Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.JournalofSurgicalResearch.com

Intelligent, Autonomous Machines in Surgery

Tyler J. Loftus, MD,[a] *Amanda C. Filiberto, MD,*[a] *Jeremy Balch, MD,*[a] *Alexander L. Ayzengart, MD,*[a] *Patrick J. Tighe, MD,*[b] *Parisa Rashidi, PhD,*[c] *Azra Bihorac, MD,*[d] *and Gilbert R. Upchurch Jr., MD*[a,*]

[a] *Department of Surge ry, University of Florida Health, Gainesville, Florida*
[b] *Departments of Biomedical Engineering, Computer and Information Science and Engineering, and Electrical and Computer Engineering, University of Florida, Gainesville, Florida*
[c] *Departments of Anesthesiology, Orthopedics, and Information Systems/Operations Management, University of Florida Health, Gainesville, Florida*
[d] *Department of Medicine, University of Florida Health, Gainesville, Florida*

ARTICLE INFO

Article history:
Received 21 January 2020
Received in revised form
22 February 2020
Accepted 8 March 2020
Available online 24 April 2020

Keywords:
Surgery
Machine learning
Artificial intelligence
Innovation
Automation

ABSTRACT

Surgeons perform two primary tasks: operating and engaging patients and caregivers in shared decision-making. Human dexterity and decision-making are biologically limited. Intelligent, autonomous machines have the potential to augment or replace surgeons. Rather than regarding this possibility with denial, ire, or indifference, surgeons should understand and steer these technologies. Closer examination of surgical innovations and lessons learned from the automotive industry can inform this process. Innovations in minimally invasive surgery and surgical decision-making follow classic S-shaped curves with three phases: (1) introduction of a new technology, (2) achievement of a performance advantage relative to existing standards, and (3) arrival at a performance plateau, followed by replacement with an innovation featuring greater machine autonomy and less human influence. There is currently no level I evidence demonstrating improved patient outcomes using intelligent, autonomous machines for performing operations or surgical decision-making tasks. History suggests that if such evidence emerges and if the machines are cost effective, then they will augment or replace humans, initially for simple, common, rote tasks under close human supervision and later for complex tasks with minimal human supervision. This process poses ethical challenges in assigning liability for errors, matching decisions to patient values, and displacing human workers, but may allow surgeons to spend less time gathering and analyzing data and more time interacting with patients and tending to urgent, critical—and potentially more valuable—aspects of patient care. Surgeons should steer these technologies toward optimal patient care and net social benefit using the uniquely human traits of creativity, altruism, and moral deliberation.

* *Corresponding author*. Department of Surgery, Edward R. Woodward Professor & Chair University of Florida, PO Box 100286 1600 SW Archer Road, Room 6174, Gainesville, FL 32610-0286. Tel.: 352-265-0622; fax: +1 352 265-0701.

E-mail address: gib.upchurch@surgery.ufl.edu (G.R. Upchurch).

0022-4804/$ – see front matter

https://doi.org/10.1016/j.jss.2020.03.046

Introduction

Surgeons perform two major, primary tasks: conducting operations and engaging patients and caregivers in shared decision-making. Unfortunately, human dexterity and decision-making are biologically limited. Technical errors are the leading cause of preventable harm in surgical patients; diagnostic and judgment errors follow second.[1] Individual surgeon skill is highly variable, fine motor dexterity degrades with age and fatigue, and technical skills affect patient outcomes.[2-5] Time constraints and uncertainty impose reliance on cognitive shortcuts that lead to judgment errors, which surgeons themselves identify as the most common cause of major errors.[6-8]

Innovations in minimally invasive surgery and surgical decision-making have improved surgeons' abilities to perform operations and exercise sound judgment.[9-12] As technologies improve, these innovations rely less on human input and more on intelligent, autonomous machines, that is, computer systems that learn to perform human tasks and cognitive functions with some degree of independence.[13,14] Currently, intelligent machines can perform manual tasks and make decisions with remarkable efficacy.[15-18] History suggests that these abilities will continue to improve.[19] If there comes a time when machines perform surgeon's tasks with greater efficacy and lower cost, then market and patient demand may have machines assume these roles. Rather than regarding this possibility with denial, ire, or indifference, surgeons should seek to understand and steer these technologies toward optimal patient care and net social benefit.

Innovation curves

Innovations in minimally invasive surgery and surgical decision-making follow classic S-shaped curves with three phases: (1) introduction of a new technology, (2) achievement of a performance advantage relative to existing standards, and (3) arrival at a performance plateau, followed by augmentation or replacement with an innovation featuring greater machine autonomy and less human influence (Figure, Table).

