Respiratory Medicine (2007) **101**, 254–260

respiratoryMEDICINE

Assessing the usefulness of outcomes measured in a cystic fibrosis treatment trial

Ranjan Suri[a,*], Chris Metcalfe[b], Colin Wallis[a], Andrew Bush[c]

[a]*Respiratory Unit, Great Ormond Street Hospital for Children NHS Trust, Great Ormond Street, London WC1N 3JH, UK*
[b]*MRC Biostatistics Unit, Institute of Public Health, University Forvie Site, Robinson Way, Cambridge CB2 2SR, UK*
[c]*Department of Respiratory Paediatrics, Royal Brompton and Harefield NHS Trust, Sydney Street, London SW3 6NP, UK*

Received 13 October 2005; accepted 16 May 2006

KEYWORDS
Cystic fibrosis;
Clinical trial;
Outcome measures

Summary Forced expiratory volume in 1 s (FEV_1) is the usual primary outcome variable in clinical trials in cystic fibrosis (CF). Usually, several secondary outcomes are also measured. We assessed which secondary outcomes are likely to give additional clinically useful information about treatment effects, in order to inform the design of future studies.

The study was performed as part of a trial comparing daily rhDNase with alternate day rhDNase and hypertonic saline in CF. The primary outcome was FEV_1. Secondary outcomes were forced vital capacity (FVC), forced expiratory flow at 25–75% of forced vital capacity (FEF_{25-75}), number of pulmonary exacerbations, weight gain, quality of life (QOL), and exercise tolerance. The usefulness of each secondary outcome was investigated by assessing if the change in that outcome over the treatment period could be predicted from the primary outcome.

Change in FEV_1 correlated with changes in FVC ($r^2 = 0.76$, $P = 0.001$), FEF_{25-75} ($r^2 = 0.64$, $P = 0.001$), weight ($r^2 = 0.08$, $P = 0.001$), and change in oxygen saturation with exercise ($r^2 = 0.08$, $P = 0.001$). However, it did not correlate with changes in visual analogue score (VAS) with exercise, QOL, nor with the occurrence of pulmonary exacerbations.

Only the outcomes QOL and VAS with exercise actually provided additional information to FEV_1 in this study.

*Corresponding author. Tel.: +44 207 405 9200x5453; fax: +44 207 813 8514.
E-mail address: surir@gosh.nhs.uk (R. Suri).

0954-6111/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.rmed.2006.05.012

Introduction

In cystic fibrosis (CF), most of the morbidity and mortality is from pulmonary disease, which is characterised by obstruction of the airways by thick tenacious secretions that are difficult to clear.[1] Treatment aimed at controlling airway infection, reducing airway obstruction and improving nutritional status has been the cornerstone of the successful management of CF. Recently, anti-inflammatory therapy has been tested. Beyond these basic strategies, symptomatic treatment of the complications of the lung disease is also essential.

When considering any new therapeutic intervention, an additional benefit over the current standard therapy or placebo needs to be demonstrated. For individuals with CF, this benefit ideally would be increased survival time. However, as the estimated median predicted life span of individuals with CF approaches 40 years,[2] survival time has become an impractical measure of clinical efficacy. Surrogate markers of increased survival, such as improved lung function, need to be assessed for the proposed intervention.[3]

Most phase III clinical trials in CF[4–6] have used change in forced expiratory volume in 1 s (FEV_1) as the primary outcome, because it is easy to measure, reproducible and has a relationship with mortality in patients with CF.[7–9] Several secondary outcomes, which are based on the CF Foundation Consensus Conference recommendations,[3] are often measured as well. Time to first exacerbation has also become a popular endpoint[10] but there is no uniform agreement as to what constitutes an exacerbation[11]; many physicians now treat early without waiting for formal criteria to be fulfilled; and many patients will not exacerbate during the time period of the trial, meaning that they will not contribute to determination of efficacy. Measuring many outcomes may not merely prolong the trial visit unduly, thus likely reducing patient co-operation, but also lead to the possibility of false positive findings, unless appropriate statistical corrections are made. For example, by definition, one in 20 outcomes would be expected to be outside the 95% confidence intervals, and might be considered 'abnormal'. Hence, only those secondary outcomes which give information over and above that of the primary outcome, and where observation of change on the outcome is feasible within the trial follow up, should be measured.

