Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 198 (2022) 335–340

Procedia Computer Science

www.elsevier.com/locate/procedia

The 2nd International Workshop on Healthcare Open Data, Intelligence and Interoperability (HODII)
November 1-4, 2021, Leuven, Belgium

Business Analytics Components for Public Health Institution - Clinical Decision Area

João Abreu[a], Tiago Guimarães [a,*], António Abelha [a], Manuel Filipe Santos[a]

Algoritimi Research Center, University of Minho, Braga 4470, Portugal

Abstract

The practice of evidence-based medicine has been gaining ground in the most diverse health institutions and consequently in the work of their health professionals. Thus, the Clinical Decision Support Systems appear, which aim to ensure that health professionals are supported with the best evidence in order to improve the clinical decision-making process, reducing the occurrence of adverse events through the availability and sharing of quality information by health professionals. Business Intelligence (BI) and Business Analytics (BA) technologies are increasingly being used to extract knowledge and turn it into useful information and analytical tools for healthcare professionals. This article is part of a dissertation that has as main objective the extraction of knowledge from the hospital data of the obstetrics service in *Centro Materno-Infantil do Norte* (CMIN) from *Centro Hospitalar Universitário do Porto (*CHUP) with focus on the area of clinical decision.

Keywords: Clinical Decision Support Systems, Business Intelligence, Business Analytics, Data Warehouse, KPI

1. Introduction

Information Systems (IS) in the health field are proving to be increasingly important in the management and manipulation of an enormous volume of data from different health institutions and the clinical data of patients, in order to support health professionals [1].

* Corresponding author.
E-mail address: tsg@dsi.uminho.pt

10.1016/j.procs.2021.12.250

Thus, Clinical Decision Support Systems (SADC) appear, which aim to ensure that health professionals are supported with the best evidence to the extent of improving the clinical decision-making process, reducing the occurrence of adverse events through the availability and sharing of quality information by health professionals [2]. That said, the technologies of Business Intelligence (BI) and Business Analytics (BA), are increasingly being used regarding the extraction of knowledge and transforming it into useful information and analytical instruments for health professionals.

The ETL process is the slowest and most difficult phase of a BI process due to the large amount of data to be processed, and because it directly influences the quality and usability of the Data Warehouse (DW) that will be originated to assist in decision making. The ETL process is responsible for extracting, transforming, cleaning, and loading data. In the first stage, data is extracted from different sources of information. In the second stage there is the transformation of the data, where the cleaning, transformation and integration of the collected data occurs in order to obtain accurate, correct, complete and consistent data. This step defines the granularity of dimensions, factual tables and the DW scheme. Finally, the data loading stage, where the extracted and transformed data are loaded into the intended multidimensional structure [3]. The DW is a component of a BI system that stores and consolidates data from different sources of information and formats, in a valid and consistent format, which allows the analysis and exploration of that same data using other tools. The data stored in the DW is consistent, integrated, historical and available to be analyzed by BI tools to extract information to be applied in the decision-making process [4].

The implementation of BI in health helps in improving the quality and safety of the health services provided, in improving the financial results of health institutions, in an efficient use of resources and in the implementation of practices based on the best evidence, through process improvement. decision-making [5]. BI is composed of Information Technologies (IT) and BA. IT is responsible for consolidating and managing data, assisting in decision making through the creation of DW, providing reports and Dashboards, which serve to monitor the performance of the organization through efficient reports that address performance indicators previously defined by the organization. On the other hand, BA focuses on the detailed analysis of data and business knowledge to make decisions based on data with the purpose of supporting decision making. In BA, reports, exploratory and complex analysis of data are prepared using statistical and mathematical tools and algorithms, which are adequate in the definition and review of the organization's indicators [6].

In the case of this project, it will serve to guarantee the usefulness of a huge set of data stored in the CHUP databases, where it was possible to notice that at the moment some indicators are obtained through impractical and time-consuming techniques and tools, which result in files of Excel hardly perceived by health professionals, making it difficult to extract knowledge from them. Thus, we seek to facilitate this process in the interest of helping in the clinical decision-making process in the CMIN obstetrics service, contributing to improving the efficiency of the services provided. One of the most relevant obstetric indicators is the cesarean section rate, which in Portugal is high, with situations in which they occur unnecessarily, according to the reasons of the Directorate-General for Health (DGS). OMS recommend that this rate should be between 10 and 15%. A caesarean section has higher costs for health institutions and can lead to greater clinical complications for both mother and the baby when compared to natural birth [7]. Thus, it is important to know the number of cesarean section and why the cesarean section occurred, with a view to reducing those that occur unnecessarily, consequently reducing this rate, through better information available to health professionals so that they can make better clinical decisions.

