Real Time Biostatistics Software: Application in Acute Myeloid Leukemia Assessment

A. Bacarea[1], B.A. Haifa[2], M. Marusteri[2], M. Muji[3], A. Schiopu[2],
D. Ghiga[4], M. Petrisor[2], and V. Bacarea[4]

[1] University of Medicine and Pharmacy/Pathophysiology Department, Tirgu Mures, Romania
[2] University of Medicine and Pharmacy/ Medical informatics and biostatistics Department, Tirgu Mures, Romania
[3] Petru Maior University/ Engineering Department, Tirgu Mures, Romania
[4] University of Medicine and Pharmacy/ Medical Research Methodology Department, Tirgu Mures, Romania

Abstract— **the aim of this paper is to present an useful software in medical research. The new concept proposed is "real time biostatistics" and its application in Acute Myeloid Leukemia. For this purpose open source software (wxWidgets, R and SQLLite3) are used. The cases were patients with AML from Hematological Department, County Emergency Clinical Hospital Tirgu Mures. We created a friendly interfaced software that allows appropriate data collection and real time update of statistical parameters as each case is introduced.**

The medical importance derives from the possibility to have valid study and to see in each moment a change in the evolution of patients.

Collaboration between specialists (hematologist, PC programmer, biostatistician and methodologist) was really important for accomplishing our goal.

Keywords— **real time biostatistics, acute myeloid leukemia, open source.**

I. Introduction

Acute myeloid leukemia (AML) is a malignant disease with very standardized treatment and diagnosis techniques with a complex patient management. The incidence of AML is low, but it is increasing. Given the fact that the disease is rare one of the major concerns of the medical researcher is to be certain that each patient is accurately registered. Any data loss can reduce the level of significance of the study results.

The authors experience says that a data collection using a database management system doubled by statistic software in front of it could be a successful tool in order to obtain early warnings concerning data quality or, worse, treatment misconduct.

We have already developed a software application for data collection and analysis in AML. [1] Studying the results from previous similar studies we have found that an important tool for this application is missing: the data error warning in the very moment of data collection. Though the fact that the patient management is very well standardized, the evolution of the disease has a high variability and this reflects in the recorded parameters. Our application performs real time biostatistics and offers a snapshot of the results in every moment of the patient data entry allowing the formulation of hypotheses even if the number of cases is rather small.

Our aim is to extend the existing data collection software in order to incorporate the entire statistical protocol that will allow us to have a better real time view on the patient's status.

II. Material and methods

The software mentioned above has a user friendly design aiming to help the investigator to achieve a proper and complete data collection. In order to accomplished these tasks we used wxWidgets [2] library creating the user interface, and R functions [3] in order to estimate the level of statistical significance (p value) in the same time with the data entry. SQLite3 [4] was used as Database Management System. We have also use an open source software in developing these applications, in order to avoid legal issues (software license) and to provide to a potential investigator with low or no founding a more useful PC assisted tool for his research.

In this context we aim to propose a real time biostatistics software with practical application in AML patient management, concerning the survival analysis curves (Kaplan Meyer) and log rank test. The factors implied in stratifying the survival can be selected in the same time.

Each processed patient can modify the results in real time concerning the statistical outcomes (p value and Kaplan Meyer curves).

The patients with AML selected and included in this study were hospitalized in the Hematological Department, County Emergency Clinical Hospital Tirgu Mures.

III. Results

We created a GUI in wxWidgets to collect the desired medical data. We collected information regarding: age, date

S. Vlad and R.V. Ciupa (Eds.): MEDITECH 2011, IFMBE Proceedings 36, pp. 32–35, 2011.
www.springerlink.com

of diagnosis, date of death, blood count values, the CD leukocyte markers detected by flow cytometry. Survival period is automatically calculated in months, years or days of investigator's free will. You can enter and edit in real-time database cases, changes that are made are saved in real time and statistical results are automatically updated. Standardizing the entry modes of data does not allow incomplete data entry and the interference of data with subsequent processing of data. For binary data, it will also enable the introduction of binary values and also a variant with lack of value.

We have created a very friendly interface in order to asses in real time Kaplan Meyer curves (right window). Also we can evaluate a significant difference between the survival curves applying log rank test (left window). (Fig. 1) Each patient recorded can produce effects in drawn curves at once he/she is introduced in the database.