The aim of this report was to assess which secondary outcomes gave clinically useful information about treatment effects in a trial of rhDNase in CF, additional to that obtained from the primary outcome (FEV_1), in order to inform the design of future treatment trials. A secondary outcome measure was considered clinically useful if a treatment comparison based on that measure could not be predicted from the same comparison based upon FEV_1 measurements; that is, different conclusions would be reached as to treatment efficacy based on FEV_1 and the secondary outcome measure under consideration.

Methods

This study was conducted within a prospective open, randomised cross-over trial comparing daily rhDNase with alternate day rhDNase and hypertonic saline (HS) in children with CF, that has been reported in detail elsewhere.[12] The study was approved by the ethics committees of both institutions. Each patient was allocated to receive, in random order, consecutive 12 week treatments of once-daily 2.5 mg rhDNase, alternate day 2.5 mg rhDNase and twice-daily 5 mls of 7% HS. There was a 2-week washout period between treatments.

The primary clinical outcome was change in FEV_1. Secondary outcomes that were measured at the beginning and end of each treatment period were forced vital capacity (FVC), forced expiratory flow between 25% and 75% of forced vital capacity ($FEF_{25–75}$), weight gain, exercise tolerance, quality of life (QOL), and the number of pulmonary exacerbations. Subjects were blinded to the results of objective outcome measures (such as FEV_1) when completing subjective outcome measures (such as the QOL questionnaire).

Lung function

At each visit, lung function was assessed by standard spirometry using a compact spirometer (Vitalograph, UK).[13] FEV_1, FVC and $FEF_{25–75}$ were measured in accordance with the American Thoracic Society guidelines.[14]

Exercise tolerance

The ''3-min step test'' was performed at each study visit.[15] The patients stepped up and down a single 15 cm step at 30 steps per minute for 3 min (regulated by a metronome). Oxygen saturation (SaO_2) was recorded continually during the exercise test using a Biox 3700 pulse oximeter (Ohmeda, USA). Change in SaO_2 was calculated as the lowest percentage oxygen saturation during exercise minus the pre-exercise percentage oxygen saturation.

A visual analogue score (VAS) for dyspnoea was recorded before and after the exercise.[16] This consisted of a 10 cm horizontal line with two anchor points, one at each end. On the left (zero) it was labelled, ''not at all short of breath'', while at the other end (10 cm) it was labelled ''the most breathless I have ever felt''. Patients put a mark through the line where they thought their breathlessness fitted on this scale, which was then measured (in cm) from the zero point. Change was calculated as post-exercise rating minus pre-exercise rating, positive changes indicating an increase in breathlessness.[17]

Quality of life

The Quality of Well-Being Scale self-administered form 1·04 (QWB-SA UCSD Health Outcomes Assessment Program) was used to assess QOL.[18] The self-administered form is a questionnaire that was filled out by the patient and their carer together at each study visit.

Pulmonary exacerbation

No universal definition exists for a pulmonary exacerbation so a previously outlined protocol for respiratory tract infections was used.[19] A pulmonary exacerbation was said to have occurred when a patient was treated with parenteral antibiotics for any four of the following 12 signs and symptoms: change in sputum; new or increased haemoptysis; increased cough; increased dyspnoea; malaise, fatigue, or lethargy; temperature above 38 °C; anorexia or weight loss; sinus pain or tenderness; change in sinus discharge; change in physical examination of the chest; decrease in pulmonary function by 10% or more from a previously recorded value; or radiographic changes indicative of pulmonary infection.

Statistical analysis

The sample size for the trial was calculated on the basis of change in FEV_1 (primary clinical outcome).[12] The analysis focused on two separate pair wise comparisons of the treatments: daily rhDNase vs. HS, and daily vs. alternate day rhDNase, on the basis of within-subject differences by analysis of covariance. Between treatment comparisons of the number of pulmonary exacerbations were made using McNemar's test of paired proportions. Adjustment for additional covariates (treatment period, quarterly season of the year and the child's age at the beginning of the period) was undertaken using multiple regression. A marginal model was used, based on general estimating equations with robust standard errors, which allows for the correlation between observations taken on the same subject over time.[20]

Comparison of outcomes

Using data from all treatment periods together, Pearson's correlation coefficient was used to compare percentage change in FEV_1 with percentage change in FVC and FEF_{25-75}, and absolute change in measures of exercise tolerance, weight, QOL, and pulmonary exacerbations. Percentage change was calculated as the difference between the post- and pre-treatment measures, divided by the pre-treatment measure. To accommodate within-patient correlation between the two or three treatment periods a patient contributed to, bootstrap techniques were employed in the calculation of *P*-values and 95% confidence intervals for the correlation coefficient.[21]

Results

Forty-eight children were randomised, eight to each of the six possible treatment orders. Eight children were unable to complete all three treatment periods.[12] Consequently 43 children were included in the comparison of daily and alternate day rhDNase, and 40 children in the comparison of daily rhDNase and HS.

