

SAD_BaSe: A Blood Bank Data Analysis Software

Augusto Ramoa, Salomé Maia, and Anália Lourenço

Abstract. The main goal of this project was to build a Web-based information system – SAD_BaSe – that monitors blood donations and the blood production chain in a user-friendly way. In particular, the system keeps track of several data indicators and supports their analysis, enabling the definition of collection and production strategies and, the measurement of quality indicators required by the Quality Management System of blood establishments. Data mining supports the analysis of donor eligibility criteria.

Keywords: Decision support system, blood donations, process monitoring, quality management.

1 Introduction

The first assignment of blood establishments is to educate society for the need to donate blood. Therefore, they set up marketing campaigns so people constantly hear about it and schedule regular collection sessions not only in their facilities, but also through mobile collection sessions around their geographic area.

In the collection of blood from the donors, several safety measures must be implemented to protect the donor and the receiver. On the donor's side, a set of requirements has been established to ensure a safe procedure, accounting for the considerable amount of blood that is to be drawn (approximately 0.45 liters).

Augusto Ramoa · Salomé Maia
Instituto Português do Sangue,
Centro Regional de Sangue do Porto,
Rua de Bolama, 133, 4200-139 Porto, Portugal
e-mail: {augustoramoa,maia.salome.m}@gmail.com

Anália Lourenço
Institute for Biotechnology and Bioengineering,
Centre of Biological Engineering,
Campus de Gualtar, University of Minho, 4710-057 Braga, Portugal
e-mail: analia@deb.uminho.pt

On the recipient's side, the donor's behavior regarding drug use and sexual activity is assessed during a clinical interview to minimize the risk of infecting the patient. In parallel, laboratory tests are performed to determine the blood group and to check for a number of blood-borne diseases.

Most blood units are collected from donors as Whole Blood units (WB) and then must be divided in the laboratory. There are different therapeutic products such as Red Cell Concentrates (RCC), Fresh Frozen Plasma (FFP) and Pooled Platelet Concentrates (PPC) [2]. The blood components remain in quarantine until all lab tests are validated and the components are considered safe for use. Then, the blood establishment distributes blood components balancing hospital requests and stock thresholds. Due to the small shelf life of some blood components (e.g. only 5 days for platelets) a good planning is required to avoid blood wastage and, more importantly, blood shortage.

This work proposes a blood bank data analysis software that provides a novel contribution to the monitoring of blood collection and component production and distribution. The remainder of the paper describes the work as follows. In Section 2, current software systems in use are addressed. In Section 3, SAD_BaSe is presented and results presented. Some conclusions are provided in section 4.

2 Blood Bank Software Systems

From the moment of collection to the administration of the blood to a patient, a complex process takes place, mostly based on computer-aided procedures, to ensure the safety of the transfusion and the traceability of all steps. Indeed, the Council of Europe's Guide to the preparation, use and quality assurance of blood components, one of the most important guides of a blood bank, dedicates a sub-chapter ("Data processing systems") to the design, implementation, testing and maintenance of adequate software.

Currently, blood bank software systems are almost exclusively focused in assuring safety (Does this unit meet the necessary requirements to be issued?) and traceability (To whom the unit was collected? Who collected it? To which hospital was it issued?), however process monitoring is also very important.

The variability associated with the source product (e.g. weight and the values of hemoglobin and platelet), collection (e.g. you cannot have a good collection if the donor does not have a good vein or if the nursing staff is not well trained to proceed with the collection) and production (e.g. the considerable human manipulation inherent to the whole process) leads inevitably to wastes that need to be minimized. A unit of blood cannot be reprocessed if something goes wrong. So, by learning where the process can fail we have the opportunity to improve the process and minimize blood losses.

3 SAD_BaSe Blood Bank Data Analysis Software

According to a survey on the Portuguese Blood Establishments, one of the major user complaints to existing blood bank software systems concerns report capabilities [3]. To assist blood banks in process monitoring, we have developed a complementary information system, named SAD_BaSe. This software suite includes

the following modules: a data processing module that gathers and manipulates data generated throughout the collection, processing and distribution stages; a report module that enables query customization (e.g. detailed reports on the different processes for time periods, hospitals, collection sessions or blood components of interest); and a data mining module that enables advanced data analysis, namely to monitor and control donor rejection.

