Potential benefits of the UK Cystic Fibrosis Database

G Mehta MA MPhil E J Sims PhD F Culross BSc J D McCormick MRCPCH A Mehta FRCP (Edin) FRCPCH

J R Soc Med 2004;**97**(Suppl. 44):60–71

Cystic fibrosis (CF) is a complex condition involving genetic, nutritional, inflammatory, pulmonary, gastroenterological, psychological, renal, rheumatological and obstetric complications that arises in a quarter of the offspring from a population of around 2.4 million randomly distributed carriers in the UK. Over the last decade, a generic approach to track the progress of this chronic disease has been established by a multidisciplinary team of clinicians, paramedical staff, computer specialists and experienced clinical data entry personnel at the University of Dundee (see www.cystic-fibrosis.org.uk). The resultant UK CF Database (UKCFD) currently runs within specialist CF clinics in over 65 UK hospitals and aims to facilitate patient care, audit and research by providing customized software to each clinical team after appropriate training.

The hypothesis underpinning our approach was that robust software design[1,2] could standardize patient care while the resulting multicentre dataset would facilitate local, regional and national audit as well as cross-sectional and longitudinal research. As numbers of CF patients on the database have continued to increase with over 92% of the estimated 7500 patients[3] from all the specialist CF Centres across the UK now registered, these benefits to the CF community are beginning to accrue.

The software within the UKCFD produces graphical output, and an example showing the age profile of the CF population registered with the database by the end of 2002 is illustrated in Figure 1. Progress in CF care is illustrated by increasing numbers of adults and longevity into old age in some countries. When such survival is common-place, it is anticipated that there will be over 20 000 CF patients in the UK, approximately 3 times the current population.[3,4]

BACKGROUND

The UKCFD originated as an audit project in 1992–1994 to enumerate and characterize the Scottish CF population on an annual basis and to integrate the collection of this audit information into routine clinical practice. The pilot work was funded by the Clinical Resource and Audit Group (CRAG) of the Scottish National Health Service (NHS). A multidisciplinary group of health care professionals was formed that included representation from CF patients. The Scottish CF Specialists Group (SCFG) constituted the data steering committee. The SCFG Audit Centre was formed in Dundee and submitted the report 'Cystic Fibrosis in Scotland' in 1994 to the Scottish Office.

Following a critical evaluation of the CRAG report, it was decided the most cost-effective method for data collation required a standard paper-based clinic form coupled with centralized data entry in Dundee. Further support was provided by the National Services Division (NSD) of the Scottish NHS that funds small-volume, high-cost clinical services. A new clinically focused approach led to the appointment of a computer programmer to develop a database infrastructure that would encapsulate both patient demography (the Register function of the database) and disease severity. Once paediatric and adult CF specialists had identified the core data fields, the forms from the pilot were radically redesigned to ensure that data capture was both unambiguous and feasible within the short time available for a clinic consultation. The criterion applied was that form completion should take less than 10% of the available clinic time, with those fields that could be completed from the patient's notes being entered either prospectively or retrospectively, dependent on local staffing constraints within participating clinics. The NSD continues to support this Scottish initiative with an annual grant and today over 95% of Scottish CF patients are on the resulting database.

A second pilot, incorporating the lessons learned, was funded in 1995 by the CF Trust (www.cftrust.org.uk) and led to their longer-term support to implement a UK-wide CF database that would disseminate data capture and analysis software to all regional specialist centres and several peripheral hospitals with smaller numbers of patients. The new, millennium-compliant and Data Protection Act-compliant UKCFD design enabled data capture and analysis both locally in individual clinics and nationally for the entire UK population without the need for a centralized patient-specific PIN number allocation service (see Appendix 1 operational details).

The main functions of the data capture module are illustrated in Figure 2 and the critical operational aspects are now described.

UK Cystic Fibrosis Database, Division of Maternal and Child Health Sciences, Ninewells Hospital and Medical School, University of Dundee, Dundee DD1 9SY, Scotland, UK

Correspondence to: A Mehta FRCPH

E-mail: G.mehta@dundee.ac.uk

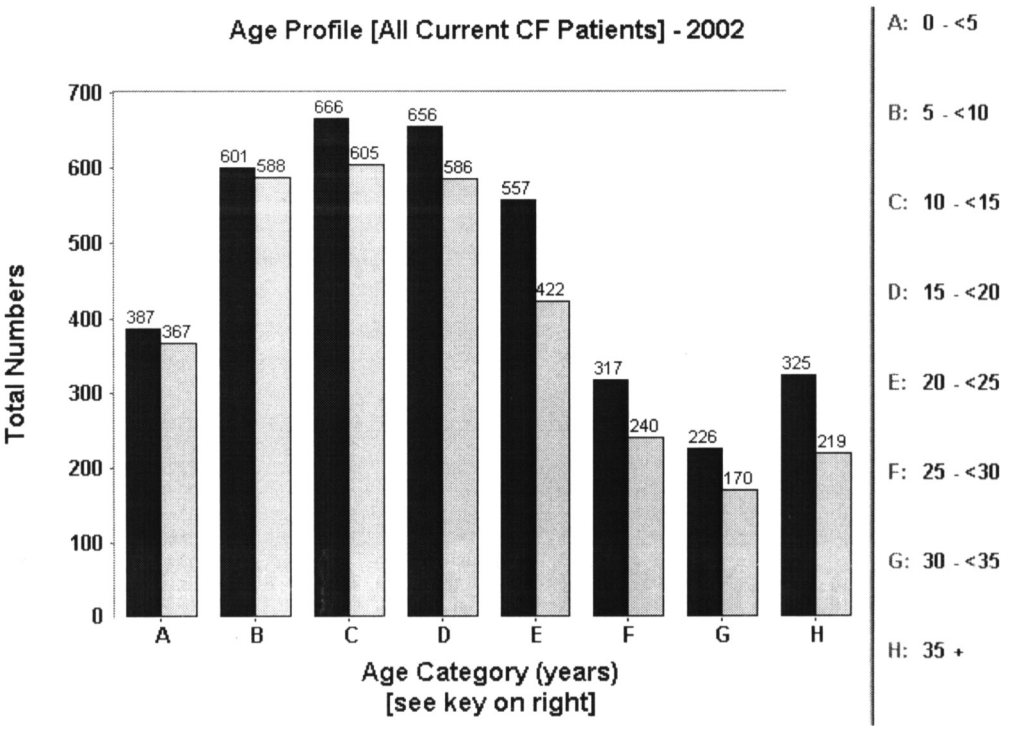


Figure 1 **Output from the software showing demographics of the UK cystic fibrosis (CF) population in the year 2002.** At present only 8% of the population are over 35 years of age (ordinate category H). The shortfall in the pre-school population (category A) results from late diagnosis but will largely disappear by 2009 when neonatal screening is universal in the UK (due to start in 2004). The male:female ratio increases with age (1.5:1 above the age of 30 years (Ref. 4). Black bar: Male; Grey bar: female

