Clareus Scientific Medical Sciences
Volume 1 Issue 1 July 2024

Data Sources for use in a Healthcare Interoperability Resources (HL7-FHIR) Platform. A Decision Guide for Clinicians and Data Scientists in Public ICUs

Citation: Bernardo Chávez P., et al. "Data Sources for use in a Healthcare Interoperability Resources (HL7-FHIR) Platform. A Decision Guide for Clinicians and Data Scientists in Public ICUs". Clareus Scientific Medical Sciences 1.1 (2024): 35-42.

Article Type: Research Article

Received: June 18, 2024

Published: June 29, 2024

Bernardo Chávez P[1,2]*, Luis Chicuy Godoy[2], Mario Cuellar Martínez[2], Mauricio Cuellar Martínez[2], Rodrigo Covarrubias Ganderats[2,3]

[1]*UCI, Hospital El Salvador. Santiago, Chile*

[2]*Hub de Inteligencia Artificial en Salud. Santiago, Chile; Bogotá, Colombia*

[3]*Neurocirugía, Hospital Naval. Viña del Mar, Chile*

***Corresponding Author:** Bernardo Chávez Plaza, ICU Service Adult, Avenida El Salvador 364, Providencia, Santiago, Chile.

Summary

Clinical information systems record large amounts of data from multiple sources. This data could be used to model the process and learn how a decision-making algorithm (CDSS) was actually executed.

We have developed an open source interface that uses data from different patient repositories for the integration of large data mining in a Fast Healthcare Interoperability Resources (FHIR) format. An interoperable solution that has emerged as a multi-layer architecture to manage critical medicine data. The layered architecture focuses on sharing processed information and making it easily available in a control panel for users of the Intensive Care Unit, improving the security and quality of medical care, meeting cybersecurity requirements and under standards. very strict ethics. The introduction of these techniques will allow health care in a more predictive way, through new algorithms.

The results demonstrate that the data mining obtained is a constant flow of highly dynamic and unstructured information; The above goes against the dogma that only EHRs are useful when making decisions. Only 4% of the data is used for this purpose and between 0.002% comes from the EHR, while even new algorithm designs will require structured data to exceed a greater percentage of all data.

We conclude that any clinical process requires a volume of structured data, robust on the basis of data mining and is not feasible with the use of background information only from the EHR, in a highly demanding unit.

Keywords: Data sources; Intensive Care Units; data mining; Artificial Intelligence; Interoperability; Electronic Health Records

Introduction

When analyzing ROI data from an ICU, the analysis is affected by the ICU admission criteria, which vary from one unit to another, and even within the same hospital at different times [1, 4, 8]. But we must also take into account sample selection biases, imprecise definitions of variables, implementation limitations, frequency of measurement of variables, subjective assignment of treatments and overfitting of models [8]. Similarly, many other clinical decisions are subject to subjectivity, such as when a patient is ready to be discharged (selective censoring) [6] or when a patient is readmitted from the ward [7].

This was evident during the COVID-19 pandemic, when demand exceeded ICU capacity [5]. For this, we developed a robust solution that captures anonymized data from different silos [9, 34] in a dynamic and highly stressful environment of a multipurpose and undifferentiated public service.

When analyzing or predicting outcomes of treatment decisions, such as clinical decision support systems (CDSS), a greater amount of data must be incorporated from different sources. But ultimately, we need to know how we will make decisions and what these data sources were and their relative importance.

Importance of methodological soundness

Although EHR data offer enormous research opportunities, their analysis is challenging for several reasons. First, the large feature space introduces more opportunities for variation in data collection with respect to the technology used and sampling frequency compared to repositories and clinical trials. Second, the cohort, exposure, and outcome are defined post hoc and therefore provide an opportunity for data wrangling and p-hacking. Third, the type and quality of data, including which patients are captured, are highly dependent on clinical practice patterns that introduce sampling selection biases leading to spurious associations. It should be noted that EHR data is not primarily collected for research purposes, but rather represents the digital escape from clinical care or, in some cases, serves an administrative function such as billing. But, we have a great variety and quantity of information sources distributed in multiple systems (repositories) that, in addition; They are in different interfaces, with different formats [12] and we can use them in decision making in ICUs. In recent years, with the increasing use of EHR data, we have observed that EHR-based studies are sometimes plagued by similar problems that affect their validity or applicability in clinical practice and this is due to their importance. relative between different data sources. Furthermore, although FHIR is widely used by health systems around the world, there remains a paucity of health system-derived data publicly available in FHIR.