Minimally invasive surgery innovations

Fatigue, imprecision, and variability in technical skill can adversely affect surgeons and their patients.[2-5] Technological advances in minimally invasive surgery improve surgeons' abilities to perform manual dexterity tasks and harbor the potential for autonomous robotic surgery.[9,10,20,21]

Rigid endoscopy

Endoscopy was first used to inspect the cervix more than 1000 y ago.[22] Following a long period of technological stagnation, Phillip Bozzini used a wax candle to illuminate a urologic endoscope, which was branded a "toy" by his contemporaries.[23] Problems with thermal injuries from light sources were overcome through use of platinum wires heated with electric currents or light sources encased in metal catheters with ice water cooling. Subsequent development of separate ocular and sheath components allowed of insertion of instruments to perform diagnostic procedures.[24,25] However, interventions were limited by the inability to triangulate instruments and vision, and the intra-abdominal contents could not be inspected. When Hans Christian Jakobaeus disseminated his work regarding the use of a trocar to establish pneumoperitoneum, the transition to laparoscopic surgery began.

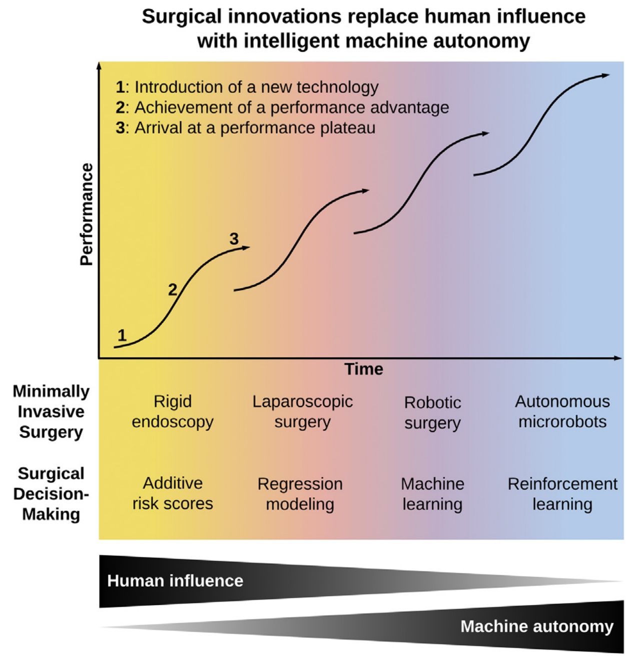

Fig – Past, present, and projected future innovations in surgery use progressively more computer autonomy and less human influence, augmenting or replacing previous methods once a cost-effective performance advantage is achieved with a new innovation. (Color version of figure is available online.)

Laparoscopic surgery

Kurt Semm described laparoscopic management of gynecologic disorders in the 1970s.[25] These techniques were applied to general surgery when Erich Mühe performed a laparoscopic cholecystectomy in 1985. He obtained pneumoperitoneum with a Veress needle, introduced pistol grip instruments though a large trocar with side-view optics and other small incisions, and removed the gallbladder though the large trocar.[26] He was ridiculed for performing "Mickey Mouse surgery", and his technique was summarized as "small brain–small incision."[27] Philippe Mouret, another pioneer of laparoscopic cholecystectomy, remarked that he felt the "weight of medico-legal responsibility for having innovated in a classic operation, which had reached a stage of near perfection."[28] Mouret's concerns were valid. In an early prospective observational study, the incidence of common bile duct injury was 5.5%, compared with 0%-0.25% for open cholecystectomy during the same era.[29] The learning curve was short and steep.[30] As laparoscopic cholecystectomy gained acceptance and adoption, its safety and efficacy improved, as evidenced by decreased mortality, pneumonia, wound

Table – Phases of innovation in minimally invasive surgery and surgical decision-making.

Surgical innovations	Phases of innovation		
	Introduction of new technology	Achievement of a performance advantage	Arrival at a performance plateau
Minimally invasive surgery			
Rigid endoscopy	Visualization of internal structures through natural orifices	Light sources, sheaths for instrument insertion	Inability to triangulate, limited working space
Laparoscopic Surgery	Trocar used to establish pneumoperitoneum	Improved outcomes for select procedures, higher costs than open surgery	Minimal advantages for natural orifice laparoscopy
Robotic Surgery	Computed tomography-guided brain biopsy	Improved outcomes for select procedures, higher costs than laparoscopic and open surgery	Requires skin and fascial defects to insert instruments
Autonomous microrobots	Ingestible robot repairs a gastric defect in 5 min	Has not yet been demonstrated	Have not yet been observed
Surgical decision-making			
Additive risk scores	Single static variable thresholds can yield high sensitivity	Risk scores using multiple variables can achieve good accuracy	Can underestimate risk for adverse outcomes among high-risk patients
Regression modeling	Estimates relationships between inputs and outputs	Patient-specific predictions may affect preoperative risk reduction strategies	Inability to accurately represent complex, nonlinear associations
Machine learning	Computer learns from data rather than conforming to rules	Improved predictive accuracy, opportunities for phenotype discovery	Predictions and phenotypes indirectly inform decision-making
Reinforcement learning	Recommends optimal actions for discrete choices and states	Has not yet been demonstrated	Have not yet been observed

infection, and hospital length of stay.[9] However, attempts to make further clinically significant improvements on modern laparoscopic surgery have had limited success.[31] As laparoscopic surgery reached a performance plateau, robotic surgery gained acceptance and adoption.