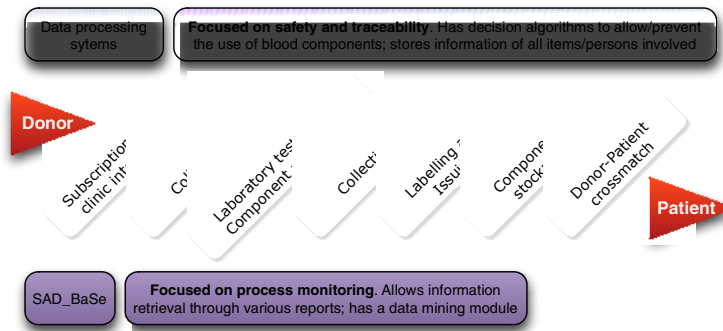


Fig. 1 Blood chain from donor to patient and the differences between existing data processing systems and SAD_BaSe

3.1 Data Modeling and Database Population

The information flow from the blood establishment system to SAD_BaSe database is fully implemented (Fig. 2). Currently, the database covers for information on the collection sessions, donors, lab analysis results and hospitals.

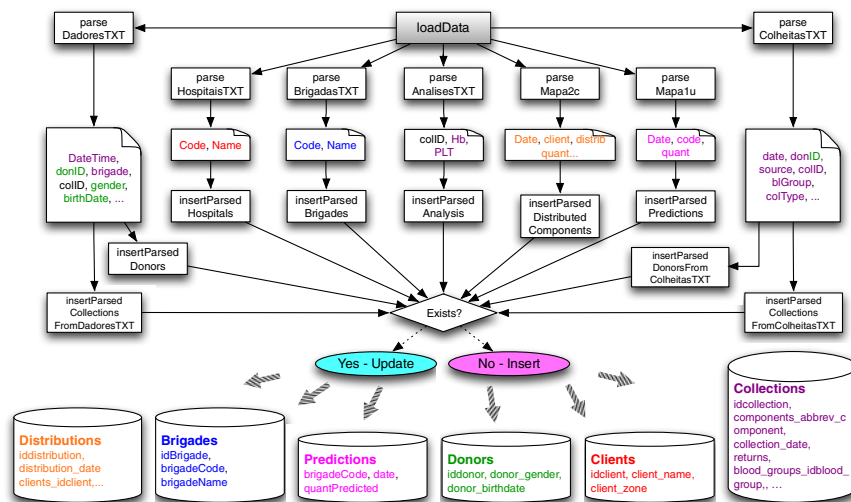


Fig. 2 SAD_BaSe information flow.

SAD_BaSe was implemented and tested in a Portuguese blood establishment. A snapshot was retrieved from its information system encompassing data from January 2009 to December 2011.

For validation purposes, the data related to 2009 was compared with the reports obtained from the blood establishment software. Some data incoherencies were detected. However, they were related to the way data is being used in the two systems. For example, the blood establishment software reports the number of units discarded in a specific time interval, whereas SAD_BaSe reports the number of units discarded that were collected in that time interval.

3.2 Reporting

The tool supports on-demand reporting and querying facilities to enable searches by component, time series, discard cause and collection site.

