

A specific database for providing local and national level of integration of clinical data in cystic fibrosis[☆]

T. Leal^a, G. Reyckler^b, P. Mailleux, J. Gigi^c, V. Godding^d, P. Lebecque^{d,*}

^a *Clinical Chemistry, Université Catholique de Louvain, Brussels, Belgium*

^b *Rehabilitation and Physical Medicine, Université Catholique de Louvain, Brussels, Belgium*

^c *Bacteriology, Université Catholique de Louvain, Brussels, Belgium*

^d *Pneumology, Université Catholique de Louvain, Brussels, Belgium*

Received 4 May 2006; received in revised form 6 July 2006; accepted 17 July 2006

Available online 5 October 2006

Abstract

It has recently been stated that a database is an essential tool in the management of CF. The purpose of this work is to create a specific database allowing optimal performance of storage, search and retrieval functions on patients with CF. A specific database was developed using a Windev licence, for application via Microsoft supported platforms or Intranet system. The database allows real-time point of care data management of medical, investigational and administrative data. It is currently being used in the 6 Belgian reference centres. It represents a useful tool for gathering information on routine clinical and lab data, bacteriology, treatments, complications and specific outcomes for clinical and research purposes. The ongoing evolution of the database includes enhancements toward research data orientation including comparison of patient data between different centres and completeness of the National CF registry questionnaire. A complimentary copy of the software can be provided to multidisciplinary accredited CF centres worldwide upon request.

© 2006 European Cystic Fibrosis Society. Published by Elsevier B.V. All rights reserved.

Keywords: Cystic fibrosis; Database; Registry

1. Introduction

Cystic fibrosis (CF) is the most common autosomal recessive inherited disease with fatal outcome in Caucasian populations. CF disease involves pulmonary, gastrointestinal, nutritional, psychological, urological and obstetric complications [1]. Despite impressive advances in understanding the molecular basis of the disease, these have not been translated into better clinical care (other than neonatal screening) and life expectancy is still limited in CF. Pulmonary inflammation and repeated respiratory infections culminating with colonisation by *Pseudomonas aeruginosa*

are responsible for the progressive loss of lung function which is the main cause of morbidity and mortality in CF. Over the last decade, a generic approach to track the progress of this chronic disease and to allow improving life expectancy and quality of life has been established by the emergence of a multidisciplinary group of health care professionals including clinicians and specialist paramedical staff such as nurses, physiotherapists, nutritionists, social workers, psychologists, pharmacists and clinical biologists [2,3]. A recent European consensus for standards of care for patients with CF has summarized functions of each member of the CF team and has stated that database is an essential tool in the management of CF and that every patient with CF should be part of a local and national database [4]. The work described here was aimed at creating a specific clinical and administrative database [5] which provides support of patient data management for multidisciplinary CF teams. The developed software should facilitate optimizing performance

[☆] Supported by grants from the Belgian CF Association.

* Corresponding author. Pediatric Pneumology, Cystic Fibrosis Reference Center, St Luc University Hospital, 10 Av Hippocrate; B-1200 Brussels, Belgium. Tel.: +32 2 764 19 39; fax: +32 2 764 89 06.

E-mail address: Lebecque@pedi.ucl.ac.be (P. Lebecque).

of storage, search and retrieval functions on patients with CF and contribute to fulfil the needs of the datasheet completion of the national registry questionnaire.

2. Technical issues and requirements of the CF database

The CF database (CFdbase) is a databank system built for matching specific needs of a multidisciplinary and accredited CF team, to which computer specialists and database managers have advantageously joined favouring the development of the software infrastructure. It was developed using Windev as a general programming language for applications via Microsoft supported platforms or Intranet system. By integrating a set of tools within the data system, the software can be expanded from a single-user workstation to large multiuser networks. The database is open to interoperability with Windows-based applications and Access supports. The CFdbase recommended system requirements are listed in Table 1. The program runs in a multilingual– English, Dutch and French– application.