Robotic surgery

In 1985, a robotic brain biopsy platform, using stereotactic coordinates derived from computed tomography brain scans, successfully navigated a robotic arm to its target.[32,33] Ten years passed before results from human studies were reported.[34] Subsequent technological improvements offered high magnification three-dimensional views, minimization or elimination of hand tremors, instruments that articulate to extreme angles, comfortable ergonomics, and platforms allowing surgeons to operate more than two robotic arms thus obviating the need for skilled assistants. Robotic surgery has a short, steep learning curve—similar to laparoscopy—and surgeons have reported lower blood loss, shorter hospital length of stay, fewer complications, and earlier return to work relative to laparoscopic and open approaches across several surgical specialties, but with higher operative costs and limited high-level evidence demonstrating significant performance advantages.[10,20,21,35,36] In a large randomized trial, robotic-assisted rectal cancer resection yielded no significant advantages over laparoscopic resection.[37] Machine learning models can assess robotic operative performance and predict patient outcomes.[38,39] Further technological advances could offer haptic feedback, eye-tracking cameras, visualization of subsurface anatomy, predictive navigation, and virtual constraints that protect anatomic structures such as vessels and nerves, offering potential advantages for the safe, effective performance of technically demanding tasks.[40] A recent pilot randomized trial demonstrated the feasibility of robot-assisted lymphovenous microanastomosis (8 mm diameter or less) for women with breast cancer–related lymphedema.[41] Compared with manual techniques, there were no significant differences in lymphedema-related outcomes at 1- and 3-mo follow-up. Autonomous robots can perform end-to-end sutured bowel anastomoses with significantly higher leak pressures than laparoscopic and open anastomoses sewn by surgeons.[18] However, in addition to cost constraints, many of the factors that hinder laparoscopic surgery also hinder robotic surgery, such as the need to create skin and fascial defects to insert instruments, incurring risk for injury during trocar and instrument insertion, wound infection, and hernia. Autonomous microrobots could mitigate these risks.

Autonomous microrobots

In the 1966 film *Fantastic Voyage*, scientists shrink a submarine and drive it through blood vessels to remove clot from an injured colleague's brain, popularizing a notion credited to Albert Hibbs: "it would be interesting in surgery if you could swallow the surgeon." Emerging technologies suggest that autonomous surgical microrobots are feasible. In 2016, an ETH Zurich team described a hydrogel microrobot that propels itself through viscous solutions with corkscrew motions by whipping a flagellum-like tail.[42] The same year, a MIT team described a biodegradable origami-like robot that folds into an ingestible pill, unfolds in the body, sticks to tissues by friction, and moves in response to external magnetic fields by redistributing its weight.[15] In a 3D-printed silicone representation of a human esophagus and stomach, the microrobot dislodged a battery embedded in the stomach wall and patched the defect in approximately 5 min. Other groups have used bull sperm and cardiac myocytes for propulsion, magnetic field-guided steering, DNA-protein orientations that allow robots to maneuver autonomously in response to their environment, and magnetotactic bacteria loaded with nanoliposomes that hone to hypoxic signals.[43-46] The authors are unaware of any studies reporting the use of autonomous microrobots for surgery on humans, much less a performance advantage over current technologies. However, history and emerging evidence suggest that as technologies improve, autonomous microrobots have the potential to transform surgery.[47]

Surgical decision-making innovations

Surgeons frequently engage patients in high-stakes shared decision-making under both time constraints and uncertainty imposed by acute surgical disease and busy clinic schedules. These circumstances promote reliance on dogma and heuristics, which can lead to bias, cognitive errors, and preventable harm.[8,48] Innovations in surgical decision-making can mitigate these challenges.

Additive risk scores

One of the simplest ways to support decisions is risk stratification by additive scores using static variable thresholds. High blood levels of C-reactive protein are associated with anastomotic leak following colorectal surgery. Postoperative day 3 C-reactive protein levels less than 172 mg/L has 97% negative predictive value for anastomotic leak, ruling out leak in nearly all cases, but a positive predictive value of only 21%, such that high levels lack clinical utility.[49] Incorporating multiple variables can improve predictive performance. Strate *et al.*[50,51] used seven risk factors to predict severe acute lower intestinal bleeding (0 risk factors = low risk [9%], 1-3 factors = moderate risk [43%], $\geq$4 factors = high risk [84%]). External validation demonstrated good discrimination with area under receiver operating characteristic curve of 0.75.[51] Clinicians can use these predictions to guide decisions regarding the urgency of diagnostic testing and the utility of close patient monitoring. Low-risk patients may be appropriate candidates for outpatient management, avoiding unnecessary use of inpatient resources. However, additive risk scores can underestimate risk for adverse outcomes among high-risk patients. Regression modeling techniques were used to identify static variable thresholds and generate scoring systems for many additive risk scores; direct application of regression modeling may be less prone to prediction errors among high-risk patients.[52]

