



PUESTA AL DÍA EN MEDICINA INTENSIVA: ACTUALIZACIÓN EN METODOLOGÍA EN MEDICINA INTENSIVA

Big data and machine learning in critical care: Opportunities for collaborative research



Organizing Committee of the Madrid 2017 Critical Care Datathon[◇]

Received 8 March 2018; accepted 9 June 2018

Available online 2 August 2018

KEYWORDS

Big data;
Machine learning;
Artificial intelligence;
Clinical databases;
MIMIC III;
Datathon;
Collaborative work

Abstract The introduction of clinical information systems (CIS) in Intensive Care Units (ICUs) offers the possibility of storing a huge amount of machine-ready clinical data that can be used to improve patient outcomes and the allocation of resources, as well as suggest topics for randomized clinical trials. Clinicians, however, usually lack the necessary training for the analysis of large databases. In addition, there are issues referred to patient privacy and consent, and data quality. Multidisciplinary collaboration among clinicians, data engineers, machine-learning experts, statisticians, epidemiologists and other information scientists may overcome these problems.

A multidisciplinary event (Critical Care Datathon) was held in Madrid (Spain) from 1 to 3 December 2017. Under the auspices of the Spanish Critical Care Society (SEMICYUC), the event was organized by the Massachusetts Institute of Technology (MIT) Critical Data Group (Cambridge, MA, USA), the Innovation Unit and Critical Care Department of San Carlos Clinic Hospital, and the Life Supporting Technologies group of Madrid Polytechnic University. After presentations referred to big data in the critical care environment, clinicians, data scientists and other health data science enthusiasts and lawyers worked in collaboration using an anonymized database (MIMIC III). Eight groups were formed to answer different clinical research questions elaborated prior to the meeting.

The event produced analyses for the questions posed and outlined several future clinical research opportunities. Foundations were laid to enable future use of ICU databases in Spain, and a timeline was established for future meetings, as an example of how big data analysis tools have tremendous potential in our field.

© 2018 Elsevier España, S.L.U. y SEMICYUC. All rights reserved.

* Corresponding authors.

E-mail addresses: anunezreiz@gmail.com (A.N. Reiz), miguel.sanchez@salud.madrid.org (M. Sánchez García).

[◇] The names of all authors of this article are listed in [Appendix](#).

PALABRAS CLAVE

Big data;
Machine learning;
 Inteligencia artificial;
 Bases de datos
 clínicos;
 MIMIC III;
 Datathon;
 Trabajo colaborativo

***Big data y machine learning* en medicina intensiva: oportunidades del trabajo colaborativo**

Resumen La aparición de los sistemas de información clínica (SIC) en el entorno de los cuidados intensivos brinda la posibilidad de almacenar una ingente cantidad de datos clínicos en formato electrónico durante el ingreso de los pacientes. Estos datos pueden ser empleados posteriormente para obtener respuestas a preguntas clínicas, para su uso en la gestión de recursos o para sugerir líneas de investigación que luego pueden ser explotadas mediante ensayos clínicos aleatorizados. Sin embargo, los médicos clínicos carecen de la formación necesaria para la explotación de grandes bases de datos, lo que supone un obstáculo para aprovechar esta oportunidad. Además, existen cuestiones de índole legal (seguridad, privacidad, consentimiento de los pacientes) que deben ser abordadas para poder utilizar esta potente herramienta.

El trabajo multidisciplinar con otros profesionales (analistas de datos, estadísticos, epidemiólogos, especialistas en derecho aplicado a grandes bases de datos), puede resolver estas cuestiones y permitir utilizar esta herramienta para investigación clínica o análisis de resultados (*benchmarking*).