3. Working with the CF database

3.1. Data input and processing

Fig. 1 outlines the data management diagram from a new patient data up to the creation from further exploitation of the generated databank. As a data collection tool, the CFdbase allows a real-time data entry of multiple administrative, clinical and investigational information. In our centre, 154 patients were entered into the CFdbase.

The main functions of the data capture patient file and different aspects of the record are illustrated in the subsections of Fig. 2. Completion of all fields of the patient file occurs prospectively by the physician during outpatient visits or hospitalizations and it takes less than 2 min. In order to illustrate the multiple functionalities of the CFdbase while recording or editing individual patient data, a real anonymised patient from our CF centre was selected. The patient is a 20-year-old girl homozygous for the $\Delta F508$ mutation. She is attending our CF centre since diagnosis at 1 month when she first presented with chronic respiratory symptoms and failure to thrive (Fig. 2.7). Last name, first name, date of birth and sex represent essential data required for recording a

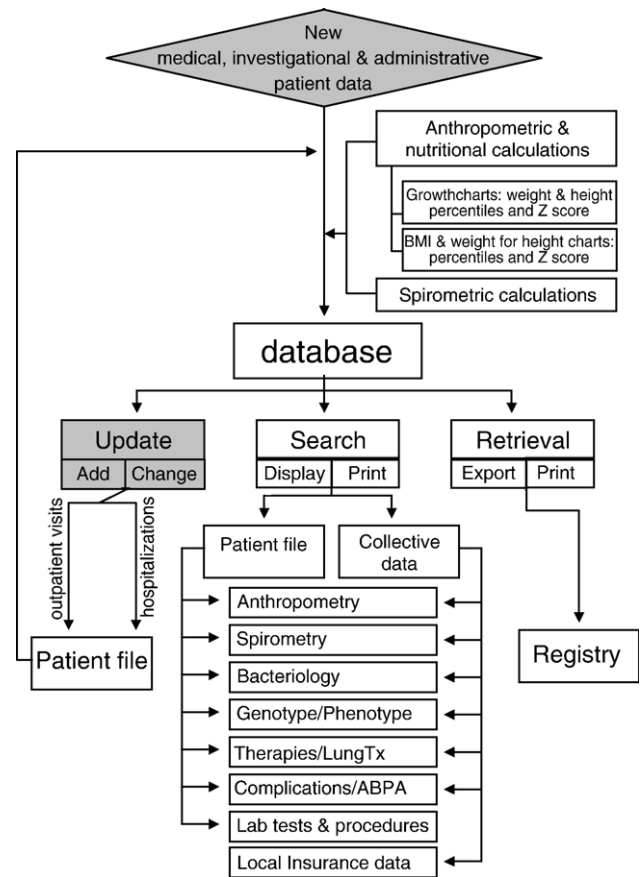


Fig. 1. Data management diagram used in the CFdbase software.

new patient in the databank system (Fig. 2.1). Data capture occurs using integrated operational modules accessed by differentiated buttons. Each button opens windows (Fig. 2.2) containing data classified into different sections, namely medical and administrative data, outpatient visits and hospitalizations, therapies, complications, lab tests and procedures including those examining the eventuality of allergic broncho-pulmonary aspergillosis (ABPA), bacteriologic data and antibiotic therapy information. A separate button is available to export anonymous data to the national registry. In the medical datasheet section, button presentations (Fig. 2.5) display modules including initial disease profile at the time of CF diagnosis or circumstances leading to diagnosis; phenotype–genotype information including pancreatic status and details of specific *CFTR* mutations; complementary diagnostic tests performed such as sweat test and nasal potential difference; family history details; evaluation of patient and/or family decision process regarding listing for organ transplantation; enlisted circumstances (cardiac, respiratory, hepatic, trauma, suicide or other causes linked to cancer, to an organ transplantation or unknown causes) in which patients die; tobacco habits and atopy history; and the presence of chronic colonization by *P. aeruginosa* and recent isolation of *Burkholderia cepacia*. Data processing of these two latter features generates particular alarm signs (Fig. 2.6). Other generated alarm icons include those for allergy to medications,