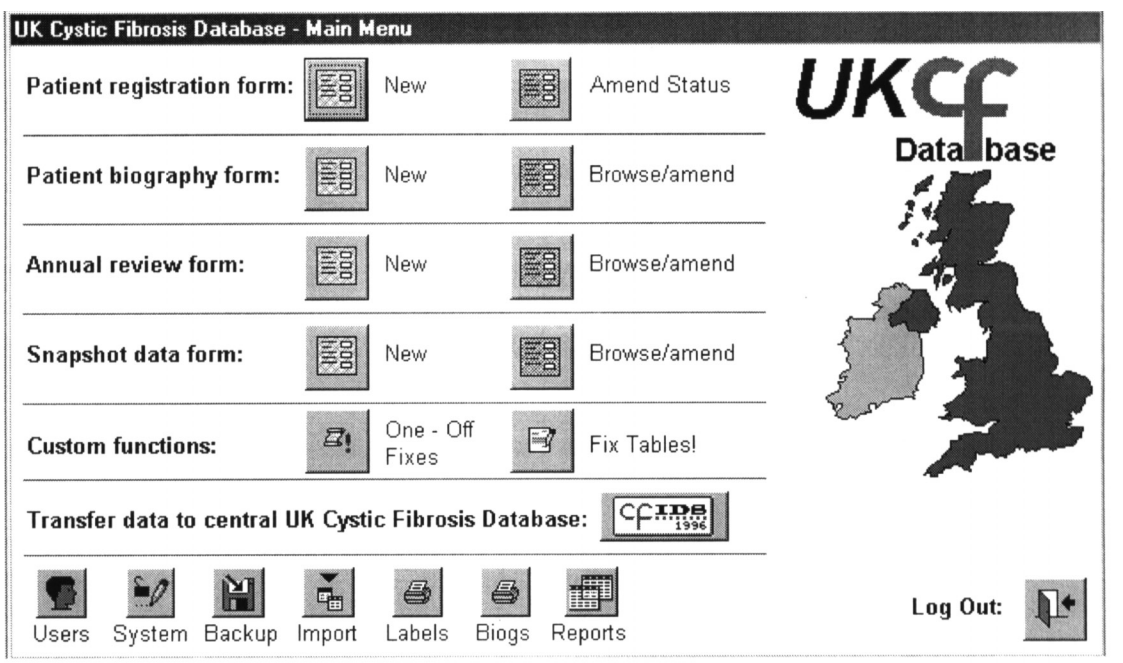


Figure 2 **Data capture occurs using the 'new' buttons and each entry can be amended.** Encrypted transfer of data occurs via the CF-IDS button and data are backed up locally using Backup. Customized reports and the means to generate patient labels are available. Password protection and system administration functionality are also provided. The software has a separate data analysis module to analyse the data. Appendix 1 shows the mode of database operation

The final flow of data at a patient level is described in Appendix 1.

OPERATIONAL STRENGTHS OF THE UK CF DATABASE

Compliance with the Data Protection Act

The approach outlined in Appendix 1 is fully compliant with the principles of the Data Protection Act and has been commended by the National Data Protection Compliance Manager (Health) as an example of excellent practice. Comprehensive information about the database together with a summary of its uses is provided to all patients (Appendix 2), with an opportunity to discuss any aspect of this further with a member of the clinical team. If a patient is willing to participate, explicit consent is recorded, otherwise a refusal of consent form is filed with the patient's details. In practice less than 1% of patients have refused consent.

Patient anonymity and the clinic-based PIN number generator

Patient anonymity is paramount within the structure of the database. This is achieved by the application of software to create a unique alpha-numeric PIN number from time-invariant patient-specific details noted by the clinic at registration (Appendix 1, stage 1). The number is generated by a 'hash algorithm' that scrambles details such as day and month of birth, gender, etc., using computer code based on large prime numbers. The PIN number identifies the patient within the database and remains with him or her for life. A great advantage of this technique is that the same number is generated in each clinic where the patient is seen without the need for an expensive central number-generating authority. Since the identical PIN number is generated should the patient transfer to another CF clinic, this feature facilitates anonymous longitudinal patient tracking. The mechanism also ensures that patient names are known only to their immediate carers, as normally a name and number are only coincident within a given clinic.

Choice of local versus national data entry

For ease of data capture in busy clinics, accuracy and rapid data entry, clinical information is recorded on paper forms and subsequently entered into the software resident on each clinic's computer. However, there is sufficient design flexibility to enable the Dundee team to perform data entry if required, e.g. due to a shortage of clinic resources, and the patient consent explicitly allows for this eventuality.

Data accuracy—a legal requirement of the Data Protection Act

Data accuracy and completeness is considered central to the database protocol and all data submitted for merging with the national UK dataset are cross-verified in Dundee against what has been recorded on the paper forms. Further range and accuracy checks are available as standard reports within each clinic. The latest version of the software being implemented provides extensive reporting of data completeness for key parameters. All queries or errors identified at the verification stage in Dundee are notified back to the clinics within 2 weeks of receipt. These must be addressed prior to the next data submission to Dundee.

