

Databases' Design and Implementation in File Maker for the Cervical Screening: Data Quality Control

LUCIANA NEAMȚIU

ABSTRACT. The paper presents the design and implementation in File Maker for the cervical cancer screening databases. The data are collected on paper form. The interface for the data collection of the database is designed to allow easy computer entry. Some checks are required in order to validate the data. The main emphasis is on quality control of the information recorded.

1. INTRODUCTION

The cervical screening program consists in testing of asymptomatic women every three years. This has proved to be an efficient means for reducing the incidence and the mortality from cervical cancer; on a population basis, the most efficient use of given resources is therefore achieved, if all women in the relevant age groups are offered screening every three years. Cancer screening is a multidisciplinary activity involving clerks, nurses, midwives, cytotechnologists, pathologists, gynecologists, surgeons, GP's, epidemiologists, economists, etc. All data provided by them should be monitored. Before cervical screening can be implemented mechanism for gathering essential data for the day to day operation of the program and for statistical purposes must be in place. A comprehensive registration system for women at risk is a prerequisite for an organized screening program as is a system for registering the Papanicolau smear reports. The system should be computerized. A register of biopsies and the histology reports from women referred for treatment is also required.

2. THE MEDICAL PROBLEM

We need to solve the following problem: how to design and implement an application for the monitoring of the data for the cervical screening program. Women between 25 and 65 years old are invited for the test. They go to their GP to receive the screening request form. (The screening request form has 4 parts: I. woman identification data, including her GP's identification, II. clinical data, identification about the specialist who took the smear and about smear, III. cytological report including the identification of the laboratory and the anatomopathologist, IV. information about the disease management in case of positive test result). After that women can be tested by their GP or by a gynecologist. The smear and the form are collected to the cytological laboratory. The pathologist interprets the smear and gives the result on the same form. The gynecologist and/ or the GP are informed

Received: 14.09.2004. In revised form: 27.11.2004.

2000 *Mathematics Subject Classification.* 92B10, 92B15.

Key words and phrases. *Databases, quality control, screening.*

about the result and the suggested management. If the test result is positive the patient goes to a gynecologist or an oncologist.

3. QUALITY CONTROL

The cervical screening databases are a source of information. The quality of information is a product of the quality of the data and the quality of their presentation. The validation of data at entry is necessary to provide complete and correct data. However, accuracy is influenced not only by erroneous data but also by missing data. So two important type of check can be made: validation checks and consistency checks.

Validation checks: the computer on each data item to ensure that no invalid codes are fed into the database carries these out. These may take the form of range checks - for example, that no patient's age can be less than 0 and greater than, say 100. The format of the data item can be checked, for example, to ensure that the patient's name contains only alphabetical characters and the age only numerical codes.

Consistency checks: These checks compare the values of certain data items against others. It is impossible to determine which of two opinions about a data is correct unless there are firm rules.

Missing information

As stated, validity is the proportion of cases recorded with given characteristics (e.g. age) which truly has this attribute. However, accuracy is influenced not only by erroneous data but also by missing data. Completeness of detail is the proportion of cases registered for which a specific item is known. Evidently it is not always possible to complete each data item on every patient, so the database will always have items recorded as "unknown". Certain items are deemed essential, for example personal identification number (CNP). Others, such as phone number, are not. The percentage of missing values for non-essential data items can vary. Our application includes the obvious traditional data checks.

4. FILEMAKER PARTICULARITIES

FileMaker is a relational databases system. Each FileMaker file contains records and informations about the file structure (fields and its definition, access privileges and passwords, layouts and scripts).

Fields and its definition: Fields can store many types of data: text, numbers, dates, times, pictures, sounds, movies, calculations, and summaries. A container field can store a picture, movie, sound, or OLE object (Windows only). A calculation field stores the result of a calculation formula. The result can be text, a number, date, time, or container. A summary field is the result of a summary formula that uses field values from more than one record in the same file. The result can be a number, date, or time. Global field: one value to be used in all records of the file. The value can be text, number, date, time, or container.

Script: One or more instructions (script steps) that there are defined to automate repetitive or difficult tasks. The scripts are defined using ScriptMaker. Clicking its button, choosing its menu command, or calling it from another script can run the script. A FileMaker databases can be used in a network environment.

Sharing (or hosting) a FileMaker Pro database is independent of file sharing that can be set up on the computer. Access Rights from Windows 95, 98, 2000 or Windows NT are not associated with access privileges set up in FileMaker Pro, and vice versa. If the file is set up to be shared, the first person to open is the host. Users who open a shared file after are guests.

5. FILEMAKER FIELD VALIDATION

In FileMaker you can set entry options to ensure that data is entered into a field correctly. For example, you might want the data to be a date instead of text, or a value to be within a specified range. When data verification options are set, FileMaker Pro displays a message if data is entered incorrectly. We can verify that: the value entered is a number, date, or time; the field is not left blank; the value entered matches another value in the same field in any other record; the value entered does not duplicate a value found in this field in other records; the value entered matches a value in a specified value list; the value entered is within a specific range of letters (alphabetically), numbers, dates, or times; the value entered is valid, based on a calculation that returns the value True or False (Non-zero values are True; zero and empty values are False.). If one or more validation options are selected for a field it is required that an entered value meets the validation requirements. FileMaker display a custom message if an entered value does not meet the validation requirements.

