accurately predicted inpatient death (AUC, 0.94)	Alesky Belcher; Intensix, Advanced ICU
Dybowski et al ²²	Retrospective	ICU patients	258	An artificial neural network predicted in-hospital mortality more accurately than logistic regression (AUC 0.86 vs 0.75)	Special Trustees for St Thomas' Hospital
Ellis et al ²³	Prospective	Volunteers	1948	Induction of fear and anger had unique and significant influences on decisions to take hypothetical medications	NCI
Gage et al ²⁴	Prospective	Patients with atrial fibrillation	2580	Models commonly used to predict risk of stroke were moderately accurate with AUC ranging from 0.58-0.70	AHA, NIH, Danish and Netherlands Heart Foundations, Zorg Onderzoek Nederland Prevention Fund, Bayer, UK Stroke Association
Gijsberts et al ²⁵	Retrospective	Patients with no baseline cardiovascular disease	60 211	Associations between risk factors and development of atherosclerotic cardiovascular disease were different across racial and ethnic groups	Netherlands Organization for Health Research and Development, NIH
Hao et al ²⁶	Retrospective	ICU patients	15 647	Deep learning models predicted 28-d mortality with 84%-86% accuracy	None reported

(continued)

Table 1. Summary of Included Studies (continued)

Source	Study Design	Population	Sample Size	Major Findings Pertinent to This Scoping Review	Sources of Funding; Conflicts of Interest
Healey et al ¹	Retrospective	Surgical inpatients	4658	Behind technical errors, diagnostic and judgment errors were the second most common cause of preventable harm	None reported
Henry et al ²⁷	Retrospective	ICU patients	16 234	A machine learning early warning score accurately predicted the onset of septic shock (AUC 0.83), identifying approximately two-thirds of all cases prior to the onset of organ dysfunction	NSF; Google Research, Gordon and Betty Moore Foundation
Hubbard et al ²⁸	Prospective	Trauma patients	980	A machine learning ensemble predicted mortality more accurately than logistic regression (5% gain)	US Army Medical Research and Materiel Command, NIH
Hyde et al ⁷	Prospective	Patients undergoing colorectal resections	288	The likelihood of a serious complication was underestimated by the ACS Surgical Risk Calculator (AUC, 0.69), but the calculator accurately predicted postoperative mortality (AUC, 0.97)	None reported
Kim et al ²⁹	Retrospective	ICU admissions	38 474	A decision tree model predicted in-hospital mortality more accurately than APACHE III (AUC, 0.89 vs 0.87)	NCRR
Knops et al ³⁰	Systematic review	Studies about decision aids in surgery	17	Decision aid use was associated with more knowledge regarding treatment options and preference for less invasive treatment options with no observable differences in anxiety, quality of life, or complications	None reported
Komorowski et al ³¹	Retrospective	Septic ICU patients	96 156	A reinforcement learning model recommending intravenous fluid and vasopressor strategies outperformed human clinicians; mortality was lowest when decisions made by clinicians matched recommendations from the reinforcement learning model	Orion Pharma, Amomed Pharma, Ferring Pharma, Tenax Therapeutics; Baxter Healthcare; Bristol-Myers Squibb; GSK; HCA International
Koyner et al ³²	Retrospective	Hospital admissions	121 158	A machine learning algorithm accurately predicted development of acute kidney injury within 24 h (AUC, 0.90) and 48 h (AUC, 0.87)	Satellite Healthcare; Philips Healthcare; EarlySense; Quant HC
Leeds et al ⁸	Observational	Surgery residents	124	Residents reported that lack of electronic and clinical workflow integration were major barriers to routine use of risk communication frameworks	NCI; ASCRS; AHRQ
Légaré et al ³³	Systematic review	Studies about shared decision-making	38	Time constraints impair the shared decision-making process among providers, patients, and caregivers	Tier 2 Canada Research Chair
Loftus et al ³⁴	Retrospective	Patients with lower intestinal bleeding	147	An artificial neural network predicted severe lower intestinal bleeding more accurately than a traditional clinical prediction rule (AUC, 0.98 vs 0.66)	NIGMS, NCATS
Lubitz et al ⁵	Retrospective	Patients undergoing colorectal surgery	150	The ACS Surgical Risk Calculator accurately predicted morbidity and mortality for elective surgery but underestimated risk for emergent surgery	None reported
Ludolph et al ³⁵	Systematic review	Articles about debiasing in health care	68	Many debiasing strategies targeting health care clinicians effectively decrease the effect of bias on decision-making	University of Lugano Institute of Communication and Health
Lundgren-Laine et al ³⁶	Observational	Academic intensivists	8	Academic intensivists made approximately 56 ad hoc patient care and resource use decisions per day	Finnish Funding Agency for Technology and Innovation; Tekes; Finnish Cultural Foundation
Morris et al ³⁷	Interviews	Academic surgeons	20	Younger surgeons felt uncomfortable defining futility and felt pressured to perform operations that were likely futile	AHRQ
Pirracchio et al ³⁸	Retrospective	ICU patients	24 508	A machine learning ensemble predicted in-hospital mortality (AUC, 0.85) more accurately than SAPS-II (AUC, 0.78) and SOFA (AUC, 0.71)	Fulbright Foundation; Doris Duke Clinical Scientist Development Award; NIH
Pirracchio et al ³⁹	Observational	Simulated data sets	1000	A machine learning ensemble predicted propensity scores more accurately than logistic regression and individual machine learning algorithms	Fulbright Foundation; Assistance Publique-Hôpitaux de Paris; NIH
Raymond et al ³	Prospective	Preoperative clinic patients	150	After reviewing ACS Surgical Risk Calculator results, 70% would participate in prehabilitation and 40% would delay surgery for prehabilitation	GE Foundation; Edwards Lifesciences; Cheetah Medical

(continued)

Table 1. Summary of Included Studies (continued)