Specialist CF clinics meeting these stringent standards are awarded funding from the CF Trust to complete data entry. Such clinics are given 'gold status', a quality standard that is fed back at the annual meeting of the Directors of specialist CF clinics. Clinics failing to meet the standard are asked to submit reasons for their failure and remedial training is undertaken where needed. Where clinics

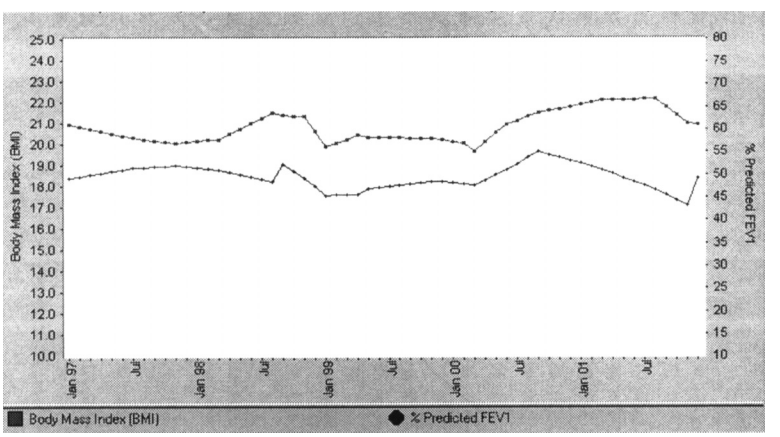


Figure 3 **Tracking patient progress in terms of body mass index (left) or lung function (right) was requested by users and is planned for the next version (3) of the database.** This feature is planned to increase the frequency of data entry because of feedback to clinic staff and the patient

Patient Summary Report

Patient ID: XXXX-XXXX-XXXX-XXXX-X **Name:** SMITH, A **DOB** 01 Apr 1983

Sex: Female

Visit Date:	Age:	(Centiles) Height:	Weight:	BMI:	(LF % Predicted) FEV1	FVC
06/08/97	14.3	67.7	0.5	0.0	28.8	31.1
15/10/97	14.5	73.6	1.9	0.0	30.2	36.7
19/11/97	14.6	72.2	1.6	0.0	28.5	31.8
21/01/98	14.8	75.9	0.6	0.0	16.9	15.0
18/03/98	14.9	77.8	0.1	0.0	22.7	26.3
06/08/98	15.3	74.2	0.2	0.0	18.3	24.0
15/10/98	15.5	65.3	0.0	0.0	18.8	30.3
18/11/98	15.6	64.3	0.0	0.0	16.2	24.0
07/01/99	15.8	57.6	0.0	0.0	18.4	25.0
18/02/99	15.8	64.8	0.0	0.0	19.3	26.2
01/04/99	16.0	68.6	0.0	0.0	21.0	29.8
17/06/99	16.2	61.9	0.0	0.0	17.1	21.7
15/07/99	16.3	67.7	0.0	0.0	15.9	20.8
16/09/99	16.4	61.0	0.0	0.0	16.5	22.3
04/11/99	16.6	66.5	0.0	0.0	15.0	21.7

Figure 4 **The new version of the UKCFD will have enhanced patient-specific reporting as requested by the database users' group and feedback from training days**

repeatedly fail to meet gold status, an accreditation visit is arranged with a report to local management highlighting areas of concern. Where such visits have taken place, increased local funding has been successfully negotiated. At the time of writing, all the specialist CF clinics in the UK are participating and over 80% have maintained gold status for more than 2 years.

IMPROVED PATIENT CARE

Needs of the individual

When a patient is seen in a specialist CF clinic, he or she has often travelled many miles, time is short and the needs of that individual often override the needs of groups of CF individuals with similar characteristics. Thus for those treating the patient, single patient 'care' is more valued than data capture to a national format that might ultimately benefit that individual's care via audit and research. To redress the balance, the software has the capability of producing individual patient reports (Figure 3) that can help the clinician to provide feedback to each patient about their growth, per cent predicted lung function, nutritional *Z*-score or infection status. This provides a further incentive to use the forms and clinics have requested more detail in such patient-specific reports. These are being programmed at present in the next version of the database (Figure 4). These reports can be printed out before each clinic and viewed by the patient alongside the graphical output.

Needs of the group

CF care is also about the needs of groups of individuals, often with different types of CF. The fields within the UKCFD can be used to construct severity scores and hence 'band' the patients into cost categories.[5] This is important for funding authorities and resource planning. For example, the database has been asked to provide anticipated numbers of new adults, advise on the siting of new CF clinics and calculate staff provision to cope with excessive patient loads. When calculating disease burden in terms of lung disease, Figure 5 shows the sequence of steps that enable clinic staff to analyse the prevalence of lung infection with a common CF pathogen using simple menu-driven analysis tools. These tools are integrated into the software and analyse the database in quasi-real time.

ACCESS TO DATA AND AUDIT OUTCOMES FROM THE UKCFD

Access to data is regulated by the Data Steering Committee of the CF Trust, which meets quarterly to consider requests submitted on a simple data request form that can be downloaded from the CF Trust web site [www.cftrust.org.uk]. At the time of writing, over 50 requests have been received for information from the UKCFD. These have come from a variety of disciplines and, for brevity, only key examples are summarized below.

Better paediatric CF care produces more adults,[3–5] and babies identified through newborn screening have the