6. FILEMAKER IMPLEMENTATION

The first step is to collect the data from the paper form. For this purpose a data application in FileMaker was designed and implemented in Cancer Institute "Prof. Dr. I.Chiricuta". The main application FileMaker file is Citologii.fp4. We also have the following catalog files: medfam.fp4 (contains the general practitioners names), ginec.fp4 (contains the gynecologists names) which are used as lists in Citologii.fp4. The main layout in Citologii.fp4 was designed similar to the paper form to facilitate the pick up work for the users. The user has some buttons for the actions (new patient, search, exit, reports). Analyzing the data wave and the paper request form the next items are needed: The patients identification data (name, personal identification number -CNP, address, phone number); The sender (doctor who take the smear - gynecologist or GP) name and address; The name and address of the patients general practitioner if not the sender; Date of smear; Source of smear; Date of last smear test; First day of the last menstrual period duration of cycle; Reason for smear i.e. screening or diagnosis; Specimen type (cervical scrape, endocervical brush, vaginal pool); Condition of patient (pregnant, post natal, post menopausal); Relevant clinical data; Cytology report; Suggested management; Laboratory; identification; Name of pathologist and date. The computer form for the data introduction (collection) is design very similar to paper form for an easy register work. The purpose of the application is to check for validity (individual data item edits) and consistency (combinations between items). The data items checked by the application are: Serial number looks for duplicate registration. All dates must be valid dates according to the format specified. Data combination edits: Date/ Birth date: the birth date must be before the test date. The date

of cytology laboratory must be the same or after the date of test. The personal identification number (CNP) must be a valid one (13 digit format, the first digit is 2, etc.). The general practitioners, the gynecologists, the pathologists must be from their respective lists. The lab must be from the lab list. The age calculated from CNP must be between 25 and 65 years old. The county (from addresses) must be from the Romanian county list. All the lists are catalog files. Some examples about validation in FileMaker definition of fields:

Field Name	Field Type	Formula/ Entropy Option
Form number	number	Do not allow user to override validation; Required value; Unique values only; Message: "Form Number ?"
GP	text	Value List; Do not allow user to override validation; Required value; Message: "GP ?"
Form Date	date	Of type date; Required value; Do not allow user to override validation; Validation calculation = If (FormDate <= Today) and (FormDate >= BirthDate) and (FormDate >= Today-60); Message: "Form Date ?"

Access privileges

The access privileges for the users of databases are essential. The rights must be specified for each action and user. Passwords and groups are related. The system administrator can define a password for each employee and then associate passwords with one or more groups that correspond to different departments.

7. CONCLUSION

The application was made for the Cancer Institute and it is in use since 2002. Advantages using FileMaker: validation of data to reduce the risk of errors; possibility to recover files; access restriction by password. The data quality control is very important and during the testing time some new validation was imposed in order to have a better quality.

Entry Options for Field "DataComplearii"

Auto-Enter Validation

☒ Of type Date

☒ Not empty ☐ Unique ☐ Existing

☐ Member of value list <unknown>

☐ In range to

☒ Validated by calculation Specify...

☒ Strict: Do not allow user to override validation

☒ Display custom message if validation fails

data completarii?

☐ Repeating field with a maximum of 2 repetitions

Storage Options... OK Cancel

REFERENCES

- [1] Anahory, S. and Murray D., *Data Warehousing in the Real World*, Addison-Wesley, 1997
- [2] Bnsca, O., *Baze de date*, Editura ALL, Timișoara, 1997
- [3] Bowers, D.S., *From data to database*, Chapman and Hall, London - New York - Tokio - Melbourne - Madras, 1990
- [4] Gardarin, G., *Maîtriser les bases de données. Model et langages* (troisième tirage), Éditions Eyrolles, Paris, 1997
- [5] Gardarin, G., *Bases de Données objet and relationnel*, Éditions Eyrolles, Paris, 1990

- [6] Irime Al., Nicula F. Al., Șuteu O., Coza D. and Neamțiu L., *Programul de prevenție și control în patologia oncologică*, Editura Medicală Universitară "Iuliu Hațieganu", Cluj-Napoca, 2002
- [7] Lupșa Luciana, *The Design and the Implementation of a Distributed Multi-User, Multi-Access Database for Supplying with Drugs a Hospital Pharmacy*, in *Bulletins for Applied and Computer Mathematics*, PAMM - Centre Budapest, BAM - 1912/ 2001 (XCVI-B), pp. 213-220
- [8] Neamțiu L., *Proiectarea bazelor de date pentru trialurile clinice*, in *Analele Universității "Aurel Vlaicu" din Arad, Seria Matematică, Fascicola Matematică Informatică*, 2002, pp. 94-99
- [9] Reboul G., *Informatique de gestion. Analyse et modèle relational*, Dunod, Paris, 1997
- [10] Șuteu O., Lazăr L., Irime A., Nicula F., Coza D., Duma R., Păis R. and Neamțiu L., *Organization and Implementation of Cervical Screening Pilot Program in County of Cluj, Transilvania*, in *5th International Multidisciplinary Congress EUROGIN 2003*, Monduzzi Editore, International Proceedings Division, Bologna, Italia, 2003, pp 235-239
- [11] *** *File Maker. User's Guide*, Claris Corporation, Santa Clara, 1997

CANCER INSTITUTE "I. CHIRICUȚĂ" CLUJ-NAPOCA
EPIDEMIOLOGY AND BIOSTATISTICS DEPARTMENT
REPUBLICII NO 36-38, 410015 CLUJ-NAPOCA, ROMANIA
E-mail address: luciana@iocn.ro