Regression modeling

Regression modeling estimates relationships between predictor and outcome variables to predict outcomes or explain associations. The National Surgical Quality Improvement Program Surgical Risk Calculator uses data from over four million surgeries—including procedure type, demographics, and comorbidities—to predict outcomes such as morbidity, mortality, hospital length of stay, and discharge disposition within 30 d of surgery.[11] The calculator makes accurate, patient-specific predictions and may increase the likelihood that patients will participate in risk reduction strategies, for example, prehabilitation.[12] Among 150 preoperative patients who reviewed their surgical risk calculator results, 70% stated that they would participate in prehabilitation and 40% stated that they would delay surgery to participate. Patients often want to be knowledgeable, engaged members of the healthcare team; without the use of decision-support tools, such as the National Surgical Quality Improvement Program calculator, this desire is often unfulfilled, and an opportunity to augment shared decision-making is lost.[53-55] Despite these advantages, data from 4 million surgeries may be insufficient to represent rare but important pathophysiology in a cohort of more than 60 million patients undergoing surgery in the United States each year, and regression model accuracy may suffer from an inability to accurately represent the complex, nonlinear associations among predictor variables.[56] Machine learning techniques are adept at this task.

Machine learning

In 1970, Dr. William Schwartz wrote in the *New England Journal of Medicine*, “Computing science will probably exert its major effects by augmenting and, in some cases, largely replacing the intellectual functions of the physician.”[57] Schwartz held that human disease is too broad and complex to be explained and interpreted by rules; machine learning algorithms learn from data rather than conforming to rules.[58] Fifty years later, computers have not replaced physicians' intellectual functions but have demonstrated potential to augment decisions with varying levels of autonomy. Machine learning models can predict risk for several postoperative complications with accuracy greater than that of physicians but often lack electronic and clinical workflow integration, limiting their use in routine clinical practice.[59,60]

Supervised algorithms learn from data labeled by humans, then classify or make predictions on new unseen data; unsupervised algorithms create their own output categories—often agnostic of any human-attributed labels—allowing discovery of patterns and associations. Supervised algorithms can predict sepsis more than 24 h before onset with area under receiver operating characteristic curve 0.83.[61] However, predictions are only as useful as the outcomes they predict. Seymour *et al.*[62] suggest that the overly broad definition of sepsis impairs the development of targeted interventions. They used unsupervised learning to phenotype sepsis patients, assigning points on a scatterplot as cluster centroids, assigning all other points to the nearest centroid, then iteratively recalculating centroids and cluster assignments to form the tightest clusters possible. This method identified four unique sepsis phenotypes, potentially representing subgroups with different responses to targeted therapies. These techniques require time-intensive hand-crafted feature engineering using human domain knowledge whereas deep models autonomously learn feature representations from raw data. Deep models can use electronic health record data to predict mortality among intensive care unit patients with greater accuracy than the sequential organ failure assessment score, even when limited to the same input data used to calculate sequential organ failure assessment.[63] Deep learning and statistical modeling can also use characters, words, and other expressions of natural language as model inputs. This technique, termed natural language processing, can generate oncologic decision-support tools predicting germline mutations.[64,65] This approach can leverage the availability of large volumes of genetic data and medical literature to produce personalized cancer prevention management strategies.[66] Deep model interpretation mechanisms elucidate the relative importance of individual input features in determining model outputs, providing opportunities to assess whether associations between inputs and outputs are biologically plausible.[63,67] Despite these advantages, predictions and classifications can only indirectly inform discrete choices facing clinicians, limiting their clinical utility. Reinforcement learning directly informs discrete choices.

Reinforcement learning

In reinforcement learning, an agent learns that specific actions under certain conditions lead to rewards and penalties, using this knowledge to identify actions that achieve an ultimate goal. Two characteristics distinguish reinforcement learning from machine learning: (1) trial-and-error search to identify the best action and (2) delayed reward, that is choosing actions that achieve the ultimate goal rather than short-term rewards.[68] For example, a model developed by Komorowski *et al.*[16] recommends vasopressor doses and intravenous fluid volumes for septic patients, assigning rewards and penalties relative to 90-d survival. The model favored higher vasopressor doses and lower intravenous fluid volumes, consistent with evidence that volume overload harms sepsis patients and that a one-size-fits-all approach to resuscitation is suboptimal.[69,70] On retrospective analysis, when actions taken by clinicians were concordant with model recommendations, mortality was slightly less than 20%. As clinician actions deviated from model recommendations, mortality significantly increased, up to 60%. Notably, clinicians may have deviated from model recommendations based on data not available to the model (e.g. physical exam findings, symptoms), and the same findings contributed to a worse prognosis, making clinician decision-making seem less effective. Therefore, available evidence does not support causal inference between model recommendations and decreased mortality.