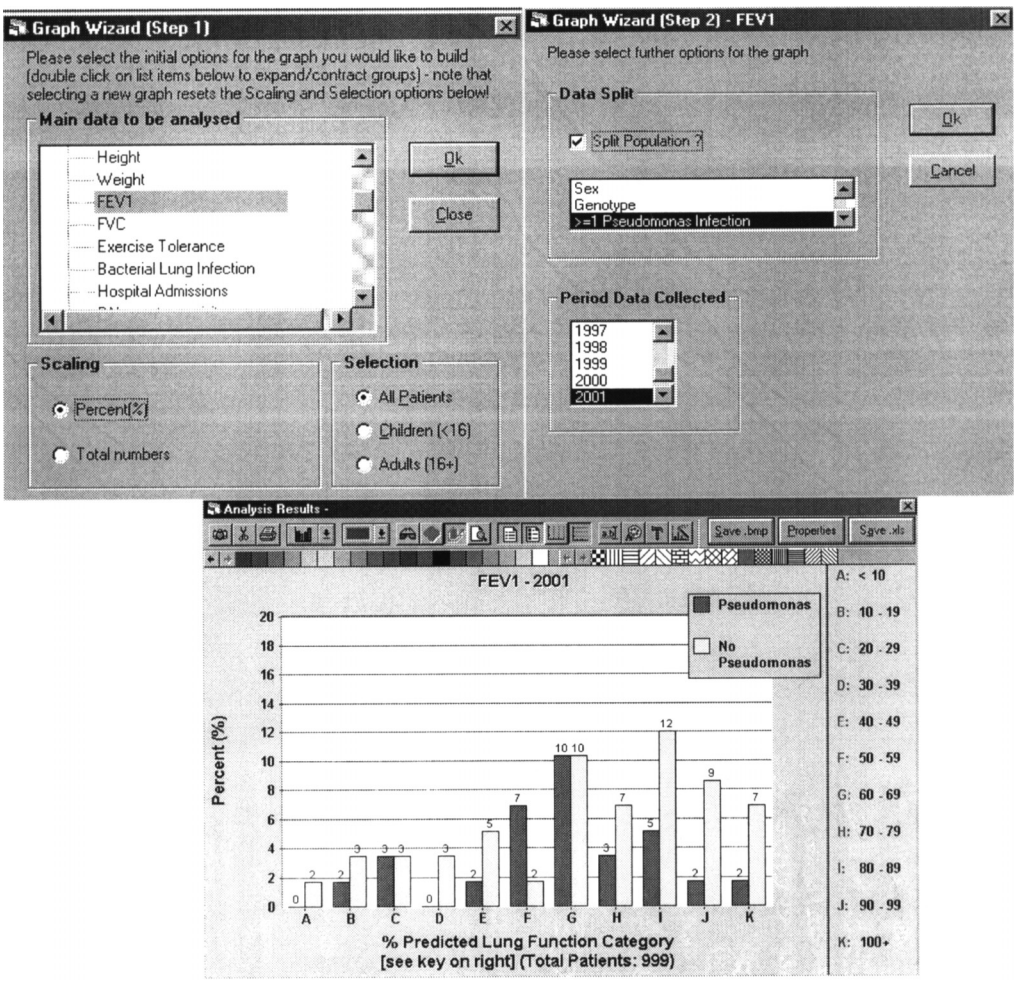


Figure 5 **The upper panel shows two sequential screens from the analysis tools.** This menu-driven interface permits the creation of audit reports for the most clinically relevant data in the database. The first stage (left) selects which graph is required (e.g. per cent predicted lung function) and whether analysis should be restricted to children, adults or all patients. Output is available showing either total numbers of patients or percentages. The second stage permits further refinement of the data presented (e.g. whether it is required to be split further by, for example, age groups), which year is required for analysis, etc. **The lower panel shows the computed output once all the selection criteria have been specified from the upper panels.** A graph is displayed with the facility to change its display and save either the graph or the underlying table of data to a Word document, PowerPoint presentation, Excel spreadsheet or bitmap file. As the same program can be run on any UKCF database within a given specialist CF clinic, it is a straightforward task to compare the performance of a clinic with the regional or national picture

opportunity to receive optimal nutrition[6] and treatment.[7] Both changes provide challenges to their respective types of healthcare. Such changes in demography between children and adults can be tracked regionally and nationally using the UKCFD.[4] The data can be used to ensure that appropriate numbers of adult physicians with CF training are in the correct geographical location. Equally, for the screened babies, the increased need for dietitians with training in neonatal and paediatric nutrition can be planned as neonatal screening becomes universal from 2004. Two examples of this use of the UKCFD are given below.

Figure 6 anticipates the resources needed to cope with approximately 250 children who are about to become young UK CF adults within the next 3 years. This is important information for the NHS as a whole because such patients can consume over £10 000 each in annual drug costs alone,[5] producing an annual bill in excess of £2.5 million that will fall on the adult medicine budget. In Scotland, where adult (but not paediatric) CF care is centrally funded by the NSD, the Scottish NHS already uses this information for budgetary purposes with quarterly reports from the audit centre in Dundee. Within Figure 6, further detail about the severity of CF for these 15-year-old patients in category D can also be calculated using the principles outlined in Figure 5. This analytical feature will be embedded as a customized report into the new version of the software once comprehensive banding criteria have been finalized.[5]

At a national level, the software has been pivotal in warning Directors of specialist CF clinics about potentially

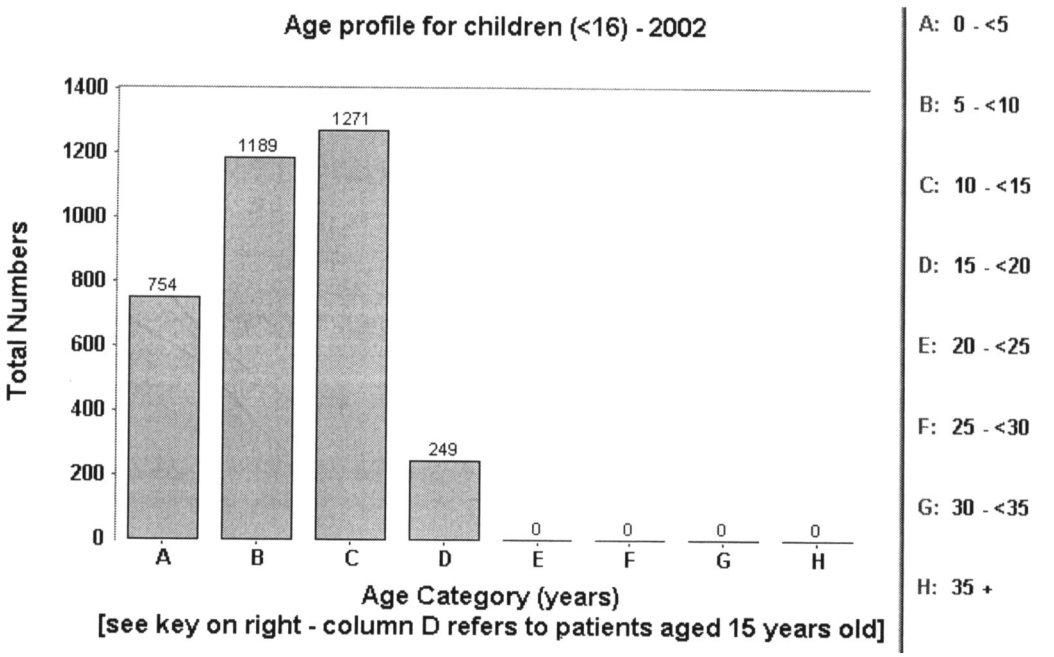
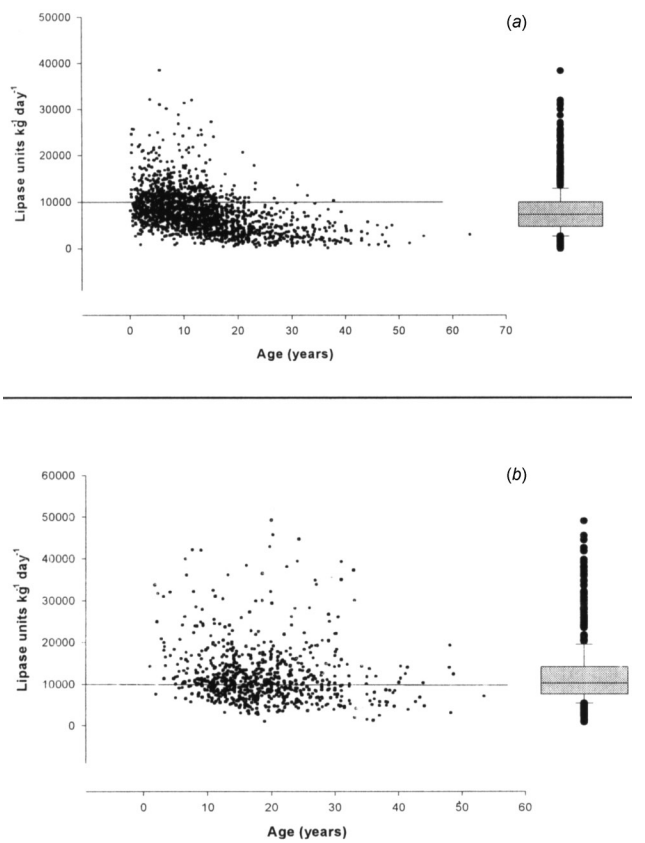


