

# Critical Care, Critical Data

Christopher V Cosgriff<sup>1,2</sup>, Leo Anthony Celi<sup>1,3</sup> and David J Stone<sup>1,4</sup>

<sup>1</sup>MIT Critical Data, Massachusetts Institute of Technology, Cambridge, MA, USA. <sup>2</sup>New York University School of Medicine, New York, NY, USA. <sup>3</sup>Division of Pulmonary, Critical Care and Sleep Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA. <sup>4</sup>Departments of Anesthesiology and Neurosurgery, University of Virginia School of Medicine, Charlottesville, VA, USA.

Biomedical Engineering and  
Computational Biology  
Volume 10: 1–7  
© The Author(s) 2019  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/1179597219856564



**ABSTRACT:** As *big data*, *machine learning*, and *artificial intelligence* continue to penetrate into and transform many facets of our lives, we are witnessing the emergence of these powerful technologies within health care. The use and growth of these technologies has been contingent on the availability of reliable and usable data, a particularly robust resource in critical care medicine where continuous monitoring forms a key component of the infrastructure of care. The response to this opportunity has included the development of open databases for research and other purposes; the development of a collaborative form of clinical data science intended to fully leverage these data resources, and the creation of data-driven applications for purposes such as clinical decision support. Most recently, data levels have reached the thresholds required for the development of robust artificial intelligence features for clinical purposes. The systematic capture and analysis of clinical data in both individuals and populations allows us to begin to move toward precision medicine in the intensive care unit (ICU). In this perspective review, we examine the fundamental role of data as we present the current progress that has been made toward an artificial intelligence (AI)-supported, data-driven precision critical care medicine.

**KEYWORDS:** Data analytics, database, critical care, intensive care units, machine learning

**RECEIVED:** March 5, 2019. **ACCEPTED:** May 21, 2019.

**TYPE:** Review

**FUNDING:** The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**DECLARATION OF CONFLICTING INTERESTS:** The author(s) received no financial support for the research, authorship, and/or publication of this article.

**CORRESPONDING AUTHOR:** Leo Anthony Celi, MIT Critical Data, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. Email: lceli@mit.edu

## Introduction

Thinking machines have long occupied and fascinated the human mind, captivating us with their potential to radically alter society. As the digitization of society has progressed and massive amounts of data have accrued, a new vocabulary has emerged. The scale of data collection has given rise to *big data*, and the field of *data science* has sprung from the marriage of statistics and computer science necessary to exploit this new resource. Statistical modeling became machine learning, and as new, sophisticated machine learning approaches achieved unprecedented performance, *artificial intelligence* (AI) permeated our lexicon. Whereas truly thinking machines remain beyond our reach, there is no doubt that we have entered a new era with these technologies transforming every facet of our lives.

Health care is no exception, and the prospect of transforming care delivery by way of these technologies is a vibrant and rapidly growing area of research.<sup>1–5</sup> Dissatisfied with besting humans at jeopardy, IBM's Watson has set its sights on health and disease (and learned so far that these are very difficult domains, indeed).<sup>6</sup> DeepMind, having wowed the world with a reinforcement learning agent that achieved superhuman performance at *Go*, now has a health care division.<sup>7</sup> Geoffrey Hinton,<sup>8</sup> the inventor of backpropagation, the fundamental mechanism by which deep neural networks are trained, has now been published in the *Journal of the American Medical Association*. However, as many have commented, the challenges of bringing these technologies to bear in the care of human beings are myriad.<sup>1,9</sup>

The potential success of these new technologies rests largely on 2 key drivers: affordable, accessible high-performance

computer hardware and an explosion of data. The latter is often taken for granted. The medical literature has been quick to embrace *big data*, but out-of-date privacy laws, competitions based on the profit motive, a culture wary of innovation and collaboration, and disparate data representations continue to hinder efforts to truly benefit from the fruits of a big data revolution. However, there has been and continues to be much work directed at overcoming these challenges.

Critical care medicine concerns itself with the care of unstable, high-acuity patients, particularly those with multi-organ failure; continuous physiologic monitoring is consequently the hallmark of the intensive care unit (ICU). With the near-total digitization of health care, the ICU represents an incredibly fertile ground for the proliferation of big data technologies. Advancements that take advantage of this wealth of data promise to fortify our currently relatively fragile evidence base by providing large cohorts for knowledge discovery and causal inference, and will provide the substrate for the next generation of clinical decision support tools.<sup>10</sup> Reliable clinical data, whether digital or not, have always been the basis for caring for our patients, but digitization provides the opportunity to leverage the troves of data generated in the ICU to advance the field into a new era of medicine. In this perspective review, we examine the fundamental role of data as we present the current progress that has been made toward a data-driven precision critical care medicine.