Effectiveness

Comparing mean FEV_1 (primary outcome) between the treatments, there was an 8% (95% CI: 2%–14%, $P = 0.01$) advantage for daily rhDNase over HS but none for daily compared to alternate day rhDNase (2%, 95% CI: −4% to 9%, $P = 0.55$). There was no evidence of treatment differences for any of the secondary clinical outcomes.[12] During the HS, daily rhDNase, and alternate day rhDNase treatment periods 15, 18, and 17 children experienced one or more pulmonary exacerbations, respectively. There was no evidence of a difference for either treatment comparison ($P = 1.00$ in each case).

Comparing the outcomes

Any potential association between the outcomes measured was analysed to assess which secondary outcomes gave information, additional to that obtained from FEV_1 about the effects of each

treatment. For the study group as a whole, change in FEV_1 correlated with changes in FVC ($r^2 = 0.76$, $P = 0.001$, 95% CI: 0.78–0.92) (Fig. 1), as expected, and also with changes in $FEF_{25–75}$ ($r^2 = 0.64$, $P = 0.001$, 95% CI: 0.73–0.85), weight ($r^2 = 0.08$, $P = 0.001$, 95% CI: 0.12–0.51) (Fig. 2) and change in SaO_2 with the 3 min step test ($r^2 = 0.08$, $P = 0.001$, 95% CI: 0.12–0.49). However, importantly it did not correlate with changes in VAS with exercise ($r^2 = 0.01$, $P = 0.144$, 95% CI: −0.34 to 0.06), QWB score ($r^2 < 0.01$, $P = 0.411$, 95% CI: −0.15 to 0.26), nor with the occurrence of one or more pulmonary exacerbations in the treatment period ($r^2 < 0.01$, $P = 0.76$, 95% CI: −0.17 to 0.21).

Assessing the data for individual patients, change in weight was only weakly associated with change in FEV_1 for individuals (Fig. 2). In some patients there was an increase in weight after 12 weeks despite there being a fall in FEV_1. As expected, change in VAS with exercise and QWB score also did not match change in FEV_1 for all individuals.

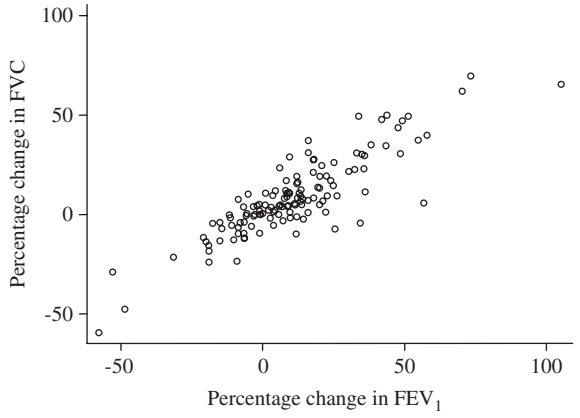

Figure 1 Positive correlation between percentage change in FEV_1 and percentage change in FVC ($r^2 = 0.76$, $P = 0.001$).

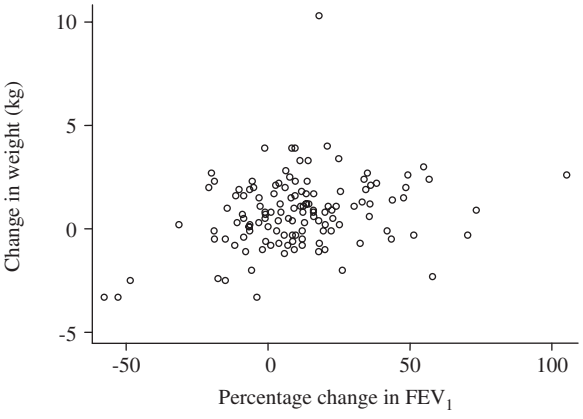

Figure 2 Weak positive correlation between percentage change in FEV_1 and change in weight ($r^2 = 0.08$, $P = 0.001$).

