# A Rational Framework for Production Decision Making in Blood Establishments

**Augusto Ramoa[1], Salomé Maia[1] and Anália Lourenço[2*]**

[1]Instituto  Português do Sangue, Centro Regional de Sangue do Porto, Rua de Bolama, 133, 4200-139 Porto, Portugal

[2]Institute for Biotechnology and Bioengineering, Centre of Biological Engineering, Campus de Gualtar, University of Minho, 4710-057 Braga, Portugal

{augustoramoa, maia.salome.m}@gmail.com, analia@deb.uminho.pt

**Summary**

SAD_BaSe is a blood bank data analysis software, created to assist in the management of blood donations and the blood production chain in blood establishments. In particular, the system keeps track of several collection and production indicators, enables the definition of collection and production strategies, and the measurement of quality indicators required by the Quality Management System regulating the general operation of blood establishments.

This paper describes the general scenario of blood establishments and its main requirements in terms of data management and analysis. It presents the architecture of SAD_BaSe and identifies its main contributions. Specifically, it brings forward the generation of customized reports driven by decision making needs and the use of data mining techniques in the analysis of donor suspensions and donation discards.

## 1       Introduction

Although blood transfusions are nowadays a safe medical practice, adverse reactions can occur, and in extreme cases, they may compromise the life of the patients. For example: patients can get infected blood if the analytical results of a syphilis infected donor are not correctly interpreted; patients can suffer from kidney failure (or even die) due to the erroneous determination of blood group; patients can get a bacterial infection if blood collection, processing and administration are not undertaken in a clean environment and using sterile procedures.

Safety measures are thus implemented to protect the patient as well as the donor (given the considerable amount of blood collected, approximately 0.45 liters). In fact, surveillance and data collection procedures covering the whole transfusion chain are mandatory by law [1]. The medical condition and social behavior of the donors (e.g. drug use and sexual activity) are observed closely during the clinical interview, prior to donation, and laboratory tests are issued to determine the blood group and to check for a number of blood-borne diseases.

The Council of Europe's Guide to the preparation, use and quality assurance of blood components [2], one of the most important guides for European blood establishments – institutions responsible for the collection, testing, processing, storing, and distributing of blood components [3] – provides background information, i.e. the *"why and how"* of blood collection, preparation and use. Specifically, the sub-chapter "Data processing systems" is

---

[*] To whom correspondence should be addressed.

1

dedicated to the design, implementation, testing and maintenance of blood bank software. Also, a set of standards outlines the steps to be taken by the quality system of blood establishments, regarding donor selection, blood collection, and component preparation, storage and distribution, among others.

Generally, the activity of a blood establishment (Figure 1) consists in promoting blood donation, collecting blood, analyzing and validating the donated blood, transforming it into blood components, and finally distributing it to hospital blood banks – institutions that store, distribute and perform compatibility tests between the blood components and the patients to whom they are destined [3].
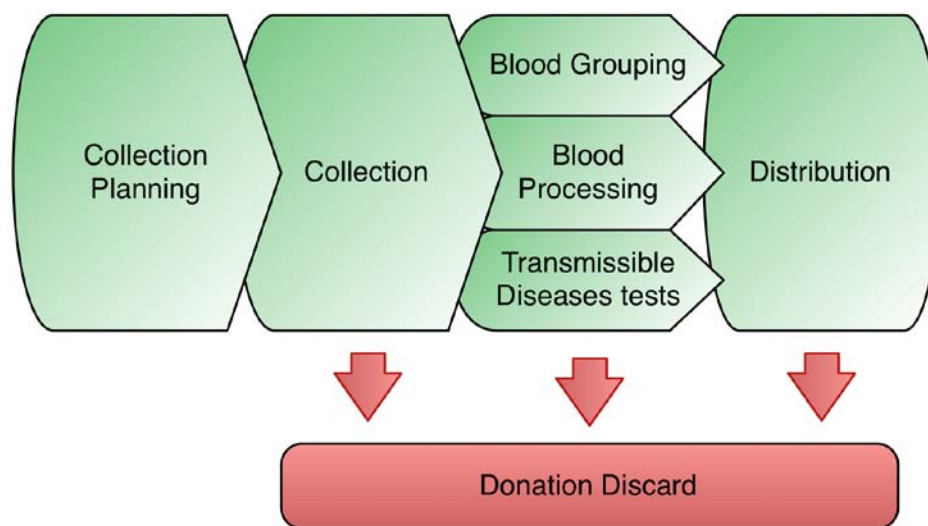


**Figure 1: Blood Establishment workflow**

The activities promoted at each step can be summarized as follows:

*Collection planning* – the first assignment of a blood establishment is to educate society about the need to donate blood. Marketing campaigns are issued on a routine basis so that people are constantly hearing about it and feel impelled to become donors;

*Collection* – to simplify the process, blood establishments have in-house facilities where regular collection sessions occur, but they also deploy many mobile collection sessions around their geographic area;

*Blood group* and *transmissible diseases tests* – a set of laboratory tests are performed to meet legal requirements [1], namely the determination the blood group and the screening for transmissible diseases;

*Blood processing* – 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]. These blood components remain in quarantine until all lab tests are validated and they are considered safe for use;

*Distribution* – the blood components are distributed to the hospitals according to their needs and the stock available in the establishment. Due to the short 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.

*Donation discard* – unfortunately, not all blood collected can be used. There is a considerable variability associated with the source product (e.g. donors have different weights, values of hemoglobin and number of platelets), collection (e.g. you cannot have a good collection if the

donor does not have a good vein or if the nursing staff in charge of collection is not well trained) and production (e.g. the human manipulation inherent to the whole process) leads to inevitable wastes that need to be minimized. So, establishments need the support of computational systems that monitor process losses, measuring a set of quality indicators, and help delineate corrective strategies.

This work presents a blood bank data analysis software, named SAD_BaSe, that provides a novel contribution to the monitoring of blood collection and component production and distribution. SAD_BaSe was tested in a Portuguese blood establishment. A data snapshot was retrieved from its information system, encompassing data from January 2009 to December 2011, and SAD_BaSe analytical outputs were compared with the reports obtained from the blood establishment software. Additionally, data was used to apply data mining techniques to the analysis of donor suspensions and donation discards.

The remainder of the paper describes the work as follows: section 2 introduces current blood establishment software systems and distinguishes between their functionalities and goals and those of SAD_BaSe; section 3 details the architecture of SAD_BaSe; section 4 shows results obtained for the case study, highlighting new analytical capabilities; last, there are some final remarks about the work, pin pointing future developments.

## 2     Blood Bank Software Systems

All blood-related institutions use a software system to manage blood collection, processing, distribution and transfusion, as well as to keep record of all tasks and their participants. According to a survey on the Portuguese blood establishments and hospital blood banks [4], in 2009, there were 9 different software systems in use in the Portuguese blood establishments and hospital blood banks: Asis (32); Sibas (28); Delphyn (13); e-Delphyn (3); ClinidataBST (3); Imaginasoft (3); Nova His (1); Pelicano (1) and Rocail (1) - in parenthesis, the number of institutions using the software is shown. To simplify, we will address these software systems as Blood Bank Software, since the same software systems are used in Blood Establishments and in Hospital Blood Banks.

Regarding the "responsibilities" of blood bank software, there are none. Portuguese law does not even enforce Blood Establishments and Hospital Blood Banks to have software systems. However, if existent, blood bank software validation is mandatory by law in Portugal [1]. Despite this legal obligation, how that validation should be performed is not yet legislated.

The information system of a blood establishment or a hospital blood bank is quite demanding in terms of information to be collected and its long-term maintenance. Specifically, any system must keep track of all donors, blood units collected and components prepared. Also, all blood components should be labeled accordingly. Unit traceability should be ensured within and outside the institution. And the most important of all requirements, a blood component that does not fulfill all legal requirements should not reach patients, i.e. should be discarded timely. For the above-mentioned motives, records are kept for up to 30 years.

Undoubtedly, these tasks are perfectly handled by a software suite and it is no surprise that 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. Indeed, one of the major user complaints to existing blood bank software systems concerns report and advanced analytical capabilities [4].

# 3    SAD_BaSe blood bank data analysis software

To complement blood bank software in process monitoring, we have developed a data analysis software named SAD_BaSe. This system aims to monitor 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. Also, the application of data mining techniques to SAD_BaSe data, specifically to the analysis of donor suspensions and donation discards, is expected to help the blood establishment minimize donor and component deferrals.

