Quality Improvement through Intensive Care Registries: The Value of Big Data in Clinical Decision Making

Ali Sanaei1,2, Mohammad Mehdi Sepehri1*

1Faculty of Industrial and Systems Engineering, Tarbiat Modares University, Tehran, Iran.
2Center of Excellence in Healthcare Systems Engineering, Tarbiat Modares University, Tehran, Iran

ABSTRACT

Background: Quality of Intensive care has got more attention in case of the high cost of healthcare and the potential for harm. Poor-quality care causes high cost and quality improvement initiatives in the ICU lead to an improvement in outcomes as well as a decrease in costs. One of the crucial tools that allow physicians and nurses to monitor change in a quality improvement effort is the development of an electronic database for data collection and reporting. The objective of Intensive Care Registries is to create a high-quality registry of patients through a collaboration of academic health centers performing uniform data collection with the purpose of improving the quality and accuracy of healthcare decisions and provide a data-driven clinical decision support system for critical care medicine.

Methods: This article reviews real-world data sources in healthcare and considers registry as the main tool to address health services and outcomes research questions in critical care, and briefly describes objective, inputs and outputs of intensive care registries. As it can be comprehended from library research, the combination of patient clinical care data, quality parameters, and ICU operating costs, integrated into an electronic database, provides a valuable tool for quality improvement and overall efficiency of offered care.

Results: Using Big Data effectively within ICUs for supporting clinical decision making can lead to predict numerous diseases and help to discover new patterns in healthcare. The ability to process multiple high-speed clinical data streams from multiple centers could dramatically improve both healthcare efficiency and patient outcomes.

Conclusion: To gain this goal, developing reliable and standardized health analytics platforms as well as quality improvement processes that translate analytical results into new clinical guidelines, is recommended.

© 2020 Tehran University of Medical Sciences. All rights reserved.

Gartner (2013) defines Big Data as ‘high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision-making’.

Big-Data Analytics is the process of collecting, organizing and analyzing large sets of data called Big Data which are archived to a data warehouse via Extract-Transform-Load (ETL) routines, to discover patterns and other useful information. Big Data Analytics can help organizations to understand better the information contained within the data and will also help in identifying the data that is most important to the business and future business decisions. Analysts working with Big Data are interested in acquiring knowledge that comes from analyzing the data [1].

Patients, providers, researchers and other healthcare constituents can benefit from big data analytics. Big data analytics can affect on how these players engage within healthcare ecosystem [2]. In the global healthcare sector,
there are three major types of digital data: clinical records, health research records, and business/organization operations records [3]. Health Registries are a great example of Big Data application in healthcare.

There are two perspectives on the use of big data in the healthcare system: 1. Big data can be a risky and costly deviation in the healthcare sector in developing countries. 2. The opposite argument is that big data can be a major developmental point in the healthcare system to improve health status.

Big Data transforms healthcare sector by improving the outcomes by applying potential healthcare analytics. Doctors can take speed decisions based on the results, which are achieved by applying Big Data Analytics. Healthcare sector business is getting profits with the advent of analytics. Healthcare data can be analyzed by choosing suitable analytical tools, data collections, data sharing through EHR, EMR and exchange of medical information. The developments of healthcare standards can improve in identifying and predicting diseases in early stage and can be cured in minimum time [4].

Therefore, Big data analysis in healthcare uses a data explosion to extract the insight needed to make better decisions. By discovering communications and understanding patterns and trends through data, data analysis has the potential for better care, saving lives and reducing costs.