Source	Study Design	Population	Sample Size	Major Findings Pertinent to This Scoping Review	Sources of Funding; Conflicts of Interest
Sacks et al ⁴⁰	Observational	Surgeons	767	Facing clinical vignettes for urgent and emergent surgical diseases; surgeons exhibited wide variability in the decision to operate (49%-85%)	Robert Wood Johnson/Veterans Affairs Clinical Scholars program
Schuetz et al ⁴¹	Retrospective	Clinical encounters in an EHR	32 787	A deep model predicted the onset of heart failure more accurately than logistic regression (AUC, 0.78 vs AUC, 0.75)	NSF; NHLBI
Shanafelt et al ²	Observational	Members of the ACS	7905	Nine percent of all surgeons reported making a major medical error in the last 3 mo, and lapses in judgment were the most common cause (32%)	None reported
Shickel et al ⁴²	Retrospective	ICU admissions	36 216	A deep model using SOFA variables predicted in-hospital mortality with greater accuracy than the traditional SOFA score (AUC 0.90 vs 0.85)	NIGMS; NSF, University of Florida CTSI; NCATS; J Crayton Pruitt Family Department of Biomedical Engineering; Nvidia
Singh et al ⁴³	Systematic review	Articles about CRP to predict leak after colorectal surgery	7	The positive predictive value of serum C-reactive protein 3-5 d after surgery was 21%-23%	Auckland Medical Research Foundation, New Zealand Health Research Council
Stacey et al ⁴⁴	Systematic review	Randomized trials about decision aids	105	Participants exposed to decision aids felt that they were more knowledgeable, informed, and clear about their values and played a more active role in the shared decision-making process	Foundation for Informed Medical Decision Making, Healthwise
Strate et al ⁴⁵	Prospective	Patients with acute lower intestinal bleeding	275	A bedside clinical prediction rule using simple cutoff values predicted severe lower intestinal bleeding (AUC, 0.75)	American College of Gastroenterology, National Research Service Award, American Society for Gastrointestinal Endoscopy
Sun et al ⁴⁶	Observational	Simulated type 1 diabetics	100	A reinforcement learning model performed as well as standard intermittent self-monitoring and continuous glucose monitoring methods, but with fewer episodes of hypoglycemia	Swiss Commission of Technology and Innovation
Van den Bruel et al ⁴⁷	Retrospective	Primary care patients	3890	Clinician intuition identified patients with illness severity that was underrepresented by traditional clinical parameters	Research Foundation Flanders, Eurogenerics, NIHR
Van den Bruel et al ⁴⁸	Systematic review	Articles about clinical parameters for serious infections	30	Traditional clinical parameters associated with serious infection were often absent among patients with serious infections	Health Technology Assessment, NIHR
Vohs et al ⁴⁹	Observational	Undergraduate students	34	Higher decision-making volume was associated with decreased physical stamina, persistence, quality and quantity of mathematic calculations, and more procrastination	NIH, Social Sciences and Humanities Research Council, Canada Research Chair Council, McKnight Land-Grant

Abbreviations: ACS, American College of Surgeons; AHA, American Heart Association; AHRQ, Agency for Healthcare Research and Quality; ASCRS, American Society of Colon and Rectal Surgeons; AUC, area under the curve; CRP, C-reactive protein; CTSI, Clinical and Translational Sciences Institute; EHR, electronic health record; GSK, GlaxoSmithKline; ICU, intensive care unit; NCATS, National Center for Advancing Translational Sciences; NCI, National Cancer Institute; NCRR, National Center for

Research Resources, Acute Physiology, and Chronic Health Evaluation; NHLBI, National Heart, Lung, and Blood Institute; NIGMS, National Institute of General Medical Sciences; NIH, National Institutes of Health; NIHR, National Institute for Health Research; NSF, National Science Foundation; SAPS, Simplified Acute Physiology Score; SCCM, Society of Critical Care Medicine; SOFA, Sequential Organ Failure Assessment.

Challenges in Surgical Decision-making

Complexity

In the hypothetical-deductive decision-making model that dominates surgical decision-making, initial patient presentations are assessed to develop a list of possible diagnoses that are differentiated by diagnostic testing or response to empirical therapy. This depends on the surgeon's ability to form a complete list of all likely diagnoses, all life-threatening diagnoses, and all unlikely diagnoses that may be considered if the initial workup excludes other causes. It also requires recognition of strengths and limitations of available tests. Once the diagnosis is established, the surgeon must recommend a plan using sound judgment. Each step introduces variability and opportunities for error.⁴⁰

Values and Emotions

Patient values are individualized by nature, precluding the creation of a criterion standard of optimal decision-making. Understanding

and incorporating these values is essential to an effective shared decision-making process.⁵⁰ This may be accomplished by simply asking patients and caregivers about their goals of care and what they value most in life. Shared decision-making improves patient satisfaction and compliance and may reduce costs associated with undesired tests and treatments. However, patients, caregivers, and clinicians often misunderstand one another, their goals may differ, and patients and caregivers are often expected to make decisions with limited background knowledge and no medical training.^{13,33,50} Surgical diseases may evoke fear and anger, which influence perceptions of risks and benefits.^{23,51} Emotions surrounding an acute surgical condition may also create a sense of urgency and pressure on surgeons to perform futile operations.³⁷

Time Constraints and Uncertainty

Surgical decision-making is often hindered by uncertainty owing to missing or incomplete data. This occurs when decisions regarding

Figure 1. Surgical Decision-making Paradigm

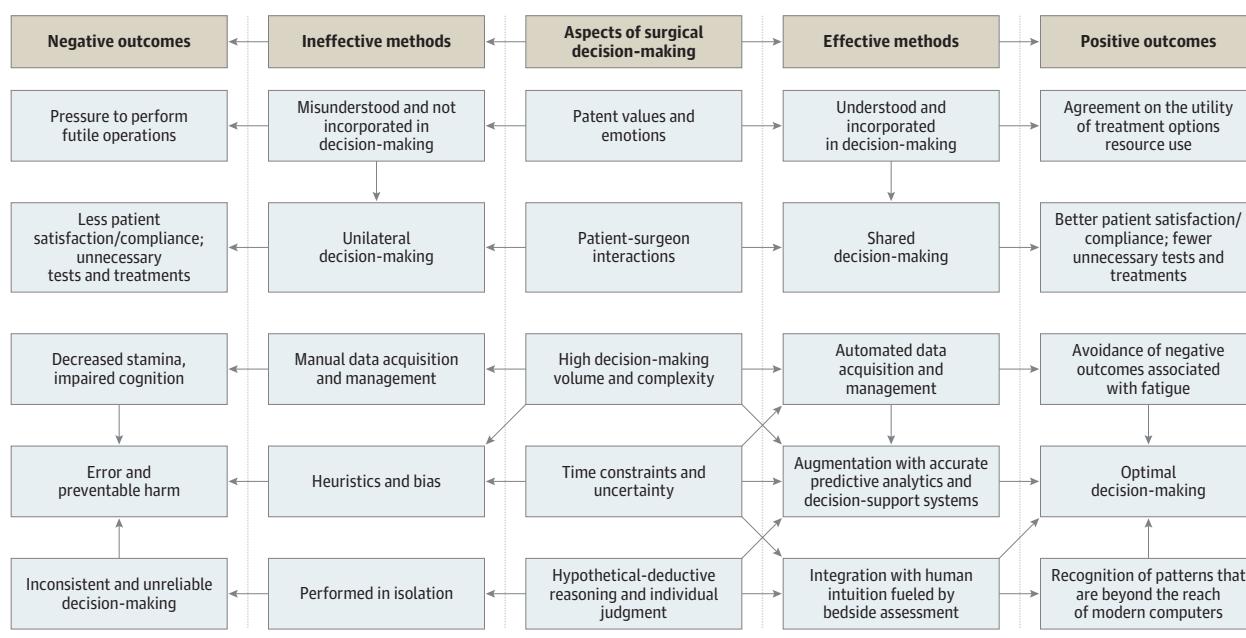


Table 2. Sources of Bias in Surgical Decision-making

Source of Bias	Examples
Framing effect	A clinician presents a clinical scenario to a surgeon in different context than the surgeon would have perceived during an independent assessment
Overconfidence bias	A surgeon falsely perceives that weaknesses and failures disproportionately affect their peers
Commission bias	A surgeon tends toward action when inaction may be preferable, especially in the context of overconfidence bias
Anchoring bias	Patients are informed of expected outcomes using data from aggregate patient populations without adjusting for their personalized risk profile
Recall bias	Recent experiences with a certain patient population or operation disproportionately affect surgical decision-making relative to remote experiences
Confirmation bias	Outcomes are predicted using personal beliefs rather than evidence-based guidelines