The software existing in blood establishments and SAD_BaSe do not overlap each other. SAD_BaSe focuses in process monitoring while Blood Bank Software is focused in assuring safety and traceability (Figure 2).
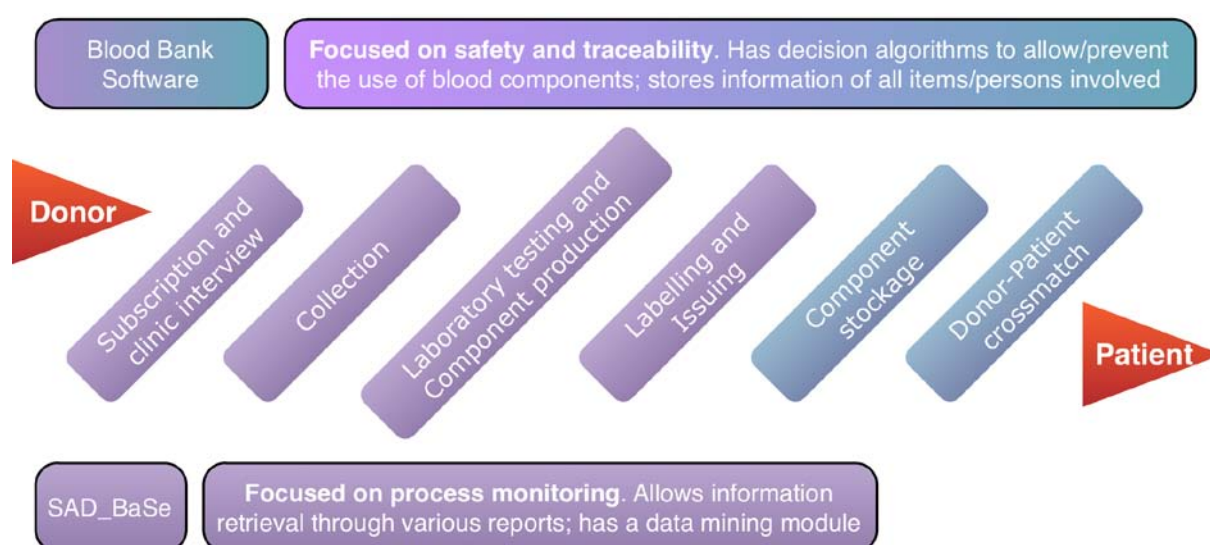


**Figure 2: Blood chain from donor to patient and the non-overlapping activities of existing data processing systems and SAD_BaSe**

To upload data from the Blood Establishment software to SAD_Base, parsers in Java® programming language where developed. A relational database running in a MySQL® server (version 5.1.32) was designed and a set of forms were created in PHP® and were made available in an Apache HTTP server (version 2.2.11). The system also supports reporting and querying facilities, enabling searches by component, time series, causes of rendered useless blood or components, among others. Finally, a PHP chart library – jpGraph® – was embedded to allow automatic chart creation.

The architecture of SAD_BaSe is modular: 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.

## 4      SAD_BaSe contents and data analysis abilities

### 4.1      Data modeling and database population

SAD_BaSe database covers for information on the collection sessions, donors, laboratory results and hospitals. The information flow from the blood establishment system to SAD_BaSe database is fully implemented (Figure 3):


- loadData – data in blood establishment software is usually exported in text files. Text files are uploaded individually, its contents are imported to a generic container and the first lines are analyzed to determine the type of the file and thus, call the most adequate parser. The parser will dump the processed data directly to the database;
- parseHospitalsTXT – parses information on hospital blood banks, such as hospital code and hospital name;
- parseCollectionSessionsTXT – parses information about mobile collection sessions, such as the mobile collection session code and name;
- parseDonorsTXT – parses the information related to the donor, namely date and time of collection, donor ID, mobile collection session, collection ID, gender, birth date, postal code, number of prior donations, maximum and minimum blood pressure, blood group, medical triage and lab classifications;
- parseCollectionsTXT – parses the file information related to the blood component, namely date, source (i.e. the unit was collected in this Blood Establishment or did it came from another one), donor ID, collection ID, blood component, collection type (if a Whole Blood donation, apheresis or autologous blood donation[1]), blood group, rejection code (if any), hospital (if any), returned unit (if any);
- parseAnalysisTXT – parses the information about analysis values, the donation date, donor ID, hemoglobin and platelets;
- parseMap2c – parses information about the date and amount (divided by blood group) of issued units to hospitals;
- parseMap1u – parses information on the mobile collection session ( date, mobile collection session code and the quantity of predicted donors).


### 4.2      Blood establishment activity monitoring

Reports are primarily intended to provide Board members and the Head of departments a summary of the blood establishment activity. The tool supports on-demand reporting and querying facilities to enable searches by component, time series, discard cause and collection site. The results can be visualized in tables, charts and a flowchart.

Figure 4 is an example of how data can be retrieved from SAD_BaSe straightforwardly. In this case, a bar chart represents the number of collections per day and the difference between workdays and weekends can be observed. This particular char aims to help define the staff's scheduling – "should more staff be working during the weekends to maximize collections?"