Table 1
CFdbase recommended system requirements

Operating system: Windows XP/2000/NT4 are the preferred platforms. Windows 95 and 98 are also supported. Administrator rights has to be enabled for Windows XP
Hardware: Intel® Pentium®-compatible processor (PII 300Mhz minimum)
Memory: 64MB (128MB RAM preferred)
2MB DirectX 6.1 compatible graphics card or above support
DirectX 6.1 compatible sound card or above support (optional)
CD-Rom drive (speed 4×) or above support
SVGA monitor (800×600) or above support
Keyboard and mouse
Up to 50MB of available hard-disk space

Fig. 2. Datasheet for input, update, search and retrieval functions from an actual CF patient registered with the CFdbase software. 1 — Patient identification; 2 — Toolbars allowing to access multiple clinical, investigational and administrative patient data modules; 3 — Toolbar allowing to select separate modules to be used by the paramedical staff (physiotherapist, nurse, social worker, nutritionist, psychologist, *varia*); 4 — Toolbar allowing to access a single screen displaying summarized information of the patient including basic medical and historical data, phenotype–genotype, longitudinal graphical representation of weight, BMI (body mass index), lung function and bacteriology data drawn as a 5-year span table; 5 — Multiple thumbnail presentations to access modules for collection data, as indicated; 6 — Area allocated for alarm icons, such as that indicating the presence of chronic colonization by *Pseudomonas aeruginosa*; 7 — Formatted data-field definitions for easy checking-off; 8 — Blank field for additional free-text annotation.

medical agreements to be renewed and patient death. In order to increase completeness, accuracy and validity, data input process is usually entered by checking-off formatted data-field definitions (Fig. 2.7). Additional blank fields are provided for free-text annotation, as appropriate (Fig. 2.8).

Absolute anthropometric and spirometric values, including weight (in Kg), height (in cm), FVC and FEV₁ (in L), entered at each clinic outpatient visit or throughout hospitalizations, are converted in percentiles, percentage of normal values for age and sex [6,7], and Z scores [8] and are therefore automatically integrated in the databank. For sequential assessment of nutritional status, body mass index (BMI) and percentage weight for height, calculated for children [9], are converted in percentile charts and Z scores. Tracking of developmental data with respect to the NCHS percentiles is drawn in the familiar graphical growth chart format. Graphical and table output of anthropometry and spirometry data are easily accessed via the longitudinal data button (Fig. 2.2).

3.2. Data update

All individual recording files can easily be edited for adding or changing data. Application updates, as new re-

coding files, are carried out in real-time in routine clinical practice by the physician during outpatient visits or throughout hospitalizations. A warning validation rule requests confirmation while adding, changing or withdrawing data.

3.3. Data search

Database search capability for individual patients and for collective data is facilitated by graphical, spreadsheets and tables output generated by the software. Assessing a button presentation (Fig. 2.4) provides a one-screen summary (Fig. 3) of the most relevant clinical data including age at diagnosis, genotype, pancreatic status, sweat chloride and nasal potential difference results, and graphical illustrations of growth, BMI chart in Z score, lung function curves and microbiology data for the 5 last years. The latter allows to easily applying Lee's classification of *Pseudomonas* infections [10] for individual patients.

For illustrating collective data capability of the software, a survey report of spirometric (Fig. 4) and microbiological (Fig. 5) data from CF patients regularly receiving full-time care at our centre at the end of the year 2005 was downloaded. Spirometric data are processed for the whole, paediatric and