Healthcare organizations ranging from single-physician offices and multi-provider associations to large hospital networks and responsible healthcare organizations benefit significantly through the digitization, integration and efficient use of big data [5]. Potential benefits include early detection of diseases where they can be treated more easily and effectively; faster and more efficient management of patient and population health and prevention of fraud in health care. Big data analytics will tackle numerous issues. Some developments or outcomes may be predicted and/or estimated based on vast historical data such as length of stay (LOS); patients choosing elective surgery; patients who are unlikely to benefit from surgery; complications; patients at risk of medical complications; and possible comorbid conditions (EMC Consulting). McKinsey reports that Big Data Analytics can save U.S. healthcare more than $300 billion a year, two-thirds of that by cutting national healthcare spending by about 8 percent. Clinical operations and R & D are two of the largest areas for potential savings with $165 billion and $108 billion in waste, respectively [6].

(Table 1) describes the platforms & tools for big data analytics in healthcare.

Real-World Data

ISPOR (International Society of Pharmacoeconomics and Outcomes Research) identifies the following real-world data sources that can be used for evaluations of health technology [7]. Each of these sources provides information tailored to different needs:

- Electronic Health Records (EHRs)
- Registries
- Information collected throughout RCTs
- Practical clinical trials
- Administrative data
- Health surveys

(Table 2) describes the various critical care databases organized by the degree of clinical detail available and describes the accessibility of each data source in a qualitative manner [9].

Often a single dataset may provide only part of the information that is necessary to conduct a successful analysis. In such situations, investigators can either supplement the data source by collecting additional data or link two or more existing data sources. The often-easier option involves the linking of two independent but preexisting data sources that together have the necessary information for the question. Occasionally, this linkage has already been done prior to obtaining the data [9].

<table>
<thead>
<tr>
<th>Platform/Tool</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Hadoop Distributed File System (HDFS)</td>
<td>HDFS allows the Hadoop cluster's underlying storage. It divides the data into smaller components and distributes it across the different servers / nodes.</td>
</tr>
<tr>
<td>Map Reduce</td>
<td>Map Reduce provides the interface for sub-task distribution and performance selection. When performing tasks, MapReduce tracks each server / node storage.</td>
</tr>
<tr>
<td>PIG and PIG Latin (Pig and PigLatin)</td>
<td>Pig programming language is set up to assimilate all data types (structured / unstructured, etc.). This consists of two main modules: the language itself, called PigLatin, and the runtime version of the execution of the PigLatin code.</td>
</tr>
<tr>
<td>Hive</td>
<td>Hive is a runtime support architecture for Hadoop that uses the Hadoop platform to leverage Structure Query Language (SQL).</td>
</tr>
<tr>
<td>Jaql</td>
<td>Jaql is a simple, declarative query language for the processing of large sets of data. Jaql converts “high-level” queries into ‘low-level’ queries consisting of MapReduce tasks to facilitate parallel processing.</td>
</tr>
</tbody>
</table>
**Zookeeper**

Zookeeper allows a centralized network with different services to provide connectivity across a database cluster. Such tools are used by large data analysis systems to organize parallel processing across large clusters.

**HBase**

HBase is a column-oriented HDFS database management system. This uses an approach that is not SQL.

**Cassandra**

Cassandra is a distributed database system which is designated as a top-level project handled big data distributed across many utility servers. It is a NoSQL system.

**Oozie**

Oozie, an open-source project, streamlines the workflow and coordination among the tasks.

**Lucene**

The Lucene project is widely used for text analysis / search and has been included in several open-source projects. The scope involves the indexing of full text and the scanning for libraries for use in a Java program.

**Avro**

Avro facilitates data serialization services. Versioning and version control are additional useful features.

**Mahout**

Mahout is another Apache project which aims to create free applications of distributed and scalable machine learning algorithms that support big data analytics on the Hadoop platform.