Figure 6 **The route to this screen is shown in Figure 5 via the main menu screen in step 1 of the analysis tools.** By restricting the analysis of demographics to children in step 1, category D now only contains patients aged 15 years who will form the emergent adult population within 3 years. This information facilitates resource planning. As shown in the lower panel in Figure 5 this population can be further subdivided by their infection status in order to permit calculations of banding and hence future costs associated with cystic fibrosis care (not shown)



toxic drug dosages.[8] Figure 7 shows the current profile of the doses of pancreatic enzyme replacement in UK CF patients following our initial publication.[8] The Committee on Safety of Medicines (CSM) has issued guidance that the dose of pancreatic enzyme replacement therapy (PERT) should not normally exceed 10 000 U of lipase per kilogram body weight per day. Figure 7 shows current practice from the UKCFD returns and reveals the degree to which the CSM guidance is routinely breached. Having found that PERT doses in excess of recommendation are commonplace, we are currently analysing possible causes and these data will be presented to UK CF Centre Directors at their annual meeting.

CF care is also variable between specialist CF clinics, and the Dundee team generates inter-clinic comparisons for circulation among the participating clinics (Box 1). An example presented to each CF Centre Director in 2003 is given in Figure 8, which shows the threefold variation between CF clinics in the prevalence of poor nutritional

Figure 7 **The latest update on a recent publication (Ref 8) on age dependence of the dose of standard/high-strength pancreatic enzyme replacement for the cystic fibrosis population.** The continuous line indicates the Committee on Safety of Medicines (CSM) advice on the upper limit above which it is thought that no further benefit accrues in terms of minimizing malabsorption. The box plots show that for (*a*) standard and (*b*) high-strength preparations about 75% and 50% of the population (respectively) receive doses above the CSM limit. The associated factors are being investigated

Box 1 **Examples of the UK Cystic Fibrosis Database benefitting individual CF centres**

CF centre 1
This is a large CF centre that needed to ascertain the number and frequency of intravenous antibiotic usage for local audit and research. Prior to the availability of the UKCFD, this would have necessitated a manual trawl of over 150 case notes. The UKCFD interrogation has reduced this exercise to just under 10 min

CF centre 2
This centre undertook an extensive study of exercise capacity.[10] These results helped validate the relationship between the UKCFD exercise tolerance score and maximal exercise capacity. This score can now be used prospectively to investigate exercise-related outcomes

CF centre 3
For many years, this CF centre has used the analysis tools supplied with the data collection software to generate their own annual report of outcomes for their management

CF centre 4
This CF centre uses the UKCFD to band their patients into different levels of severity as demanded by their management. These bands are then used to calculate levels of funding

CF centre 5
This CF centre is using their local values for pancreatic enzyme dosage to compare with those shown in Figure 7 because they currently have a policy that does not restrict enzyme dose

status in CF children. This information was fed back to each CF clinic Director in confidence as per the consent form (Appendix 2). Each CF Centre Director can determine their relative ranking. Such data will facilitate local/regional/national/international comparisons that stimulate both debate and research into the factors that might explain such variability. A similar approach is being adopted in the USA and it is of interest that similar variability is present across US CF clinics (see annual data reports from the US CF Foundation, www.cff.org). Such information needs to be used with care to avoid a 'league table' or blame culture. For example, if certain clinics have larger than expected numbers of patients from deprived backgrounds, then their population can be adjusted by postcode score to compensate for the high prevalence of poor nutrition. Equally, staffing issues might come into play because an audit of medical and paramedical CF staff within the larger clinics undertaken recently shows gross under-provision when compared with national guidance based on staff/patient ratios (data not shown).

RESEARCH INTO CF USING THE DATABASE

What is cystic fibrosis?

CF remains a clinical diagnosis but the constellation of symptoms and signs that constitute the diagnosis have come under serious challenge from the cloning of the CF gene.[11] Unfortunately, the potential advantages have blurred the edges of the diagnosis.[12–14] The key problem lies in the mismatch between those CF patients who do not have a common genetic form of CF—and have an apparently normal CF gene in initial screening for the commonest 30 mutations—yet have clinical CF. The problem with further understanding lies in the geographical spread of such patients. A large CF clinic with over 200 patients might only have two or three such patients. Many