Discussion

The principle findings reported here are that only the addition of measurement of QOL and VAS after the step test added any useful information to measurement of FEV_1 in this medium-term trial. Several secondary outcomes were measured in this study to assess and compare the benefits of the three treatments. Due to the number of outcomes measured, each study visit lasted about 1 h. The main findings that daily rhDNase was more effective than HS and that there was no evidence of difference between daily and alternate rhDNase were based on the primary outcome FEV_1. However, the important finding of this report was that only measuring FEV_1 may result in important benefits being missed, and changes in VAS with exercise and QWB score gave additional, important information. We also showed that FEV_1 did not correlate with the occurrence of a pulmonary exacerbation. However, only a larger study than ours, with longer follow-up, would be expected to show treatment differences in this endpoint.

We cannot define from our data what is an adequate r^2 or P-value upon which to accept or reject a clinically significant difference between endpoints, and in any case, such an attempt would be somewhat artificial and arbitrary. For any association with $r^2 < 1$, there will be, by definition, a less than perfect relationship between variables. Furthermore, we acknowledge that the decision as to whether a secondary outcome measure is useful is more complex than just assessing whether it has a high or low correlation with the primary endpoint. Poor correlation with the primary endpoint could be due to patient idiosyncracy or inability of the test to discriminate between patients with CF. The point of this report is, however, that in no patient would we have altered our efficacy conclusions by adding another endpoint to the three we have highlighted, namely FEV_1, VAS and QWB. Thus, although as expected there is no perfect agreement between these endpoints and other secondary endpoints, the differences would not result in a patient's response being wrongly categorised, and thus treatment being given or withheld inappropriately. Nor do the imperfections in the relationships alter the conclusions of the trial, based on analyses of group data.

For the study group, changes in FVC and $FEF_{25–75}$ correlated with change in FEV_1. The change in FVC and $FEF_{25–75}$ could be predicted from FEV_1 and

therefore provided no additional information. However, measuring several spirometry parameters did not prolong the study visit, as they are included in the spirometry print-out and may be useful in assessing patients with mild lung disease.

As expected from the low (if significant) correlation between change in FEV_1 and change in weight, FEV_1 measures could not be used to predict changes in weight in individual children. Although growth failure and malnutrition are associated with pulmonary morbidity in patients with CF,[7,22,23] there are disadvantages in measuring weight as an outcome in short or medium-term trials. Confounding variables such as dietary intake and the extent of gastrointestinal dysfunction play a role in determining weight outcomes. During the study, one child had a gastrostomy inserted and his weight increased far in excess of any other weight gain recorded. Weight changes may therefore be independent of the trial medication. However, we acknowledge that in other trials change in weight may be a useful endpoint.

Change in VAS with exercise did not correlate with FEV_1. VAS is a subjective measure of breathlessness and is a measure of how breathless the patients themselves feel. Previous studies have shown it not to correlate with objective measures of breathlessness, such as the 15 count score.[24] A problem with the VAS is that the anchor points (''not at all short of breath'' or ''the most breathless I have ever felt'') are specific to the individual, making comparisons between different people harder to interpret, although it may be a valid measure of change over time within the same individual. The problems with the VAS highlight the difficulties of measuring a subjective feeling or symptom. However, it does not invalidate the use of this score, as it is important to know how breathless the patient feels, since this is the symptom causing them concern. Objective measures of breathlessness or lung function will not take into account the intensity of the ''unpleasantness'' of breathlessness, which is something only the patient can know.

Change in SaO_2 with exercise is an objective measure of exercise response and tolerance, and also breathlessness. The correlation between the change in SaO_2 with exercise and FEV_1 suggests that the 3-min step test is sensitive to a deterioration in pulmonary function. Therefore for younger patients unable to perform spirometry, performing the 3-min step test over time may give the clinician a measure of change in lung function. In older children who can perform spirometry, any effect on exercise tolerance due to the trial drugs, as assessed by change in SaO_2 during the 3-min step test, could be predicted by changes in FEV_1. However, we accept that in other trials change in SaO_2 may be a useful endpoint. Furthermore, if patients are performing the step test to measure changes in VAS, which has been shown to be a useful outcome measure, monitoring SaO_2 as well will not take any extra time.

Change in QOL, measured by the QWB scale, also did not correlate with FEV_1. At the time of our study, no CF specific QOL measure was available, therefore the QWB scale, which is a utility measure of QOL, was used. Quittner and colleagues[25] have recently developed a disease-specific QOL measure for children, adolescents and adults with CF.