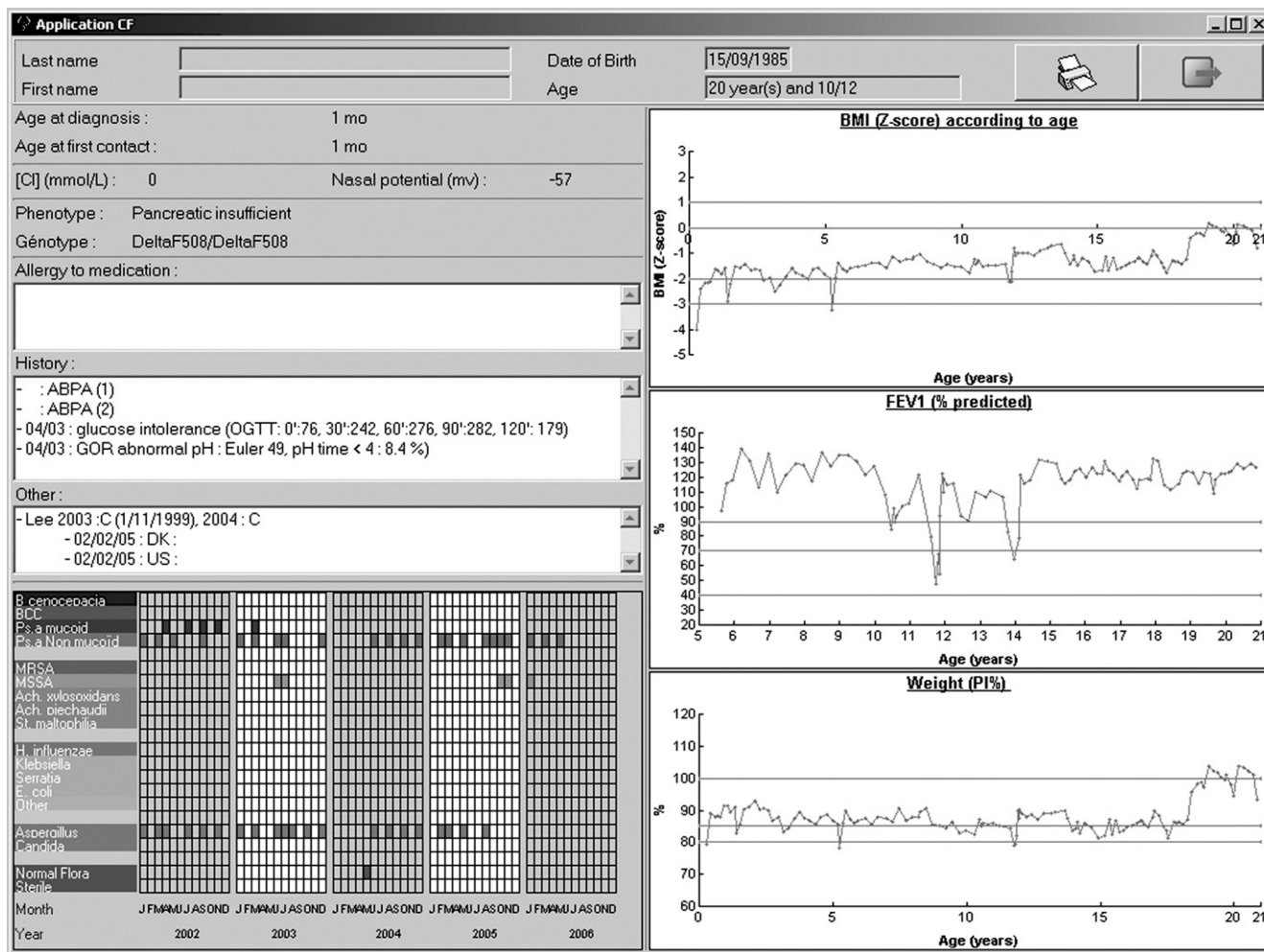


Fig. 3. One-screen summary of clinical and laboratory data of an actual CF patient registered with the CFdbase software. BMI is expressed in Z score according to age. FEV1 is expressed as percentage of predicted values for age and sex according to Knudson et al [6]. Weight is expressed as percentage of ideal weight. Results from cultures performed in every outpatient visit are drawn as a 5-year span table in which each successive month is defined.

adult population, also assigned according to the degree of severity of the lung disease [11]. Analysis of global frequency and prevalence of respiratory pathogens is determined annually taking into account results of cultures entered in every outpatient visit. As illustrated in Fig. 5, the prevalence of each listed pathogens, including mucoid and non-mucoid *P. aeruginosa* strains, appears clearly in the report.

