Technical Report

How I Do It:

A Practical Database Management System to Assist Clinical Research Teams with Data Collection, Organization, and Reporting

Howard Lee, BS, Julius Chapiro, MD, Rüdiger Schernthaner, MD, Rafael Duran, MD, Zhijun Wang, MD, PhD, Boris Gorodetski, BS, Jean-François Geschwind, MD, MingDe Lin, PhD

Rationale and Objectives: The objective of this study was to demonstrate that an intra-arterial liver therapy clinical research database system is a more workflow efficient and robust tool for clinical research than a spreadsheet storage system. The database system could be used to generate clinical research study populations easily with custom search and retrieval criteria.

Materials and Methods: A questionnaire was designed and distributed to 21 board-certified radiologists to assess current data storage problems and clinician reception to a database management system. Based on the questionnaire findings, a customized database and user interface system were created to perform automatic calculations of clinical scores including staging systems such as the Child-Pugh and Barcelona Clinic Liver Cancer, and facilitates data input and output.

Results: Questionnaire participants were favorable to a database system. The interface retrieved study-relevant data accurately and effectively. The database effectively produced easy-to-read study-specific patient populations with custom-defined inclusion/exclusion criteria.

Conclusions: The database management system is workflow efficient and robust in retrieving, storing, and analyzing data.

Key Words: Database; query; statistical analysis; graphical user interface; organization.

With the growing amount of clinical research studies in the field of interventional oncology, selective patient data are becoming more difficult to store and organize effectively. Existing hospital electronic medical record (EMR) systems store patient data in the form of reports and data tables. Our institution's EMR system placed our researchers in a position where time-consuming methods are needed to search for suitable patients for clinical studies. Researchers had to manually read through the reports and data tables to filter patients and gather data. For most studies, spreadsheet programs such as Microsoft Excel (Microsoft, Washington) are often used as a data repository similar to a database to record and organize patient data for research. Once the spreadsheet is populated, it is manually filtered by set study parameters and then pushed to statistical analysis software for further analysis. For statistical analysis, columns containing text are translated into binary values (1 or 0) to be in a format acceptable by statistical analysis software. For example, each tumor entity is assigned a new column. Patient histologic reports are read manually to assign a value of 1 or 0 to each tumor entity column, 1 for positive, and 0 for negative. Under a tumor entity column, researchers would write a value of 1 for all patients with the tumor and a 0 for all patients without the tumor.

This method of data storage has limitations in the organization and the quality of the data. Data input and analysis without a database run a higher risk of incorrect data entry, patient exclusion, and a higher risk of introducing duplicates. Furthermore, data selection and calculation are time consuming. An alternative could be the clinical research database that Meineke et al. proposed (1). However, it is too unspecific for interventional oncology research and would need additional optimization, for example, the capability to automatically calculate various variables such as tumor staging systems and to record information about multiple treatment sessions.

The purpose of this study was to provide an improved workflow-efficient tool through the use of a clinical research database management system (DBMS) optimized for interventional oncology clinical research.

MATERIALS AND METHODS

This was a single-institution prospective study. The study was compliant with the Health Insurance Portability and

Acad Radiol 2015; ■:1–7

From the Russell H. Morgan Department of Radiology and Radiological Science, Division of Vascular and Interventional Radiology, The Johns Hopkins Hospital, Sheikh Zayed Tower, Ste 7203, 1800 Orleans St, Baltimore, MD 21287 (H.L., J.C., R.S., R.D., Z.W., B.G., J.F.G.); and U/S Imaging and Interventions (UII), Philips Research North America, Briarcliff Manor, NY (M.L.). Received July 30, 2014; accepted December 6, 2014. Funding Sources: Funding and support for this study has been provided by National Institutes of Health/National Cancer Institute R01 CA160771, P30 CA006973, and Philips Research North America, Briarcliff Manor, NY. **Address correspondence to:** J.F.G. e-mail: jfg@jhmi.edu