an urgent or emergent condition must be made before all relevant data can be gathered and analyzed. Nonurgent decisions may be hindered by time constraints and uncertainty owing to sheer decision-making volume, the time-consuming nature of manual data acquisition, and team dynamics. Academic intensivists make approximately 56 patient care and resource use decisions per day.³⁶ In an assessment of medical student and resident intensive care unit (ICU) patient presentations, potentially important data were omitted from 157 of 157 presentations.¹⁰ Even when data collection and analysis are complete, high decision-making volume begets decision fatigue, manifesting as procrastination, less persistence when facing adversity, decreased physical stamina, and lower quality and quantity of mathematic calculations.⁴⁹ These impairments are exacerbated by acute and chronic sleep deprivation, which occurs in as many as two-thirds of all acute care surgeons taking in-house call.^{52,53} For a surgical oncologist with a busy outpatient clinic, automated production of prognostic data from artificial intelligence

models could improve efficiency and preserve face-to-face patient-surgeon interactions by obviating manual data acquisition and entry into prognostic models.

Heuristics and Bias

When facing time constraints and uncertainty, decision-making may be influenced by heuristics or cognitive shortcuts.^{54,55} Heuristics may lead to bias or predictable and systematic cognitive errors, as described in Table 2.^{16,35}

Traditional Predictive Analytics and Clinical Decision Support

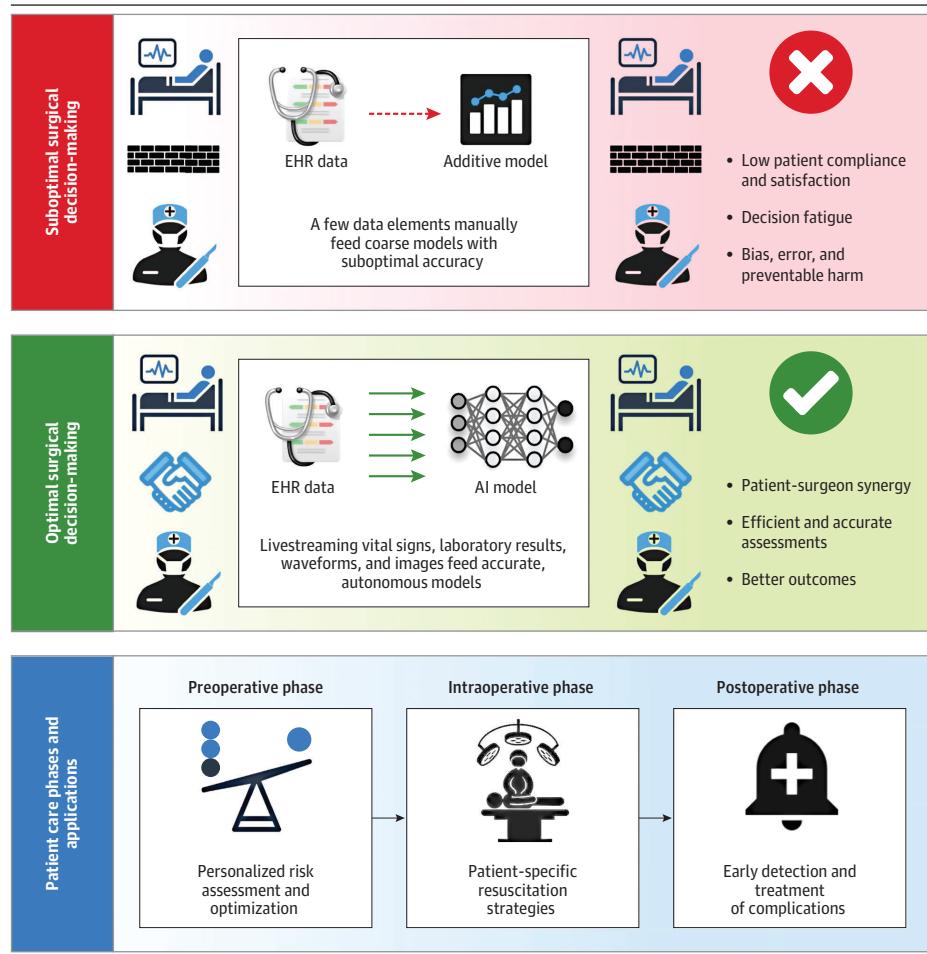
Decision Aids

Decision aids provide specific patient populations with background information, options for diagnosis and treatment, risks and benefits for each option, and outcome probabilities. In a systematic review⁴⁴ including 31 043 patients facing screening or treatment decisions, patients exposed to decision aids felt more knowledgeable and played a more active role in the decision-making process. In a systematic review of 17 studies investigating decisions made by surgical patients, decision aids were associated with more knowledge regarding treatment options, preference for less invasive treatments, and no observable differences in anxiety, quality of life, morbidity, or mortality.³⁰ However, because decision aids apply to heterogeneous patient populations with 1 common clinical presentation or choice, they do not consider individual patient physiology and risk factors.

Prognostic Scoring Systems

Traditional prognostic scoring systems use regression modeling on aggregate patient populations to identify static variable risk factor thresholds, which are applied to individual patients. For example, elevated serum levels of C-reactive protein (CRP) are associated with anastomotic leak after colorectal surgery. A meta-analysis⁴³

Figure 2. Optimal and Suboptimal Approaches to Surgical Decision-making



AI indicates artificial intelligence; EHR, electronic health record.

found that the optimal postoperative day 3 CRP cutoff value was 172 mg/L (to convert to nanomoles per liter, multiply by 9.524). This is easy to apply at the bedside but does not accurately reflect pathophysiology. Serum CRP has a relatively constant half-life, and its production is directly associated with inflammation along a continuum.⁵⁶ If 4 different patients have CRP levels of 10 mg/L, 171 mg/L, 173 mg/L, and 1000 mg/L 3 days after a colectomy, few clinicians would group these patients according to the 172 mg/L cutoff. The negative predictive value was 97%, such that a low value usually indicates no leak, but the positive predictive value was 21%.