---

[1] An autologous blood donation is a donation to be used only by the person to whom the blood was collected, e.g. before a scheduled surgery blood is collected to the patient and if necessary it can be transfused during or after the surgery.
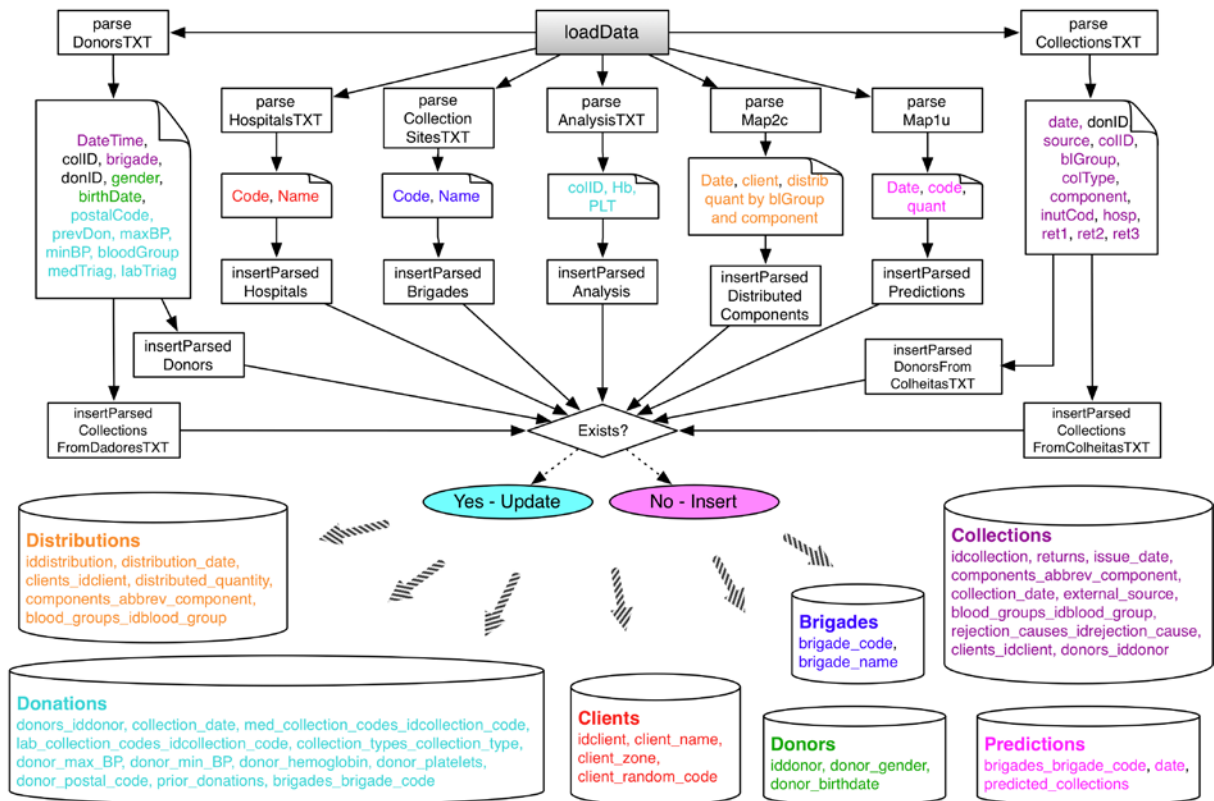
**Figure 3: SAD_BaSe information flow. Several files are exported from the blood bank software, the file content is parsed and the obtained information inserted/updated into SAD_BaSe.**
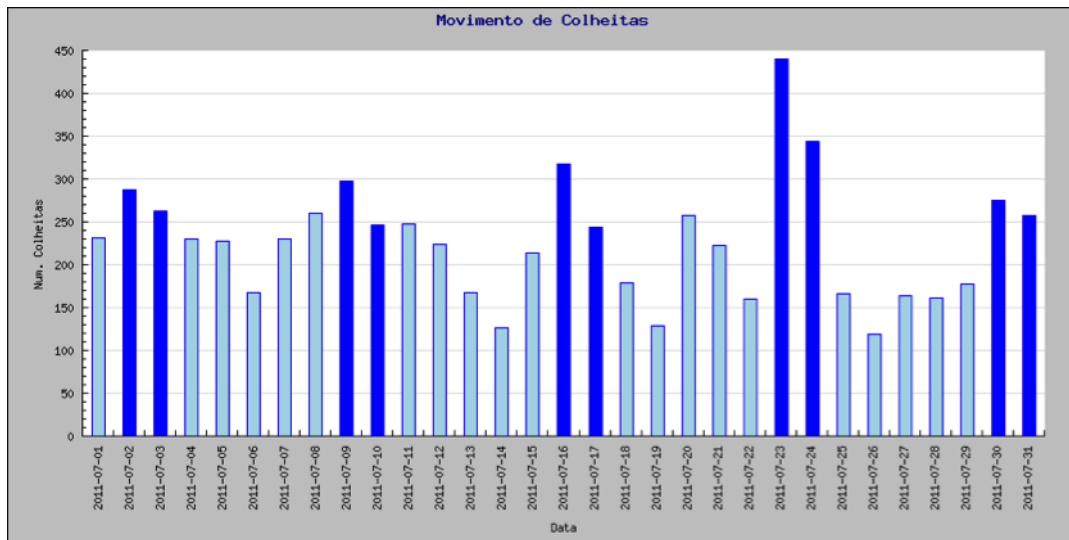