3.4. Data retrieval

Data retrieval from the CFdbase meet clinical routine practice, epidemiological research purposes and the needs of the national CF registry that documents diagnosis and annual follow-up evaluations of CF patients who are seen at national accredited centres. Thanks to interoperability with Microsoft Office software components, individual or collective data can be easily printed or exported as spreadsheets, tables or graphs. Exporting data to an Excel format file is used for transferring data to the national registry.

As originally developed in the UK CF database [12], for completion of the national CF registry, a unique clinic-based PIN number is automatically generated for each patient by a standard hashing algorithm that retains personal identifying details such as the patient date of birth and sex, his(her) position amongst siblings and his(her) mother's date of birth. The PIN code allows patient anonymity and identifies the patient within the database. The PIN code is kept private even the patient is seen in another CF clinics. After written informed consent has been obtained, patient data are annually assembled to complete the database protocol and for merging with the national CF dataset.

4. Discussion

We developed an advanced, specialised, yet simple to use, CF database system allowing day-to-day patient data management in routine clinical practice during outpatient visits or throughout hospitalizations. The software contains a help file

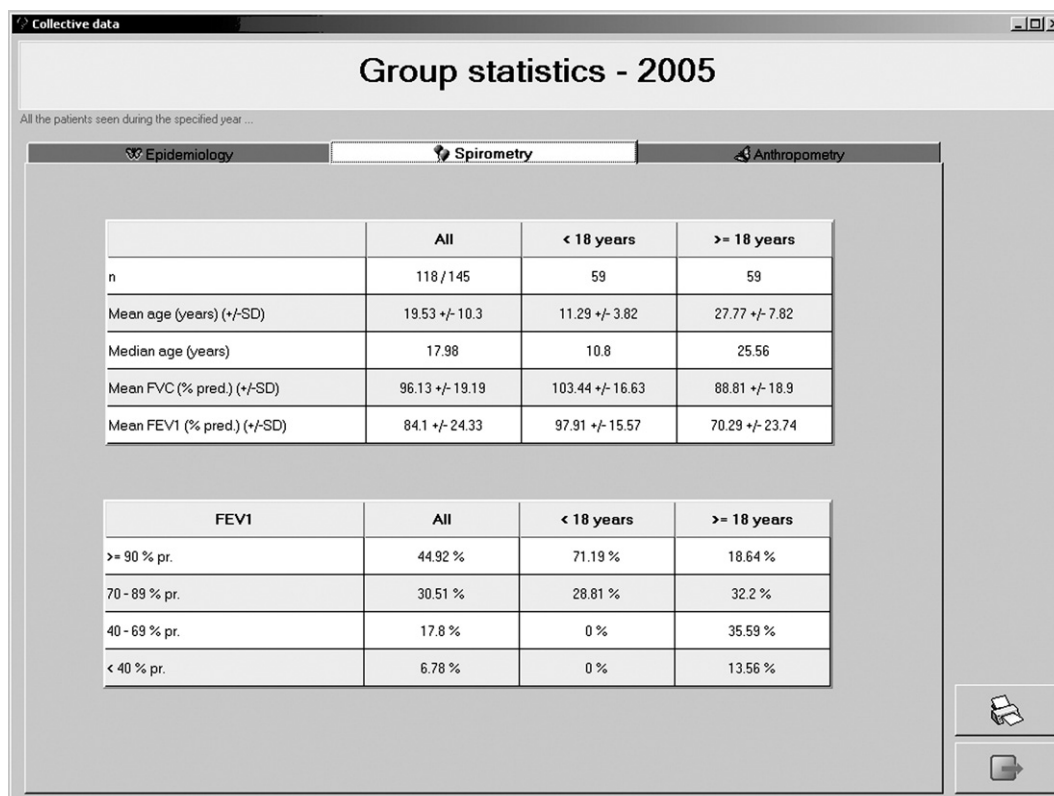


Fig. 4. Collective spirometric data of 118 CF patients attending our centre at the end of 2005. Data are processed for the whole, paediatric and adult population, also assigned according to the degree of severity of the lung disease [11].