Most diseases are not driven by a single physiologic parameter; therefore, prognostic scoring systems often incorporate multiple parameters for tasks such as measuring illness severity and predicting stroke and severe gastrointestinal bleeding.^{24,45,57} Parametric regression prognostic scoring systems assume that relationships among input variables are linear.^{22,29} When the relationships are nonlinear, the scoring system is similar to a coin toss.¹¹

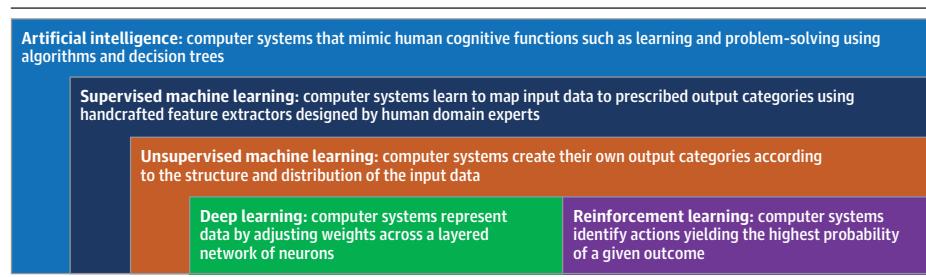
To facilitate clinical adoption, prognostic scoring systems have been implemented as online risk calculators. The NSQIP Surgical Risk Calculator is a prominent example. Calculator use may increase the likelihood that patients will participate in risk-reduction strategies such as prehabilitation.³ However, input variables must be entered

manually, and its predictive accuracy is suboptimal, especially for nonelective operations, representing opportunities for improvement.⁴⁻⁷

Artificial Intelligence Predictive Analytics and Augmented Decision-making

In 1970, William B. Schwartz published a Special Article in the *New England Journal of Medicine* stating, "Computing science will probably exert its major effects by augmenting and, in some cases, largely replacing the intellectual functions of the physician."⁵⁸ Despite extraordinary advances in computer technology, this vision has not been realized. Several factors may contribute. Traditional clinical decision-support systems require time-consuming manual data acquisition and entry, which impairs their adoption.^{8,33} Even the most successful and widely used static variable cutoff values do not accurately represent individual patient pathophysiology, as reflected by their suboptimal accuracy.^{34,43,56} Parametric regression equations also fail to represent the complex, nonlinear associations among input variables, further limiting the accuracy of traditional multivariable regression models.^{22,29} The weaknesses of traditional approaches may be overcome by artificial intelligence models fed with livestreaming intraoperative and EHR data to augment surgical decision-making through preoperative, intraoperative, and postoperative phases of care (Figure 2).

Figure 3. Summary of Artificial Intelligence Techniques



Artificial intelligence refers to computer systems that mimic human cognitive functions such as learning and problem-solving. In the broadest sense, a computer program using simple decision tree functions can mimic human intelligence. However, artificial intelligence usually refers to computer systems that learn from raw data with some degree of autonomy, as occurs with machine learning, deep learning, and reinforcement learning (Figure 3). Whereas traditional clinical decision-support systems use rules to generate codes and algorithms, artificial intelligence models learn from examples. Herein lies the strength of artificial intelligence for predictive analytics in medicine: human disease is simply too broad and complex to be explained and interpreted by rules.^{59,60}

Machine Learning

Machine learning is a subfield of artificial intelligence in which a computer system performs a task without explicit instructions. Supervised machine learning models require human domain expertise and computer engineering to design handcrafted feature extractors capable of transforming raw data into desired representations. The algorithm learns associations between input data and prescribed output categories. Once trained, a supervised model is capable of classifying new unseen input data. With unsupervised techniques, input data have no corresponding annotated output categories; the algorithm creates its own output categories according to the structure and distribution of the input data. This approach allows discovery of patterns and phenotypes that were unrecognized prior to model development.

Machine learning has been used to accurately predict sepsis, in-hospital mortality, and acute kidney injury using intraoperative time-series data.^{9,21,27,32} Each machine learning algorithm has distinct advantages and disadvantages for different tasks such that performance depends on fit between algorithm and task. To capitalize on this phenomenon, SuperLearner ranks a set of candidate algorithms^{20,28,38,39} by their performance and applies an optimal weight to each, creating ensemble algorithms that can accurately predict transfusion requirements and mortality among trauma patients.^{20,28,38,39} Supervised and unsupervised machine learning input features must be handcrafted using domain knowledge. In deep learning, features are extracted by the model itself.

Deep Learning

Deep learning is a subfield of machine learning in which computer systems learn and represent highly dimensional data by adjusting weighted associations among input variables across a layered hierarchy of neurons or artificial neural network. Early warning systems that alert clinicians to unstable vital signs illustrate data dimension-

ality. As the number of vital sign data sources increases linearly, the combinations of alarm parameters that trigger early warning system alarms increase exponentially, resulting in frequent false alarms. Even without a corresponding exponential increase in observations, data are highly dimensional when many variables are used to represent a single patient or event, especially when the number of patients or events in the data set is relatively low, producing unique and rare mixtures of data. Prediction models are less effective when classifying mixtures of data that are rare or absent in the development or training data set. The ability of deep models to represent highly dimensional data is important to their application to surgical decision-making.

In deep models, the initial input and final output layers are connected by hidden layers containing hidden nodes. Each hidden node is assigned a weight that is influenced by previous layers, affects the output from that neuron, and has the potential to affect the outcome classification of the entire network. An algorithm optimizes and updates weights as the model is trained to achieve the strongest possible association between input and output layers. This structure allows accurate representation of chaotic and nonlinear yet meaningful relationships among input features. Deep models automatically learn optimal feature representations from raw data without handcrafted feature engineering, providing a logistical advantage over machine learning models that require time-intensive feature engineering.⁶¹ Automatic feature extraction also promotes discovery of novel patterns and phenotypes that may have been overlooked by handcrafted feature selection techniques.

Clinical applications of deep learning benefit from the ability to include multiple different types and sources of data as inputs for a single model, including wearable sensors and cameras capturing patient movements and facial expressions with computer vision, an artificial intelligence subfield in which deep models use pixels from images and videos as inputs.^{60,62,63} Deep models have successfully performed patient phenotyping, disease prediction, and mortality prediction tasks.^{19,26,41,64} When applied to the same variable set used to calculate SOFA scores, deep models outperform traditional SOFA modeling in predicting in-hospital mortality for ICU patients.⁴² Preliminary data suggest that deep models are theoretically capable of accurately predicting risk for perioperative and postoperative complications and augmenting recommendations for operative management and the informed consent process. Despite their utility for predictive analytics, deep learning only provides outcome probabilities that loosely correspond to specific decisions and actions. In contrast, reinforcement learning is well suited to support specific decisions made by patients, caregivers, and surgeons.