## Critical Care Databases

Knowledge discovery, decision support model development, and the education of the next generation of clinician



data scientists all require health data to be available and easily accessible. In fact, we envision a future in which all clinicians will be data scientists to a certain degree. Prior commercial databases developed primarily for the development of benchmark models and national registries lack the resolution and volume required to support the breakthroughs of this new era.<sup>11</sup> However, over the last 2 decades, we have witnessed the emergence of large-scale, highly granular, critical care databases for use in observational research and predictive model development.

The Multiparameter Intelligent Monitoring in Intensive Care (MIMIC) database was the first resource of this kind.<sup>12</sup> Developed and maintained over the past 2 decades by the MIT Laboratory for Computational Physiology (LCP), the database is now its third iteration as the Medical Information Mart for Intensive Care (MIMIC-III) database.<sup>13</sup> The MIMIC-III database contains high-resolution and multi-modal de-identified data from the electronic health record (EHR) associated with 53 342 distinct hospital admissions to the Beth Israel Deaconess Medical Center (BIDMC) in Boston, Massachusetts. The data include, but are not limited to, vital sign recordings and waveforms, laboratory data, clinical notes, diagnostic reports, and administered interventions, including medications. Some of the data are quantitative or structured, but much requires extraction from text format.

In addition to MIMIC-III, the LCP partnered with Philips to release de-identified data from the Philips eICU Research Institute. The eICU Collaborative Research Database (eICU-CRD), now at version 2.0, is a multi-center critical care database containing data from more than 200 000 ICU admissions from across the United States that were archived from Philips' ICU telehealth platform.<sup>14</sup> This resource allows for the development of models with populations more representative of the entire United States, ascertaining the generalizability of findings and models.

Long believing that open access to data spurs innovation and accelerates progress, the LCP makes MIMIC-III and eICU-CRD publicly available to any individual who completes a standard course on human subject research and signs a data use agreement. In doing so, these data have allowed for countless projects in academia and industry, and the availability of MIMIC-III has made the BIDMC ICU population the most intensely studied critically ill cohort to date. In addition, the data use agreement for MIMIC-III requires that the code for projects developed with MIMIC-III be publicly shared. This has led to the rapid development of reusable concepts and their respective codes and queries, with the LCP maintaining a large, publicly available code repository.<sup>15</sup> The availability of this code accelerates research and promotes reproducibility by ensuring that common concepts are implemented consistently across studies.

By way of international collaborations that will be discussed further below, MIMIC-III has inspired the development of

similar critical care databases in Spain, Brazil, China, Australia, and Switzerland. The existence of these databases drives similar progress in those respective countries and should lead to an international system of data sharing capable of supporting the development of large international observational cohorts and generalizable predictive models. Unfortunately, there remain major barriers to data sharing endeavors.

From a technical perspective, international data sharing represents a complex challenge. There is currently no widely accepted database structure for critical care databases. Should such a structure be developed, disparate concepts between centers would require harmonization, and we currently lack a common standard for representing various sources of clinically relevant information.<sup>16</sup> As medical centers have substantial differences in the way care is delivered, with variable access to medical technologies, and cultural differences in the way care is documented, the development of a system for cross talk between critical care databases would be a major engineering feat.

Varying perspectives on privacy and data sharing represent an even greater barrier, and the prospect is increasingly limited by complex legal frameworks.<sup>17</sup> For example, the European Union (EU) General Data Protection Regulation (GDPR) applies to all data controllers and processors of personal data for subjects in the EU regardless of whether the processing occurs in the EU or not, and thus databases based outside of the EU must comply with GDPR if residents from the EU are included in the data. Therefore, the linking of MIMIC-III which contains subjects not requiring explicit consent, to a database from the EU for a larger, more broadly applicable analysis would require explicit consent be obtained by the researchers if anonymization is deemed inadequate. There are also numerous opponents to public data sharing on the grounds of missed financial opportunities to monetize these intrinsically valuable data.<sup>18</sup> Together, these barriers hinder progress toward the development of a global network for health data exchange, and legal and ethical frameworks must evolve for us to make progress toward this ultimate goal.

## Collaborative Data Science

Deriving insight from large EHR databases is a non-trivial task requiring skills and expertise that span multiple disciplines from clinical intensive care to sophisticated statistical methods. The methods by which data are explored, processed, harmonized, transformed, and modeled fall well beyond the purview of traditional medical training and can lead to misunderstandings of what can and cannot be accomplished with these tools. As implementation of these methods often requires acumen with programming languages like SQL, Python, and R, clinicians may find themselves overwhelmed, even when they have a relatively sound understanding of complicated biostatistical approaches.