including application description and biometric and spirometric references used. It was built for answering specific needs of multidisciplinary and accredited CF teams [13–16] following the recommendations of the recent European consensus [4], and for contributing to solve eventual tensions between local clinicians and the needs of the national registry. New recording, update, display and retrieval functions of clinical, investigational and administrative patient data are tightly integrated for an unlimited number of registered patients. Real-time point of care data entry does not need to be previously recorded on paper forms. For resourced CF centres, this represents a time saving advantage as others in clinical routine practice which also advantageously favours homogeneity in data acquisition (Table 2). Indeed, electronic data sources have proven to be a great help in achieving complete and validated data, in reducing data variability and in minimizing handoffs, human errors or bias potential that enter into retrospective data collection [17]. By integrating a set of tools within the data system, the database can be utilized from a single-user workstation to large multiuser networks. A transition period for transferring historical data from paper-based files to electronic database is required.

The software brings advanced search capability of producing individual and longitudinal patient reports that can help the clinician to provide feedback to each patient about their growth, lung function, nutritional or infection status. These

graphical outputs can be shared with the patient or his/her family during outpatient visits, in the hope that this could improve their adherence to follow-up. Moreover, this dataset can facilitate epidemiological, cross-sectional and longitudinal research. Ready access to longitudinal analysis of individual patients is provided by graphical, worksheet and table output generation including growth charts, drawn in absolute values or Z scores, and plots of lung function according to age. An original graphical display of the software is that of the microbiological data, which plots results of respiratory cultures performed at each outpatient visit, every 3-month interval, over a 5-year selected period. This allows to easily using Lee's classification [10] for defining the prevalence of *P. aeruginosa* colonization. In the absence of agreement on what is best practice, percentage weight for height was chosen for assessment of nutritional status in children though its reliability has been questioned because of the wide inter and intraexaminer variation of the measure [9].

At a national level, the database has been pivotal in making measurable improvements in the quality of data and in favouring reporting completeness of the national registry questionnaire, which is quite similar to that of the CF Foundation Patient Registry. The database developed in 2003 is regularly running in Belgian CF reference centres facilitating national registry of 865 CF patients [18]. It is now widely recognized that CF patient registry serves as a vital resource to access the