Reinforcement Learning

Reinforcement learning is an artificial intelligence subfield in which computer systems identify actions yielding the highest probability of an outcome. Reinforcement models can be trained by series of trial and error scenarios, exposing the model to expert demonstrations, or a combination of these strategies. This occurs in a Markov decision process framework, consisting of a set of states, a set of actions, the probability that a certain action in a certain state will lead to a new state, and the reward that results from the new state. Using this framework, the system creates a policy that identifies the choice or action with the highest probability of a desired outcome, assessing total rewards attributable to multiple actions performed over time and the relative importance of present and future rewards, facilitating application of reinforcement learning to clinical scenarios that evolve over time.

Reinforcement learning has been used to recommend optimal fluid resuscitation and vasopressor administration strategies for patients with sepsis.³¹ Ninety-day mortality was lowest when care provided by clinicians was concordant with model recommendations. Reinforcement learning has also been used to recommend basal and bolus insulin administration for virtual type 1 diabetics.⁴⁶ The algorithm performed as well as standard intermittent self-monitoring and continuous glucose monitoring methods, but with fewer episodes of hypoglycemia. Similar methods could be applied to augment the decision to operate.

Implementation

Automated Electronic Health Record Data

The Health Information Technology for Economic and Clinical Health Act of 2009 incentivized adoption of EHR systems.⁶⁵ Within 6 years, more than 4 of 5 US hospitals adopted EHRs.⁶⁶ The volume of data generated by EHRs is staggering and will likely increase over time. Approximately 153 billion GB of data were generated in 2013, with projected growth of 48% per year.⁶⁷ This data volume is ideal for artificial intelligence models, which thrive on large data sets.

Because EHRs are continuously updated as patient data become available, artificial intelligence models can provide real-time predictions and recommendations. Works published within the last year demonstrate the feasibility of this approach. The *MySurgeryRisk* platform uses EHR data for 285 variables to predict 8 postoperative complications with an area under the curve (AUC) of 0.82-0.94 and to predict mortality at 1, 3, 6, 12, and 24 months with an AUC of 0.77-0.83.¹⁵ Electronic health record data feed the algorithm automatically, obviating manual data search and entry and overcoming a major obstacle to clinical adoption. In a prospective study, the algorithm predicted postoperative complications with greater accuracy than physicians.¹⁷

Mobile Device Outputs

To optimize clinical utility and facilitate adoption, automated model outputs could be provided to mobile devices. This would require several elements that communicate with one another reliably and efficiently, including robust quality filters, a public key infrastructure, and encryption that can only be deciphered by the intended receiver.⁶⁸ Model outputs could be provided to mobile devices equipped with the appropriate RestAPI client-server relationship and security clearance or through Google Cloud Messaging. To our knowledge, automated surgical risk predictions with mobile device out-

puts have not yet been reported. However, efforts to use manual data entry to feed machine learning models for surgical risk prediction on mobile devices have been successful.¹⁴

Human Intuition

Human intuition seems to arise from dopaminergic limbic system neurons that modify their connections with one another when a certain pattern or situation leads to a reward or penalty such as pleasure or pain.^{69,70} Subsequently, similar patterns or situations evoke positive and negative emotions, or *gut feelings*, which are powerful and effective decision-making tools. In a sentinel investigation¹² of intuitive decision-making, participants drew cards from 1 of 4 decks for a cash reward. Two decks were rigged to be advantageous and 2 were rigged to be disadvantageous. Participants could explain differences between decks after drawing 80 cards, but demonstrated measurable anxiety and perspiration when reaching for a disadvantageous deck after drawing 10 cards and began to favor the advantageous deck after 50 cards before they could consciously explain what they were doing or why they were doing it. Similar phenomena occur in fight-or-flight survival responses, naval warfare, and financial decision-making.^{71,72} Intuition can also identify patients with life-threatening conditions that would be underappreciated by traditional clinical parameters alone.^{47,48}

Challenges to Adoption

Data Standardization and Technology Infrastructure

To produce models that may be integrated with any EHR in any setting, data must be standardized. The Fast Healthcare Interoperability Resources framework establishes standards for health information exchange using a set of universal components assembled into systems that facilitate data sharing across EHRs and cloud-based communications. In addition, the Epic EHR that dominates the market has exclusive rights to develop new functions. To avoid legal conflicts, virtual models can live outside the EHR.¹⁵ However, this requires technology infrastructure that is not currently available in all clinical settings.

Interpretability

Diligent clinicians and informed patients will want to know why a computer program made a certain prediction or recommendation. Several techniques address this challenge, including attention mechanisms that reveal periods during which model inputs contributed disproportionately to the output, plotting pairwise similarities between data points to display phenotypic clusters, and training models on labeled patient data and then a linear gradient-boosting tree so that the model will assign relative importance to patient data input features.^{18,42,73}

Safety and Monitoring

If model inputs are flawed or model outputs are not carefully monitored by data scientists and interpreted by astute clinicians, many patients could be harmed in a short time frame. Artificial intelligence models trained on erroneous or misrepresentative data are likely to obscure the truth. Because studies with positive results are more likely to be submitted and published, artificial intelligence literature may be overly optimistic. Prior to clinical implementation, machine and deep learning models must be rigorously analyzed in a

retrospective fashion and externally validated to ensure generalizability. Performing a stress test of artificial intelligence models by simulating erroneous and rare model inputs and assessing how the model responds may allow clinicians to better understand how and why failures occur. Initial prospective implementation should occur on a small scale under close monitoring, similar to phase 1 and 2 clinical trials for experimental medications, with analysis of how decision-support tools affect decisions across populations and among individual patients.⁷⁴ In cooperation with the International Medical Device Regulators Forum, the US Food and Drug Administration created the Software as Medical Device category and developed a voluntary Software Precertification Program to aid health care software developers in creating, testing, and implementing Software as Medical Device. Medicolegal regulation of Software as Medical Device is not rigidly defined.

Ethical Challenges

When algorithms are trained on data sets that are influenced by bias, algorithm outputs will likely reflect similar bias. In 1 prominent example, a model designed to augment judicial decision-making by predicting the likelihood of crime recidivism demonstrated predilection for racial/ethnic discrimination.⁷⁵ When data used to train an algorithm are predominantly derived from patient populations with different demographics than the patient for whom the algorithm is applied, accuracy may suffer. For example, the Framingham heart study primarily included white participants. A model trained on this data may reflect racial and ethnic bias because associations between cardiovascular risk factors and events differ by race and

ethnicity.²⁵ Accountability for errors poses another challenge. Our justice system is well-equipped to address scenarios in which an individual clinician is responsible for making an errant decision, but it may prove difficult to assign blame to a computer program and its developers.

Conclusions

Surgical decision-making is impaired by time constraints, uncertainty, complexity, decision fatigue, hypothetical-deductive reasoning, and bias, leading to preventable harm. Traditional decision-support systems are compromised by time-consuming manual data entry and suboptimal accuracy. Automated artificial intelligence models fed with livestreaming EHR data can address these weaknesses. Successful integration of artificial intelligence with surgical decision-making would require data standardization, advances in model interpretability, careful implementation and monitoring, attention to ethical challenges, and preservation of bedside assessment and human intuition in the decision-making process. Artificial intelligence models must be rigorously analyzed in a retrospective fashion with robust external validation prior to prospective clinical application under the close scrutiny of astute clinicians and data scientists. Properly applied, artificial intelligence has the potential to transform surgical care by augmenting the decision to operate, the informed consent process, identification and mitigation of modifiable risk factors, recognition and management of complications, and shared decisions regarding resource use.