**Figure 4: The collections of July 2011 grouped by day. The blue bars represent collections on Saturdays and Sundays and cyan bars represent collections in workdays.**

To provide a global and more comprehensive view of activities, a flowchart is generated (Figure 5). This highlights a set of quality indicators, 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.
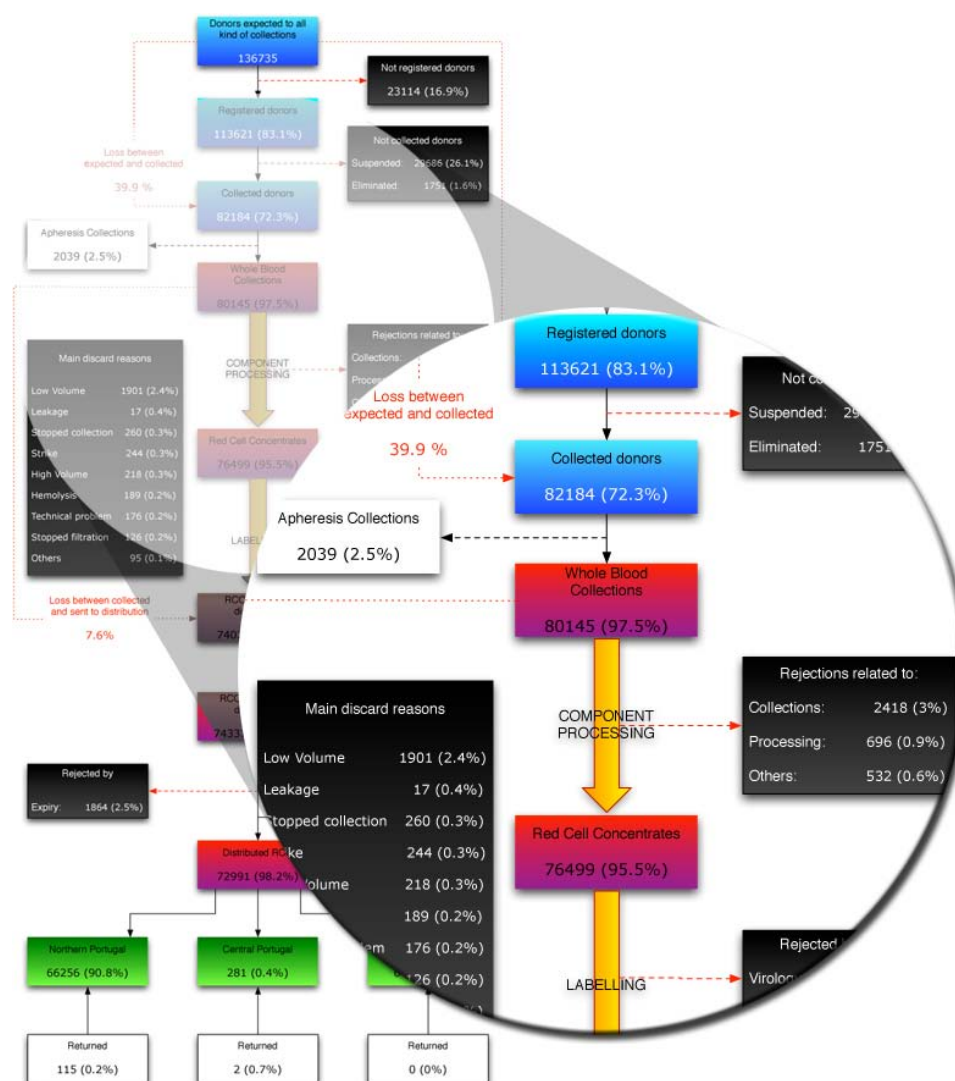
**Figure 5: SAD_BaSe Flowchart. The zoom is set on the losses after component processing. In this example, from a total of 80145 Whole Blood Units, 95.5% of were processed into Red Cell Concentrates 3% discarded by collection-related problems, 0.9% discarded by Processing issues and 0.6% by other causes**