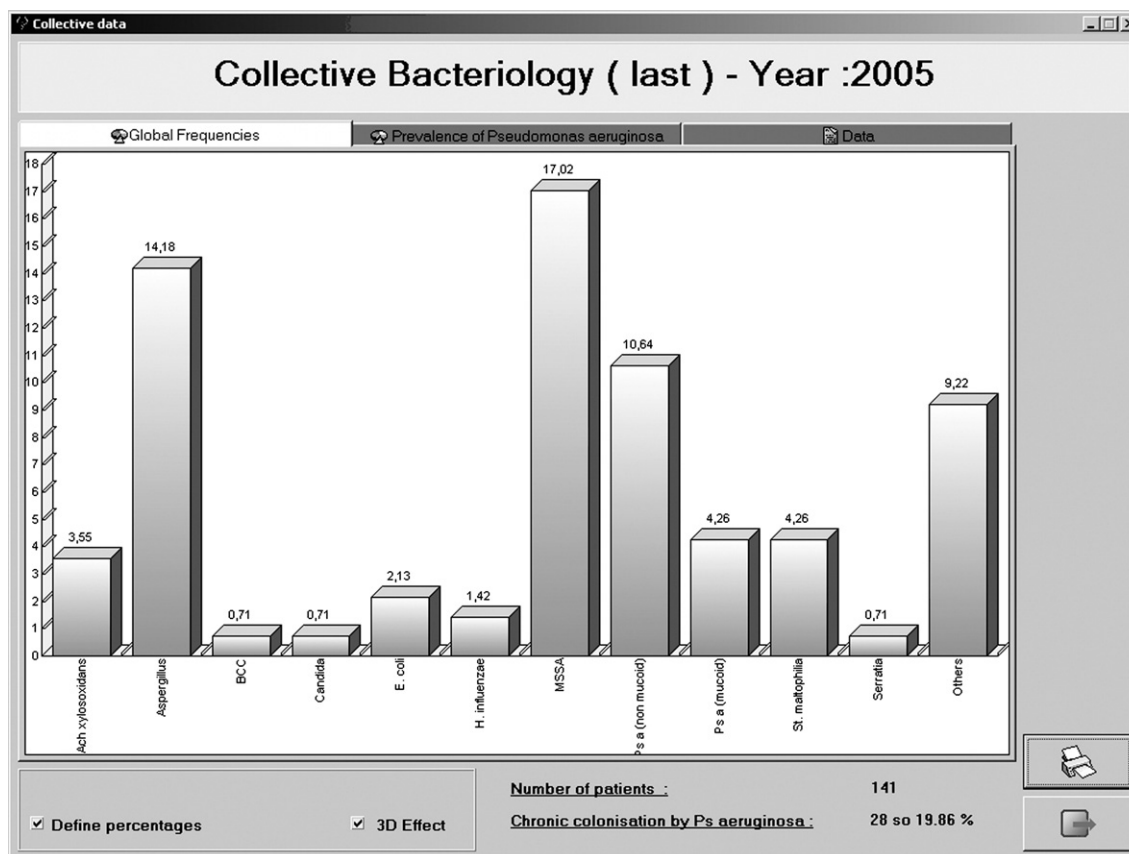


Fig. 5. Collective microbiological data of 141 CF patients attending our centre at the end of 2005. Y-axis indicates the number of infected patients for the listed pathogens. Corresponding percentages are indicated in the top of columns.

course of the disease and how the disease has changed over time [12]. Assessment of electronically available patient information is of paramount importance for improving data accuracy while reducing the needs for manual retrospective review of medical records for the national CF registry records. The overall trend will be that of the fields with missing values in the national registry questionnaire, which has yet to achieve 100% com-

pleteness, will disappear and all CF patients, remaining linked and confidential, will be a part of the national registry.

Finally, creating and having a database should not be an end goal but rather a source of valid data and a means for generating information by which to readily assess care process, performance, and outcomes quality. As CF is a genetic disease common across many nations that derive their population from Caucasians and as the software was designed to cope with a multi-system disease such as CF and to integrate information about various aspects of the disease, we agree the CF database we developed could be attractive to be used in other CF centres worldwide. The use of a specific standardised tool for data gathering could facilitate the identification of the influences of the environment and of differences in therapies on the progression of CF disease. Moreover, as proposed for standardisation of national registry reports [19,20], it could favour comparisons between countries of patient data [21].

In conclusion, our Belgium CF database provides an ongoing record of disease progression and a useful tool for gathering information on routine clinical and lab data, bacteriology, treatments, complications and specific outcomes for clinical routine and research purposes. The ongoing evolution of the database includes enhancements toward research data orientation including comparison of patient data between different centres and completeness of the National CF registry

Table 2

Main advantages of electronic version of CF database over paper-based files

For individual data

- Easy accessibility to administrative and medical data
- Standardization of data acquisition
- Rapid visualization of recent or historical clinical data
- Summary and graphical representation of longitudinal outcomes (spirometry, anthropometry, bacteriology)
- Easy application of Lee's classification [10]

For collective data

- Annual (or specified period) review of main patient parameters
- Warning list including patients requiring particular follow-up
- Possibility of exportation of data to research and analysis

Registry

- Automatic generation of majority of data
- Few risk of misencoding or losing data

questionnaire. A complimentary copy of the software can be provided free of charge to multidisciplinary accredited CF centres worldwide upon request.