ARTICLE INFORMATION

Accepted for Publication: August 24, 2019.

Published Online: December 11, 2019.

doi:10.1001/jamasurg.2019.4917

Author Affiliations: Department of Surgery, University of Florida Health, Gainesville (Loftus, Filiberto, Efron, Brakenridge, Mohr, Upchurch); Departments of Anesthesiology, Orthopedics, and Information Systems/Operations Management, University of Florida Health, Gainesville (Tighe); Departments of Biomedical Engineering, Computer and Information Science and Engineering, and Electrical and Computer Engineering, University of Florida, Gainesville (Rashidi); Department of Medicine, University of Florida Health, Gainesville (Bihorac).

Author Contributions: Dr Bihorac had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Loftus, Tighe, Efron, Mohr, Rashidi, Bihorac.

Acquisition, analysis, or interpretation of data: Filiberto, Brakenridge, Mohr, Upchurch.

Drafting of the manuscript: Loftus, Efron, Rashidi.

Critical revision of the manuscript for important intellectual content: Tighe, Filiberto, Efron, Brakenridge, Mohr, Upchurch, Bihorac.

Obtained funding: Brakenridge, Rashidi.

Administrative, technical, or material support: Loftus, Tighe, Efron, Brakenridge, Mohr, Upchurch.

Supervision: Tighe, Filiberto, Efron, Rashidi, Upchurch, Bihorac.

Conflict of Interest Disclosures: Dr Tighe reported grants from the National Institutes of Health during the conduct of the study. Dr Rashidi reported patents to Method and Apparatus for Pervasive Patient Monitoring pending and Systems and Methods for Providing an Acuity Score for Critically Ill or Injured Patients pending. Dr Bihorac reported grants from the National Institutes of Health and the National Science Foundation during the conduct of the study; in addition, Dr Bihorac has a patent to Systems and Methods for Providing an Acuity Score for Critically Ill or Injured Patients pending. No other disclosures were reported.

Funding/Support: Dr Efron was supported by R01 GM113945-01 from the National Institute of General Medical Sciences (NIGMS). Drs Bihorac and Rashidi were supported by R01 GM110240 from the NIGMS. Drs Bihorac and Efron were supported by P50 GM-111152 from the NIGMS. Dr Rashidi was supported by CAREER award NSF-IIS 1750192 from the National Science Foundation, Division of Information and Intelligent Systems, and by the National Institute of Biomedical Imaging and Bioengineering (grant R21EB027344-01). Dr Tighe was supported by R01GM114290 from the NIGMS. Dr Loftus was supported by a postgraduate training grant (T32 GM-008721) in burns, trauma, and perioperative injury from the NIGMS.

Role of the Funder/Sponsor: The National Institute of General Medical Sciences, National Science Foundation, and the National Institute of Biomedical Imaging and Bioengineering had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the

manuscript; and decision to submit the manuscript for publication.

REFERENCES

1. Healey MA, Shackford SR, Osler TM, Rogers FB, Burns E. Complications in surgical patients. *Arch Surg*. 2002;137(5):611-617. doi:10.1001/archsurg.137.5.611
2. Shanafelt TD, Balch CM, Bechamps G, et al. Burnout and medical errors among American surgeons. *Ann Surg*. 2010;251(6):995-1000. doi:10.1097/SLA.0b013e3181b7dab3
3. Raymond BL, Wanderer JP, Hawkins AT, et al. Use of the American College of Surgeons National Surgical Quality Improvement Program Surgical Risk Calculator during preoperative risk discussion: the patient perspective. *Anesth Analg*. 2019;128(4):643-650. doi:10.1213/ANE.0000000000003718
4. Clark DE, Fitzgerald TL, Dibbins AW. Procedure-based postoperative risk prediction using NSQIP data. *J Surg Res*. 2018;221:322-327. doi:10.1016/j.jss.2017.09.003
5. Lubitz AL, Chan E, Zarif D, et al. American College of Surgeons NSQIP risk calculator accuracy for emergent and elective colorectal operations. *J Am Coll Surg*. 2017;225(5):601-611. doi:10.1016/j.jamcollsurg.2017.07.1069
6. Cohen ME, Liu Y, Ko CY, Hall BL. An examination of American College of Surgeons NSQIP surgical risk calculator accuracy. *J Am Coll Surg*. 2017;224(5):787-795.e1.
7. Hyde LZ, Valizadeh N, Al-Mazrou AM, Kiran RP. ACS-NSQIP risk calculator predicts cohort but not individual risk of complication following colorectal

resection. *Am J Surg.* 2019;218(1):131-135. doi:10.1016/j.amjsurg.2018.11.017

8. Leeds IL, Rosenblum AJ, Wise PE, et al. Eye of the beholder: risk calculators and barriers to adoption in surgical trainees. *Surgery.* 2018;164(5):1117-1123. doi:10.1016/j.surg.2018.07.002

9. Adhikari L, Ozrazgat-Baslanti T, Ruppert M, et al. Improved predictive models for acute kidney injury with IDEA: Intraoperative Data Embedded Analytics. *PLoS One.* 2019;14(4):e0214904. doi:10.1371/journal.pone.0214904

10. Artis KA, Bordley J, Mohan V, Gold JA. Data omission by physician trainees on ICU rounds. *Crit Care Med.* 2019;47(3):403-409. doi:10.1097/CCM.0000000000003557

11. Bagnall NM, Pring ET, Malietzis G, et al. Perioperative risk prediction in the era of enhanced recovery: a comparison of POSSUM, ACPGBI, and E-PASS scoring systems in major surgical procedures of the colorectal surgeon. *Int J Colorectal Dis.* 2018;33(11):1627-1634. doi:10.1007/s00384-018-3141-4

12. Bechara A, Damasio H, Tranel D, Damasio AR. Deciding advantageously before knowing the advantageous strategy. *Science.* 1997;275(5304):1293-1295. doi:10.1126/science.275.5304.1293

13. Bertrand PM, Pereira B, Adda M, et al. Disagreement between clinicians and score in decision-making capacity of critically ill patients. *Crit Care Med.* 2019;47(3):337-344. doi:10.1097/CCM.0000000000003550

14. Bertsimas D, Dunn J, Velmahos GC, Kaafarani HMA. Surgical risk is not linear: derivation and validation of a novel, user-friendly, and machine-learning-based Predictive Optimal Trees in Emergency Surgery Risk (POTTER) calculator. *Ann Surg.* 2018;268(4):574-583. doi:10.1097/SLA.0000000000002956

15. Bihorac A, Ozrazgat-Baslanti T, Ebadi A, et al. MySurgeryRisk: development and validation of a machine-learning risk algorithm for major complications and death after surgery. *Ann Surg.* 2019;269(4):652-662. doi:10.1097/SLA.0000000000002706

16. Blumenthal-Barby JS, Krieger H. Cognitive biases and heuristics in medical decision making: a critical review using a systematic search strategy. *Med Decis Making.* 2015;35(4):539-557. doi:10.1177/0272989X14547740

17. Brennan M, Puri S, Ozrazgat-Baslanti T, et al. Comparing clinical judgment with the MySurgeryRisk algorithm for preoperative risk assessment: a pilot usability study. *Surgery.* 2019;165(5):1035-1045. doi:10.1016/j.surg.2019.01.002

18. Che Z, Purushotham S, Khemani R, Liu Y. Interpretable deep models for ICU outcome prediction. *AMIA Annu Symp Proc.* 2017;2016:371-380.

