Specialized Software Engineering in Clinical Trials. GANfort Study

Bacârea V[1], Haifa B[2], Petrișor M[2], Ghiga Dana[1], Bacârea Anca[3], Muji M[4], Măruşteri M[2]

[1] Research Methodology Department, University of Medicine and Pharmacy, Tîrgu Mureş, Romania
[2] Medical Informatics and Biostatistics Department, University of Medicine and Pharmacy, Tîrgu Mureș, Romania
[3] Physiopathology Department, University of Medicine and Pharmacy, Tîrgu Mureș, Romania
[4] Engineering Department, Petru Maior University, Tîrgu Mureș, Romania

Objective: We aim to develop and implement a personalized software to accomplish data quality management in real time, reducing the chance of error in data collection and a "real time biostatistics" software linked to a collector datasheet.
Material and methods: We used C++ for programming, R for statistics and JavaScript (AJAX) for the interface. This application was developed for phase 3 GANfort study. This is a multicentric study. The results presented are simulated.
Results: The application presented below has a datasheet collection view with three tabs and a general presentation of the study and patient. The first tab collects data from the first visit (study inclusion and initiating the treatment), the second tab is for surveillance visit and the third tab generates real time statistic parameters.
Discussions: Using this type of software many methodological problems concerning data management can be avoided. "Missing data" and "outliers" or writing and typing errors become non-existent; typing constraints issued by datasheets and real time biostatistics eliminate them. The data can be introduced in the same time in different places and the matching data is performed simultaneously.
Conclusions: The time consuming data quality management is automatically solved using the software we proposed. Statistical parameters are calculated in real time. The end of data collection coincides with a final report of the study.

Keywords: statistical software, clinical trial, methodology, real-time biostatistics

Introduction

Data quality is a very important issue in "data analysis plan" in Clinical Trials management. There are some rules which should be taken into consideration concerning PC assisted data processing: doubling data collection by two independent operators, matching the two databases and verifying by sampling some patient information.

In order to produce new drugs many millions are invested in medical research. After a new drug is created comes the study of the conditioned drug. This study is very well standardized by the ICH (International Committee of Harmonization) in two parts (preclinical and clinical part) and four phases concerning the human administration. The first two phases consist in first human healthy volunteers administration in order to assess pharmacological features concerning the target tissue or organ (C_{max}, T_{max}) and the existence of serious, life threatening side effects. A phase three study presumes the administration of the tested drug to diseased patients seizing the expected effect. The number of patients enrolled in this phase of clinical trial is higher than phase I (9 healthy volunteers) or phase II (12–24 healthy volunteers), about hundreds.

The purpose of this study is to assess the efficacy of the newly discovered drug. An investigator brochure is made, containing a data collection form, a side effects report form, a detailed drug description, etc. Data collection form is a preprinted paper which contains open fields where the investigators (medical doctors) fill in patient information. Those filled forms are sent via special mail to the main centre where the data are analyzed. This process includes digitalization and statistical processing of the data. The newly created database needs to be accurate, respecting the data quality management: two independent PC operators, matching control of these two obtained databases, paper versus database matching extracting random samples.

Considering all of the above and the necessity of using a PC in this part of the study, we aim to provide a fully computerized method to accomplish data processing and statistic calculations.

Material and methods

In order to allow the data entry from multiple locations and to keep the system portable the same time, we choose a hybrid programming design based on the philosophy of Chrome OS. We used a client server approach. The server was written in C++ using Boost and wxWidgets libraries with a SQLite3 database backend [1,2]. The frontend graphical interface was written in JavaScript with the Qooxdoo AJAX framework, running in a browser environment and communicating with the server through HTTP protocol [3]. This design allows for multiple concurrent data access and at the same time the embedded SQL database together with the homebrew HTTP server represents a portable system that can be easily run from a regular PC with no need for installing or for a complex management required by a dedicated HTTP and SQL server.