---

### Table 2 - Description of features for example data sources containing critically ill patients [9]

<table>
<thead>
<tr>
<th>Type of Data</th>
<th>Clinical Detail</th>
<th>Example Data Sources</th>
<th>Represented population</th>
<th>Accessibility/Cost</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Publicly-available studies</td>
<td>High</td>
<td>Acute respiratory</td>
<td>Specific to individual</td>
<td>High/Free</td>
<td>Application for data available online</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Distress Syndrome Network(52)</td>
<td>studies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Electronic medical record</td>
<td>High</td>
<td>Multiparameter Intelligent Monitoring in Intensive Care Unit(44)</td>
<td>&gt;30,000 ICU patients at Beth Israel Deaconess since 2002</td>
<td>High/Free</td>
<td>Minute-to-minute data on physiologic variables; time-stamped treatments</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>Veterans Affairs, Inpatient Evaluation Center(42)</td>
<td>&gt;100 Veterans hospitals</td>
<td>Low/variable</td>
<td>Allows for risk adjustment on par with APACHE; Use requires partnering with VA investigator</td>
</tr>
<tr>
<td>Quality improvement, benchmarking</td>
<td>High</td>
<td>Epic, Cerner, local systems</td>
<td>Usually limited to local health system unless EMR system widely disseminated</td>
<td>Low/variable</td>
<td></td>
</tr>
<tr>
<td>Registries</td>
<td>Moderate</td>
<td>Michigan Hospital Association Keystone initiative(30)</td>
<td>10 participating hospitals in Michigan</td>
<td>Low/-</td>
<td>Use usually requires Keystone investigator</td>
</tr>
<tr>
<td></td>
<td></td>
<td>APACHE(75)</td>
<td>45 voluntarily hospitals</td>
<td>Low/-</td>
<td>Use requires partnering with Cerner</td>
</tr>
<tr>
<td></td>
<td></td>
<td>eICU Research Institute (formerly VISCU)(79)</td>
<td>&gt;1million patients in &gt;100 non-federal voluntary hospitals</td>
<td>Low/-</td>
<td>Use requires partnering with University of Maryland</td>
</tr>
<tr>
<td>Administrative/utilization claims</td>
<td>Moderate</td>
<td>University Health System Consortium(31)</td>
<td>90% academic and affiliated hospitals in US</td>
<td>Moderate/-</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>National Trauma Data Bank(82)</td>
<td>Nationally representative sample of trauma</td>
<td>Moderate/2000/yr</td>
<td>Relatively inexpensive; obtained from the American College of Surgeons</td>
</tr>
<tr>
<td>Registries</td>
<td>Moderate</td>
<td>American Heart Association Get With the Guidelines Resuscitation(87)</td>
<td>&gt; 400 voluntarily participating hospitals</td>
<td>Moderate/Free</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Premier Perspective(105)</td>
<td>Discharges of all payers from over 600 hospitals</td>
<td>Moderate/$50,000-$100,000</td>
<td>Nation's largest patient drug utilization database and unique includes timestamps for labs, tests, procedures</td>
</tr>
<tr>
<td></td>
<td>Moderate/low</td>
<td>MarketScan(102)</td>
<td>&gt;56 million discharges; &gt;100 insurers</td>
<td>Low/Thousands</td>
<td>Inpatient + skilled nursing combined file (MedPAR) = $900-$2,500</td>
</tr>
<tr>
<td></td>
<td>Moderate/low</td>
<td>Medicare(93)</td>
<td>Currently enrolled beneficiaries</td>
<td>Moderate/5000-$20,000/file/year</td>
<td></td>
</tr>
</tbody>
</table>

---

**Methods**

**Registry as a source of Real-World Data**

In the field of health, several definitions of the term registry or register have been provided. In 1949, marjorie [10] defined register as “system of recording frequently used in the general field of public health which serves as a device for the administration of programs concerned with the long-term care, follow-up or observation of individual cases.” In 1974, the Brooke and Organization [11] defined a register as a “file of documents containing uniform information about individual persons, collected..."
in a systematic and comprehensive way, in order to serve a predetermined purpose.” Another definition was provided by Workman [12] who defined a registry as a “database of identifiable persons containing a clearly defined set of health and demographic data collected for a specific public health purpose.” A slightly different definition of a registry is proposed by Polygenis [13], which describes a registry as a “prospective an observational study of subjects with certain shared characteristics, which collects ongoing and supporting data overtime on well-defined outcomes of interest for analysis and reporting.” A more specific definition is provided by the US National Committee on Vital and Health Statistics, which defines a registry as “an organized system for the collection, storage, retrieval, analysis, and dissemination of information on individual persons who have either a particular disease, a condition (e.g., a risk factor) that predisposes (them) to the occurrence of a health-related event, or prior exposure to substances (or circumstances) known or suspected to cause adverse health effects.”