Similarly, data scientists rarely have the clinical insights to know what questions are relevant to medical care and how the

data themselves were generated in practice and how they should be interpreted. Consider also that patterns of missing data in the EHR are rarely uninformative: a serum lactate level is ordered when physicians are concerned about the adequacy of organ perfusion, and thus the very presence of this laboratory test in a patient's data tells us something about the clinical context. This small insight is obvious to physicians, but the apparent "missingness" of lactate values might perplex an uninformed data scientist and lead to an incorrect modeling decision. Similarly, a key step in model building processes is feature engineering and selection. Considering the breadth of data available in an electronic medical record, when should, for example, a serum phosphate level be included in a predictive model? Certainly, it will be more useful when modeling a population of patients with kidney disease, but likely less useful in a population of patients with acute trauma.

There has been no dearth of literature arguing for changes in medical education such that the next generation of clinicians can understand and work with complex statistical methods, and grasp the computational approaches that will undoubtedly be incorporated into their practices.<sup>19–21</sup> That said, whereas the clinician data scientist will surely emerge (in a manner akin to the translational scientist) to bridge the computational and clinical science realms, the future of medical research and health care delivery will progressively rely on collaboration between clinicians and data scientists. The term *datathon* has thus been introduced as a powerful tool in linking these disparate worlds.<sup>22–24</sup> Pioneered by MIT Critical Data, a consortium founded by members of the LCP and the Computer Science and Artificial Intelligence Laboratory (CSAIL), datathons pair clinicians and data scientists to challenge them to work together to solve a clinical problem.<sup>22,23,25–27</sup> Clinicians learn the nuances of data extraction and model development, and data scientists are provided invaluable insights into clinical data capture and decision making.<sup>28</sup>

The success of the datathon model has relied heavily on the availability of data. The MIMIC-III and the eICU-CRD databases serve as the substrates on which clinicians can learn to ask questions amenable to secondary analysis, and data scientists can begin wrangling real health care data. The events often begin with physicians from local hospitals pitching their research questions to the audience. Teams are formed and immediately get to work to parse the question into a study design, extract the cohort, and build models with the support and guidance of clinical data scientists from MIT Critical Data. With the publicly available code repository containing many of the common concepts required for critical care research, projects can be rapidly performed in the span of a weekend.<sup>15,28</sup> Mentors provide feedback throughout the entire process and ultimately evaluate the clinical relevance, technical implementation, and reproducibility of the final projects.

MIT Critical Data has hosted more than 20 datathon events in 10 countries across 5 continents jumpstarting numerous

international collaborations. Many of the projects pitched and initiated at datathon events are eventually published in the scientific literature.<sup>29–32</sup> In addition, as mentioned in the previous section, these international collaborations have demonstrated the value of secondary EHR analysis to countless decision makers at health care institutions across the world and have led to the development of similar critical care databases. Despite the aforementioned barriers, this trend is laying the groundwork for a network of EHR data sharing that will ultimately allow for multi-national and multi-institution analyses.

This collaborative format, in which clinicians propose research questions and work with teams of data scientists to address them, has also given rise to a course at the Harvard-MIT Division of Health Science, and Technology (HST). The course "Collaborative Data Science for Medicine" introduces students to MIMIC-III and the eICU-CRD, and features lectures on database querying, statistics and epidemiology, data exploration and visualization, machine learning, and causal inference. The course, now in its third year, produces numerous abstracts, presentations, and publications and will serve as a model for other courses around the world.<sup>33–37</sup> To promote such efforts, MIT Critical Data has published a textbook for the course, *Secondary Analysis of Electronic Health Record Data*, and made it freely available as an eBook online.<sup>38</sup>

All of these efforts seek to build a bridge between clinician and data scientist that works to improve understanding of health and disease, and ultimately impact patient outcomes. Working side by side, clinicians and data scientists provide a skill set far greater than the sum of their parts. This partnership is the only way medicine can hope to navigate the *big data* age successfully, with the education and training of clinicians at every level interfaced with data scientists.