References

- [1] Davis PB. Cystic fibrosis since 1938. *Am J Respir Crit Care Med* 2006;173:475–82.
- [2] Mahadeva R, Webb K, Westerbeek RC, Carroll NR, Dodd ME, Bilton D, et al. Clinical outcome in relation to care in centres specializing in cystic fibrosis: cross sectional study. *BMJ* 1998;316:1771–5.
- [3] Johnson C, Butler SM, Konstan MV, Morgan W, Wohl ME. Factors influencing outcomes in cystic fibrosis. A center based analysis. *Chest* 2003;123:20–7.
- [4] Kerem E, Conway S, Elborn S, Hejerman H. Consensus Committee. Standards of care for patients with cystic fibrosis: a European consensus. *J Cyst Fibros* 2005;4:7–26.
- [5] Reyckler G, Mailleux P, Gigi J, Leal T, Godding V, Lebecque P. A CF-specific database as a clinical and epidemiological tool. *J Cyst Fibros* 2004;3:S115.
- [6] Knudson R, Lebowitz M, Holberg C, Burrows B. Changes in the normal maximal expiratory flow-volume curve with growth and aging. *Am Rev Respir Dis* 1983;127:725–34.
- [7] Eigen H, Bieler H, Grant D, Christopher K, Terril D, Heilman D, et al. Spirometric pulmonary function in healthy preschool children. *Am J Respir Crit Care Med* 2001;163:619–23.
- [8] Cole TJ. The LMS method for constructing normalized growth standards. *Eur J Clin Nutr* 1990;44:45–60.
- [9] Poustie VJ, Watling RM, Ashby D, Smith RL. Reliability of percentage ideal weight for height. *Arch Dis Child* 2000;83:183–4.
- [10] Lee T, Brownlee K, Conway S, Denton M, Littlewood J. Evaluation of a new definition for chronic *Pseudomonas aeruginosa* infection in cystic fibrosis patients. *J Cyst Fibros* 2003;2:29–34.
- [11] Cystic Fibrosis Foundation, Patient Registry 2004 Annual Report, Bethesda, Maryland.
- [12] Mehta G, Sims EJ, Culross F, McCormick JD, Mehta A. Potential benefits of the UK cystic fibrosis database. *J R Soc Med* 2004;97 Suppl 44 :60–71.
- [13] Caldwell EJ, Dziodzio JT. CF-specific computerized clinical patient data management system. *Pediatr Pulmonol* 1993;S9:285–6.
- [14] Caldwell EJ, Dziodzio JT, Cairns AM. Team-oriented clinical management system for cystic fibrosis. *Pediatr Pulmonol* 1996; S13:345.
- [15] Caldwell EJ, Dziodzio JT, Cairns AM. Integrated cystic fibrosis clinical data management system for Microsoft® Windows. *Pediatr Pulmonol* 1998;S17:409.
- [16] Caldwell EJ, Dziodzio JT, Cairns AM, Zuckerman JB. Quality assurance in the clinical record: tracking clinical practice via an electronic medical record system. *Pediatr Pulmonol* 2000;S20:341–2.
- [17] Haan CK, Adams M, Cook R. Improving the quality of data in your database: lessons from a cardiovascular center. *Jt Comm Qual Saf* 2004;30:681–8.
- [18] Belgish Mucoviscidose Register-Registre Belge de la Mucoviscidose-1999, 2000, 2001, 2002. Annual Data Report, Belgium.
- [19] Sims E, McCorminck J, Green MW, Mehta G, Culross F, Mehta A. Call for standardisation of national registry reports — the problem outlined. *J Cyst Fibros* 2003;2:S383.
- [20] Sims EJ, McCorminck J, Green MW, Mehta G, Culross F, Mehta A. Call for standardisation of national CF database reports — an interim solution. *J Cyst Fibros* 2003;2:S384.
- [21] McCorminck J, Sims EJ, Green MW, Mehta G, Culross F, Mehta A. Comparative analysis of cystic fibrosis registry data from the UK with USA, France and Australia. *J Cyst Fibros* 2005;4:115–22.