For more complex decision-making scenarios in high-volume, high-dimension datasets, exhaustive searches for optimal actions can be prohibitive or impossible, but deep representation of the agent's environment can mitigate these challenges. The Go board game has 32,490 possible first moves.[71] A deep reinforcement model first learned from a human Go expert, then defeated the European Go champion five games to zero. Subsequently, a completely autonomous model trained on self-play only defeated the human input model 100 games to zero.[17] Similar approaches have the potential to transform surgical decision-making, but in the absence of high-level evidence for medical applications, this potential remains theoretical.[72] In addition, reinforcement learning models require large training data sets to maintain effective sample sizes in sequential decision-making tasks, and such data are not available for many surgical diseases, especially rare ones.[73]

Lessons learned from automotive innovations

The automotive industry adopts intelligent, autonomous machine innovations that achieve performance advantages according to consumer demands and business advantages. Similar market forces will likely drive surgery toward machine autonomy. Currently, there is no level I evidence demonstrating that intelligent, autonomous machines improve patient outcomes compared with existing standards for performing operations or surgical decision-making tasks, specifically (see the Supplement describing an Embase, MEDLINE, and PubMed search performed by the authors 10/30/2019). These technologies remain on the initial, flat portion of the innovation S-curve (Figure). However, if future hospitals can purchase robotic surgical platforms that autonomously perform operations with lower costs and higher quality than human surgeons, or deep reinforcement learning models that consistently make better decisions than clinicians, then it seems likely that these technologies will gain adoption. History suggests that intelligent, autonomous machines will initially be used for simple, common, rote tasks under close human supervision and then for complex tasks with minimal human supervision. Automation of programmable tasks may allow surgeons to spend less time gathering and analyzing data and more time interacting with patients and tending to urgent, critical—and potentially more valuable—aspects of patient care.

Lessons learned from automotive innovations reveal opportunities to capitalize on the performance advantages of new technologies without disenfranchising the people that use and benefit from them. When robotic arms largely replaced human assembly line workers in performing rote mechanical tasks, automobile prices fell within reach of the middle class, but many assembly line workers lost their jobs. As in the industrial revolution, there was a lag time between incorporation of autonomous machines and redistribution of the human workforce. Perhaps if this transition were anticipated, a smoother transition could be achieved. Anticipating a similar transition in surgery seems prudent. Eventually, the automotive industry evolved to use human effort and expertise in designing and overseeing robotic arm assembly lines. Automotive workforces also pivoted toward tasks requiring creativity, long-term planning, and moral deliberation, which are especially relevant in designing self-driving cars that sense the environment and respond accordingly. Responses are programmable and have important moral implications. Awad *et al.*[74] created online simulations in which participants identify preferences for how self-driving cars should behave when distributing harm in unavoidable collisions, for example, the car can maintain its course and hit a jaywalking teenager or swerve and crash, harming its elderly passenger. The authors collected data on nearly 40 million decisions by participants in 233 countries and found significant cross-cultural variation in preferences for moral dilemmas facing self-driving cars, precluding a one-size-fits-all approach to morally sound programming.

Moral and ethical dilemmas also challenge the adoption of intelligent, autonomous machines in surgery. Management of a patient with both pulmonary edema and prerenal azotemia could proceed with either diuresis or volume resuscitation. The tradeoff between respiratory failure requiring mechanical ventilator support *versus* renal failure requiring renal replacement therapy depends in part on the desires and values of the patient and their caregivers. An autonomous reinforcement learning platform trained to optimize an arbitrary endpoint such as 90-d mortality could make a recommendation or decision that is medically sound, but contrary to patient values. Also, algorithms trained on biased data sets are likely to produce biased outputs, as demonstrated for crime recidivism predictions.[75] Similar problems could occur in machine learning healthcare applications. For example, associations between cardiovascular risk factors and adverse cardiovascular events differ by race and ethnicity; a model trained on data from the Framingham Heart Study, which primarily included white subjects, could produce racially and ethnically biased outputs.[76] Algorithms used for allocating liver transplants may disenfranchise female organ recipient candidates by prioritizing serum creatinine, which is lower among women.[77] Therefore, investigators must align training data set and target population demographics and other characteristics that have potential to introduce bias. In addition, judicial systems have limited experience assigning liability for errors made by intelligent machines and differentiating between human and machine error. In making a critical management decision for a life-threatening postoperative complication, a surgeon could be privy to history and physical exam information that is unavailable to an autonomous decision-support platform, take a different course of action than recommended by a model with proven efficacy, and be subject to unwarranted scrutiny when the patient suffers a poor outcome. Similarly, robotic surgical platforms with virtual constraints intended to protect anatomic structures could delay or prevent a surgeon from gaining control of an injured blood vessel, harming a patient and pitting human *versus* machine in assigning liability. Surgeons must meet these challenges with creativity, altruism, moral deliberation, and emotional intelligence, that is, the ability to recognize emotional states and act accordingly. These traits remain inaccessible to machines. The surgeon's role may evolve to interpreting decision-support tools and offering wisdom for patients and caregivers facing complex, high-stakes surgical decisions, using and overseeing semi-autonomous and fully autonomous surgical instruments and robotic platforms in the operating room and ensuring the safe and effective integration of intelligent, autonomous machines with surgical care.