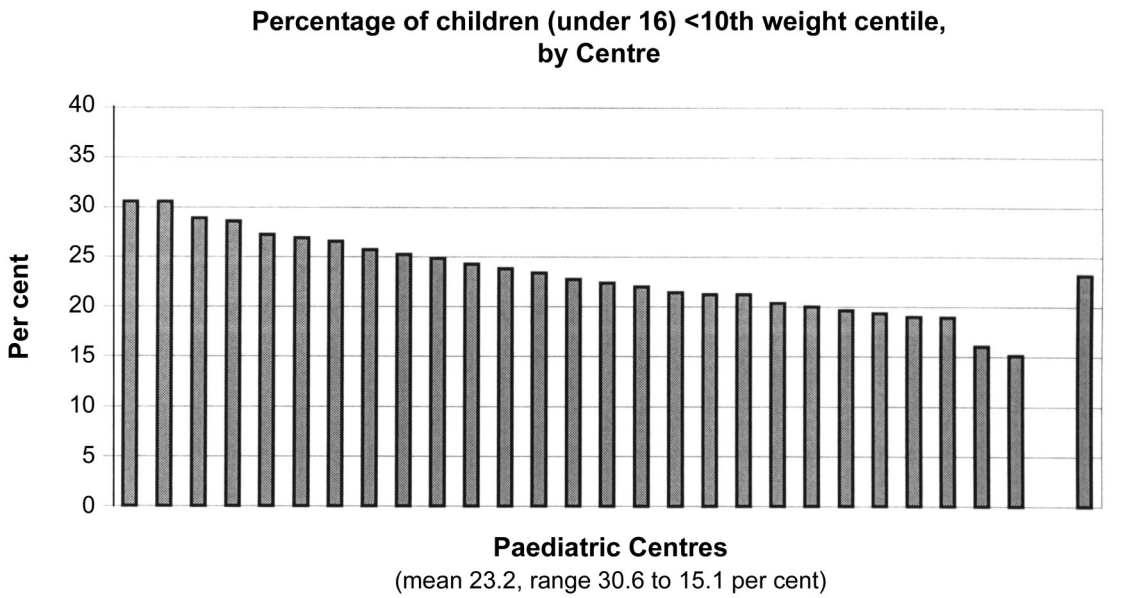


Figure 8 **Better nutrition is thought to improve survival (Ref 9) and yet there is a considerable variation in outcomes between paediatric cystic fibrosis centres, with those on the left having twice the expected number of patients with poor weight compared with the best-performing centres on the right.** The bar at the far right indicates averaged national performance. The crude data can be further refined for case mix and social deprivation using postcodes as surrogate markers

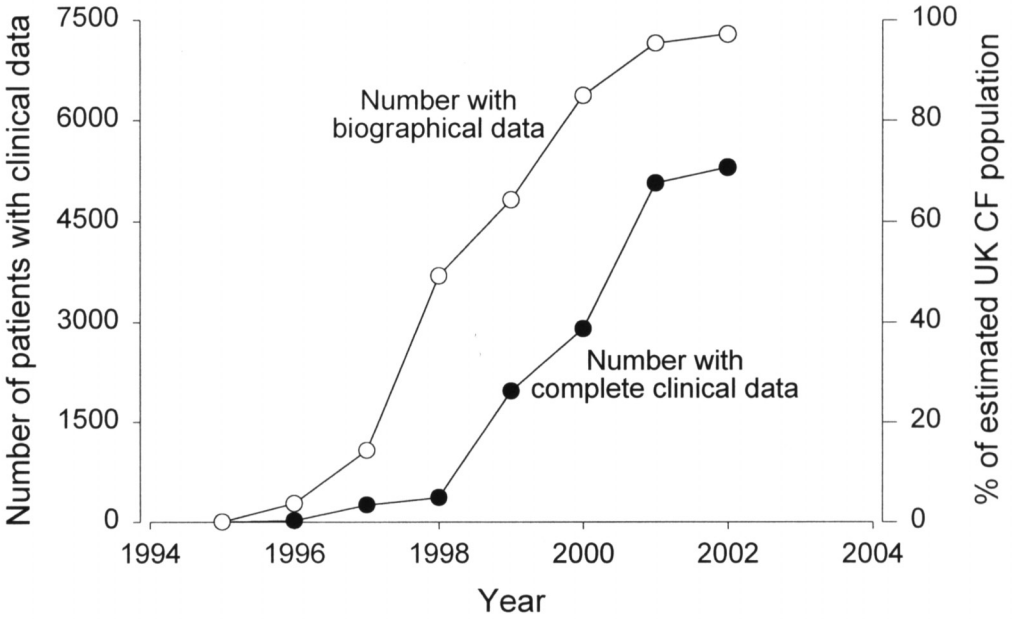


Figure 9 **Ontogeny of the UKCFD showing the rise in registered cystic fibrosis (CF) patients (open circles) following the Scottish pilot scheme in 1995.** The number of these patients has almost reached the expected number of UK CF patients projected from the original register developed by Dodge *et al*. (Ref 3). The registration process generated an anonymous patient number from the biographical data in each patient. This number can follow the patient for life, even if they move to a new CF clinic, provided the UKCFD software is used to re-register the patient. The number of patients with complete clinical data (filled circles) is increasing but the gap between the two populations will only be closed by a targeted approach to poorly staffed clinics outside the larger CF centres

clinics have none at all, but the UKCFD can amalgamate these patients thus providing a research platform.

MREC submissions are under way to analyse the CF gene in detail in such patients (estimated to be around 150 in total in the UK). This project will define whether these patients have some other form of CF, perhaps unrelated to the gene thought to cause the common form of the disease. Alternatively, they might have very rare defects in the CF gene that cannot be detected by routine methods (e.g. intron disease). Whatever the outcome, such patients will be very informative in further defining genotype–phenotype interactions using the UKCFD as the vehicle to undertake research.

Post-neonatal screening follow-up

The data contained within the UKCFD were central in the submission to the UK Government to introduce screening for all newborn infants. How should such infants be followed up? A screening data form is being piloted in Scotland with data returns to the Dundee centre. This form has attempted to track the 13 babies identified to date since the start of screening in Scotland in February 2003. This number of CF babies is almost exactly equal to the expected number from *a priori* calculations. Where CF clinics fail to submit data returns for their screened babies, the gaps in their service will be investigated by the appointment of an audit facilitator. Thus the Scottish NHS accrues benefit from the audit function of the database and is able to raise standards by identification of gaps in service provision.

Agreement has been reached with the NHS in Scotland that the data recorded on the form following a positive diagnosis will be incorporated into the electronic birth record as an episode of care. It is anticipated that those CF carriers detected as a by-product of screening because they happened to have a falsely positive high immuno-reactive trypsinogen (IRT) level will also be recorded as an episode of care. False-positive raised IRT values are a common occurrence due to either faecal contamination of the blood spot card or ischaemia/hypoxia of the pancrease at birth unrelated to CF. These unintentionally diagnosed CF carrier babies are currently not on the UKCFD but some of these will present as variant CF in future years.[13,14] The Scottish birth record will incorporate carrier status as an episode of care so that these babies can be tracked in future years. It is hoped to extend this principle to the rest of the UK via the new National Newborn Screening Programme Centre based at Great Ormond Street Hospital, London.

The first aim for the new data collection form is to create a post-screening CF-positive dataset to facilitate research, and the CF Trust has recently asked Directors of CF clinics to submit proposals for clinical trial research on such babies. The UKCFD infrastructure will facilitate this research.

Complications and the paucity of clinical trials in CF

CF has had few clinical trials compared with some other diseases. The UKCFD is assisting current clinical trials by assigning controls for clustered case–control studies and has also provided information for randomized trials by calculating anticipated patient numbers.