## Machine Learning and Decision Support

Clinical decision making is rife with uncertainty: we seek to leverage the evidence derived from clinical trials and observational studies, but often the specific study we require does not exist, and when it does, it is usually insufficient in one or more respects. Furthermore, as the ground truth is constantly shifting in medicine, even a perfectly performed and applicable study from a few years prior may no longer apply as new tests and treatments are incorporated into practice and patient demographics change. Information gaps are one of the drivers of variation in care as physicians rely on their prior experiences and training as well as institutional culture to guide decisions. A process of continually using routinely collected clinical data to update knowledge and guide practice, intimately linking knowledge generation and care delivery, represents a new paradigm that promises to bring us closer to a true evidence-based care.<sup>39</sup> This concept has often been referred to as the *Learning Healthcare System*.<sup>40</sup>

The emergence of sophisticated machine learning methods has inched us closer to this vision, and we have recently seen a

variety of exciting implementations of machine learning applications in critical care medicine.<sup>1,3</sup> It should be noted that this is not a completely novel concept in critical care as approaches like multivariate logistic regression, a form of machine learning, have long been applied in this specialty. For example, illness severity scores such as the APACHE (Acute Physiology and Chronic Health Evaluation) system represent an early form of machine learning in health care, although APACHE and similar models have generally not been used to guide clinical decision making.<sup>11,41,42</sup> More recent applications of machine learning with EHR data have included gradient boosted decision trees that can forecast acute kidney injury and predict readmission; convolutional neural networks that can diagnose diabetic retinopathy; recurrent neural networks that can prognosticate directly from clinical time series data; and a reinforcement learning agent that can make treatment decisions in sepsis.<sup>43–49</sup> This last example encapsulates the essence of a vast collective experience: the agent was trained on the management decisions of clinicians caring for more than 100 000 sepsis patients and learned to tailor treatment to each individual patient with the goal of reducing 90-day mortality.<sup>46</sup>

Many of these more complex methods befuddle clinicians. Rooted in intricate mathematical concepts and proofs, their correct application to clinical problems is not trivial. However, formatting data for model training and fitting a model correctly to minimize the generalization error represent the easiest steps in the creation and deployment of clinical decision support tools. As has been stressed above, the first requirement in this process is the data. Data preparation for machine learning—which includes aggregation, integration, and harmonization—requires substantial effort and buy-in from health care administrators, hospital information technologists, data engineers, and data scientists. The challenge of navigating these barriers dwarfs that of model development. Should these 2 steps be successfully traversed, bringing the model to the bedside presents an equally monumental challenge. Model safety with attention to identification of algorithm bias must be considered and clinical validation is crucial; usability and information overload must also be considered.<sup>50,51</sup> We will focus the remainder of this section on specific challenges to developing models that can be effectively incorporated into routine care.

George EP Box famously stated, “all models are wrong, but some are useful.”<sup>52</sup> The question then is how do we determine which are useful. Classification models are frequently described by their ability to discriminate. Discrimination is most often measured by the area under the receiver operating characteristic curve (AUROC).<sup>53</sup> However, the AUROC is a less appropriate measure of performance when a model’s task is detection of rare events, as is common in the critical care context, because, for rare events, specificity disproportionately drives accuracy.<sup>54</sup> The area under the precision-recall curve (AUPRC) should be used in these instances, as it provides a more accurate measure in the face of rare events.

Neither of these metrics captures a model’s performance regarding the quantification of absolute risk, which is often of greater clinical value than discrimination of event from non-event.<sup>53</sup> A classification model’s ability to adequately quantify absolute risk probabilities is termed calibration. Calibration may be examined visually with reliability curves and may be quantified by way of observed-to-predicted ratios; null hypothesis tests such as the Hosmer-Lemeshow goodness-of-fit test are not recommended.<sup>53</sup> However, calibration has been less emphasized in the literature and has recently been described as the “Achilles Heel” of clinical predictive model development.<sup>55</sup> Model calibration is sensitive to shifts in measured and unmeasured covariates, and thus if a patient is not drawn from a population similar to the cohort the model was trained on, the model may provide an incorrect risk estimate.<sup>56</sup> We have broached this problem in illness severity score development, but ultimately deployed models will need to have calibration continuously evaluated, requiring regular re-calibration, as well as users who have the ability to tell when a model does not apply to the patient in front of them.<sup>57</sup>

Although correct metric selection should drive the development of a well-performing model, there remains an important, and as of yet addressed, caveat: causation.<sup>58</sup> Machine learning approaches have demonstrated incredible performance in fitting the associations inherent in the underlying data generating process while avoiding overfitting the random noise that threatens generalization. Nevertheless, these models do not grasp causal structure. As such, they optimize for metrics which ensure prediction based on the associations within the data, but sometimes these associations are spurious and the model relies on an association from a “backdoor path” provided by an unmeasured variable or variables. For example, Caruana et al<sup>59</sup> found that asthma was protective of death from pneumonia when building predictive models for pneumonia outcomes. In fact, in the institution where the data were obtained for the model, an asthma attack triggers a higher level of care. A similar issue has been noted in the application of illness severity scores to morbidly obese patients who are critically ill: their physiology is altered at baseline so that they appear sicker than they truly are based on cut-off values established in a cohort with few morbidly obese patients.<sup>33</sup> This blindness to causality has recently been discussed more broadly within the context of training models on data wrought with human biases. Racial and sex underrepresentation within datasets as a result of structural biases may lead to models that misclassify underrepresented groups, causing misallocations in care that ultimately amplify health care disparities.<sup>60</sup> Model developers must therefore take care because current machine learning approaches are blind to causal structure. Schulam and Saria<sup>61</sup> have approached this with some success by attempting to model not only the factual Gaussian processes present in data, as current approaches do, but also the counterfactual Gaussian processes. However, there is much work to be done