Conclusions

As technologies improve, intelligent, autonomous machines may gain the capacity to augment or outperform humans in operative and decision-making tasks. History suggests that intelligent, autonomous machines will be used in surgery initially for simple, common, rote tasks under close human supervision and then for complex tasks with minimal human supervision. Automation of programmable tasks may allow surgeons to spend less time gathering and analyzing data and more time interacting with patients and tending to urgent, critical—and potentially more valuable—aspects of patient care. This process poses ethical challenges in assigning liability for errors, distributing harm, and displacing human workers. Surgeons should assume active roles in guiding these technologies toward optimal patient care and net social benefit, channeling human creativity, moral deliberation, and altruism.

Acknowledgment

Author contributions: T.J.L., A.C.F., and J.B. designed the study, reviewed the literature, and drafted the manuscript. All authors analyzed and interpreted available literature. A.L.A., P.J.T., P.R., A.B., and G.R.U. provided critical revisions.

Disclosure

A.B. and P.R. were supported by R01 GM110240 from the National Institute of General Medical Sciences (NIGMS) and by Sepsis and Critical Illness Research Center Award P50 GM-111152 from the NIGMS. P.R. was supported by CAREER award, NSF-IIS 1750192, from the National Science Foundation (NSF), Division of Information and Intelligent Systems (IIS) and by NIH NIBIB R21EB027344-01. P.T.J. was supported by R01GM114290 from the NIGMS. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jss.2020.03.046.

REFERENCES

1. Healey MA, Shackford SR, Osler TM, Rogers FB, Burns E. Complications in surgical patients. *Arch Surg*. 2002;137:611–617 [discussion 617-618].
2. Birkmeyer JD, Finks JF, O'Reilly A, et al. Surgical skill and complication rates after bariatric surgery. *N Engl J Med*. 2013;369:1434–1442.
3. Begg CB, Riedel ER, Bach PB, et al. Variations in morbidity after radical prostatectomy. *N Engl J Med*. 2002;346:1138–1144.
4. Furuya S, Tominaga K, Miyazaki F, Altenmuller E. Losing dexterity: patterns of impaired coordination of finger movements in musician's dystonia. *Sci Rep*. 2015;5:13360.
5. Carmeli E, Patish H, Coleman R. The aging hand. *J Gerontol A Biol Sci Med Sci*. 2003;58:146–152.
6. Shanafelt TD, Balch CM, Bechamps G, et al. Burnout and medical errors among American surgeons. *Ann Surg*. 2010;251:995–1000.
7. *Groopman JE How doctors think*. Boston: Houghton Mifflin; 2007.
8. Wolf FM, Gruppen LD, Billi JE. Differential diagnosis and the competing-hypotheses heuristic. A practical approach to