Se describe la reunión multidisciplinar (Critical Care Datathon) realizada en Madrid los días 1, 2 y 3 de diciembre de 2017. Esta reunión, celebrada bajo los auspicios de la Sociedad Española de Medicina Intensiva, Crítica y Unidades Coronarias (SEMICYUC) entre otros, fue organizada por el Massachusetts Institute of Technology (MIT), la Unidad de Innovación y el Servicio de Medicina Intensiva del Hospital Clínico San Carlos, así como el grupo de investigación «Life Supporting Technologies» de la Universidad Politécnica de Madrid. Tras unas ponencias de formación sobre *big data*, seguridad y calidad de los datos, y su aplicación al entorno de la medicina intensiva, un grupo de clínicos, analistas de datos, estadísticos, expertos en seguridad informática de datos realizaron sesiones de trabajo colaborativo en grupos utilizando una base de datos reales anonimizada (MIMIC III), para analizar varias preguntas clínicas establecidas previamente a la reunión.

El trabajo colaborativo permitió establecer resultados relevantes con respecto a las preguntas planteadas y esbozar varias líneas de investigación clínica a desarrollar en el futuro. Además, se sentaron las bases para poder utilizar las bases de datos de las UCI con las que contamos en España, y se estableció un calendario de trabajo para planificar futuras reuniones contando con los datos de nuestras unidades.

El empleo de herramientas de *big data* y el trabajo colaborativo con otros profesionales puede permitir ampliar los horizontes en aspectos como el control de calidad de nuestra labor cotidiana, la comparación de resultados entre unidades o la elaboración de nuevas líneas de investigación clínica.

© 2018 Elsevier España, S.L.U. y SEMICYUC. Todos los derechos reservados.

One of the fundamental challenges intensivists face in the care of the critically ill is the management of a huge amount of information gathered from patients. In intensive care units equipped with electronic information systems, hundreds of individual elements of structured information are collected every day in an electronic chart, including vital signs, basic and advanced monitoring records, laboratory data, drug prescriptions and their administration, as well as nursing records, among others. A lot of textual information, like nurse's and physician's comments and progress notes, is also generated, as well as images of different formats and heterogeneous structure.¹

It is impossible for the human brain to fully assimilate this abundance of data, so one of the fundamental qualities of the good intensivist is knowing how to prioritize and filter the most relevant information for a given patient. The exponential growth in computerized information storage and processing systems, however, enables a different approach in which all the available information is taken into account and processed.

Big data analysis (BDA) can be applied to an individual patient, but also be extended to population management

of critically ill patients, or to subgroups of patients meeting certain criteria.² The large volume of data may partially address the lack of "formal" randomization of tests and treatments by presenting instrumental variables with which patients are "pseudo-randomized" to receive or not receive a particular intervention. Analysis of observational data can also generate hypotheses and inform the study design of prospective trials to evaluate those hypotheses. BDA is already routinely employed in other fields as varied as marketing, strategic business decision-making, insurance, banking, transport and logistics services and fraud detection in electronic commerce. Additionally, BDA requires that high quality data are entered by clinicians and nurses.

Apart from the data, the other component of BDA is what is referred to as "Machine Learning" (ML), which fuels what is collectively known as artificial intelligence (AI). This technology uses different methodologies (Table 1) to identify patterns from the data which typically involve either classification or event prediction and subsequently apply the software output to a very specific task, such as automatic interpretation of medical images.³ This technology has been used successfully in domains that, until recently, seemed

Table 1 Machine learning methodologies (Refs. 4–7).