toward the development of models that can identify causal structure.

The problem of causality speaks to the greater problem of model interpretation. As models grow more sophisticated, they also tend to become less interpretable. Deep learning models boast a remarkable ability to examine complex non-linear interactions between inputs, exploiting patterns within the data beyond what the human mind could identify alone, but these models are difficult to interpret and currently represent predictive black boxes.<sup>8,62</sup> A lack of interpretability is a non-starter for physicians in practice and a significant barrier to incorporation of such models into clinical decision making.

Shortliffe and Sepulveda<sup>63</sup> highlighted this issue and emphasized the need to consider how these tools will actually be deployed in practice. In a recent publication, they provide a series of criteria for the development of clinical decision support tools. Their insightful considerations reflect the real-life barriers to uptake including clinician workload and application usability. As most models end up in a graveyard of citations that are never deployed, future efforts should focus on usability, interpretability, and, most importantly, impact on relevant population or health system outcomes.

### Toward a Precision Critical Care

Precision medicine seeks to tailor care to the individual, and precision critical care has become an active research area.<sup>64</sup> Whereas clinicians have always individualized care based on their interpretation of clinical data, the term “precision” has come to mean the use of genomics, expression analyses including proteomics, metabolomics, and other data sources to target the mechanisms which define specific disease phenotypes as well as therapeutic responsiveness. This philosophy has flourished in oncology, where driver mutations and pathway-specific therapies have emerged. However, the critically ill are defined by multi-organ failure occurring via a complex interplay of exposure, host response, and genomic substrate and expression along with innumerable other dimensions of variation that challenge the successful application of such approaches.<sup>65</sup> These -omics data are inherently *big data*, and when this domain is eventually merged with the clinical data, the aforementioned methods, as well as approaches not yet invented, will be essential in detecting true signal from background noise.<sup>66</sup>

The availability of large EHR databases and sophisticated modeling approaches bring us closer to the promise of a precision critical care. The granularity afforded by access to all of the data a patient generates presents an opportunity to examine nuances of care previously inaccessible to the unaided individual clinician. This ability to capture and analyze all the available data will allow clinicians to continuously trend signals to support the iterative formulation of assessments and plans.<sup>67</sup> These data at the population level should also assist in the future creation of more precise therapeutic interventions than

currently available in critical care. For example, individualized differences in the nature or timing of the immune response to sepsis or trauma would inform the selection of treatment A rather than treatment B for a particular constellation of insult, host state of instability, and immunologic response.<sup>68</sup> Artificial intelligence will grow to fill in the gaps in this process out of necessity as the volume and dynamics of the data inputs exceed even the abilities of a clinician dedicated to the bedside care of a single patient.

One example of a potential use for clinical data analytics is in the area of laboratory test interpretation. The idea that a *normal* range based on healthy individuals for most physiologic and laboratory data applies acceptably well to all critically ill patients is increasingly questioned.<sup>69</sup> In practice, clinicians frequently encounter large, but apparently clinically insignificant, deviations from these so-called *normal* ranges. A recent exploratory analysis performed in MIMIC-III examined this very concept, finding that whereas the distribution of laboratory values differed but also overlapped between those who had good outcomes and those who had poor outcomes, accepted reference ranges seemed irrelevant to differentiating these 2 cohorts.<sup>69</sup> Although only a proof-of-concept, this work marks a first step toward patient-individualized, context, and outcome-based reference ranges. The work represents a diagnostic example, but precision medicine involves combined use of diagnostic and therapeutic precision. The latter may be represented by potential treatments for sepsis that are based more specifically on etiologic, organ dysfunction, and immunologic factors.<sup>64,68</sup>

More abstractly, for any given patient in the ICU, there exist a set of recorded variables; the values of these variables, as well as the presence or absence of such data, and even the time of the data collection collectively define an interaction between the patient and the caregivers. Furthermore, each of these collections of patient variables is an element of a data mart which defines the interaction between the ICU's population, the system within which the ICU exists, and the ICU's caregivers. Formalizing these as the ingredients that define the substrate for understanding collective experience, we envision the next generation of EHR to support “dynamic clinical data mining” (DCDM).<sup>70</sup> Specifically, DCDM would enable examination of any single ICU encounter within the context of *similar* encounters, where *similarity* is defined by some metric for grouping. This could be done as simply as applying certain exclusion and inclusion criteria from one ICU course to ascertain others, or by sophisticated unsupervised machine learning approaches that can identify clusters in large multidimensional data.