Despite differences in definition, it is clear that a registry involves a long-term, systematic and organized data collection process driven by specific, pre-defined objectives [14].

A registry is a system based on information technology that aims to set up, sustain, and extend a dataset for a particular subject (patients, illnesses, healthcare providers or any others who receive any healthcare services). The variety, acceptability, and use of a registry determine its effectiveness and significance. Data Analytics can deepen researches and be employed for policy making, planning, improving processes, increasing quality and reducing costs.

**History of Intensive Care (IC) Registries**

The collection of information on intensive care patients, their treatment and their outcomes began in the 1950s in order to communicate and exchange experiences [15]. Several initiatives emerged from these early attempts to collect and analyze comprehensive information sets from a large number of intensive care units (ICUs), with an initial focus on documenting and learning from daily practice. From collecting data from the charts of discharged patients, registries then formed to collect more reliable data that was previously defined. The importance of registries increased by shifting from ICU-accumulated information to individual patient-level data [16].

In the mid-1980s and 90s, interest was broadened beyond organ support techniques and disease entities to include the performance of ICUs in general. The focus was, in the beginning, mainly directed at mortality addressing effects of, for example, case mix and resource use. The arrival of risk adjustment models specific for intensive care was an important development promoting audit and benchmarking [17]. In parallel, intensive care resource use and costs climbed significantly, strongly demanding that good arguments could be produced to support this type of care. Resource consumption, as an important management factor, was already addressed in 1980 by Thibault and co-workers [18]. In 1991, a task force was launched by ESICM to improve ICU management and quality of intensive care provided by European ICUs [19]. It was later stated that the highest quality treatment should be aimed for in order to obtain the best outcome (‘to perform is to achieve’; Reis Miranda 1998, in the foreword to the EURICUS study) [17].

Words and phrases that were highly thought of at the time were related to the quality of care and quality assurance, and significant contribution to the techniques and theories that were applied in intensive care came from industry. Although the interest and enthusiasm for quality assurance was strong, it was also appreciated at that early time that the processes of quality improvement are complex, hard, not straightforward, and thereby making it difficult to progress [20]. One important step forward was the joint collection of clinical data describing important aspects, including outcomes of intensive care and quality of care. Although national as well as international registries for separate diseases and in different specialties have existed long before, as exemplified by, that is, the amniocentesis registry from 1970, [21] at this time, that is, in the late 1980s, increasing efforts were being made to create national intensive care registries to be used for the purpose of describing intensive care and outcome after intensive care and supporting the medical profession to provide high quality and safe care has, since then, become a main mission of intensive care registries. For intensive care in Scandinavia, the Finnish registry was the first to launch in 1994 (Finnish Intensive Care Consortium), [22] and this was subsequently followed by the Norwegian (Norskt intensive register), the Swedish (SIR) and the Danish (Dansk Intensive Database) registries.

While sketchy data that characterized ICUs by describing levels of activity (i.e. number of admissions, length of stay, workload, etc.) were useful at the start, registries holding detailed individualized information on consecutive patients (i.e. characteristics, diseases, interventions, outcomes) for long periods of time became powerful tools, which generated important observations from the ‘real world’ [23]. The registry’s range has grown over the years and many of the current agenda includes performance and accountability issues, often in partnerships with national intensive care societies or similar general-purpose professional bodies to improve the quality of intensive care [16].