With respect to patient mortality, MREC permission has been granted to compare deaths reported to the Offices for National Statistics (ONS) throughout the UK against UKCFD records (MREC/03/0/02). This information has already identified ONS coding errors in the national statistics relating to CF and these have been fed back to the ONS. Thus the multiple source ascertainment function of the database will increase the accuracy of research into survival.[3,4,15]

PROBLEMS WITH THE UKCFD

While all CF centres are now participating in the UKCFD, over 92% of the estimated UK CF patient population of 7500 have registered biographical information (Figure 9). At present clinical data are only available for approximately 70% of this population. This coverage compares favourably with other national databases (USA, France, Australia, Germany all have a broadly similar degree of coverage for their clinical data). Following pilot work,[2] UKCFD staff are working closely with the Cystic Fibrosis Society for Europe (ECFS) to standardize a minimum dataset that could form the basis for a Europe-wide repository of CF data. Such a Registry is the only means to investigate very rare forms of CF. However, no funding has yet been identified for this venture.

The discrepancy between the numbers registered with the UKCFD and those with complete clinical data results from a variety of factors. These include the presence of shared-care patients who often live far away from the CF centre and are seen elsewhere. A number of these, particularly in the case of paediatric patients, have yet to be registered by the peripheral CF clinic. Even for those registered, CF centres sometimes find it difficult to obtain all the clinical data required on the database forms, given the constraints of geography and their local shared-care arrangements. Whilst most have registered all their patients, two centres have only recently overcome local funding issues to collect their data and continue to add new patients during 2003. In a small minority of CF centres, some patients are seen in clinic but the database forms are not completed due to staffing problems. A few patients fail to attend and are included only in the registered population. There are a number of patients who are likely never to be on the database because they are only seen at District General Hospitals and never at a specialist CF centre. These patients will only enter the database when the local clinic size warrants the appointment of clinicians with an appropriate interest or if they choose to travel to a larger CF centre. Information on these patients is therefore not available and is a matter of some concern on clinical governance grounds alone. Finally, less than 1% of patients have withheld consent for their details to be included on the database.

At present the UKCFD does not provide enough information to comprehensively record and track newly screened CF babies. Work is under way to correct this deficiency drawing on the experience of Scotland where screening commenced in 2003. The necessary resource to implement the work has not been identified to date.

Data quality has to be continuously monitored because of staff turnover, which necessitates re-training of clinic personnel. All clinics are given a quality mark and at present all participating clinics are providing gold-quality data. Training is provided to new staff in 10–15 clinics per annum as staff turnover occurs. Extra training is also needed as users become more sophisticated and require more detailed feedback from the database.

Changes to the forms used to collect the data are being evaluated as users request more sophisticated analyses and also provide feedback on the difficulties in completing problem fields. However, as the quantity of information to be recorded and hence the time taken to do so increases, any proposed changes must ensure that data quality is not compromised. These proposals are fed back to the CF centres annually after careful evaluation by a pilot user group. Paramount in this discussion is the consideration that as data density increases, data quality has the propensity to decline exponentially.

THE FUTURE

To date, the UKCFD has concentrated on CF patients cared for in large CF clinics, but increasingly, smaller emerging CF clinics are wishing to participate as their patient numbers increase. These clinics are often poorly resourced and have the additional problem that they often share patient care with one or more of the larger CF clinics. The new version of the software is currently being modified to cope with this trend and the UKCFD is now a core project within the CF Trust that guarantees funding for at least 3 further years.

The UKCFD has attracted international attention and we have proposed standards for CF data collection to learned CF societies (EU, Canada, Australia, South Africa, etc). These standards are not currently agreed but are needed to make valid comparisons between countries.[15] Such comparisons are important because CF is an example of a genetic mutation that is common across many nations

that derive their populations from Europe. However, should clinical outcomes be truly different, then CF will illuminate the interaction between genotype and environment to yield clues about the causes of the variability of the resultant phenotype. Standards for data gathering are crucial for this understanding.

Finally, the ability of the UKCFD to cope with a multisystem disease suh as CF raises the prospect of the application of the software to other chronic diseases. It remains to be seen whether we have created a generic solution for patient care, audit and research outside CF and many other groups have expressed an interest in our approach. The key limitation to the spread of our infrastructure to other diseases is the lack of ring-fenced funding to design data collection to suit other specific needs and to train staff within the clinics caring for those patients. Few of these disease groupings have the combined financial strength within their equivalent patient organizations (the CF Trust in our case) and the NSD that acts as the 'top-slicing' authority for rare conditions within the Scottish NHS. We believe that such a partnership is essential to ensure continuity and clarity of purpose. We commend our clinician-led approach to other disciplines and look forward to meeting the challenges of CF in future years.

Acknowledgments: This project was initiated by a grant from the Clinical Resource and Audit Group of the Scottish NHS and has been generously supported by the Cystic Fibrosis Trust and the National Services Division of the NHS in Scotland for many years. The expert data validation was undertaken by Margaret Fraser, Kath Nichol and Sheila Krawczyk. Many clinicians, programmers, students and analysts have worked on the database since its inception and we gratefully acknowledge their contribution. There are no conflicts of interest and we dedicate this article to those whose lives have been terminated prematurely from the complications of this disease.

APPENDIX 1

UK Cystic Fibrosis Database operation

The UK CF Database was designed to collate information about various aspects of the disease from CF clinics all over the country. The forms used to collect this information can be viewed at www.cystic-fibrosis.org.uk. A summary of a typical patient journey through the database is shown below:

(1) Patient registration

- Clinics complete a Patient Registration Form for each patient consisting of time-invariant information that is used to generate a unique PIN number. This identifier is used henceforth to 'tag' all patient records. This number cannot be reverse-engineered
- Clinics complete a Patient Biography Form for each patient
- The Registration and Biography Forms are sent to the national team in Dundee where the patients are registered into the database as a one-off exercise, a patient number generated and a copy of the software configured for the clinic. Once the software is released to the clinic, all new patients are registered by the local clinic. The requirement for this procedure is explained in the patient consent form.

(2) Gathering clinical information

- When patients are seen in clinic, Clinic Vist/Annual Review forms are completed, tagged by the patient's PIN number as generated in stage 1 above
- These details are entered by clinic staff into their local computer system
- Periodically, data captured via the Clinic Visit, Annual Review and Biography forms, tagged by the patient number alone, is encrypted and forwarded on zip disk to Dundee for merging with the national database.