As critical care does not yet possess gene-based therapies, precision medicine in this area rests on a data-driven capacity to make more individualized decisions in a greater variety of clinical contexts. To begin to approach this task, we must start to store all pertinent data on individual patients, as we are doing, and develop open, de-identified population databases as

we are only beginning to do. Appropriate software, including a variety of machine learning applications, will be required to harness, analyze, and apply the data necessary to ensure that precision medicine can be practiced in this especially complex domain.

## Conclusions

In the short term, with the emergence of powerful machine learning approaches, and data volumes that allow patients to be mapped across an expanding dimension of physiologic variations, we stand at the precipice of a new era of critical care that will be individualized in a data-driven manner. The barriers are myriad, but if clinicians, data scientists, and policy-makers can work together, the vision of a learning health care system may be realized. To achieve this vision of personalized care, physician collaboration with relevant experts such as data scientists is crucial. The training of physicians will require some very fundamental overhauls that may be poorly understood by and even in conflict with the educational hierarchy already in place in medicine. And most importantly, having the most complete, reliable, and interoperable data to work with represents a necessary if insufficient goal for the infrastructure of a digital, learning health system for acutely ill patients.

## Author Contributions

DJS and LAC outlined the scope of the review, and CVC developed the initial manuscript draft. All authors participated in developing the final manuscript.