19. Chen-Ying Hung, Wei-Chen Chen, Po-Tsun Lai, Ching-Heng Lin, Chi-Chun Lee. Comparing deep neural network and other machine learning algorithms for stroke prediction in a large-scale population-based electronic medical claims database. *Conf Proc IEEE Eng Med Biol Soc.* 2017;2017:3110-3113.

20. Christie SA, Hubbard AE, Calcutt RA, et al. Machine learning without borders? an adaptable tool to optimize mortality prediction in diverse clinical settings. *J Trauma Acute Care Surg.* 2018;85(5):921-927. doi:10.1097/TA.0000000000002044

21. Delahanty RJ, Kaufman D, Jones SS. Development and evaluation of an automated machine learning algorithm for in-hospital mortality risk adjustment among critical care patients. *Crit Care Med.* 2018;46(6):e481-e488. doi:10.1097/CCM.0000000000000301

22. Dybowski R, Weller P, Chang R, Gant V. Prediction of outcome in critically ill patients using artificial neural network synthesised by genetic algorithm. *Lancet.* 1996;347(9009):1146-1150. doi:10.1016/S0140-6736(96)90609-1

23. Ellis EM, Klein WMP, Orehek E, Ferrer RA. Effects of emotion on medical decisions involving tradeoffs. *Med Decis Making.* 2018;38(8):1027-1039. doi:10.1177/0272989X18806493

24. Gage BF, van Walraven C, Pearce L, et al. Selecting patients with atrial fibrillation for anticoagulation: stroke risk stratification in patients taking aspirin. *Circulation.* 2004;110(16):2287-2292. doi:10.1161/01.CIR.0000145172.55640.93

25. Gijsberts CM, Groenewegen KA, Hoefer IE, et al. Race/ethnic differences in the associations of the framingham risk factors with carotid IMT and cardiovascular events. *PLoS One.* 2015;10(7):e0132321. doi:10.1371/journal.pone.0132321

26. Hao Du, Ghassemi MM, Mengling Feng. The effects of deep network topology on mortality prediction. *Conf Proc IEEE Eng Med Biol Soc.* 2016;2016:2602-2605.

27. Henry KE, Hager DN, Pronovost PJ, Saria S. A targeted real-time early warning score (TREWScore) for septic shock. *Sci Transl Med.* 2015;7(299):299ra122. doi:10.1126/scitranslmed.aab3719

28. Hubbard A, Munoz ID, Decker A, et al; PROMMTT Study Group. Time-dependent prediction and evaluation of variable importance using superlearning in high-dimensional clinical data. *J Trauma Acute Care Surg.* 2013;75(1)(suppl 1):S53-S60. doi:10.1097/TA.0b013e3182914553

29. Kim S, Kim W, Park RW. A comparison of intensive care unit mortality prediction models through the use of data mining techniques. *Healthc Inform Res.* 2011;17(4):232-243. doi:10.4258/hir.2011.17.4.232

30. Knops AM, Legemate DA, Goossens A, Bossuyt PM, Ubbink DT. Decision aids for patients facing a surgical treatment decision: a systematic review and meta-analysis. *Ann Surg.* 2013;257(5):860-866. doi:10.1097/SLA.0b013e3182864fd6

31. Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA. The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. *Nat Med.* 2018;24(11):1716-1720. doi:10.1038/s41591-018-0213-5

32. Koyner JL, Carey KA, Edelson DP, Churpek MM. The development of a machine learning inpatient acute kidney injury prediction model. *Crit Care Med.* 2018;46(7):1070-1077. doi:10.1097/CCM.0000000000003123

33. Légaré F, Ratté S, Gravel K, Graham ID. Barriers and facilitators to implementing shared decision-making in clinical practice: update of a systematic review of health professionals' perceptions. *Patient Educ Couns.* 2008;73(3):526-535. doi:10.1016/j.pec.2008.07.018

34. Loftus TJ, Brakenridge SC, Croft CA, et al. Neural network prediction of severe lower intestinal bleeding and the need for surgical intervention. *J Surg Res.* 2017;212:42-47. doi:10.1016/j.jss.2016.12.032

35. Ludolph R, Schulz PJ. Debiasing health-related judgments and decision making: a systematic review. *Med Decis Making.* 2018;38(1):3-13.

36. Lundgrén-Laine H, Kontio E, Perttilä J, Korvenranta H, Forsström J, Salanterä S. Managing daily intensive care activities: an observational study concerning ad hoc decision making of charge nurses and intensivists. *Crit Care.* 2011;15(4):R188. doi:10.1186/cc10341

37. Morris RS, Ruck JM, Conca-Cheng AM, Smith TJ, Carver TW, Johnston FM. Shared decision-making in acute surgical illness: the surgeon's perspective. *J Am Coll Surg.* 2018;226(5):784-795. doi:10.1016/j.jamcollsurg.2018.01.008

38. Pirracchio R, Petersen ML, Carone M, Rigon MR, Chevret S, van der Laan MJ. Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): a population-based study. *Lancet Respir Med.* 2015;3(1):42-52. doi:10.1016/S2213-2600(14)70239-5

39. Pirracchio R, Petersen ML, van der Laan M. Improving propensity score estimators' robustness to model misspecification using super learner. *Am J Epidemiol.* 2015;181(2):108-119. doi:10.1093/aje/kwu253

40. Sacks GD, Dawes AJ, Ettner SL, et al. Surgeon perception of risk and benefit in the decision to operate. *Ann Surg.* 2016;264(6):896-903. doi:10.1097/SLA.0000000000001784

41. Choi E, Schuetz A, Stewart WF, Sun J. Using recurrent neural network models for early detection of heart failure onset. *J Am Med Inform Assoc.* 2017;24(2):361-370.