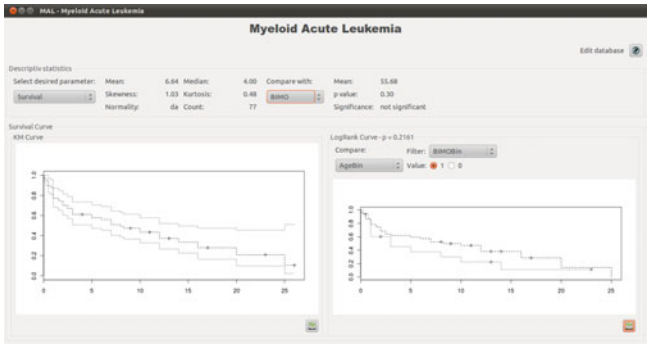

Fig. 1 Real time analysis interface with automatic survival curve (Kaplan Mayer) with 95% Confidence Interval (right window) and log rank test (left window)

For assessment of and testing for survival according to blood count values at diagnosis is needed binarization of quantitative results. This operation is specifically permitted in the program presented by changing the threshold value of binarization. It is known that there is no threshold value, for which the survival prognosis can be said to be good or bad (with some exceptions). In order to clarify these challenges the program provides the solution (Fig. 2).

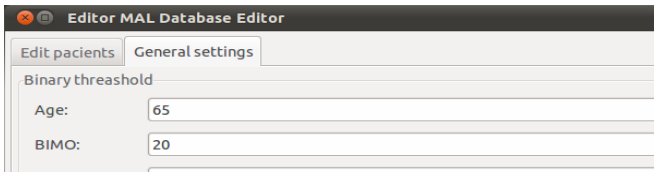

Fig. 2 Menu "General Settings" for establishing threshold values for laboratory measurements

The variable of interest which could generate differences in survival can be selected using a "combo box" (Fig. 3).

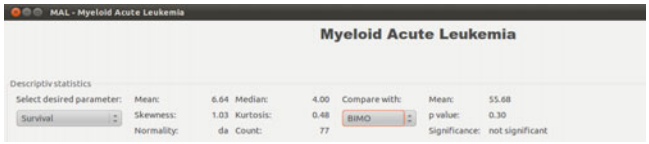

Fig. 3 Selecting the interest variable

Descriptive statistical estimators are calculated instantly and the results of log rank test as inferential analysis are listed above. (Fig. 4)

Fig. 4 Real time p value for log rank test

The time factor. In AML the incidence is low, so each new case is important to be carefully managed. By introducing records (patients with full data) the outcome is changing. In real time we can evidence of changes that can be really important for the treatment and the prognosis of the disease.

According to "real time biostatistics" the concept that we proposed, our software can provide a step by step update of information.

When introducing real cases in the AML database in order to compare survival periods, the program will draw a Kaplan Meyer curve for the studied variables including each new value recorded in. Also, using the "combo box" "Compare:" for selection of the binary variable which is suspected to modify the survival, a log rank test is performed and the outcome is displayed above.

To achieve the goal we have tested differences issued by adding new records in two different points in time. The results concerning descriptive statistics are shown below (Fig. 5).

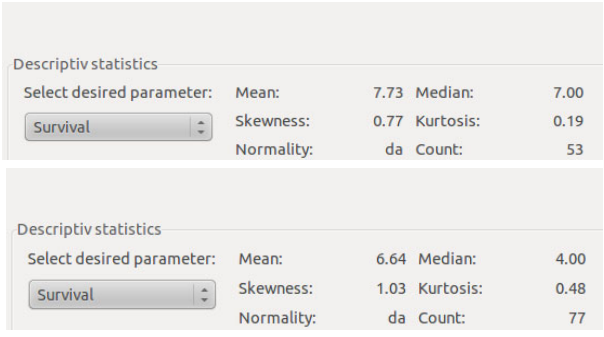

Fig. 5 Real time descriptive statistics

The program calculates descriptive statistics parameters as shown above (mean, skewness, normality test, mean median, kurtosis).

The software we propose gives visual information to the physician, at two or more points in time, the Kaplan Meyer curves, and the log rank test results for the variables tested. (Fig. 6)

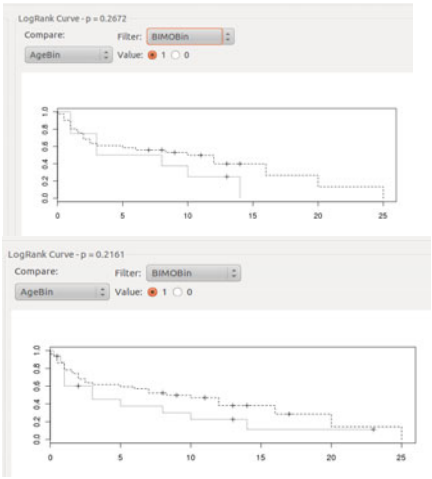

Fig. 6 Real time Survival curves (log rank test)

IV. Disscusions

For writing this program we used an open source GUI library called wxWidgets. This library is an open source construction kit for writing complex C++ interfaces and applications for a variety of platforms (Windows, Linux, Mac OS X and even Pocket PC). This library has been around for a long time now, contains mature code with a lot of functionality that helps a programmer in its work for building quality software. Its cross platform intrinsic nature allows us to create a software that behaves identically in all major operating systems and that is a big plus in the research activity, allowing for all involved researcher to use the operating system that suites them best without fear that it will create incompatibilities in the collected data.