## REFERENCES

- Beam AL, Kohane IS. Big data and machine learning in health care. *JAMA*. 2018;319:1317–1318.
- Celi LA, Davidzon G, Johnson AE, et al. Bridging the health data divide. *J Med Internet Res*. 2016;18:e325.
- Naylor CD. On the prospects for a (deep) learning health care system. *JAMA*. 2018;320:1099–1100.
- Beam AL, Kohane IS. Translating artificial intelligence into clinical care. *JAMA*. 2016;316:2368–2369.
- Bailey S, Meyfroidt G, Timsit JF. What's new in ICU in 2050: big data and machine learning. *Intensive Care Med*. 2018;44:1524–1527.
- Ferrucci DA. Introduction to, "This is Watson." *IBM J Res Dev*. 2012;56:235–249.
- Silver D, Schrittwieser J, Simonyan K, et al. Mastering the game of Go without human knowledge. *Nature*. 2017;550:354.
- Hinton G. Deep learning—a technology with the potential to transform health care. *JAMA*. 2018;320:1101–1102.
- Stead WW. Clinical implications and challenges of artificial intelligence and deep learning. *JAMA*. 2018;320:1107–1108.
- Ridgeon EE, Young PJ, Bellomo R, Muchetti M, Lembo R, Landoni G. The fragility index in multicenter randomized controlled critical care trials. *Crit Care Med*. 2016;44:1278–1284.
- Zimmerman JE, Kramer AA, McNair DS, Malila FM, Shaffer VL. Intensive care unit length of stay: benchmarking based on Acute Physiology and Chronic Health Evaluation (APACHE) IV. *Crit Care Med*. 2006;34:2517–2529.
- Saeed M, Villarroel M, Reisner AT, et al. Multiparameter Intelligent Monitoring in Intensive Care II: a public-access intensive care unit database. *Crit Care Med*. 2011;39:952–960.
- Johnson AEW, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. *Sci Data*. 2016;3:160035.
- Pollard TJ, Johnson AEW, Raffa JD, Celi LA, Mark RG, Badawi O. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. *Sci Data*. 2018;5:180178.
- Johnson AEW, Stone DJ, Celi LA, Pollard TJ. The MIMIC Code Repository: enabling reproducibility in critical care research. *J Am Med Inform Assoc*. 2018;25:32–39.
- Haendel MA, Chute CG, Robinson PN. Classification, ontology, and precision medicine. *N Engl J Med*. 2018;379:1452–1462.
- McLennan S, Shaw D, Celi LA. The challenge of local consent requirements for global critical care databases. *Intensive Care Med*. 2019;45:246–248.
- Blumenthal D. Realizing the value (and profitability) of digital health data. *Ann Intern Med*. 2017;166:842–843.
- Moskowitz A, McSparron J, Stone DJ, Celi LA. Preparing a new generation of clinicians for the era of Big Data. *Harv Med Stud Rev*. 2015;2:24–27.
- Obermeyer Z, Lee TH. Lost in thought—the limits of the human mind and the future of medicine. *N Engl J Med*. 2017;377:1209–1211.
- Wartman SA, Combs CD. Medical education must move from the information age to the age of artificial intelligence. *Acad Med*. 2018;93:1107–1109.
- Abobaj J, Celi LA, Charlton P, et al. A "datathon" model to support cross-disciplinary collaboration. *Sci Transl Med*. 2016;8:333ps8.
- Celi LA, Lokhandwala S, Montgomery R, et al. Datathons and software to promote reproducible research. *J Med Internet Res*. 2016;18:e230.
- Lyndon MP, Cassidy MP, Celi LA, et al. Hacking hackathons: preparing the next generation for the multidisciplinary world of healthcare technology. *Int J Med Inform*. 2018;112:1–5.
- Li P, Xie C, Pollard T, et al. Promoting secondary analysis of electronic medical records in China: summary of the PLAGH-MIT critical data conference and health datathon. *JMIR Med Inform*. 2017;5:e43.
- Nunez Reiz A. Big data and machine learning in critical care: opportunities for collaborative research. *Med Intensiva*. 2019;43:52–57.
- Neto AS, Kugener G, Bulgarelli L, et al. First Brazilian datathon in critical care. *Rev Bras Ter Intensiva*. 2018;30:6–8.
- Piza FMT, Celi LA, Deliberato RO, et al. Assessing team effectiveness and affective learning in a datathon. *Int J Med Inform*. 2018;112:40–44.
- Bose S, Johnson AEW, Moskowitz A, Celi LA, Raffa JD. Impact of intensive care unit discharge delays on patient outcomes: a retrospective cohort study. *J Intensive Care Med*. 2018;2:800276.
- Lokhandwala S, McCague N, Chahin A, et al. One-year mortality after recovery from critical illness: a retrospective cohort study. *PLoS ONE*. 2018;13:e0197226.
- Marshall DC, Saliccioli JD, Goodson RJ, et al. The association between sodium fluctuations and mortality in surgical patients requiring intensive care. *J Crit Care*. 2017;40:63–68.
- Severson KA, Ritter-Cox L, Raffa JD, Celi LA, Gordon WJ. Vasopressin administration is associated with rising serum lactate levels in patients with sepsis. *J Intensive Care Med*. 2018;12:794925.
- Deliberato RO, Ko S, Komorowski M, et al. Severity of illness scores may misclassify critically ill obese patients. *Crit Care Med*. 2018;46:394–400.
- Gehrmann S, Dernoncourt F, Li Y, et al. Comparing deep learning and concept extraction based methods for patient phenotyping from clinical narratives. *PLoS ONE*. 2018;13:e0192360.
- Moskowitz A, Chen KP, Cooper AZ, Chahin A, Ghassemi MM, Celi LA. Management of atrial fibrillation with rapid ventricular response in the intensive care unit: a secondary analysis of electronic health record data. *Shock*. 2017;48:436–440.
- Sauer CM, Dong J, Celi LA, Ramazzotti D. Improved survival of cancer patients admitted to the intensive care unit between 2002 and 2011 at a U.S. teaching hospital [published ahead of print October 10, 2018]. *Cancer Res Treat*. doi:10.4143/crt.2018.360.
- Serpa Neto A, Deliberato RO, Johnson AEW, et al. Mechanical power of ventilation is associated with mortality in critically ill patients: an analysis of patients in two observational cohorts. *Intensive Care Med*. 2018;44:1914–1922.
- MIT Critical Data. *Secondary Analysis of Electronic Health Records*. New York: Springer; 2016.
- Celi LA, Mark RG, Stone DJ, Montgomery RA. "Big data" in the intensive care unit. Closing the data loop. *Am J Respir Crit Care Med*. 2013;187:1157–1160.
- Budronis A, Bellika JG. The learning healthcare system: where are we now? A systematic review. *J Biomed Inform*. 2016;64:87–92.
- Breslow MJ, Badawi O. Severity scoring in the critically ill: part 2: maximizing value from outcome prediction scoring systems. *Chest*. 2012;141:518–527.
- Breslow MJ, Badawi O. Severity scoring in the critically ill: part 1—interpretation and accuracy of outcome prediction scoring systems. *Chest*. 2012;141:245–252.
- Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316:2402–2410.
- Wong TY, Bressler NM. Artificial intelligence with deep learning technology looks into diabetic retinopathy screening. *JAMA*. 2016;316:2366–2367.
- Cosgriff CV, Celi LA, Sauer CM. Boosting clinical decision-making: machine learning for intensive care unit discharge. *Ann Am Thorac Soc*. 2018;15:804–805.