This project has as main objective the extraction of knowledge from obstetric data with a focus on supporting clinical decisions, through BI and BA techniques: analyzing, extracting, transforming, and loading obstetric data in order to create the necessary Dashboards, promoting indicators (KPI) of quality and efficiency in the area of clinical decision.

2. Kimball Methodology

For the realization of this project, the Kimball methodology was followed, because it fits the objectives of this project, such as the development and implementation of ETL processes, the development and implementation of a DW, the development of dashboards and the definition of indicators (KPI). This methodology, also called Business Dimensional Lifecycle, argues that a successful implementation of DW depends on the proper integration of the organization's tasks and components and will allow guidance throughout the development process of this project [8].

Initially, a search was made of all the indicators needed for the CHUP obstetrics service, where it was possible to notice that at that moment some indicators are obtained through impractical and time-consuming techniques and tools, which result in Excel files that are barely noticeable to health professionals making it difficult to extract knowledge from them, namely:

- Total number of deliveries;
- Number of deliveries broken down by type, cesarean section (surgical procedure in which the baby is removed through the mother's abdomen), elective cesarean section (previously programmed), eutocic delivery (normal delivery, in which there are no changes and occurs spontaneously ending with the exit of the baby through the vagina), aided by forceps (an instrument used when the mother's natural contraction is not sufficient for the passage of the fetus in a cephalic position), aided by a suction cup (an instrument that uses the vacuum in order to pull the fetus, in cephalic position, over the head) and assisted pelvis;
- Total number of twin births;
- Number of twin deliveries broken down by type (cesarean section, elective cesarean section, eutocic delivery, forceps, suction cup and assisted pelvis);
- Number of vaginal deliveries with epidural analgesia;
- Number of vaginal deliveries after cesarean section;
- Number of deliveries by cesarean section with epidural anesthesia;
- Number of vaginal births in which labor was induced;
- Number of episiotomies in vaginal deliveries (incision made in the perineum to enlarge the birth canal);
- Number of perineal lacerations per degree (1st degree laceration, 2nd degree laceration, 3rd degree laceration, 4th degree laceration and cervical laceration);
- Apgar score of the newborn (assessment of heart rate, breathing, muscle tone, reflex irritability and skin color) in the first and fifth minutes after birth;
- Number of cesarean sections performed according to their urgency, absence or stage of labor, main reason and main characteristics of the pregnancy;
- Number of cases of severe maternal morbidity during pregnancy and up to 6 weeks postpartum.

Therefore, it is intended to make better use of this data, so that is possible to generate knowledge and facilitate the entire clinical decision-making process through an easy and intuitive consultation of them. Besides this, some more performance indicators have been defined, which are mentioned by the DGS to enrich the study and draw conclusions regarding the quality of CHUP's midwifery service, namely:

- Percentage of deliveries by cesarean section;
- Rate of vaginal deliveries after caesarean delivery.

2.1. Architecture and Tools

Figure 1 shows the architecture divided into two environments, the backend and the frontend. Initially, the data collection from the CHUP obstetrics service takes place, followed by the extraction, transformation and loading (ETL) of that same data for the fact and dimension tables created in the DW. Finally, the treated and useful information is made available intuitively to the end user through the visualization tool.

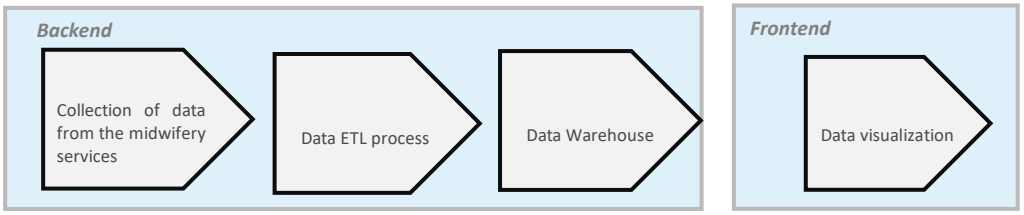

Fig. 1. Architecture

Oracle SQL Developer was the tool used for the ETL process, DW development and creation of the dimensional model. This tool allows the extraction of information from large databases and the implementation of ETL processes through PL/SQL commands, in addition to compliance with Oracle databases.