Methodology	Definition
Decision trees	Uses a decision tree as a predictive model with a value assigned to each branch.
Association rules	Serves to discover interesting relationships between variables of large databases.
Neural networks	Based on the operational characteristics of biological neural networks, with different nodes and inferences stored in their interconnections. Non-linear, they serve to explore complex relationships between input and output variables of the system.
Intensive learning (Deep learning)	Neural networks of multiple layers, using hardware with great processing capacity. They try to simulate the model of the human brain and have been used for visual analysis and language recognition.
Inductive logic programming	Uses logical and functional programming to derive software models that replicate the knowledge and relationships that exist in the system.
Support vector machines	Used to create predictive models between two possible outcomes (for example, survival or death). Uses classification and regression methods.
Clustering	Used to allocate observations to subsets of similar study subjects.
Bayesian networks	Represents the interdependence between variables through directed acyclic graphs. Used, for example, to compute the probability of the presence of a disease, given a set of symptoms.
Reinforcement-based learning	Serves to create agents that perform actions in a system, maximizing long term benefits, even if conditions are suboptimal.
Representation learning	Intended to improve the way in which the entries are presented to the system, discovering their general properties.
Learning similarity and metrics	Pairs of similar examples and a similarity metrics are used to generate a model of the system.
Learning of dispersed terms	Used in situations where there is information that is only scarcely represented in the examples.
Genetic Algorithms	A heuristic search using mechanisms of natural selection, evolution, mutation and crossing to progressively improve algorithms.
Rule-based learning	Translates the characteristics of the system into rules that can then be applied in new examples.
Boosted Decision Trees (Ensemble Model) ⁴	Supervised learning method that learns by adjusting the residual of the previous trees.
Random Forests (Ensemble Model) ⁵	Supervised classification algorithm that creates a forest with a direct relation between the number of trees and the results the model can get.
Adaboost (Ensemble Model) ⁶	It is a boosting algorithm developed for binary classification. Boosting algorithm aims to create a strong classifier from a number of weak classifiers
Natural Language Processing ⁷	Artificial intelligence technique to interpret and manipulate human language to computer understanding.

unlikely use cases, such as Jeopardy and the board game Go. Traditional computer systems differ from AI in that in the former, rules are agreed upon by experts and fed into the computers as algorithms. With AI, data is fed into the computers, and the computers “discover” and then implement the rules. In addition, the computers continuously assess the rules and re-calibrate them as needed.

Critical care is particularly well suited for machine learning technology, because technology is a large component of the care provided in the ICU with data as an exhaust of the technology.^{8,9} In fact, there have been ML applications in the ICU, such as in natural language processing, pharmacovigilance and decision support.¹⁰ Providers, including intensivists, however, are typically not trained to develop, implement and evaluate such technologies. The next section deals with how to solve this problem.

Collaborative work

Since it would be extremely unusual that one individual would have both the clinical training and the computer science background for BDA and ML, teams with professionals across different fields are required, composed of data scientists including statisticians and epidemiologists,

clinicians and other biomedical researchers. The coordination of a group with different schedules, the establishment of clear objectives and an effective leadership to manage members with different cultures and who “speak different languages” are all individual challenges in themselves,¹¹ although in critical care medicine teamwork is routinely applied in the day-to-day workflow.

This article describes an initiative event organized by the Massachusetts Institute of Technology (MIT) Critical Data group, the Life Supporting Technologies (LifeSTech) research group at Universidad Politécnica de Madrid (UPM) and the Innovation Unit and Critical Care Department of Hospital Clínico de San Carlos (HCSC) in Madrid.

Databases of critically ill patients

The use of BDA and ML techniques creates a new avenue for clinical research that includes non-traditional, non-academic investigators, especially when large anonymized databases of critical patients are shared under pre-specified data user agreement. The pioneers in this field are Beth Israel Deaconess Medical Center and MIT, whose partnership created the MIMIC database (Medical Information Mart in Intensive Care, <http://mimic.physionet.org>) for the

research community. The MIMIC-III version contains data from more than 50,000 admissions to Beth Israel ICUs from the years 2001 to 2012 and is available online under a data user agreement. It includes vital signs and other monitoring data, laboratory results, data of prescription and administration of drugs, as well as progress notes, diagnostic codes, physiologic waveforms, both in the ICU and for some variables, from throughout the entire stay in the hospital.