**Results**

**A Data-Driven Decision Support for Critical Care**

Medical research’s landmark studies have generally focused on investigating the effect of a single drug, treatment, or diagnostic technique. And while important, at this point, work mutes the complexity and interconnectedness that defines the medicine’s modern-
day nature. This mode of research does not catch issues with processes such as essential interactions in a complex physiological environment with concomitantly applied therapies. A high degree of variation in practice is therefore unavoidable because physicians are often left with contradictory and imperfect medical knowledge derived from a patient population or clinical environment that may not represent their own [24].

Perhaps tempering enthusiasm for large-scale data archiving systems are the concerns of cost and efficiency. A recent study [25] indicates that the cost of electronic health records implementation is high, while the benefits remain uncertain. This conclusion is understandable and highlights the fact that EHRs are clearly a critical, although costly, first step in the process of reorganizing health care into a closed-loop system that can reliably and contiguously produce and integrate feedback to enhance and increase efficiency. To construct a data substrate, it is necessary to create means for recording and archiving individual clinical experiences. Such a data substrate, if available freely, will provide a means for clinicians and data scientists to resolve knowledge gaps and mistakes, and promote a form of crowdsourcing for clinical practice evidence-building [26].

The intensive care unit (ICU) presents an especially compelling case for clinical data analysis. The value of many treatments and interventions in the ICU is unproven, and high-quality data supporting or discouraging specific practices are embarrassingly sparse [27]. Guidelines developed to standardize practice are dependent on an evidence base that is surprisingly thin, considering the copious data generated in the ICU. A knowledge gap of this magnitude is unacceptable for a medical discipline comprising 1% of U.S. gross domestic product, and for which ongoing demand is rising sharply [27]. In a systematic review of multicenter randomized controlled trials evaluating the effect of ICU interventions [28], only one in seven studies showed benefit; the rest either had no measurable value or were found to be harmful. The supposed explanations behind this perplexing finding are that the results of ICU procedures are subject to this environment’s exceptional variability and are especially vulnerable to variance across patient subsets and medical contexts. In addition, human physiological variations are not nearly as troublesome as the variations imposed, some unexplained and even irrational [29]. Some of this practice variability is due to lack of adherence to best practices, but the vast majority occurs simply because no evidence has been established for the issue in question [30]. The traditional approach to evidence creation, therefore, needs to change and take advantage of the technical feasibility of creating complete, highly detailed critical care databases. These databases could motivate clinical investigations, support the development of clinical decision support tools, and permit the testing and perfecting of algorithms with the use of real-world data. The oncoming clinical use of “big” data sets, such as genomics and proteomics, will clearly require data management at this level [27].

Intensive Care Registries in Quality Improvement

The focus on quality of intensive care is increasing because of the high cost of healthcare and the potential for harm. Poor-quality care is costly, and quality improvement initiatives in the ICU have been shown to improve outcomes as well as decrease costs [31]. The development of an electronic database for data collection and reporting is one of the key tools that allow physicians and nurses to monitor change in a quality improvement effort [32].

Electronic formats for data collection in ICU have been continuously appraised and found much feasible. ICU outcomes have been the subject of controversy and depend on a variety of factors including clinical presentations, heterogeneity of the population with respect to age, gender, underlying comorbidities, trajectory and severity of illness, time to presentation and infecting organisms. The type and number of affected organs adds further complexity and outcomes depend upon patient-centered care. Stringent implementation and continuous evaluation of patient-centered care using an automated data management system is crucial to the improvement in ICU outcomes [33].

Several electronic medical registries have already been developed at the national level, such as the United Kingdom Intensive Care National Audit and Research Centre Case Mix Program Database, the United States National Registry of Cardiopulmonary Resuscitation, and the Australian and New Zealand Intensive Care Society Adult and Pediatric Databases. Other similar databases mentioned in the international literature have also been developed as either a national registry for special patient populations (burn injury, rheumatology, or neonatal or pediatric intensive care, or for research and further analyses on specific fields such as ICU incident report or infection registries [34].