Therefore, an HL7 FHIR-based system provides unprecedented opportunities for broader dissemination and sharing in the logic of a CDSS, with innovative capabilities and instruments that facilitate the development of predictive models in the critical environment.

The exchange of clinical information is not continuous due to several technical difficulties, with information interoperability being one of the main technical challenges. If interoperability standards allow the transfer and exchange of health data mining, we must also consider authentication, authorization, registration and auditing mechanisms that must guarantee the privacy, integrity and confidentiality of personal information.

Objectives of this work

We all know that enormous amounts of data are generated in the intensive care environment (a true data mining). Given its potential role in CDSS. In this article we review the current set of CDSS standards that are based on HL7-FHIR in the development of a solution for an ICU of a public health system. Ultimately, large sample sizes may provide greater statistical power to perform subgroup analyzes and reduce the risk of type II errors.

Our objective is to demonstrate the large data mining of an ICU and what are the different sources of clinical data that could be effectively used in decision making [5, 6] when using a data capture solution, estimating the relative importance of the different datasets.

Methodology

First part: Selection of literature

The literature search was carried out in PubMed and Embase, we systematically searched for observational studies published between January 2019 and July 2021 and narrated reviews before November 24, 2023. Abstracts without full text were excluded. Search terms used to find literature included: ("machine learning" or "deep learning" or "neural network" or "artificial intelligence") and ("ICU", "Critical care", "Electronic Clinical Record", "Take of Clinical Decisions"). Initially, a total of 74 articles were identified with these search terms, of which 15 non-full-text abstracts were excluded, leading to a final count of 59. Given the narrative nature of this review, the final cohort of articles to provide the reader with the best overview of the topic and were not intended to be exhaustive. We selected some other research manuscripts, systematic reviews and doctoral theses, and referenced a number of narrative reviews. This article is based on previous studies and only contains studies with human participants.

Second part: Application of the principles of Artificial Intelligence

The original prediction system is mainly based on CDSS, usually using data collected at the bedside in the ICU by a Smart UPC system [9]. These massive and constant streams of data include heart rate, respiratory rate, blood pressure (systolic, diastolic and mean, differential pressure), temperature, urine output. Information from the electronic medical record (EHR), monitoring of mechanical ventilators, coverage of visual images from closed circuit video systems.

It is extracted through the various application programming interfaces (API); Image data is used via DICOM data capture protocols. Or through gateways, from the same suppliers for various clinical devices. Likewise, we collect and analyze more and more data from different sensors in critical units, through biosurveillance monitors and other Internet of Things (IoT) data sources in real-time transmission to a solution based on different layers. This information timed on timelines is displayed on a dashboard for the decision-making work of doctors and paramedical staff in the command and control rooms that are incorporated into the ICU [9].

The harmonization and standardization of digital medical information for research purposes is an ongoing challenge and collaborative effort. Current research data repositories typically require extensive efforts to harmonize and transform original clinical data. The FHIR standard provides a framework for structuring health data and supporting data exchange between disparate systems and providers. FHIR documentation is very detailed and openly available. Briefly, FHIR consists of a set of resources that describe the most common entities in healthcare. As local contexts are expected to deviate from the global standard, FHIR introduces a mechanism called "profiling." Profiles are modifications of core FHIR resources intended for use in a local context, such as a hospital system or primary care clinic. These profiles allow the flexibility to capture data as it is exchanged on the local system, while also allowing for standardization, as the majority of fields present in each resource will be consistent with the global standard (the base resource). FHIR also provides significant details on all aspects of information sharing and has a large ecosystem of tools, particularly around data sharing. The FHIR format was designed primarily to represent processes; therefore, it closely resembles the clinical data model and is more available in modern electronic health systems. However, no common standardized data format is directly suitable for statistical analyses, and data must be pre-processed before statistical analysis.