Although the QWB scale has been validated for patients with CF, it is not disease specific. Its main advantage is its weighting of societal preferences for various symptoms and functional states. However its main limitations are a lack of sensitivity to clinically meaningful changes in CF and its uncertain applicability to children and adolescents. Despite this, it was interesting that QWB score did not correlate with FEV_1. Although FEV_1 can be tracked with time and various interventions,[26] it does not evaluate the impact of CF on the patient's overall health status and level of daily functioning. For instance, it does not take into account other pulmonary and non-pulmonary problems associated with the disease or its treatment.

Measures of QOL provide information about the impact of an illness and its treatment that may be more meaningful to patients with CF and their families than other conventional outcomes. They assess benefits and side effects of treatment and can express outcome in terms of net benefit since they are not limited to one organ system. QOL reflects an individual's subjective evaluation of his or her daily functioning and well-being because it centres on the whole patient rather than just a single measurement of physiological function. However, there is a risk that non-pulmonary events unrelated to the trial drug, which can severely affect a patient (e.g. diarrhoea) may influence the QOL measure and lead to inaccurate conclusions of how the trial drug is affecting the patient. This may be less of a problem if a CF specific QOL measure is used.

Pulmonary exacerbations cause progressive lung damage in CF and affect QOL and are often assessed as a secondary outcome. Furthermore, a comparison of the occurrence of pulmonary exacerbations is important to assess for unexpected adverse effects of the drugs. At present there is no standardised and universally accepted definition for pulmonary exacerbation.[3] The definition used in this study was taken from the large rhDNase trial by

Fuchs and colleagues.[19] Dakin and colleagues[11] have shown that there is a lack of agreement in diagnostic criteria to identify a pulmonary exacerbation among physicians caring for patients with CF. This disagreement involves their approach to diagnosis and management of a pulmonary exacerbation, and until this and other issues (above) are resolved, using either pulmonary exacerbations or time to first exacerbation as an outcome is likely to be difficult.

In this study, for each trial drug the number of treatment periods where an individual had one or more pulmonary exacerbations were compared. Although each treatment period was only 3 months, this outcome gave an assessment about how well the patient remained throughout the treatment period and how this compared to the other treatment periods. Although there was no increased risk of pulmonary exacerbation during HS and alternate day rhDNase, as compared with daily rhDNase, the study duration may not have been long enough (nor the patient numbers large enough) for any significant differences to become apparent. Fuchs and colleagues[19] have shown that daily rhDNase statistically reduces the risk of pulmonary infection compared to placebo, with patients on once daily rhDNase spending a mean of 1.3 fewer days in hospital for a pulmonary exacerbation, compared to those on placebo. However even in this 6 month, 968 patient study, these effects were of little clinical significance. There might be differences in the exacerbation rate between the treatments in this study, but these are not likely to be very clinically important on the basis of the data of Fuchs and colleagues.[19]

We have reported on the results from a single trial, and therefore this paper must be considered hypothesis generating, rather than definitive. Our trial had a small sample size with short follow-up, so outcomes based on events (e.g. pulmonary exacerbations), which may only be affected by treatment indirectly, may not be sensitive enough to demonstrate a treatment effect. The examination of data on secondary endpoints from future trials is needed to confirm or refute whether our results are generalisable.

In summary, in disagreement with the CF Consensus Conference recommendations,[3] only QOL and VAS with exercise actually provided additional information to FEV_1 regarding the effectiveness of the trial drugs in this study. The data raise the possibility that redundant endpoints are being recorded and reported, but this requires confirmation. Only large trials, which enrol several hundred subjects for a prolonged period of time, will be able to detect a change in the rate of pulmonary exacerbation. In planning future trials in CF there is a need to assess carefully which outcomes will provide useful information and to ensure that the study is sufficiently powered to detect differences in all outcomes. This will then minimise the study visit duration and inconvenience to participants. Any additional outcomes measured need to be justified, as performing several tests runs the risk of finding false positive outcomes due to multiple comparisons.

Acknowledgements

This study was funded by the NHS Health Technology Assessment Programme.