Other databases such as the Canadian “BC ICU Database” and the Brazilian Epimed Monitor System (Epimed Solutions, Rio de Janeiro, Brazil) accumulate routine data from a great number of ICUs for epidemiology and research purposes and/or aiming at quality improvement. The Brazilian Epimed ICU Monitor in particular, a commercial cloud-based registry for quality improvement, performance evaluation, and benchmarking purposes, manages clinical and epidemiological information and generates clinical reports from approximately 800 ICUs [34].

A few areas wherein participation in a registry may be particularly helpful are discussed briefly below and outlined in (Table 3) [16].
Table 3- Possible Roles for Intensive Care Registries in Quality Management [16]

<table>
<thead>
<tr>
<th>Role</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Collect data for comparative audit</td>
<td></td>
</tr>
<tr>
<td>2. Select data, provide definitions and data collection guidelines</td>
<td></td>
</tr>
<tr>
<td>3. Provide mechanisms for accurate and complete data capture</td>
<td></td>
</tr>
<tr>
<td>4. Maintain a customized risk adjustment model</td>
<td></td>
</tr>
<tr>
<td>5. Produce timely reports and provide easy access to data</td>
<td></td>
</tr>
<tr>
<td>6. Support advanced analysis of data</td>
<td></td>
</tr>
<tr>
<td>7. Organize audits and coach improvement programs locally</td>
<td></td>
</tr>
<tr>
<td>8. Arrange meetings for participants to discuss and develop intensive care quality</td>
<td></td>
</tr>
</tbody>
</table>

A normally applied framework to outline medical care quality uses 5 domains (www.qualitymeasures.ahrq.gov). These areas unit the classical 3 domains of the Donabedian model (structure, method and outcome) [35] and 2 further domains: access to medical care and patient expertise. Structure indicators represent organization, resources and equipment; method indicators are a unit concerning the method of care between caregiver and patient, what we tend to do or fail to try to do for patients and their families; and outcome indicators represent the results that we tend to bring home the bacon at the patient level. Access to medical care is that the ability to supply timely and applicable care. Patient expertise within the context of medical care might not solely embrace patients’ experiences however additionally embrace family members’ observations. The structure, process, access and knowledge domains should be connected to a clinically relevant set of outcomes [16].

**Discussion**

**IC Registry: Objective, Inputs and Outputs**

**Objective**
- To create a high-quality registry of mechanically ventilated patients through a citywide collaboration of academic health centers performing uniform data collection
  - Minimum data set – all patients
  - Modular data sets
  - Centrally-trained/quality-assured study assistants
  - Central data management/security

**Inputs**
- Daily collection of minimum data sets in ICU
  - Demographics, ICU admission diagnosis, the severity of illness
    - Modular data sets
      - Investigator-initiated
      - Hypothesis-driven
      - Time-limited

**Outputs**
- Epidemiology
  - MV, ARDS, specific therapies/interventions, outcomes
- Prevention
  - Risk factors for progression/development of ARDS
  - Linkage to population-level administrative data
- Biomarker/Genetic Analyses
  - Link the phenotypical information collected in a subset of patients to biological specimens (BAL, blood)
- Pilot Data
  - Concept/hypothesis proof for future grants/clinical trials
- Quality Improvement
  - Adherence/compliance with "best practices" to identify local or system-wide QI goals

**Conclusion**

In this era of “Big Data”, researchers realize the potential of Real World Data stored in clinical registries such as Intensive Care. The combination of patient clinical care data, quality parameters, and ICU operating costs, integrated into an electronic database, provides a valuable tool for quality improvement and overall efficiency of offered care. IC Registries could
revolutionize Intensive Care medicine, leading to creating decision alternatives in the field of intensive care and reduce errors and improve quality and security in the mentioned area. With advances, there is a need to ensure that the potential behind clinical registries is unlocked and better registries and better data capture tools are developed.

References