Demographic data, comorbidities, vital signs, medications infused, and test results are included in the training data set in the different algorithmic models of future use in an ICU [14]. For example, a model based on deep learning can advance in the prediction of ventilatory asynchronies or sepsis [11] it will have better performance and its analysis of the area under the curve will be more optimal if the training data increases. In addition to the deep learning mentioned above, other studies have developed new explainable AI models for the prediction of cardiovascular, neurological pathologies, etc. [6, 11] in ICU environments.

With sufficient data, we are in a position to perform normal artificial intelligence (AI) with training data and control for algorithms according to the different requirements that we have been and will build for the different pathologies [4].

With the models developed based on public data from the UCI [24], we can correlate results with our system. The explainable AI model extracts time-varying features and is capable of real-time prediction of different pathologies based on clinical requirements.

The influence of each feature on real-time forecasting will be further analyzed to show their interoperability in future reports.

This model not only has superior performance in estimating the risk of isolated pathologies in real time, but also provides interpretable information to understand the different risks of patients with multiple pathologies [7]. In our work proposal, to evaluate this function, for example, the performance of deep learning was compared with other methods in the early prediction of different pathologies and which we will make explicit in future reports.

Part Three: The security and challenges of using AI as a decision-making system (CDSS). Data bias errors

The potential to create AI-based healthcare applications can match or exceed the capabilities of doctors in specific diseases. However, health care is a complex and changing topic. When the AI system makes a decision, doctors and safety engineers essentially cannot control the process, and it is difficult to fully understand how the AI system makes the decision accurately. Compared to standard clinical practice, AI-based tools also require generating ethical restrictions and safety standards.

The clinical context of an ICU is very complex and many variables (new therapies, new diagnoses, different intervention times and intervention methods, trained personnel and own protocols) will affect the results. However, the requirements of all clinical environments shown in the computational model are difficult to achieve at a technical design stage. Therefore, the behavior of our solution in the system may not adequately reflect clinical intentions [34]. This issue currently resolves some aspects of the process, without limiting the number of entries, but may have unintended consequences.

It is the doctor who interprets the data, he must reason about the most important content in the different stages of a disease. For example, doctors may choose to ignore highly abnormal test results, which may be due to errors in sampling, testing, or recording. In this way, the definitions are prone to errors and are not fixed. Additionally, if reported, model performance should be compared to existing algorithms, whether based on expert opinions or based on machine learning; If machine learning-based algorithms are used to compare, differences. Differences in definitions should also be explored. In recent years, this analysis has been carried out using the EHRs as a basis. However, EHRs often do not capture the relevant clinical determinants of health in such a short space of time and their multiple clinical parameters that are being generated in continuous, real-time monitoring of an ICU.

Figure 1 shows a list of data sources and their total data load/hour; total data/day and capture percentage in an undifferentiated public ICU.

On the other hand, there are problems in the different AI models, for example, many studies have only trained and validated the model in the same cohort of patients with static databases, but have not yet tested its generality in other populations in dynamic times. These models need to undergo further prospective testing to demonstrate their clinical benefits or other outcomes (the reason for our upcoming publications).

AI will also face many implementation difficulties when used in clinical practice in an ICU. Currently, many organizations are ill-equipped to implement AI in clinical practice, requiring considerable AI expertise and mature information technology or IT capabilities, such as assessment, fusion, continuous monitoring, and AI recalibration. The security and reliability of digital data collection and use must also be addressed.