Regarding the statistical environment, we used an open source software, highly appreciated by the academic community, called R [4]. In order to integrate R with our system, we wrote an R script which acquired the data through RSQLite package, processed and sent back the results to our server. The statistical results were integrated in a web page and displayed in the frontend. For generating and displaying charts we used Open Flash Chart, a system

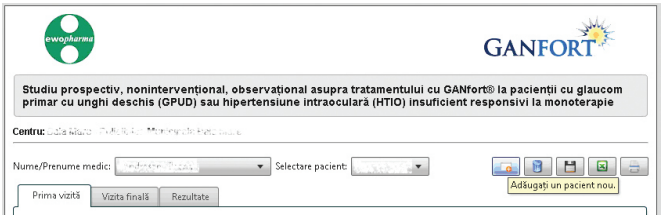

Fig. 1. The frontend header

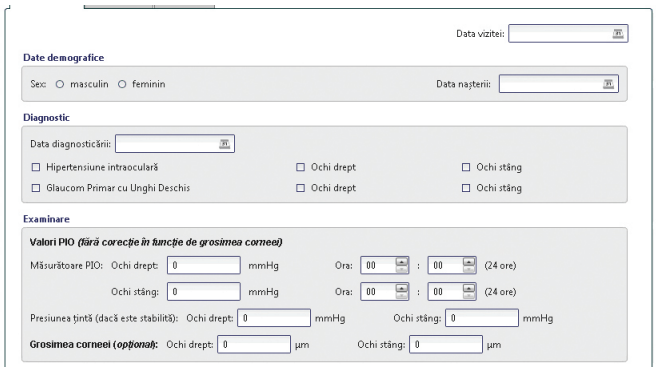

Fig. 2. First visit tab

Fig. 3. Type input validation

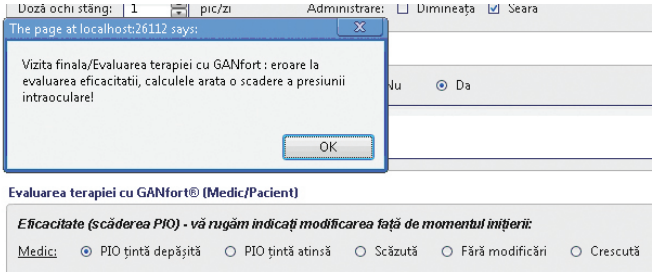

Fig. 4. Item control checking

that allows real-time user interaction with the generated charts.

This software was created for phase III study of GANfort, eye drops, used for lowering intraocular pressure.

Results

The study is prospective with a three months period between the two visits. The first visit was performed in order to assess the baseline parameters. All the data concerning the ocular health status was recorded on collection data form on paper. In this visit the doctor also recommended the new treatment with GANfort. Only the patients who agreed with the treatment change were included in the study. Those patients were called again after three months to assess the effect of the new treatment. All the previously recorded parameters were measured again and inserted in the data form collection paper.

All the paper data collection forms were sent for statistical processing. In order to digitalize the data, two inde-

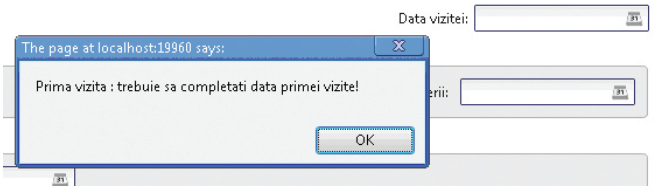

Fig. 5. Missing data alert

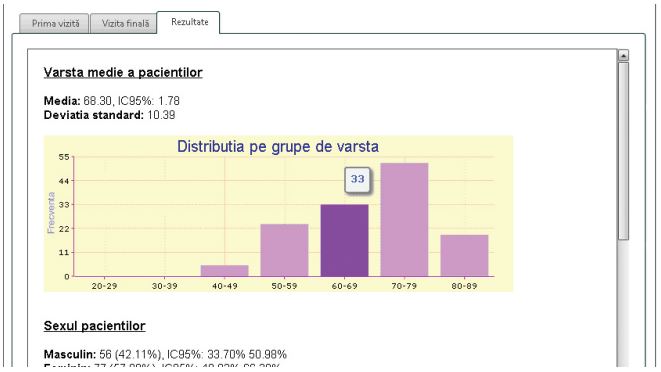

Fig. 6. Age histogram

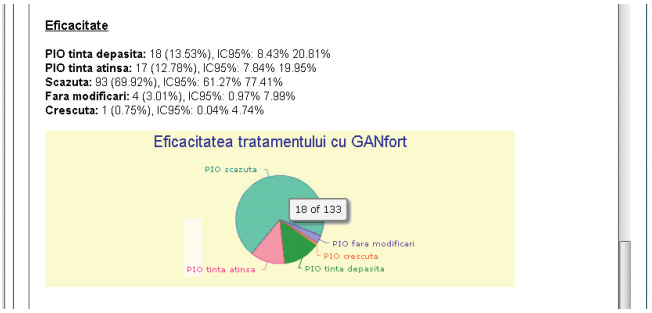

Fig. 7. Treatment efficacy

pendent PC operators were given our software. They filled in the information from the papers in the newly created database. The frontend of the software was modeled based on the study design with a short information concerning the study, sponsor and the studied drug followed by a complex form composed of three tabs (Fig. 1).