Power Bi Desktop was the tool used to visualize data, to present indicators and results. It was chosen for its ease of handling and versatility in the creation of Dashboards, reports and indicators, being a tool with increasing relevance in the BI, BA and data visualization market. Additionally, it allows a quick visualization of the data facilitating the decision-making process. Gartner recognizes the tool as a Leader for the 13th consecutive year at Gartner's Magic Quadrant 2020 for BI Platforms and Data Analysis [9].

2.2. Dimensional Modeling

The data from the CHUP obstetrics service were made available in a set of datasets in Excel format and later loaded into the Oracle database made available by the advisors through SQL Developer. The available data contained information regarding the births of the year 2016, such as the number of episodes, the semester, the date of birth, the month, the year, and the value of the newborn's Apgar Index (heart rate assessment, breathing, muscle tone, reflex irritability, and skin color) in the first and fifth minutes after birth and the type of delivery. Regarding cesarean section, it contained information about the reasons, characteristics, doses of epidural, stage of labor in which it occurred and its urgency. Regarding vaginal births, it contains information on whether labor induction, laceration, episiotomy, epidural doses, and whether the mother had a c-section terminated birth in the past. Finally, it contained information on diagnoses of severe maternal morbidity during pregnancy and up to 6 weeks after birth.

Based on the available data and the indicators previously defined, the dimensional model was defined, consisting of 3 fact tables and 9 dimension tables.

It is through the time dimension that all the de fact tables of this model are linked, and for this reason it is a schema in a constellation of facts because the de fact tables share the link to a dimension table.

2.3. ETL process

Through the exploration of the data, it was possible to identify some incongruities that needed correction right from the data sources, namely null values, repeated values, orographically errors and special characters. After treatment, the data was then loaded into the dimension and fact tables as described in the dimensional model.

Table 1. ETL Process.

Incongruities	Correction
Null values.	**Eliminate null value or assign an irrelevant value (such as 0 and unknown when appropriate).**
Orographically errors.	**Correction of errors.**
Special characters.	**Removing special characters.**
Same episode registered as eutocic and suction cup/ forceps/assisted pelvis.	**The eutocic delivery was assumed to be wrong, having eliminated that record because in principle it would be an outdated record since they did not achieve normal delivery.**
Same episode with more than an APGAR value.	**The highest APGAR value was assumed to be correct.**
Episiotomy recorded in caesarean deliveries.	**Replace the value, as an episiotomy only occurs in vaginal deliveries.**
Episodes wrongly registered as eutocic delivery because they have values in the columns referring to deliveries by caesarean deliveries.	**Replace the registry for Cesarean.**

2.4. Data Visualization

In Power BI Desktop, a development and data visualization environment were created in order to obtain DW data and generate reports and Dashboards with indicators in an intuitive way for health professionals in order to contribute to the improvement of the process decision-making by CHUP health professionals. Concerning this, the connection to the Oracle database was made through Power BI Desktop. The Power BI tool allows direct access to the information present in the DW. Furthermore, it allows the creation of hierarchies, for example in "DIM_TEMPO", a hierarchy was created from the date of birth, that is, year, quarter, month and day, which allows searches through a drill down to the day. Therefore, it was not necessary to create an OLAP cube, as a result of taking advantage of these Power Bi functionalities.

For the purpose of complementing the study, some measures were created that allow a better analysis of the data and allow to satisfy the indicators previously defined, such as the cesarean section rate. This was obtained through the features of Power BI and the DAX language (Data Analysis Expressions) incorporated in the tool.