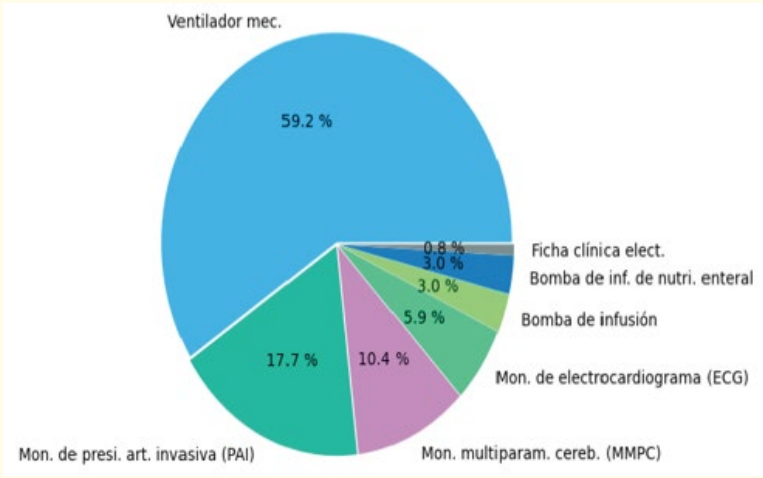

Figure 1: Data loading table and graph, used in Smart UPC system.

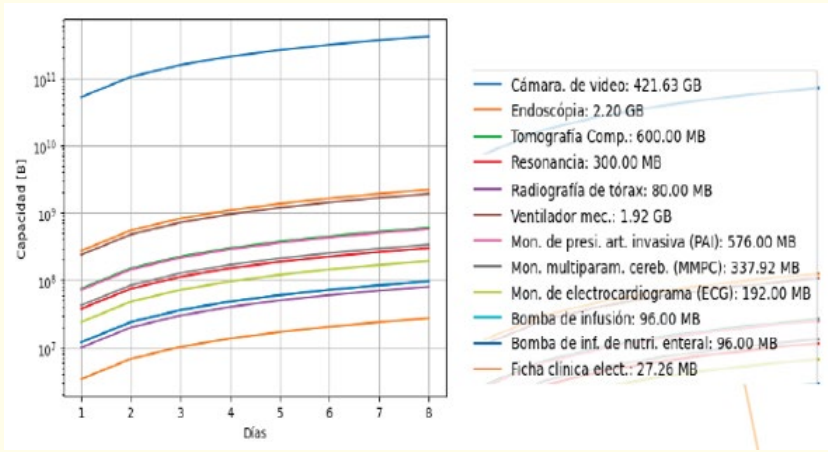

Figure 2: Shows a logarithmic graph of the data collected in 1 patient in 8 days of ICU with video control.

Results

More than 3,000 hours of work have been necessary between medical teams and specialized technicians from the UPC to obtain a data management platform scalable hybrid system to easily collect, process, protect, and analyze all of this data to evaluate ways to improve patient outcomes [9, 34].

With data visualization and processing capacity for medical decision making and also in nursing, pharmacology and kinesiology clinics. We integrate different systems in Standard HL7v3 FHIR formats on our platform [9].

The situational analysis of each patient through the visualization layer of the Integrated Command Center and deployable on any mobile device in the future Figure 3 [34].

Decision-making and predictive models are not yet widely developed with the current known health systems and many improvements will be necessary to serve as a tool in our healthcare centers. Here we present the ability of a solution to maintain different datasets and their relative importance in clinical decision support systems (CDSS).

Here we demonstrate that in general the data available through a Smart UPC solution reaches a value of 2244 MB/hr, that is, an average of 53.34GB/day, which is available for decision making (CDSS).

Of the total available data; the EHRs only reach 0.8% of the data used. For 1 patient, during 8 days of ICU with video control (average) we can collect data according to what is expressed in the graph in Figure 2. Structured data represents only 4% of the total. Unstructured data, 46.5% and metadata, 34%. There is a percentage of data: 14.5% that correspond to different variable sources of unstructured, non-daily data use in the ICU (dialysis machines, ultrasounds, echocardiography, plasmapheresis equipment, neurological multiparametric monitoring equipment, interconsultant evaluations, etc.).

Our results, using this Smart UPC solution, show that the data volume of an EHR reaches 0.002% and the current decision making is carried out with only 3.964% of the total.