## 4.3    Hospital consumptions

Usually, blood establishments are also interested in monitoring the demand of Red Cell Concentrates by blood groups. Knowing the normal distribution of the blood groups on the Portuguese population, hospital blood banks are expected to demand Red Cell concentrates in the same proportion. Unfortunately, not all hospital blood banks follow the best practices, preferring blood groups that are largely compatible. For example, while the O blood group can normally be transfused to any patient, the AB group can only be given to patients of that blood group. Some hospitals that do not demand AB Red Cell Concentrates, causing losses of due to expiration date, and creating shortage on O units, which represents a risk to patients of this group that can only receive O units.

As it is not the blood establishment responsibility to oversee these issues, the best it can do is present these data to the hospital and appeal to their consciousness. Figure 6 represents the

consumptions of one hospital in 2011, where it can be observed that only 2 B+ and no AB unit were sent (Figure 6).

| Componente | Grupo | Quant Enviados | % Enviados | % na População |
|---|---|---|---|---|
| CE | A + | 1010 | 38.4 | 40.0 |
| CE | A - | 176 | 6.7 | 8.0 |
| CE | O + | 1173 | 44.6 | 34.0 |
| CE | O - | 266 | 10.1 | 7.0 |
| CE | B + | 2 | 0.1 | 6.0 |
| CE | B - | 5 | 0.2 | 1.0 |
| CE | AB + | 0 | 0 | 3.0 |
| CE | AB - | 0 | 0 | 1.0 |
| Total: | | 2632 | 100% | 100% |

**Figure 6: Red Cell Concentrates issued to a Hospital in 2011 by blood group. The number of issued units is in column 3, the percentage of issued units is in column 4 and the percentage in the Portuguese population is in column 5.**

### 4.4    Mining unsuitable donors and donation discards

Data mining techniques have been applied to the analysis of donor suspensions and donation discards by insufficient blood quantity drawn. The population targeted for blood donation is generally in good health, thus, the aim is 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. universities or factories) are suspected to have more donors suspended. Also, first time donors unaware of donation requirements are presumed to be more ineligible for donation.

Despite such common knowledge and the application of some "rules of thumb", neither the blood establishment had a study to corroborate these suspicions nor other databases stored data on such features. So, a dataset containing all donor interviews in the first semester of 2011 (56666 records) was prepared to be further analyzed. The description of the selected attributes is presented below:

- Age of the donor (age) –donations were grouped into donors with less than 24 years old, donors aged between 24 and 60, and donors above 60;
- Gender (gender) – often, 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 donors;
- 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 distinguishes collection sessions 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.

The Waikato Environment for Knowledge Analysis (WEKA 3.6.0) [5] supported dataset mining [6]. Even though several classification techniques have been tested (rules: ZeroR and

PRISM; trees: Id3; Nearest-neighbour: IB1), the well-consolidated decision tree algorithm C4.5 [7] (j48, under WEKA's implementation) was chosen due to its discriminative abilities as well as the visual ability to inspect and interpret the outputted trees. The algorithm was run with the default configuration and as options a 10-fold cross-validation was chosen. Also, no dataset balancing was preformed [8].