Fig. 2. Example of a Dashboard

As an example, in this Dashboard related to cesarean sections, it allows the visualization of the indicators "Number of cesarean sections performed according to the absence or stage of labor" by the circular graph and "Percentage of deliveries by cesarean". The rate of cesarean sections is the main indicator of the quality of an obstetrics service, so it is possible to observe that it is well above the reference value according to WHO, which states that the cesarean rate must be between 10 and 15%, and for that reason the value of 26.90% can be seen as a bad indicator that needs attention. Cesarean section is a surgery that causes complications for both mother and the baby, so it is vital to reduce this rate with a view to diminish these complications and improve the quality of health care provided by the CHUP obstetrics service. It is also possible to search for a specific episode, observe the variation in the number of deliveries by cesarean section over the months of the year 2016, select between cesarean section or lethal cesarean section and the stage of labor. Therefore, health professionals through the other indicators developed and presented referring to cesarean sections as the reason, characteristics, urgencies, among others, will be able to draw conclusions and make better decisions in order to lower this rate.

3. Results and Conclusion

This project resulted in a prototype that allows knowledge extraction from existing hospital data, facilitating the acquisition of indicators and useful information to CHUP's health professionals, assisting them in the decision-making process. Previously, these data were obtained manually and in a long duration, using techniques and tools that were impractical and time-consuming, resulting in Excel reports that described these same data, which were not easily perceived by health professionals, making it difficult to extract knowledge from them. Now, as a result of the automation and improvement of the process of obtaining indicators and enrichments, these are provided to CHUP health professionals in an easy and intuitive way through the Dashboards developed in Power BI, which present the relevant indicators for analysis, helping themselves in the decision-making process.

Based on the available data, its relevance and the needs met, a set of indicators were defined and created that will be important in the decision-making process of CHUP health professionals. ETL processes were developed and implemented, resulting in useful and perceptible information for CHUP health professionals. A DW was developed and implemented and a dimensional scheme was created where the DW was allocated, with its respective factual tables and dimension tables. Dashboards and reports were developed that present the indicators created in an easy and intuitive way to CHUP health professionals, facilitating their decision-making process.

All the work carried out culminated with the fulfillment of its main objective, the extraction of knowledge for the sake of supporting the clinical decision of the CHUP obstetrics service. In short, a better use of the data stored in the CHUP databases was achieved and the result generated in this project met the defined objectives, ensuring a quality of information from the data sources, which allows health professionals to obtain scientific evidence through the indicators observed and manipulated, helping them in their clinical decision-making process.

4. Future work

Soon, it would be interesting to analyze more years beyond 2016, as a means to obtain more analysis and conclusions with the created indictors, and to try to perceive variations over the years. In addition, indicators related to the mother and the newborn can be interesting to enrich this study. Finally, the creation of a tool that allows to predict the evolution of the cesarean section rate according to the indicators already created, would reveal itself of enormous utility.

Acknowledgements

This work has been supported by FCT – Fundação para a Ciência e Tecnologia within the R&D Units Project Scope: UIDB/00319/2020.

References

[1] de Vasconcelos, José Braga, Álvaro Rocha, and Rui Gomes. (2004) "Sistemas de Informação de Apoio à Decisão Clínica Estudo de um caso de uma Instituição de Saúde.".

[2] de Vasconcelos, José Braga, Álvaro Rocha Henriques, and A. Rocha. (2006) "Modelo para o desenvolvimento de Sistemas de Apoio à Decisão Clínica para a prática da Medicina Baseada na Evidência." Anais do X Congresso Brasileiro de Informática em Saúde (CBIS 2006).

[3] El-Sappagh, Shaker H. Ali, Abdeltawab M. Ahmed Hendawi, and Ali Hamed El Bastawissy. (2011) "A proposed model for data warehouse ETL processes." Journal of King Saud University-Computer and Information Sciences **23(2):** 91-104.

[4] Silva, Eva, et al. 2015 "Business intelligence and nosocomial infection decision making." Integration of Data Mining in Business Intelligence Systems. IGI Global,193-215.

[5] Foshay, Neil, and Craig Kuziemsky. (2014) "Towards an implementation framework for business intelligence in healthcare." International Journal of Information Management **34(1)**: 20-27.

[6] Bartlett, R. (2013). "A PRACTITIONER'S GUIDE TO BUSINESS ANALYTICS: Using Data Analysis Tools to Improve Your Organization's Decision Making and Strategy." *McGraw Hill Professional*.

[7] Programme, World Health Organization Human Reproduction. (2015) "WHO statement on caesarean section rates.": 149-150.

[8] Kimball, Ralph, and Margy Ross. (2013) "The data warehouse toolkit: The definitive guide to dimensional modeling." *John Wiley & Sons*.

[9] Microsoft. (2020). "What is power bi?".