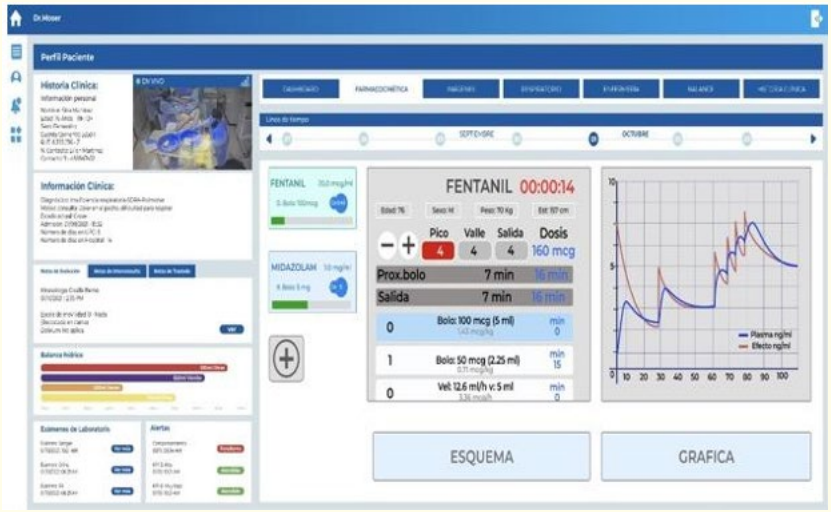

Figure 3: Dashboard solution.

Conclusions

Databases are essential for innovative and disruptive solutions. Any clinical process for Decision Making supported (CDSS) by Artificial Intelligence requires a robust volume of structured data to carry out validly effective algorithms.

A CDSS system, such as Smart UPC, is supported by structured data from multiple sources, which is evident in a highly demanding unit such as ICUs.

The data represented by the EHRs represent 0.002% and only 4% is useful as structured data that can be used in predictive models.

Recognition

To the medical teams and specialists of the UPC of Hospital El Salvador, Santiago de Chile. To all the engineers and professionals of

the Artificial Intelligence Hub in Health. Santiago, Chile - Bogotá, Colombia.

Conflict of interests

The authors (Bernardo Chávez P. Rodrigo Covarrubias G.) have no conflict of interest. (Luis Chicuy Godoy; Mario Cuellar Martínez; Mauricio Cuellar Martínez; Germán Torres) Partners and data scientists of technology companies (smart cities and cybersecurity).

Contribution of own private capital through the Project

"*Strengthening Patient Safety through Technological Solutions for Critical Medicine Units*"; of the Artificial Intelligence in Health Hub group. Santiago, Chile and Bogotá, Colombia.

Informed consent

This study belongs to a bibliographic review and does not require patient consent.

Ethical approval

Institutional Review Board approval was obtained exclusively for the project.

References

1. Tambone V., et al. "Ethical Criteria for the Admission and Management of Patients in the ICU Under Conditions of Limited Medical Resources: A Shared International Proposal in View of the COVID-19 Pandemic". Frontiers in Public Health 8 (2020): 284.
2. Curtis JR and Vincent J-L. "Ethics and end-of-life care for adults in the intensive care unit". The Lancet 376.9749 (2010): 1347-53.
3. Piers RD., et al. "Perceptions of Appropriateness of Care Among European and Israeli Intensive Care Unit Nurses and Physicians". JAMA 306.24 (2011): 2694-703.
4. Lapsley I and Melia K. "Clinical actions and financial constraints: the limits to rationing intensive care". Sociology of Health & Illness 23.5 (2001): 729-46.
5. Trentini F., et al. "The pressure on healthcare system and intensive care utilization during the COVID-19 outbreak in the Lombardy region: a retrospective observational study on 43,538 hospitalized patients". Am J Epidemiol 191.1 (2022): 137-146.
6. Thoral PJ., et al. "Explainable Machine Learning on AmsterdamUMCdb for ICU Discharge Decision Support: Uniting Intensivists and Data Scientists". Critical Care Explorations 3.9 (2021): e0529.
7. Cosgriff CV, Celi LA and Sauer CM. "Boosting Clinical Decision-making: Machine Learning for Intensive Care Unit Discharge". Annals ATS 15.7 (2018): 804-5.
8. Sauer Christopher M., et al. "Leveraging electronic health records for data science: common pitfalls and how to avoid them". The Lancet Digital Health 4.12 (2022): e893-e898.
9. Chávez Plaza B., et al. "HL7 FHIR Platform, Scalable, Reliable and Comprehensive of Clinical Databases Analyzed with Machine Learning for ICU Public Healthcare Center". Springer Nature Switzerland AG 2024. E. Pino et al. (Eds.): ICBHI 2022, IFMBE Proceedings 108 (2024) 1-6.
10. Tobias Gentner., et al. "Data Lakes in Healthcare: Applications and Benefits from the Perspective of Data Sources and Players". Procedia Computer Science 225 (2023): 1302-1311.
11. Sutton RT., et al. "An overview of clinical decision support systems: benefits, risks, and strategies for success". npj Digit. Med 3 (2020): 17.
12. Hripcsak G. "Arden Syntax for Medical Logic Modules". MD Comput 8.2 (1991): 76-8.
13. HL7 International. Arden Syntax v 2.10 (Health Level Seven Arden Syntax for Medical Logic Systems, Version 2.10).
14. Johnson PD., et al. "A virtual medical record for guideline-based decision support". Proc AMIA Symp (2001): 294-8.
15. Topol EJ. "High-performance medicine: the convergence of human and artificial intelligence". Nat Med 25 (2019): 44-56.