Similar high-resolution ICU databases exist, like the Critical Care – Health Informatics Collaborative (CC-HIC) Database,¹² but MIMIC is unique because it is completely open-access. In addition, the MIT group published a textbook that is available at no cost and teaches a course to guide investigators on the analysis of the database.^{3,13–15} Since 2016, they have also organized several datathons¹⁶ in various regions of the world (MIT, London, Paris, Melbourne, Beijing, Sao Paulo and Singapore), to jumpstart disseminate the field of machine learning as applied to health data.

The creators of MIMIC have proposed a system of learning and continuous improvement in Intensive Care they call “closed circuit”.⁸ This strategy consists in using the patient database to answer questions of clinical interest and then taking it back to the bedside.

Initiatives in Spain

Although adoption of electronic medical records and databases of critically ill patients is still in the growth phase in our country, and not all Units have implemented an integrated digital information system, there are already some initiatives under way.

The Spanish Society of Critical Care (Sociedad Española de Medicina Intensiva, Crítica y Unidades Coronarias, SEMICYUC) has published reference standards for the implementation of Clinical Information Systems (CIS),¹⁷ and is in the process of building the national electronic record of the minimum basic data set (CMBD) registry to facilitate quality measurement, benchmarking and clinical research.

In some services and hospitals, the process of creating anonymous data bases for collaborative research has already begun. The Critical Care Department at Hospital Clínico San Carlos in Madrid, in collaboration with Philips, has already anonymized its database including its first 12,000 admissions after implementation of the IntelliSpace Critical Care and Anesthesia (ICCA™, Philips) clinical information system. Other ICUs in Spain, such as the Unit at Hospital Joan XXIII in Tarragona, are working on similar initiatives, using the data obtained automatically from their CIS.^{18,19}

The creation and research use of these patient-level databases have some important legal implications, which need to be addressed to comply with regulatory policies, particularly those under the Spanish Data Protection Law. Projects that employ BDA and ML of these patient databases also require prior evaluation and approval by local Ethics Committees.

Critical Care Datathon Madrid

In order to promote the use of BDA and ML in healthcare in our country, the seventh Critical Care Datathon, organized jointly by the MIT Critical Data Group, the HCSC in

Madrid and the LifeSTech from UPM, took place on December 1, 2 and 3, 2017 at Impact-Hub co-working space in Madrid. The event brought together clinicians, computer scientists and other professionals interested in health data science. HCSC and LifeSTech-UPM group collaborated in the event diffusion informing participants about the content and meaning of the Datathon.

Prior to the event, a group of clinicians from the Critical Care Department of Hospital Clínico San Carlos prepared a list of eight research questions of clinical interest, by way of brain storming, discussion and reaching internal agreement, to be explored during the event: (1) the choice of empirical antibiotics in the ICU on hospital mortality, (2) risk factors for prolonged ICU stay and mortality after cardiac surgery, (3) prediction of poor prognosis after admission for resuscitated out-of-hospital cardiac arrest, (4) prediction of mortality in patients with non-elective ICU admission, (5) prognostic implications of hyperglycemia in critically ill patients with and without diabetes, (6) the epidemiology of massive transfusion of packed blood cells, (7) understanding the association of positive fluid balance and mortality, and (8) seasonality of subarachnoid hemorrhage.

After brief presentations of the background and the rationale behind the research questions with consideration of the non-healthcare professionals among the Datathon attendees, eight groups were formed, one for each topic, with the stipulation that all the necessary expertise (clinical, data extraction and processing, machine learning) is represented in each group.

The MIT Critical Group with the support of LifeSTech from UPM set up for the event virtual servers with the MIMIC-III database with access to platforms and software for data analysis (Python, R, Jupyter notebook). LifeSTech evaluated using natural language processing techniques to extract information from clinician’s reports written in free text, with the purpose of codifying them through SNOMED-CT terminology.²⁰ Mentors from the MIT team and LifeSTech were on hand throughout the weekend to help troubleshoot issues, both technical and those related to the MIMIC database or to data modeling. Finally, each group presented their methods and results to the rest of attendees and a panel of judges from MIT, UPM and HCSC.