42. Shickel B, Loftus TJ, Adhikari L, Ozrazgat-Baslanti T, Bihorac A, Rashidi P. DeepSOFa: a continuous acuity score for critically ill patients using clinically interpretable deep learning. *Sci Rep.* 2019;9(1):1879. doi:10.1038/s41598-019-38491-0

43. Singh PP, Zeng IS, Srinivasa S, Lemanu DP, Connolly AB, Hill AG. Systematic review and meta-analysis of use of serum C-reactive protein levels to predict anastomotic leak after colorectal surgery. *Br J Surg.* 2014;101(4):339-346. doi:10.1002/bjs.9354

44. Stacey D, Légaré F, Lewis K, et al. Decision aids for people facing health treatment or screening decisions. *Cochrane Database Syst Rev.* 2017;4:CD001431. doi:10.1002/14651858.CD001431.pub5

45. Strate LL, Saltzman JR, Ookubo R, Mutinga ML, Syngal S. Validation of a clinical prediction rule for severe acute lower intestinal bleeding. *Am J Gastroenterol.* 2005;100(8):1821-1827. doi:10.1111/j.1572-0241.2005.41755.x

46. Sun Q, Jankovic M, Budzinski J, et al. A dual mode adaptive basal-bolus advisor based on reinforcement learning [published online Dec 17, 2018]. *IEEE J Biomed Health Inform.* doi:10.1109/JBHI.2018.2887067

47. Van den Bruel A, Thompson M, Buntinx F, Mant D. Clinicians' gut feeling about serious infections in children: observational study. *BMJ.* 2012;345:e6144. doi:10.1136/bmj.e6144

48. Van den Bruel A, Haj-Hassan T, Thompson M, Buntinx F, Mant D; European Research Network on Recognising Serious Infection Investigators.

Diagnostic value of clinical features at presentation to identify serious infection in children in developed countries: a systematic review. *Lancet*. 2010;375(9717):834-845. doi:10.1016/S0140-6736(09)62000-6

49. Vohs KD, Baumeister RF, Schmeichel BJ, Twenge JM, Nelson NM, Tice DM. Making choices impairs subsequent self-control: a limited-resource account of decision making, self-regulation, and active initiative. *J Pers Soc Psychol*. 2008;94(5):883-898. doi:10.1037/0022-3514.94.5.883

50. Kopecky KE, Urbach D, Schwarze ML. Risk calculators and decision aids are not enough for shared decision making. *JAMA Surg*. 2019;154(1):3-4. doi:10.1001/jamasurg.2018.2446

51. Ferrer RA, Green PA, Barrett LF. Affective science perspectives on cancer control: strategically crafting a mutually beneficial research agenda. *Perspect Psychol Sci*. 2015;10(3):328-345. doi:10.1177/1745691615576755

52. Coleman JJ, Robinson CK, Zarzaur BL, Timsina L, Rozycki GS, Feliciano DV. To sleep, perchance to dream: acute and chronic sleep deprivation in acute care surgeons. *J Am Coll Surg*. 2019;229(2):166-174. doi:10.1016/j.jamcollsurg.2019.03.019

53. Stickgold R. Sleep-dependent memory consolidation. *Nature*. 2005;437(7063):1272-1278. doi:10.1038/nature04286

54. Goldenson RM. *The Encyclopedia of Human Behavior; Psychology, Psychiatry, and Mental Health*. Garden City, NY: Doubleday; 1970.

55. Groopman JE. *How Doctors Think*. Boston: Houghton Mifflin; 2007.

56. Pepys MB, Hirschfield GM, Tennent GA, et al. Targeting C-reactive protein for the treatment of cardiovascular disease. *Nature*. 2006;440(7088):1217-1221. doi:10.1038/nature04672

57. Vincent JL, Moreno R, Takala J, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure: on behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med*. 1996;22(7):707-710. doi:10.1007/BF01709751

58. Schwartz WB. Medicine and the computer: the promise and problems of change. *N Engl J Med*. 1970;283(23):1257-1264. doi:10.1056/NEJM197012032832305

59. Schwartz WB, Patil RS, Szolovits P. Artificial intelligence in medicine: where do we stand? *N Engl J Med*. 1987;316(11):685-688. doi:10.1056/NEJM198703123161109

60. Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial intelligence in surgery: promises and perils. *Ann Surg*. 2018;268(1):70-76. doi:10.1097/SLA.00000000000002693

61. Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE J Biomed Health Inform*. 2018;22(5):1589-1604. doi:10.1109/JBHI.2017.2767063

62. Esteve A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. *Nat Med*. 2019;25(1):24-29. doi:10.1038/s41591-018-0316-z

63. Davoudi A, Malhotra KR, Shickel B, et al. Intelligent ICU for autonomous patient monitoring using pervasive sensing and deep learning. *Sci Rep*. 2019;9(1):8020. doi:10.1038/s41598-019-44004-w

64. Robinson PN. Deep phenotyping for precision medicine. *Hum Mutat*. 2012;33(5):777-780. doi:10.1002/humu.22080

65. Birkhead GS, Klompass M, Shah NR. Uses of electronic health records for public health surveillance to advance public health. *Annu Rev Public Health*. 2015;36:345-359. doi:10.1146/annurev-publhealth-031914-122747

66. Adler-Milstein J, Holmgren AJ, Kralovec P, Worzala C, Searcy T, Patel V. Electronic health record adoption in US hospitals: the emergence of a digital "advanced use" divide. *J Am Med Inform Assoc*. 2017;24(6):1142-1148. doi:10.1093/jamia/ocx080

67. Stanford Medicine. Stanford medicine 2017 health trends report: harnessing the power of data in health. <http://med.stanford.edu/content/dam/sm/sm-news/documents/StanfordMedicineHealthTrendsWhitePaper2017.pdf>. Accessed February 23, 2019.

68. Feng Z, Bhat RR, Yuan X, et al. Intelligent perioperative system: towards real-time big data analytics in surgery risk assessment. *DASC PiCom DataCom CyberSciTech 2017* (2017). 2017;2017:1254-1259. doi:10.1109/DASC-PiCom-DataCom-CyberSciTech.2017.201

69. Sanfey AG, Rilling JK, Aronson JA, Nystrom LE, Cohen JD. The neural basis of economic decision-making in the Ultimatum Game. *Science*. 2003;300(5626):1755-1758. doi:10.1126/science.1082976

70. Kahneman D. *Thinking, Fast and Slow*. New York, NY: Farrar, Straus and Giroux; 2013.

71. LeDoux J. Rethinking the emotional brain. *Neuron*. 2012;73(4):653-676. doi:10.1016/j.neuron.2012.02.004

72. Seymour B, Dolan R. Emotion, decision making, and the amygdala. *Neuron*. 2008;58(5):662-671. doi:10.1016/j.neuron.2008.05.020

73. van der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res*. 2008;9:2579-2605.

74. Shortliffe EH, Sepulveda MJ. Clinical decision support in the era of artificial intelligence. *JAMA*. 2018;320(21):2199-2200. doi:10.1001/jama.2018.17163

75. Angwin J, Larson J, Mattu S, Kirchner L. Machine bias. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>. Published May 23, 2016. Accessed January 24, 2019.