The first tab contains the data provided from the inclusion visit (first visit), with general information regarding the patient (the doctor who treated him, age, sex, diagnostic, etc.) followed by specific data regarding the eye health (intraocular pressure) (Fig. 2).

Similarly, the second tab contains information recorded from the final visit that reflects the patient's health status after 3 months of treatment with GANfort.

Data quality management requires several mandatory steps to avoid error in data input, such as missing data, outliers, typing error, etc. In order to prevent the time and resource wasted with data quality management regularly done by hand, we automatized most of this process.

The software checks the data entered by the operator in multiple ways. It is able to check for variable type (eg. text instead of number), it verifies if the entered value is in the acceptable range (e.g. it returns an error if the intraocu-

lar pressure seems to be an outlier) and it checks for some control items (e.g. efficacy evaluation based on the values of intraocular pressure before and after the treatment with GANfort) (Fig. 3 and 4).

Also, our solution doesn't allow for missing data in the required fields, generating an error that the operator must address before the software allows for data saving.

The database can be exported into a Microsoft Excel compatible XML format, allowing for data sharing with a third party institution.

In this way we enforce the database integrity and enable real-time statistical data processing, without fear of using incorrect data. For the ease of access, all the results are displayed on a third tab. Each data alteration (insertion, deletion, modification) triggers an automatic refresh of the results tab.

Based on the data analysis plan, the results tab includes information regarding demographic statistics, treatment efficacy and statistical inference for the parameters recorded before and after the treatment (Fig. 6 and 7).

All the results can be saved as a HTML page for later review of the statistical result at a specific moment or it can be directly printed.

Discussions

In the context of continuous development of the pharmaceutical industry which tries to provide new drugs and new therapies, the resources invested in clinical trials required for testing these new drugs are constantly increasing. The human factor is error prone [5], thus it is desirable to replace it with an automated system as much as possible, leaving only the interpretation of the outcomes to the clinical research operator.

From the authors' personal experience in developing software used in medical research [6], we can state that using such a software is very useful in medical studies, reducing the potential errors and minimizing the human and time resources required.

The current methodological procedures require at least two PC operators and a highly specialized person for statistical analysis of data. The software developed, presented above, automatizes the data quality control process, provides statistical analysis and generates the final report of the study. Thus, the proposed solution reduces the costs involved in this phase.

It is common knowledge that missing data and outliers in statistics are undesirable. We emphasize that our solution can solve the mentioned difficulties. At this time, most of the studies do not use such an automated system for data collection, validation [7,8] and processing or they are using simple forms for collecting data in Microsoft Access databases or Excel spreadsheets without any processing or complex validation.

Also, the data storage systems currently used require expensive licenses and management systems and trained personnel. The same goes for the statistical analysis software (SPSS, SAS, Graph Prism, etc), while our system is built using only open-source software.

Conclusions

Data quality management issues are automatically solved using the software we proposed, addressing missing data and outliers.

Statistical parameters are calculated in real time, providing snapshots of the statistical results throughout data entry process. The end of data collection coincides with a final report of the study.

Acknowledgement

This paper is partially supported by the Sectoral 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. *** wxWidgets Cross platform GUI library available at http://www.wxwidgets.org/, apr. 2011
2. *** SQLite Home page available at www.sqlite.org, apr. 2011
3. *** Qooxdoo Home page available at http://qooxdoo.org/, apr 2011
4. *** R project for statistical computing available at http://www.r-project.org/, apr. 2011
5. Mark Dale, Stephen Goundrey-Smith, Steve Higham and Andy Collinson – Reducing Human Error in Checking Concomitant Medications. Journal of Clinical Research Best Practices 2010, 6(10): 1–5.
6. Anca Bacârea, Bogdan Adnan Haifa, Marius Muji, Tudor Călinici, Alexandru Șchiopu – Software application for data collection and analysis in acute myeloid leukemia. Applied Medical Informatics 2011, 28(1): 16–22.
7. Curley RM, Evans RL, Kaylor J, Pogash RM, Chinchilli VM – Development and deployment of an internet-based Data Management system for use by the Asthma Clinical Research Network. Controlled Clin Trials 2001, 22: 135S–155S.
8. Trocky NM, RN, MSN and Fontinha M – Quality Management Tools: Facilitating Clinical Research Data Integrity by Utilizing Specialized Reports with Electronic Case Report Forms 2005 AMIA Symposium Proceedings: 1142.