The obtained results point to 72.89% correctly classified instances and the obtained tree illustrates the importance of the previous donations (Figure 7). These results are mainly explained by the fact that returning donors already know that some medical conditions and behaviors are incompatible with blood donations.
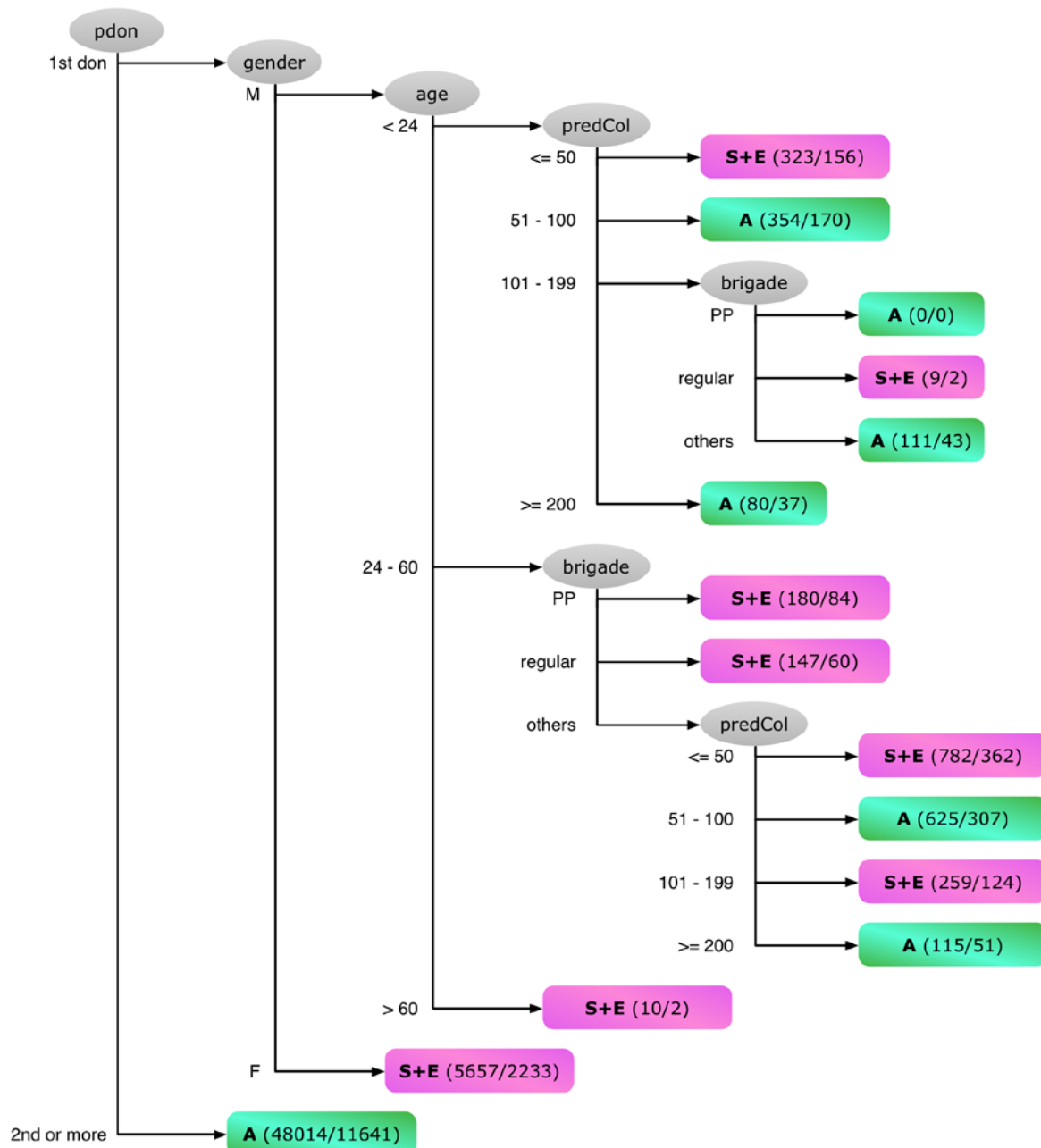


**Figure 7: A perspective of the decision tree obtained about the influence of first time donations and gender**

The gender is also a very important attribute. When looking at the data stored in SAD_BaSe of first time donors (5657 women/2995 men), a small difference between approved and not

approved men is verified, unlike what is verified amongst female first time donors where the amount of not approved donors is expressively higher than those approved (Figure 8).
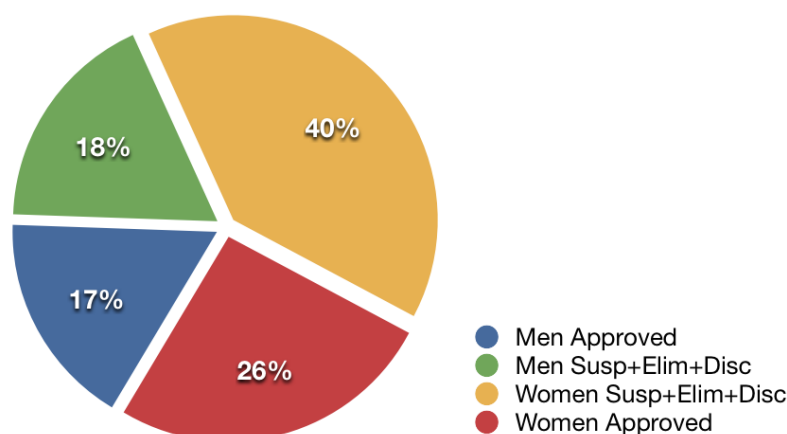