46. 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.
47. Koyner JL, Carey KA, Edelson DP, Churpek MM. The development of a machine learning inpatient acute kidney injury prediction model. *Crit Care Med*. 2018;46:1070–1077.
48. Rajkomar A, Oren E, Chen K, et al. Scalable and accurate deep learning with electronic health records. *npj Digital Medicine*. 2018;1:18.
49. Rojas JC, Carey KA, Edelson DP, Venable LR, Howell MD, Churpek MM. Predicting intensive care unit readmission with machine learning using electronic health record data. *Ann Am Thorac Soc*. 2018;15:846–853.
50. Agniel D, Kohane IS, Weber GM. Biases in electronic health record data due to processes within the healthcare system: retrospective observational study. *BMJ (Clinical Research Ed)*. 2018;361:k1479.
51. Gianfrancesco MA, Tamang S, Yazdany J, Schmajuk G. Potential biases in machine learning algorithms using electronic health record data. *JAMA Intern Med*. 2018;178:1544–1547.
52. Box GEP. Robustness in the strategy of scientific model building. *Robust Stat*. 1979;1:201–236.
53. Alba AC, Agoritsas T, Walsh M, et al. Discrimination and calibration of clinical prediction models: users' guides to the medical literature. *JAMA*. 2017;318:1377–1384.
54. Leisman DE. Rare events in the ICU: an emerging challenge in classification and prediction. *Crit Care Med*. 2018;46:418–424.
55. Shah ND, Steyerberg EW, Kent DM. Big data and predictive analytics: recalibrating expectations. *JAMA*. 2018;320:27–28.
56. Vergouwe Y, Moons KG, Steyerberg EW. External validity of risk models: use of benchmark values to disentangle a case-mix effect from incorrect coefficients. *Am J Epidemiol*. 2010;172:971–980.
57. Cosgriff CV, Ko S, Sundaresan T, et al. Developing well calibrated illness severity scores for decision support in the critically ill. (Forthcoming)
58. Hernán MA, Hsu J, Healy B. Data science is science's second chance to get causal inference right: a classification of data science tasks. <https://www.semanticscholar.org/paper/Data-science-is-science's-second-chance-to-get-A-of-Hern%C3%A1n-Hsu/d2f83aa22def149095f1dd89b4cf36d09a748a87>.
59. Caruana R, Lou Y, Gehrke J, Koch P, Sturm M, Elhadad N. Intelligible models for healthcare: predicting pneumonia risk and hospital 30-day readmission. Paper presented at: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 11–13, 2015; Sydney, NSW, Australia. <http://people.dbmi.columbia.edu/noemie/papers/15kdd.pdf>.
60. Khullar D. A.I. could worsen health disparities. *New York Times*. January 31, 2019.
61. Schulam P, Saria S. What-if reasoning with counterfactual Gaussian processes. <https://pdfs.semanticscholar.org/4a8e/692ede416d4864e15042696f53300a52089b.pdf>.
62. Wainberg M, Merico D, DeLong A, Frey BJ. Deep learning in biomedicine. *Nature Biotech*. 2018;36:829–838.
63. Shortliffe EH, Sepulveda MJ. Clinical decision support in the era of artificial intelligence. *JAMA*. 2018;320:2199–2200.
64. Maslove DM, Lamontagne F, Marshall JC, Heyland DK. A path to precision in the ICU. *Crit Care*. 2017;21:79.
65. Seymour CW, Gomez H, Chang CH, et al. Precision medicine for all? Challenges and opportunities for a precision medicine approach to critical illness. *Crit Care*. 2017;21:257.
66. Bos LDJ, Azoulay E, Martin-Loeches. Future of the ICU: finding treatable needles in the data haystack. *Intensive Care Med*. 2019;45:240–242.
67. Rush B, Celi LA, Stone DJ. Applying machine learning to continuously monitored physiological data [published online ahead of print November 11, 2018]. *J Clin Monit Comput*. doi:10.1007/s10877-018-0219-z.
68. Shankar-Hari M, Madsen MB, Turgeon AF. Immunoglobulins and sepsis. *Intensive Care Med*. 2018;44:1923–1925.
69. Tyler PD, Du H, Feng M, et al. Assessment of intensive care unit laboratory values that differ from reference ranges and association with patient mortality and length of stay. *JAMA Netw Open*. 2018;1:e184521.
70. Celi LA, Zimolzak AJ, Stone DJ. Dynamic clinical data mining: search engine-based decision support. *JMIR Med Inform*. 2014;2:e13.