As an illustration of the relevant conclusion rendered by the collaborative work of the groups during the datathon, we briefly describe the experience of the group who worked on the project looking at the prediction of mortality among patients with non-elective ICU admission.

First, the clinician in the team presented the existing critical care mortality prediction models (APACHE II, MPM, SAPS 2, SAPS 3), and interest in developing an updated score with better calibration from a larger database. A literature search was carried out to select the scores with the best performance for comparison. The OASIS²¹ and APACHE II²² scores were chosen as the current “gold standard”. OASIS, which was constructed using ML techniques, has the best performance among critically ill patients in the MIMIC III database (area under the receiver operating curve or AUROC, 0.82) (personal communication from MIMIC team), while the APACHE II score, the most frequently used for benchmarking and clinical research, has an AUROC of 0.75.²³

Likewise, the clinician identified variables in MIMIC-III potentially related to prognosis. The data scientists were

responsible for generating the table of more than 40,000 patients with all the selected variables. The entire team went over phases of the model building. In the first phase (training), Support Vector Machines (SVM) was trained using 25,000 patients and produced a model of 30 variables with an AUROC of 0.81. This performance was validated in a different subset of 15,000 patients, with a similar AUROC, confirming the robustness of the model. Presently, the group is working in refining the model and evaluate its performance in the anonymized ICCA™ (Philips) database with 12,000 patients of our Critical Care Department.

The same collaborative work between clinicians and data scientists performed in previous similar events and that we witnessed in the Madrid Datathon, has led to numerous publications,^{24–28} and we have the same expectation from the Madrid event.

The meeting, which was organized under the auspices of SEMICYUC, also served to connect Spanish ICUs working in the field of BDA and fortified partnerships among them to move this initiative forward.

What the future holds and conclusions

We are conceivably at the dawn of a new era in intensive care, in which clinical decision making will increasingly be assisted by computers that perform data integration and analysis. The clinician will therefore have to navigate a specialty of critical care, that harnesses the power of data to individualize care in order to improve population health, where collaborative work with other non-healthcare specialists in the area of data science is essential to leverage all the information that is routinely collected in the process of care, but without compromising patient rights to privacy.

We encourage all those interested in participating in this project to contact the authors, in order to establish a national collaborative network in the field of BDA and ML as applied to healthcare data collected in the ICU, under the auspices of SEMICYUC.

Author's contributions

MIMIC III experts: Celi, Armengol, Deliberato, Paik, Pollard, Raffa, Torres.

Technical logistics (servers and database access): Celi, Armengol, Deliberato, Paik, Pollard, Raffa, Torres, Mayol, Chafer, Rey, Gonzalez, Fico, Lombroni, Hernandez, Lopez, Merino, Cabrera, Arredondo.

Preparation and presentation of research questions: Blesa, Martín, Nieto, Martínez, Álvarez, Sánchez, del Pino, Gil, Núñez.

Clinical support for the research questions: Sanchez, Núñez, Álvarez, Martínez, del Pino, Gil, Bodí, Rodríguez, Gómez, Prada.

Concept and Design of the Datathon: Celi, Armengol, Deliberato, Paik, Pollard, Raffa, Torres, Mayol, Chafer, Rey, Gonzalez, Fico, Lombroni, Hernandez, Lopez, Merino, Cabrera, Arredondo.

Drafting of the manuscript: Núñez, Sánchez, Celi, Bodí, Fico, López.

Venue of the event: Gonzalez, Mayol, Chafer, Rey.

Diffusion: Chafer, Rey, González, Sánchez.