References

1. FitzSimmons SC. The changing epidemiology of cystic fibrosis. *J Pediatr* 1993;**122**:1–9.
2. Elborn JS, Shale DJ, Britton JR. Cystic fibrosis: current survival and population estimates to the year 2000. *Thorax* 1991;**46**:881–5.
3. Ramsey BW, Boat T. Outcome measures for clinical trials in cystic fibrosis. *J Pediatr* 1994;**124**:177–92.
4. McCoy K, Hamilton S, Johnson CAC. Effects of 12-week administration of dornase alfa in patients with advanced cystic fibrosis lung dusease. *Chest* 1996;**110**:889–95.
5. Ramsey BW, Pepe MS, Quan JM, et al. Intermittent administration of inhaled tobramycin in patients with cystic fibrosis. Cystic Fibrosis Inhaled Tobramycin Study Group. *N Engl J Med* 1999;**340**:23–30.
6. Equi A, Balfour-Lynn IM, Bush A, Rosenthal M. Long term azithromycin in children with cystic fibrosis: a randomised, placebo-controlled crossover trial. *Lancet* 2002;**360**:978–84.
7. Corey M, McLaughlin FJ, 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.
8. Kerem E, Reisman J, Canny GJ, Bentur L, Canny G, Levison H. Prediction of mortality in patients with cystic fibrosis. *N Engl J Med* 1992;**326**:1187–91.
9. Huang NN, Scidlow DV, Szatrowski TH, Palmer J, Laraya-Cuasay LR, Yeung W. Clinical features, survival rate and prognostic factors in young adults with cystic fibrosis. *Am J Med* 1987;**82**:871–7.
10. Saiman L, Marshall BC, Mayer-Hamblett N, et al. Azithromycin in patients with cystic fibrosis chronically infected with *Pseudomonas aeruginosa*: a randomized controlled trial. *JAMA* 2003;**290**:1749–56.
11. Dakin C, Henry R, Field P, Morton J. Defining an exacerbation of pulmonary disease in cystic fibrosis. *Pediatr Pulmonol* 2001;**31**:436–42.
12. Suri R, Metcalfe C, Lees B, et al. Comparison of hypertonic saline and alternate-day or daily recombinant human deoxyribonuclease in children with cystic fibrosis: a randomised trial. *Lancet* 2001;**358**:1316–21.
13. Polgar G, Promodhat V. *Pulmonary function testing in children: techniques and standards*. Philadelphia: WB Saunders Company; 1971.

14. American Thoracic Society. Standardization of spirometry. *Am J Respir Crit Care Med* 1994;**152**:1107–36.
15. Balfour-Lynn IM, Prasad SA, Laverty A, Whitehead BF, Dinwiddie R. A step in the right direction: assessing exercise tolerance in cystic fibrosis. *Pediatr Pulmonol* 1998;**25**:278–84.
16. Gift AG. Validation of a vertical visual analogue scale as a measure of clinical dyspnoea. *Rehabil Nurs* 1989;**14**:323–5.
17. Adams L, Chronos N, Lane R, Guz A. The measurement of breathlessness induced in normal subjects: validity of two scaling techniques. *Clin Sci* 1985;**69**:7–16.
18. Orenstein DM, Nixon PA, Ross EA, Kaplan RM. The quality of well-being in cystic fibrosis. *Chest* 1989;**95**:344–7.
19. Fuchs H, Borowitz DS, Christiansen DH, et al. Effect of aerosolized recombinant human DNase on exacerbations of respiratory symptoms and on pulmonary function in patients with cystic fibrosis. *N Engl J Med* 1994;**331**:637–42.
20. Diggle PJ, Liang KY, Zeger SL. *Analysis of longitudinal data*. Oxford: Oxford University Press; 1994.
21. Davison AC, Hinkley DV. *Bootstrap methods and their application*. Cambridge: Cambridge University Press; 1997.
22. Kraemer R, Rudeberg A, Hadorn B, Rossi E. Relative underweight in cystic fibrosis and its prognostic value. *Acta Paediatr Scand* 1978;**67**:33–7.
23. Zemel BS, Kawchak DA, Cnaan A, Zhao H, Scanlin TF, Stallings VA. Prospective evaluation of resting energy expenditure, nutritional status, pulmonary function, and genotype in children with cystic fibrosis. *Pediatr Res* 1996;**40**:578–86.
24. Prasad SA, Randall SD, Balfour-Lynn IM. Fifteen-count breathlessness score: an objective measure for children. *Pediatr Pulmonol* 2000;**30**:56–62.
25. Quittner AL, Sweeny S, Watrous M, et al. Translation and linguistic validation of a disease-specific quality of life measure for cystic fibrosis. *J Paediatr Psychol* 2000;**25**:403–14.
26. Cerny FJ, Cropp GJ, Bye MR. Hospital therapy improves exercise tolerance and lung function in cystic fibrosis. *Am J Dis Child* 1984;**138**:261–5.