**Figure 8: First time donors Approved and Not Approved (suspended, eliminated and discarded unit)**

To understand what causes this clear difference, data of first time donations was looked in detail, specifically the causes of not approval. Low Hemoglobin (the value of hemoglobin in the donor is below the established limit) was the main cause of donor deferral in the female population with 626 (11.1%) deferrals whereas in the males only 51 (1.7%) donors were deferred by this cause. Hypotension and Low Weight (the donor weights below 50 kg) also present similar differences between females and males, 220 (3.9%) versus 52 (1.7%) and 140 (3.0%) versus 0, respectively. The main cause of men deferrals was the existence of a new sexual partner in the last six months, accounting for 204 (6.8%) deferrals, whereas in the females this accounted for a smaller number of deferrals, 177 (3.1%) (Figure 9).
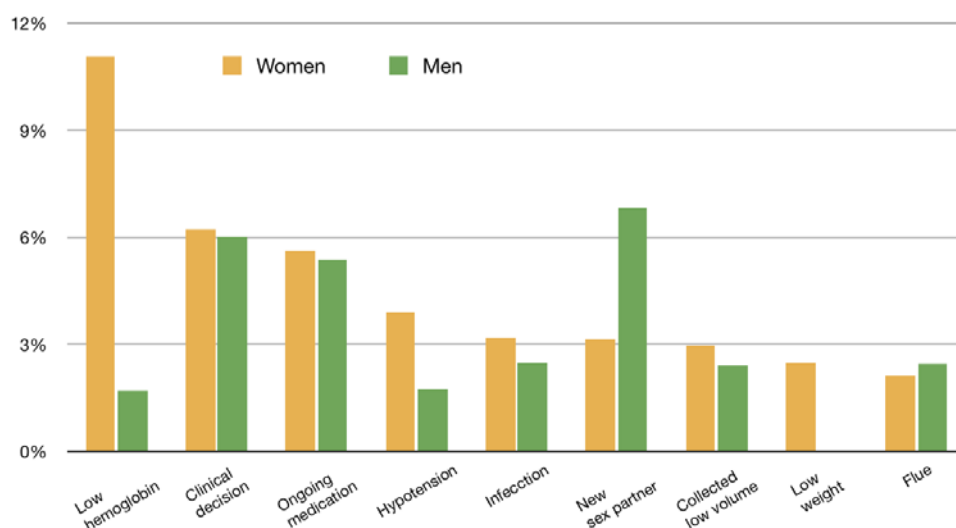


**Figure 9: Top causes (> 2%) of donor rejection by gender. Low hemoglobin, Hypotension, New sex partner and Low weight are the attributes where a difference between genders is noticed**

## 5　　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 was tested in a Portuguese blood establishment. A data snapshot from collections performed between January 2009 and December 2011 was uploaded and a coherence check performed to evaluate the data processing module. No incoherencies were found.

Staff from the blood establishment was invited to test and comment on the report module. The global opinion was very positive, especially from those involved in collecting data indicators prior to SAD_BaSe (paper reports were obtained from the blood bank software and Excel® spreadsheets were used to monitor blood bank activity).

This data analysis software complements common blood bank 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.

The application of data mining to the examination of potential donors corroborates some previous perceptions related to first time donors and gender and disclaims others like age, collection site and predicted collections. Actions can be taken by the blood establishment to try to mitigate them.

# References

[1] Decreto-Lei n.º 267/2007 de 24 de Julho, 2007. Available at: http://dre.pt/pdf1sdip/2007/07/14100/0469604717.pdf

[2] Council of Europe. Guide to the Preparation, Use and Quality Assurance of Blood Components. European Directorate for the Quality of Medicines, Recommendation No. R (95) 15, Strasbourg, 2011.

[3] DIRECTIVE 2002/98/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 January 2003, Setting standards of quality and safety for the collection, testing, processing, storage and distribution of human blood and blood components and amending Directive 2001/83/EC. *Official Journal of the European Union,* 46(L33):30-40, 2003.

[4] J. Leal. Validação de Sistemas Informáticos de Serviços de Sangue e de Medicina Transfusional. Faculdade de Medicina da Universidade do Porto, 2010.

[5] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I. H. Witten. The WEKA Data Mining Software: An Update. *ACM SIGKDD Explorations Newsletter,* 11(1):10-18, 2009.

[6] R. R. Bouckaert, E. Frank, M. A. Hall, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. WEKA – Experiences with a Java Open-Source Project. *Journal of Machine Learning Research,* 11:2533-2541, 2010.

[7] R. Quinlan. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, San Mateo, CA, 1993.

[8] I. H. Witten, E. Frank, and M. A. Hall. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, Burlington, MA, 3rd edition, 2011.