Conflict of interests

The authors declare no conflict of interests

Appendix. List of authors who have participated in the Organizing Committee of Critical Care Datathon Madrid 2017

Antonio Núñez Reiz (MD)^{a,*}, Fernando Martínez Sagasti (MD PhD)^a, Manuel Álvarez González (MD PhD)^a, Antonio Blesa Malpica (MD PhD)^a, Juan Carlos Martín Benítez (MD PhD)^a, Mercedes Nieto Cabrera (MD PhD)^a, Ángela del Pino Ramírez (MD)^a, José Miguel Gil Perdomo (MD)^a, Jesús Prada Alonso^a, Leo Anthony Celi (MD MS MPH)^b, Miguel Ángel Armengol de la Hoz (MBE)^{b,c,d,e}, Rodrigo Deliberato (MD PhD)^b, Kenneth Paik (MD MS MBA)^b, Tom Pollard (PhD)^b, Jesse Raffa (PhD)^b, Felipe Torres^b, Julio Mayol (MD PhD)^f, Joan Chafer (MD)^g, Arturo González Ferrer (PhD)^g, Ángel Rey^g, Henar González Luengo^g, Giuseppe Fico (PhD)^h, Ivana Lombroni (MSc)^h, Liss Hernandez (MSc)^h, Laura López (MSc)^h, Beatriz Merino (MSc)^h, María Fernanda Cabrera (PhD)^h, María Teresa Arredondo (PhD)^h, María Bodí (MD PhD)ⁱ, Josep Gómez (PhD)^{i,j}, Alejandro Rodríguez (MD PhD)ⁱ, Miguel Sánchez García (MD PhD)^{a,*}

^a Servicio de Medicina Intensiva, Hospital Universitario Clínico San Carlos, Madrid, España

^b MIT Critical Data, Laboratory for Computational Physiology, Harvard-MIT Health Sciences & Technology, MIT, Cambridge, Massachusetts, United States

^c Division of Clinical Informatics, Department of Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts, United States

^d Harvard Medical School, Boston, Massachusetts, United States

^e Biomedical Engineering and Telemedicine Group, Biomedical Technology Centre CTB, ETSI Telecomunicación, Universidad Politécnica de Madrid, Madrid, Spain

^f Department of Surgery, Hospital Clínico San Carlos de Madrid, Instituto de Investigación Sanitaria San Carlos, Universidad Complutense de Madrid, Madrid, Spain

^g Unidad de Innovación, Hospital Clínico San Carlos, Instituto de Investigación Sanitaria del Hospital Clínico San Carlos (IdISSC), Madrid, Spain

^h Life Supporting Technologies, departamento de Tecnología Fotónica y Bioingeniería, Universidad Politécnica de Madrid, Madrid, Spain

ⁱ Service of Intensive Care Medicine, Hospital Universitari Joan XXIII, IISPV-URV, Tarragona, Spain

^j Department of Electronic Engineering, Metabolomics Platform, Rovira i Virgili University, IISPV, Tarragona