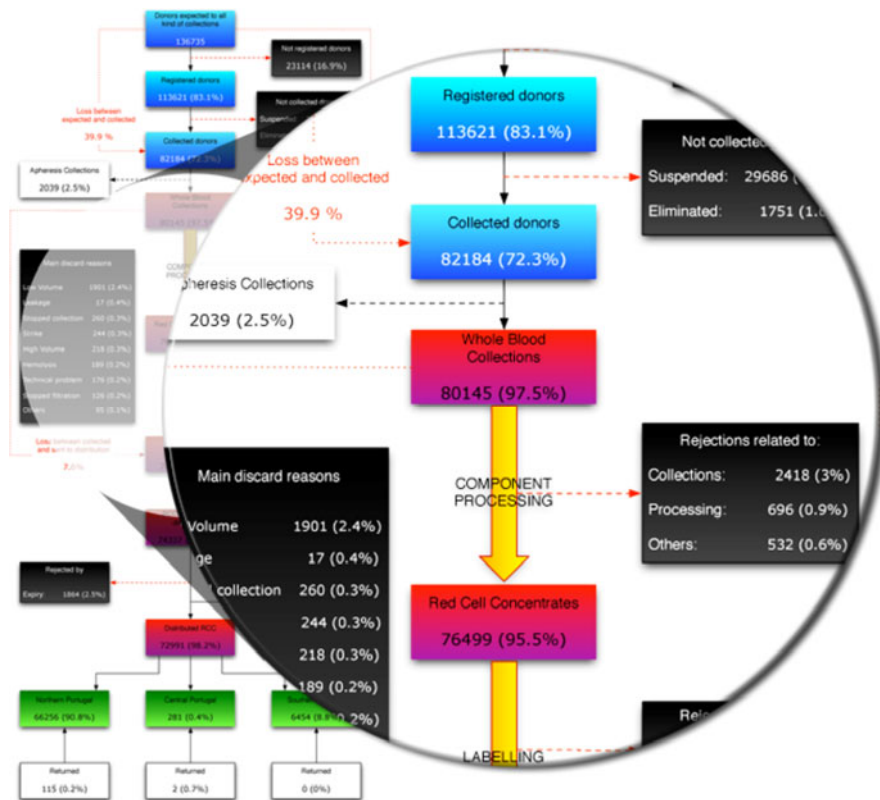


Fig. 3 SAD_BaSe Flowchart

Reports are primarily intended to provide Board members and the Head of departments a summary of the blood establishment activity. The flowchart generated (an execution example is provided in Fig. 3) highlights a set of quality indicators to be monitored, such as: predicted number of donors for all collection sessions; losses by absent donors, suspended and deferred donors, discards related to collection, component processing, serological markers, immuno-hematology problems and expired components; components available for distribution and components received from other blood establishments; and components issued to the hospitals by geographic region.

3.3 *Data Mining*

Analysis was set on the examination of potential donors and the alert for eventual donation discard. It must be mentioned beforehand that we are aiming to analyze a process where it is expected that most donors are considered suitable to undergo collection. However, young, first time donors are often suspected to be less predisposed to go through the donation without any problems and, large collection sessions (e.g. in universities) are suspected to have more donors suspended. Also, first time donors unaware of donation requirements are presumed to be more ineligible for donation. Despite this “common knowledge”, neither the blood establishment had a study to corroborate these suspicions nor other databases stored data on such features.

So, a dataset was prepared containing all donor interviews in the first semester of 2011 (56666 records). The description of the selected attributes is presented below:

- Age of the donor (age) – arranges donations in three groups, donors with less than 24 years old, donors aged between 24 and 60, and donors above 60;
- Gender (gender) – generally speaking, women have less body blood volume than men, but the volume drawn is the same in both cases;
- Previous donations (pdon) – two groups were created according to the number of previous donations, first time and recurrent;
- Collection site (brigade) – the collection sites where the donations take place are quite different. The collection site that is operated in the blood establishment facility is the only one that runs daily. Others run weekly and many run less frequently. This attribute considers three categories: daily collections (PP), weekly collections (regular) and the remaining collections (others);
- Predicted collections (predCol)– before each collection session an estimate of the number of donors is necessary for staff and materials setup. This attribute divides the dataset by the size of the collection session in small sessions (less than 50 donors), medium sessions (between 51 and 100 donors), large sessions (101 to 199 donors) and larger sessions (200 or more donors);
- Conclusion – if the donor is somewhere along the blood chain suspended or permanently deferred (eliminated) it is labeled as S+E, if the donor is considered apt, it is labeled as A.