judgment under uncertainty and Bayesian probability. *JAMA*. 1985;253:2858–2862.
9. Coccolini F, Catena F, Pisano M, et al. Open versus laparoscopic cholecystectomy in acute cholecystitis. Systematic review and meta-analysis. *Int J Surg*. 2015;18:196–204.
10. Hohwu L, Akre O, Pedersen KV, et al. Open retropubic prostatectomy versus robot-assisted laparoscopic prostatectomy: a comparison of length of sick leave. *Scand J Urol Nephrol*. 2009;43:259–264.
11. Bilimoria KY, Liu Y, Paruch JL, et al. Development and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aid and informed consent tool for patients and surgeons. *J Am Coll Surg*. 2013;217, 833-842.e1-e3.
12. 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*. 2018;128:643–650.
13. Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial intelligence in surgery: promises and perils. *Ann Surg*. 2018;268:70–76.
14. Haidegger T. Autonomy for surgical robots: concepts and paradigms. *IEEE Trans Med Robot Bionic*. 2019;1:65–76.
15. Miyashita S, Guitron S, Yoshida K, et al. Ingestible, controllable, and degradable origami robot for patching stomach wounds. *IEEE Int Conf Robot Autom*. 2016;2016:909–916.
16. 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;24:1716–1720.
17. Silver D, Schrittwieser J, Simonyan K, et al. Mastering the game of Go without human knowledge. *Nature*. 2017;550:354–359.
18. Shademan A, Decker RS, Opfermann JD, et al. Supervised autonomous robotic soft tissue surgery. *Sci Transl Med*. 2016;8:337ra364.
19. Kurzweil R. *The singularity is near: when humans transcend biology*. New York: Viking; 2005.
20. Bell MC, Torgerson J, Seshadri-Kreaden U, Suttle AW, Hunt S. Comparison of outcomes and cost for endometrial cancer staging via traditional laparotomy, standard laparoscopy and robotic techniques. *Gynecol Oncol*. 2008;111:407–411.
21. Altieri MS, Yang J, Telem DA, et al. Robotic approaches may offer benefit in colorectal procedures, more controversial in other areas: a review of 168,248 cases. *Surg Endosc*. 2016;30:925–933.
22. Spaner SJ, Warnock GL. A brief history of endoscopy, laparoscopy, and laparoscopic surgery. *J Laparoendosc Adv Surg Tech A*. 1997;7:369–373.
23. Gomella LG, Albala DM. Laparoscopic urological surgery: 1994. *Br J Urol*. 1994;74:267–273.
24. Gaskin TA, Isobe JH, Mathews JL, Winchester SB, Smith RJ. Laparoscopy and the general surgeon. *Surg Clin North Am*. 1991;71:1085–1097.
25. Blum CA, Adams DB. Who did the first laparoscopic cholecystectomy? *J Minim Access Surg*. 2011;7:165–168.
26. Muhe E. [Laparoscopic cholecystectomy]. *Z Gastroenterol Verh*. 1991;26:204–206.
27. Litynski GS. Erich Muhe and the rejection of laparoscopic cholecystectomy (1985): a surgeon ahead of his time. *JSLS*. 1998;2:341–346.
28. Mouret P. How I developed laparoscopic cholecystectomy. *Ann Acad Med Singapore*. 1996;25:744–747.
29. Kum CK, Eypasch E, Lefering R, et al. Laparoscopic cholecystectomy for acute cholecystitis: is it really safe? *World J Surg*. 1996;20:43–48 [discussion 48-49].
30. Peters JH, Ellison EC, Innes JT, et al. Safety and efficacy of laparoscopic cholecystectomy. A prospective analysis of 100 initial patients. *Ann Surg*. 1991;213:3–12.
31. Schwaitzberg SD, Roberts K, Romanelli JR, et al. The NOVEL trial: natural orifice versus laparoscopic cholecystectomy-a prospective, randomized evaluation. *Surg Endosc*. 2018;32:2505–2516.
32. Koyama H, Uchida T, Funakubo H, Takakura K, Fankhauser H. Development of a new microsurgical robot for stereotactic neurosurgery. *Stereotact Funct Neurosurg*. 1990;54-55:462–467.
33. Kwoh YS, Hou J, Jonckheere EA, Hayati S. A robot with improved absolute positioning accuracy for CT guided stereotactic brain surgery. *IEEE Trans Biomed Eng*. 1988;35:153–160.
34. Glauser D, Fankhauser H, Epitaux M, Hefti JL, Jaccottet A. Neurosurgical robot Minerva: first results and current developments. *J Image Guid Surg*. 1995;1:266–272.
35. Kane WJ, Charles EJ, Mehaffey JH, et al. Robotic compared with laparoscopic cholecystectomy: a propensity matched analysis. *Surgery*. 2019;167:432–435.
36. Janjua H, Cousin-Peterson E, Barry TM, et al. The paradox of the robotic approach to inguinal hernia repair in the inpatient setting. *Am J Surg*. 2019;219:497–501.
37. Jayne D, Pigazzi A, Marshall H, et al. Effect of robotic-assisted vs conventional laparoscopic surgery on risk of conversion to open laparotomy among patients undergoing resection for rectal cancer: the ROLARR randomized clinical trial. *JAMA*. 2017;318:1569–1580.
38. Hung AJ, Chen J, Gill IS. Automated performance metrics and machine learning algorithms to measure surgeon performance and anticipate clinical outcomes in robotic surgery. *JAMA Surg*. 2018;153:770–771.
39. Hung AJ, Chen J, Jarc A, et al. Development and validation of objective performance metrics for robot-assisted radical prostatectomy: a pilot study. *J Urol*. 2018;199:296–304.
40. Cha J, Broch A, Mudge S, et al. Real-time, label-free, intraoperative visualization of peripheral nerves and micro-vasculatures using multimodal optical imaging techniques. *Biomed Opt Express*. 2018;9:1097–1110.
41. van Mulken TJM, Schols RM, Scharmga AMJ, et al. First-in-human robotic supermicrosurgery using a dedicated microsurgical robot for treating breast cancer-related lymphedema: a randomized pilot trial. *Nat Commun*. 2020;11:757.
42. Huang HW, Sakar MS, Petruska AJ, Pane S, Nelson BJ. Soft micromachines with programmable motility and morphology. *Nat Commun*. 2016;7:12263.
43. Magdanz V, Sanchez S, Schmidt OG. Development of a sperm-flagella driven micro-bio-robot. *Adv Mater*. 2013;25:6581–6588.
44. Douglas SM, Bachelet I, Church GM. A logic-gated nanorobot for targeted transport of molecular payloads. *Science*. 2012;335:831–834.
45. Williams BJ, Anand SV, Rajagopalan J, Saif MT. A self-propelled biohybrid swimmer at low Reynolds number. *Nat Commun*. 2014;5:3081.
46. Felfoul O, Mohammadi M, Taherkhani S, et al. Magneto-aerotactic bacteria deliver drug-containing nanoliposomes to tumour hypoxic regions. *Nat Nanotechnol*. 2016;11:941–947.
47. Maier-Hein L, Vedula SS, Speidel S, et al. Surgical data science for next-generation interventions. *Nat Biomed Eng*. 2017;1:691–696.
48. Bekker HL. Making choices without deliberating. *Science*. 2006;312:1472.
49. Singh PP, Zeng IS, Srinivasa S, et al. Systematic review and meta-analysis of use of serum C-reactive protein levels to predict anastomotic leak after colorectal surgery. *Br J Surg*. 2014;101:339–346.