References

1. Morris GGR. Computer applications. In: Hall JB, Schmidt GA, Wood LDH, editors. *Principles of critical care*. New York: McGraw-Hill, Health Professions Division; 1992. p. 500–14.
2. Ghassemi M, Celi LA, Stone DJ. State of the art review: the data revolution in critical care. *Crit Care*. 2015;19:118.
3. Celi LA, Csete M, Stone D. Optimal data systems: the future of clinical predictions and decision support. *Curr Opin Crit Care*. 2014;20:573–80.
4. Takaki J, Petersen T, Ericson G. Boosted Decision Tree Regression 2018 [23/2/2018]. Available from: <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/boosted-decision-tree-regression>
5. Gu S, Wu Q. How Random Forest Algorithm Works in Machine Learning 2017. Available from: <https://medium.com/@Synced/how-random-forest-algorithm-works-in-machine-learning-3c0fe15b6674> [cited 2018].
6. Brownlee J. Boosting and AdaBoost for Machine Learning 2016. Available from: <https://machinelearningmastery.com/boosting-and-adaboost-for-machine-learning>
7. SAS. Natural Language Processing. Available from: https://www.sas.com/en_us/insights/analytics/what-is-natural-language-processing-nlp.html
8. 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;187:1157–60.
9. Iwashyna TJ, Liu V. What's so different about big data? A primer for clinicians trained to think epidemiologically. *Ann Am Thorac Soc*. 2014;11:1130–5.
10. Ross MK, Wei W, Ohno-Machado L. "Big data" and the electronic health record. *Yearb Med Inform*. 2014;9:97–104.
11. Wu JT, Gehrmann DF, Tyler S, Moseley PD, Carlson ET, Grant ET, et al. Behind the scenes: a medical natural language processing project. *Int J Med Inf*. 2018.
12. Cambridge Uo. NIHR Health Informatics Collaborative. Available from: <http://www.clinicalinformatics.group.cam.ac.uk/hic/intensive-care/>
13. Data MC. Secondary Analysis of Electronic Health Records 2016.
14. Badawi O, Brennan T, Celi LA, Feng M, Ghassemi M, Ippolito A, et al. Making big data useful for health care: a summary of the inaugural MIT critical data conference. *JMIR Med Inform*. 2014;2:e22.
15. Johnson AE, Pollard TJ, Shen L, Lehman LW, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. *Sci Data*. 2016;3:160035.
16. Aboab J, Celi LA, Charlton P, Feng M, Ghassemi M, Marshall DC, et al., A "datathon" model to support cross-disciplinary collaboration. *Sci Transl Med*. 2016;8:333ps8.
17. Gomez Tello V, Alvarez Rodriguez J, Nunez Reiz A, Gonzalez Sanchez JA, Hernandez Abadia de Barbara A, Martinez Fresneda M, et al. Technical and functional standards and implementation of a clinical information system in intensive care units. *Med Intensiva*. 2011;35:484–96.
18. Bodi M, Blanch L, Maspons R. Clinical information systems: an opportunity to measure value, investigate and innovate from the real world. *Med Intensiva*. 2017;41:316–8.
19. Sirgo G, Esteban F, Gómez J, Moreno G, Rodríguez A, Blanch L, et al. Validation of the ICU-DaMa tool for automatically extracting variables for minimum dataset and quality indicators: the importance of data quality assessment. *Int J Med Inform*. 2018;112:166–72.
20. SNOMED-CT. The Global Language of Healthcare: SNOMED International 2018. Available from: <https://www.snomed.org/snomed-ct>
21. Johnson AE, Kramer AA, Clifford GD. A new severity of illness scale using a subset of acute physiology and chronic health evaluation data elements shows comparable predictive accuracy. *Crit Care Med*. 2013;41:1711–8.
22. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med*. 1985;13:818–29.
23. Varghese YE, Kalaiselvan MS, Renuka MK, Arunkumar AS. Comparison of acute physiology and chronic health evaluation II (APACHE II) and acute physiology and chronic health evaluation IV (APACHE IV) severity of illness scoring systems, in a multidisciplinary ICU. *J Anaesthesiol Clin Pharmacol*. 2017;33:248–53.
24. Ghassemi MM, Richter SE, Eche IM, Chen TW, Danziger J, Celi LA. A data-driven approach to optimized medication dosing: a focus on heparin. *Intensive Care Med*. 2014;40:1332–9.
25. Paonessa JR, Brennan T, Pimentel M, Steinhaus D, Feng M, Celi LA. Hyperdynamic left ventricular ejection fraction in the intensive care unit. *Crit Care*. 2015;19:288.
26. Ghassemi M, Marshall J, Singh N, Stone DJ, Celi LA. Leveraging a critical care database: selective serotonin reuptake inhibitor use prior to ICU admission is associated with increased hospital mortality. *Chest*. 2014;145:745–52.
27. 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;3:42–52.
28. Shrimel MG, Ferket BS, Scott DJ, Lee J, Barragan-Bradford D, Pollard T, et al. Time-limited trials of intensive care for critically ill patients with cancer: how long is long enough? *JAMA Oncol*. 2016;2:76–83.