(3) The two levels of operation of the UK CF database

- The local clinic database: runs on the clinic's computer. Clinic staff are able to generate their own clinic audit reports as well as track the progress of any named patient (e.g. medication, lung function)
- The national database: consists of data merged from several clinics and used for regional and national audit and research. Patients cannot be identified by name. As trends in the disease will be tracked over time, longitudinal data are held
- Use of data in the national database is subject to approval from the CF Trust's Data Steering Committee. A short form is available from the CF Trust web site for those requesting access to the data [www.cftrust.org.uk].

(4) Support for the database from the national team

- The Dundee team provides technical, data entry and software support for all the clinics sending data for merging with the national dataset.
- This team has access to patient data only on a need-to-know basis, for instance in the following situations:

When the clinic's database needs to be reconfigured, for instance if it requires upgrading or gets corrupted or in the event of a computer malfunction

When clinic data entry falls behind tight deadlines and the national team gets them up-to-date

A few clinics do not have the resources for any data entry and send forms to the national team for entering into the system (by prior arrangement only)

Paper forms identifying patients are kept securely in a locked room in Dundee and used occasionally either when patients are transferred to another clinic or when the accuracy of patient details needs to be double-checked.

(5) Data output

- Annual reports containing national audit reports are sent to each participating centre
- Annual reports are created for each clinic to reference their own clinical data against the national picture
- Clinic comparison reports are produced for key outcomes, e.g. poor lung function or nutritional status
- Clinics are provided locally with analysis software to track the progress of their own individual patients.

APPENDIX 2

Patient consent information sheet

The UK Cystic Fibrosis Database was started 1995 by the Cystic Fibrosis Trust to record the number and location, the state of health and the treatment of the people in the UK who have cystic fibrosis. The information recorded is similar to that usually recorded at the CF clinic, e.g. weights, height, respiratory cultures, lung function tests, X-ray results and a few other investigations.

All the information stored in the national UK CF Database is anonymous and confidential, as no patient names are recorded. Each patient is issued with a code number under which information is sent to the UK CF Database. The patients' names are known only to their local clinic as the data are entered locally. Sometimes, a trusted third party may also perform this data entry (e.g. should clinic resources be unavailable).

The database will have many uses, some of which are as follows:

- To follow your progress and ensure you are receiving the best treatment. The data will be available for the staff in your local clinic and the staff will have your results available in an easily accessible form in tables and graphs
- To measure the performance of your particular CF clinic when compared with the results of the UK as a whole
- To measure the performance of all clinics to obtain a national or regional picture and compare with other countries or regions
- To provide information for national planning of future services required for people with CF
- To ensure that the NHS puts a fair share of resources into your clinic
- To help judge which types of treatment are of greater benefit, and how overall care is improving
- To conduct research on the data to identify new trends, e.g. an increase in a particular complication or new infection
- To co-operate with CF databases in other countries with a view to gaining a better understanding of the condition and its treatment
- To identify particular patient groups who may be suitable and could be approached to take part in particular research studies. Contact in this respect would only come through your local CF clinic.

The use of any information from the UK Database will require the approval of a Database Steering Committee made up of representatives from the CF Trust, CF Clinic Directors, UKCF Database staff and a patient representative.

Your participation in the database is entirely voluntary, and you have the right to withdraw at any time without stating a reason.

The Data Controller is The Cystic Fibrosis Trust. Our nominated representative to deal with data privacy issues is the Chairman of the Database Steering Committee, should you have any queries about how your information is being used or should you wish your information to be removed from our records. The Chairman of the Database Steering Committee can be contacted at the CF Trust, 11 London Road, Bromley, Kent BR1 1BY.

REFERENCES

1 Mehta A. Data standards in cystic fibrosis workshop. *Isr J Med Sci* 1996;**32**:S177–9

2 Duncan LS, Muhsin F, Mehta G, De Winton C, Mehta A. User friendly data analysis tools linked to a Microsoft Access database. In: Proc. Med Records Inst. Towards an Electronic Health Record. TEPR 1998;**1**:76–83 (ISBN 0-9640667-4-2)

3 Dodge JA, Morison S, Lewis PA, *et al.* Incidence, population, and survival of cystic fibrosis in the UK, 1968–95. *Arch Dis Child* 1997;**77**:493–6

4 McCormick J, Green MW, Mehta G, Culross F, Mehta A. Demographics of the UK cystic fibrosis population: implications for the national screening programme. *Eur J Hum Genet* 2002;**10**:583–90

5 Robson M, Abbott J, Webb K, Dodd M, Walsworth-Bell J. A cost description of an adult cystic fibrosis unit and cost analyses of different categories of patients. *Thorax* 1992;**47**:684–9

6 Farrell PM, Kosorok MR, Rock MJ, *et al.* Early diagnosis of cystic fibrosis through neonatal screening prevents severe malnutrition and improves long-term growth. *Pediatrics* 2001;**107**:1–13

7 Waters DL, Wilcken B, Irwig L, *et al.* Clinical outcomes of newborn screening for cystic fibrosis. *Arch Dis Child Fetal Neonatal Ed* 1999;**80**: F1–7

8 Mehta A. Further correspondence on fibrosing colonopathy. *Lancet* 2001;**358**:1546–7

9 Corey M, McLaughlin PJ, Williams M, Levison H. A comparison of survival, growth and pulmonary function in patients with cystic fibrosis in Boston and Toronto. *J Clin Epidemiol* 1988;**41**:583–91

10 Bright-Thomas RJ, McVean R, Webb AK, *et al*. Value of CF Database exercise tolerance score to assess fitness level in adults with cystic fibrosis. *Pediatr Pulmonol* 2003;**25**:317–18

11 Ratjen F, Doring G. Cystic fibrosis. *Lancet* 2003;**361**:681–9

12 Groman JD, Meyer ME, Wilmott RW, Zeitlin PL, Cutting GR. Variant cystic fibrosis phenotypes in the absence of CFTR mutations. *N Engl J Med* 2002;**347**:401–7

13 Mekus F, Ballmann M, Bronsveld I, *et al*. Cystic fibrosis-like disease unrelated to the cystic fibrosis transmembrane conductance regulator. *Hum Genet* 1998;**102**:582–6

14 Rosenstein BJ. Nonclassic cystic fibrosis: a clinical conundrum. *Pediatr Pulmonol* 2003;**36**:10–12

15 Fogarty A, Hubbard R, Britton J. International comparison of median age at death from cystic fibrosis. *Chest* 2000;**117**:1656–60