50. Strate LL, Ayanian JZ, Kotler G, Syngal S. Risk factors for mortality in lower intestinal bleeding. *Clin Gastroenterol Hepatol*. 2008;6:1004–1010.
51. 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;100:1821–1827.
52. Jin R, Grunkemeier GL, Providence Health System Cardiovascular Study G. Additive vs. logistic risk models for cardiac surgery mortality. *Eur J Cardiothorac Surg*. 2005;28:240–243.
53. Uldry E, Schafer M, Saadi A, Rousson V, Demartines N. Patients' preferences on information and involvement in decision making for gastrointestinal surgery. *World J Surg*. 2013;37:2162–2171.
54. de Mik SML, Stubenrouch FE, Balm R, Ubbink DT. Systematic review of shared decision-making in surgery. *Br J Surg*. 2018;105:1721–1730.
55. Wilson A, Ronnekleiv-Kelly SM, Pawlik TM. Regret in surgical decision making: a systematic review of patient and physician perspectives. *World J Surg*. 2017;41:1454–1465.
56. 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;33:1627–1634.
57. Schwartz WB. Medicine and the computer. The promise and problems of change. *N Engl J Med*. 1970;283:1257–1264.
58. Schwartz WB, Patil RS, Szolovits P. Artificial intelligence in medicine. Where do we stand? *N Engl J Med*. 1987;316:685–688.
59. 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;165:1035–1045.
60. Leeds IL, Rosenblum AJ, Wise PE, et al. Eye of the beholder: risk calculators and barriers to adoption in surgical trainees. *Surgery*. 2018;164:1117–1123.
61. Henry KE, Hager DN, Pronovost PJ, Saria S. A targeted real-time early warning score (TREWScore) for septic shock. *Sci Transl Med*. 2015;7:299ra122.
62. Seymour CW, Kennedy JN, Wang S, et al. Derivation, validation, and potential treatment implications of novel clinical phenotypes for sepsis. *JAMA*. 2019;321:2003–2017.
63. Shickel B, Loftus TJ, Adhikari L, et al. DeepSOFA: a continuous acuity score for critically ill patients using clinically interpretable deep learning. *Sci Rep*. 2019;9:1879.
64. Hughes KS, Zhou J, Bao Y, et al. Natural language processing to facilitate breast cancer research and management. *Breast J*. 2020;26:92–99.
65. Braun D, Yang J, Griffin M, Parmigiani G, Hughes KS. A clinical decision support tool to predict cancer risk for commonly tested cancer-related germline mutations. *J Genet Couns*. 2018;27:1187–1199.
66. Deng Z, Yin K, Bao Y, et al. Validation of a semiautomated natural language processing-based procedure for meta-analysis of cancer susceptibility gene penetrance. *JCO Clin Cancer Inform*. 2019;3:1–9.
67. Shamout FE, Zhu T, Sharma P, Watkinson PJ, Clifton DA. Deep interpretable early warning system for the detection of clinical deterioration. *IEEE J Biomed Health Inform*. 2019;24:437–446.
68. Sutton RS, Barto AG. *Reinforcement learning: an introduction*. Cambridge, Massachusetts: The MIT Press; 2018.
69. Boyd JH, Forbes J, Nakada TA, Walley KR, Russell JA. Fluid resuscitation in septic shock: a positive fluid balance and elevated central venous pressure are associated with increased mortality. *Crit Care Med*. 2011;39:259–265.
70. Maitland K, Kiguli S, Opoka RO, et al. Mortality after fluid bolus in African children with severe infection. *N Engl J Med*. 2011;364:2483–2495.
71. Silver D, Huang A, Maddison CJ, et al. Mastering the game of Go with deep neural networks and tree search. *Nature*. 2016;529:484–489.
72. Topol EJ. A decade of digital medicine innovation. *Sci Transl Med*. 2019;11.
73. Gottesman O, Johansson F, Komorowski M, et al. Guidelines for reinforcement learning in healthcare. *Nat Med*. 2019;25:16–18.
74. Awad E, Dsouza S, Kim R, et al. The moral machine experiment. *Nature*. 2018;563:59–64.
75. Angwin J, Larson J, Mattu S, Kirchner L. Machine bias. ProPublica. 2016. Available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing. Accessed January 24, 2019.
76. 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:e0132321.
77. Myers RP, Shaheen AA, Aspinall AI, Quinn RR, Burak KW. Gender, renal function, and outcomes on the liver transplant waiting list: assessment of revised MELD including estimated glomerular filtration rate. *J Hepatol*. 2011;54:462–470.