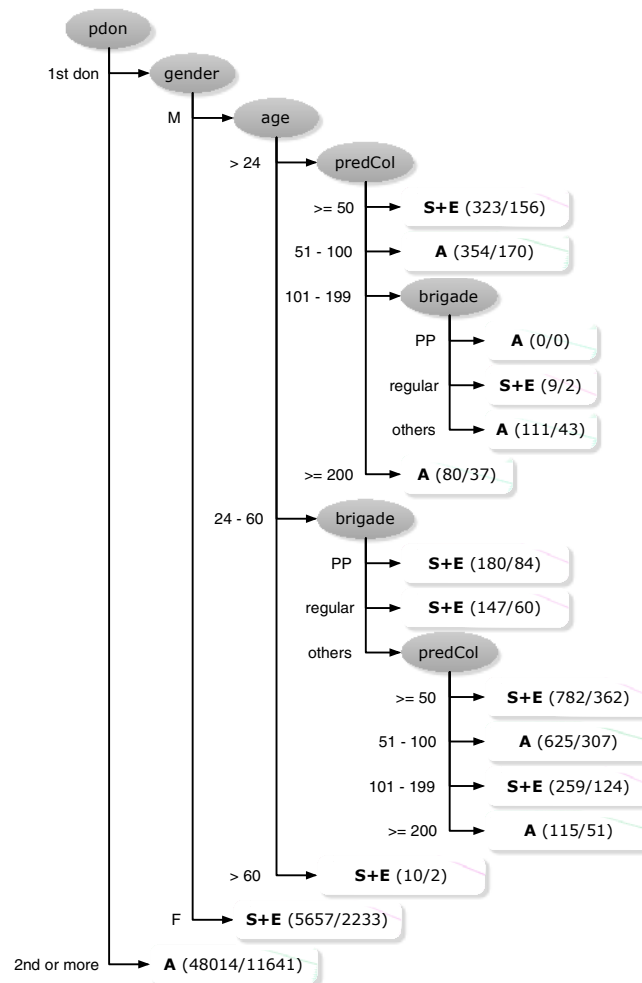


Fig. 4 Decision tree obtained using Weka j48

To perform data mining we have used the software workbench, Waikato Environment for Knowledge Analysis (WEKA 3.6.0.) [4]. The dataset was classified using the decision tree algorithm C4.5 [5] through its Java implementation, j48.

The obtained results point to 72.89% correctly classified instances and the obtained tree illustrates the importance of the previous donations (Fig. 4) which is mainly explained by the fact that returning donors already know that some medical conditions and behaviors are incompatible with blood donations. The gender is also a very important attribute. When looking at the data stored in SAD_BaSe of first time donors, low hemoglobin is the main causes of donor deferral in the female population with 626 deferrals whereas in the males only 51 donors were

deferred by the same cause. The main cause of men deferrals was the existence of a new sexual partner in the last six months accounting for 204 deferrals, whereas in the females this accounted for a similar number of deferrals, 177.

4 Conclusions

Blood transfusions are nowadays a routine procedure in hospitals. Blood establishments are equipped with suitable operational information systems but monitoring procedures is still a requirement.

SAD_BaSe complements blood establishment software in two ways: it monitors mandatory quality indicators and supports production planning, maximizing production and minimizing process costs/losses. Its monitoring and analysis abilities respond not only to current requirements, but also to upcoming ones, making new points of monitoring available, with less intervention of the users.

References

1. Council of Europe – Health Policy web page,
http://www.coe.int/t/dg3/health/themes_en.asp
2. Guide to the Preparation, Use and Quality Assurance of Blood Components, European Directorate for the Quality of Medicines & HealthCare - Council of Europe, Strasbourg (2011)
3. Leal, J.: Validação de Sistemas Informáticos de Serviços de Sangue e de Medicina Transfusional. Faculdade de Medicina da Universidade do Porto (2010)
4. Hall, M.: The WEKA Data Mining Software: An Update. SIGKDD Explorations 11(1) (2009)
5. Quinlan, R.: C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, San Mateo (1993)