http://dx.doi.org/10.1016/j.acra.2014.12.002

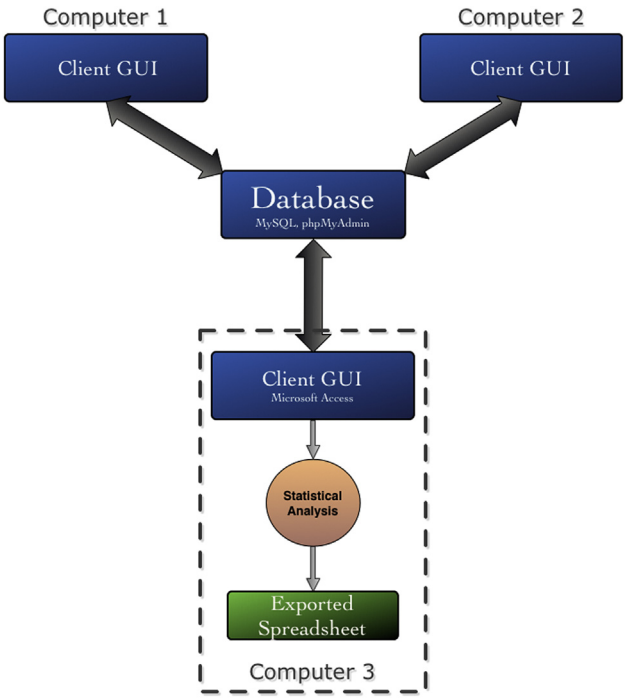

Figure 1. The dataflow chart. This chart shows a general layout of the database server and its clients. It illustrates how the database management system performs queries (*orange circle*) such as statistical analysis. Multiple computers are granted access to the database. The *blue rectangles* represent the database management system software. Researchers can use the database client graphical user interface (GUI) to import data without needing to format. Researchers also control data through the GUI. Queries are usually run through the GUI to provide wanted results. Once the results are obtained, researchers export the query to a spreadsheet, illustrated by the *green rectangle*. SQL, structured query language. (Color version of figure is available online.)

Accountability Act (HIPAA) and was waived by the institutional review board.

Database and Query Interface Design

The presented DBMS has two distinct parts, the database server and client interface, illustrated in Figure 1. The database is run by software (MySQL, Oracle Corporation, California, and phpMyAdmin, The phpMyAdmin Project, California) on a central computer server within the department (2,3). Authorized users were granted access to this password protected and encrypted secured server (HIPAA compliant). Multiple users concurrently add, edit, and query data remotely through a customized graphical user interface (GUI) using Microsoft Access (Microsoft, Washington). Any data changes are immediately logged for others to see. The database performed automatic calculations using queries, user-defined search criteria. Queries were saved, rerun, and exported to spreadsheets. Queries aid in data analysis and increase study productivity (4). They are powerful tools for filtering and sorting data sets. Figure 2 illustrates the query interface and an example of request from the database.

Graphical User Interface Design and Utility

In our research environment, the database GUI was created to facilitate patient data input. This was done by using custom user-friendly interface forms that contain textboxes and labels including demographic data, treatment information (eg, conventional transarterial chemoembolization [TACE]), tumor types, dates and types of radiologic examinations, and so forth. The GUI is used to view patient data and allows users to add/edit data (Fig 3). The database interface is not limited to one form. It can have multiple forms, shown as tabs, to assist grouping various medical data. Figure 4 shows an example of multiple tabs for groups of related data.

Automatic Calculations

Automatic calculations may be run between values, such as dates. For example, the database may calculate the time between baseline imaging, follow-up imaging, treatment dates, pretreatment and posttreatment dates, date of diagnosis, and the patient's date of death in relation to a particular treatment or event (eg, randomization), essential for survival studies. Using these queries, the database can also calculate the median overall survival automatically. The database can also automatically calculates clinical scores such as the Child-Pugh score and Barcelona Clinic Liver Cancer (BCLC) stage as shown in Figure 5 (5). For our purposes, the Child-Pugh score and BCLC stage were calculated using baseline data before a patient's first embolization as is typically done for staging. The illustrated calculators can be revised as needed. Once patient blood data are available, queries are run to produce a list of all patients with Child-Pugh scores. Researchers can then quickly retrieve them.

Statistical Output

Another powerful feature of the database is its ability to provide a first tier of statistical information. Using this GUI, the user defines the search criteria and runs queries to obtain immediate statistical information about a particular set of parameters. With this feature, the database can quickly output an accurate summary of patient data such as, for example, how many patients have colorectal carcinoma and undergo conventional TACE.

Questionnaire Assessment

A questionnaire (15 questions) was designed and distributed to 21 board-certified interventional radiologists who conduct clinical research at our academic hospital that include phase I, II, and III clinical trials and retrospective studies. The questionnaire determined how data are controlled in retrospective studies and the likelihood to use the database. The questionnaire is shown in Table 1. The purpose of the questionnaire was to 1) illustrate the general scope of where researchers were having problems within Excel and data organization,