16. Bailly S, Meyfroidt G, Timsit JF. "News at UCI in 2050: big data and machine learning". Intensive Care Med 44 (2018): 1524-1527.
17. Ghassemi M, Celi LA and Stone DJ. "State of the art review: the data revolution in critical care". Critical Care 19 (2015): 118.
18. Beam AL and Kohane IS. "Translate artificial intelligence into clinical care". JAMA 316 (2016): 2368-2369.
19. Ince C. "Intensive care medicine in 2050: the ICU in vivo". Intensive Care Med 43 (2017): 1700-1702.
20. Nemati S., et al. "An interpretable machine learning model for accurate prediction of sepsis in the ICU". Crit Care Med 46 (2018): 547-553.
21. Pirracchio R., et al. "Predicting mortality in intensive care units with the Super ICU Learning Algorithm (SICULA): a population-based study". Lancet Respir Med 3 (2015): 42-52.
22. Gårdlund B., et al. "Six subphenotypes in septic shock: latent class analysis of the PROWESS Shock study". J Crit Care 47 (2018): 70-79.
23. Antcliffe DB., et al. "Transcriptomic signatures in sepsis and a differential response to steroids: from the randomized VANISH trial". Am J Respir Crit Care Med 199 (2018): 980-986.
24. Seymour CW., et al. "Derivation, validation, and potential treatment implications of novel clinical phenotypes for sepsis". JAMA 321.20 (2019): 2003-2017.
25. Sutton RS and Barto AG. "Reinforcement Learning: An Introduction". 2nd ed. MIT Press, Cambridge (2018).
26. Guiza F, Van Eyck J and Meyfroidt G. "Predictive data mining on intensive care unit monitoring data". J Clin Monitor Computer 27 (2013): 449-453.
27. Gottesman O., et al. "Guidelines for reinforcement learning in healthcare". Nat Med 25 (2019): 16-18.
28. Komorowski M., et al. "AI doctor learns optimal treatment strategies for sepsis in intensive care". Nat Med 24 (2018): 1716-1720.
29. HL7 International. DSTU updates.
30. HL7 International. Welcome to FHIR.
31. 21st Century Cures Act: Interoperability, information blocking, and the ONC Health IT Certification Program. Fed Regist 85 (2020): 25642-961.
32. Jung CY, Sward KA and Haug PJ. "Executing medical logic modules expressed in ArdenML using Drools". J Am Med Inform Assoc 19.4 (2012): 533-6.
33. Q Mao., et al. "Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU". BMJ Open 8.1 (2018): e017833.
34. Bernardo Chávez P., et al. SUB-PROJECT: "Monitoring with command and control panels for the centralized management of data in critical patients, which guarantees the safety and quality of services, with a robust, interoperable and resilient platform of Artificial Intelligence and Machine Learning, at the Hospital El Salvador, Santiago de Chile". International Conference on Biomedical and Health Informatics 2022. IFMBE Proceedings (2022).