In order to store the information for the patients we needed a database management system. We wanted our software to be portable and easily maintainable so we decided to use an embedded DBMS, SQLite3. It supports a big part of the SQL standard, it is fast and requires low memory overhead. But the most important facility is that it keeps the data in a single file that can be moved to another computer without affecting data integrity. We avoided in this way the bloating of the software that a regular DBMS would have added, the necessity of keeping a server for the data and to have a network to access it, in the same time being able to use the flexibility that the SQL language offers in data processing.

This wxWidgets library unfortunately doesn't contain any specialized section for statistical/biostatistical processing so we had to use something else for that. We decided to go with "R", a free software environment and language for statistical computing and graphics. It is a well known statistical software in the academic community. The preference for this software is given by the fact that it is also an open source software, it is supported on most platforms and has a modular design and has a shell that allows for scripting and bulk processing. Its facilities can be extended using packages downloaded from different mirrors. Many of its statistical extensions have been written by third party developers and researchers from different universities. So, it offers solutions for a variety of different problems making it a logical choice for inclusion in our program. The main problem was how to interface the R software with the GUI written in wxWidgets. The solution has been given by two extensions of "R", Rcpp and RInside. The first one, Rcpp transforms the R specific objects in objects more similar with C++ design and implementation, allowing C/C++ code to be run from inside the R shell. The former, RInside, as its name suggests contains the headers and libraries needed to interface with Rcpp objects. Using this two packages we were able to use the R statistical abilities in order to see how our results evolved in time as we could add more and more patients to the database.

Another facility that we considered useful in the program was the way information is saved during data entry. Any modification to the patient data is saved instantly without the need for any special action from the user. This helps a lot when there is a lot of data to be entered and also prevents loose of data if the user forgets pressing a "Save" button.

Any modification to the data also triggers a recalculation of the statistical parameters involved in the study so that the researcher can always have a clean and clear picture of the results.

Medically it is important to asses a difference between the mean of "Survival" variable. This finding should alert the physician if the mean decrease or increase. He/she must reevaluate the patients in order to detect which is the cause of this event. Evaluating a cohort of patients only after a full data collection can not give such information in real time, and a perturbation given can influence in both ways the AML patient management.[5]

Our precedent studies of prognosis in AML showed the necessity of such real time software. AML is an malignant disease with poor outcome concerning the survival period and the improvement of treatment is expected to produce

real time changing of the survival curve. AML is the disease where there is no rule regarding some parameters (cell blood count); there is no typical pattern of the disease that makes the software more valuable concerning an uniform data entry and data quality control.[6]

The existence of such of software is useful from the methodological point of view because a more accurate study design is possible; possibility of bias regarding data collection is diminished and that certify the validity of the study.

We have started the development of this software since 2008, relying on the lab hematologist, methodologist and biostatician's experience and the possibility of real time interpretation of data came out like a natural necessity.

So it is a team where everybody's contribution it is considered of equal importance.

V. Conclusions

The presence of open source software helps the researcher to develop useful tools in each medical field of interest.

Team work is a very important in the development of such software that will be used in particularly medical research.

Acknowledgment

This paper is partially supported by the Sectorial Operational Programme Human Resources Development, financed from the European Social Fund and by the Romanian Government under the contract number POSDRU/89/1.5/S/60782.

References

1. Anca Bacârea, Bogdan Adnan Haifa, Marius Muji, Alexandru Şchiopu (2011) Software Application For Data Collection And Analysis In Acute Myeloid Leukemia. Applied Medical Informatics Vol. 28, No. 1/2011, Pp: 16-22.
2. ***Wxwidgets Cross Platform Gui Library [Online][Cited Decemcer 2010]. Available From : Url: Http://Www.Wxwidgets.Org
3. *** R Project For Statistical Computing [Online][Cited Decemcer 2010]. Available From : Url: Http://Www.R-Project.Org
4. *** Sqlite [Online] [Cited 2010 December]. Available From : Url: Www.Sqlite.Org
5. Davis, Bh; Holden, Jt; Bene, Mc, Et Al. (2006). 2006 Bethesda International Consensus Recommendations On The Flow Cytometric Immunophenotypic Analysis Of Hematolymphoid Neoplasia: Medical Indications Cytometry Part B-Clinical Cytometry 72: S5-S13
6. Vardiman J W, Thiele J, Arber D A , Et Al. (2009) The 2008 Revision Of The World Health Organization (Who) Classification Of Myeloid Neoplasms And Acute Leukemia: Rationale And Important Changes Blood 114: 937-951

Author: V. Bacarea
Institute: University of Medicine and Pharmacy
Street: 38 Gh. Marinescu
City: Tirgu Mures
Country: Romania
Email: bacarea@yahoo.com