Sample size estimation in studies monitoring exercise-induced bronchoconstriction in asthmatic children

W B Hofstra, J K Sont, P J Sterk, H J Neijens, M C Kuethe, E J Duiverman

Abstract

Background – The repeatability of the response to standardised treadmill exercise testing using dry air and monitoring of heart rate in asthmatic children suffering from exercise-induced bronchoconstriction (EIB) has not been well established.

Methods – Twenty seven asthmatic children with known EIB performed standardised exercise testing twice within a period of three weeks. The tests were performed on a treadmill while breathing dry air. During both tests heart rate had to reach 90% of the predicted maximum. Response to exercise was expressed as % fall in forced expiratory volume in one second (FEV_1) from baseline and as area under the curve (AUC) of the time-response curve.

Results – The intra-class correlation coefficients for % fall and AUC (log-transformed) were 0.57 and 0.67, respectively. From these data, power curves were constructed that allowed estimations to be made of sample sizes required for studies of EIB in children. These indicated that, if a drug is expected to reduce EIB by 50%, as few as 12 patients would be sufficient to demonstrate this effect (90% power) using a parallel design study.

Conclusions – Standardised exercise testing for EIB using dry air and monitoring of heart rate is adequately repeatable for use in research and clinical practice in children with asthma.

(*Thorax* 1997;**52**:739–741)

Department of Pulmonary Pediatrics, Juliana Children's Hospital, 2566 ER The Hague, The Netherlands
W B Hofstra
E J Duiverman

Department of Pulmonary Diseases, Leiden University Medical Centre, The Netherlands
J K Sont
P J Sterk

Sophia Children's Hospital, Rotterdam, The Netherlands
H J Neijens

't Lange Land Hospital Zoetermeer, The Netherlands
M C Kuethe

Correspondence to:
Dr E J Duiverman.

At present exercise-induced bronchoconstriction (EIB) is regarded as an expression of airway hyperresponsiveness to non-sensitising bronchoconstrictor stimuli, which is a common characteristic of current symptomatic asthma.[1] One of the goals of asthma treatment is to enable the patients to participate in activities and exercise without limitations.[2] Exercise testing for determining the severity of EIB can therefore be a helpful tool in the diagnosis and management of asthma, and in research studies investigating drug therapy, provided the method of exercise testing is standardised and reproducible. It is recommended that inspired dry air should be used and that the work intensity of the subject should be adjusted to achieve 40–50% of predicted maximal voluntary ventilation[3] or, alternatively, a heart rate ≥90% of maximum predicted during the last minutes of the test.[4] Although data are available on the reproducibility of the response to standardised exercise testing,[4–7] data for dry air exercise testing in children are lacking. We have therefore investigated the repeatability of EIB induced by standardised treadmill exercise testing while breathing dry air in asthmatic children with documented EIB, and used the data to estimate the sample size necessary for studies of EIB.

Methods

PATIENTS

Twenty seven asthmatic children (12 boys) aged 6–14 years with a current history of EIB were recruited from the clinic of the Juliana Children's Hospital, The Hague, and 't Lange Land Hospital in Zoetermeer. Pre-exercise forced expiratory volume in one second (FEV_1) was above 70% of predicted for all children, while post-exercise testing showed a fall of ≥15% in FEV_1 compared with baseline in all subjects. Four children used continuous inhaled corticosteroids, but these were withheld in the week before and during the study period. All children used short acting inhaled bronchodilators on demand only, which were withheld for eight hours before exercise testing.

STUDY DESIGN

After the screening exercise the children attended at approximately the same time of day on two separate days (interval range 6–24 days). At both visits a standardised treadmill exercise challenge was performed. The two tests were considered to be acceptable for analysis if the child had reached the target heart rate in both tests, regardless of the speed of the treadmill used, and if the duration of the two tests did not differ by more than 30 seconds.

EXERCISE CHALLENGE

Before exercise the baseline FEV_1 was measured in triplicate and the highest value was used for analysis. Exercise testing was performed by running on a treadmill[4] (LE 2000, Jaeger, Germany or Tunturi J880, Finland), while breathing dry air[3] (relative humidity <10%) during running. Dry air was obtained by pressurised medical air collected in a Douglas bag (contents 150 litres) and inhaled by the child through a face mask (Hans-Rudolph) with an inspiratory and expiratory port. The heart rate was monitored by a radiographic device (Polar

Sport Tester). The incline of the treadmill was set at 5–10%, depending on the physical condition of the child. During the first minute of the test the children walked at a slow speed to familiarise themselves again with the procedure. Subsequently, the speed of the treadmill was increased during the first three minutes of the test to induce a heart rate of $\geq 90\%$ of the predicted maximum (approximately 210 – age)[4] by the third minute of the test. Thereafter, the children ran for another three minutes unless dyspnoea made further running impossible. FEV_1 was measured in duplicate at 1, 3, 5, 7.5, 10, 15, 20 and 30 minutes after running, with the highest FEV_1 at each time point being retained for analysis.

ANALYSIS OF DATA

The severity of EIB was expressed as the maximal percentage decrease in FEV_1 after exercise compared with the baseline FEV_1 (% fall) and as the area under the time-response curve (AUC) between 0 and 30 minutes after exercise. The values on the test days of heart rate during the last minute of the test, % fall, and AUC were compared using the Student's *t* test (paired samples) with p values of <0.05 (two sided) being considered as statistically significant. Analysis of repeatability was performed by calculating the intra-class correlation coefficient (between subject variance/within + between subject variance) according to published guidelines.[8]

Curves for estimation of sample size were constructed using within subject variability measurements and published power functions (one sided) for predicting sample size.[9] For comparison of active versus placebo treatment the change over time in the outcome variable was chosen as the main efficacy parameter, the treatment induced change over time being a more sensitive indicator of drug effect than the absolute values of the outcome variable at each time point of measurement. Assuming absence of period or carryover effects, the approximate number of subjects required in a crossover design study was subsequently calculated using the formula:

N (number of subjects) =
$$[(Sdd \times \{Z_\alpha - Z_\beta\})/D]^2$$

where Sdd is the standard deviation of the change in outcome variable, Z_α is the standard normal deviate corresponding to a right-hand tail area of α, Z_β is the standard normal deviate corresponding to a left-hand tail area of β, and D is the expected mean change in EIB produced by treatment – that is, 40%, 50%, etc.

For a parallel group design study, comparing the change over time of the outcome variable for the two groups, the number of subjects was calculated according to the formula:

N (number of subjects) =
$$2 \times [(Sdd \times \{Z_\alpha - Z_\beta\})/D]^2 \text{ in each group.}$$

(Caveat: With estimated sample sizes of ≤ 30, Z values should be replaced by Student's *t*
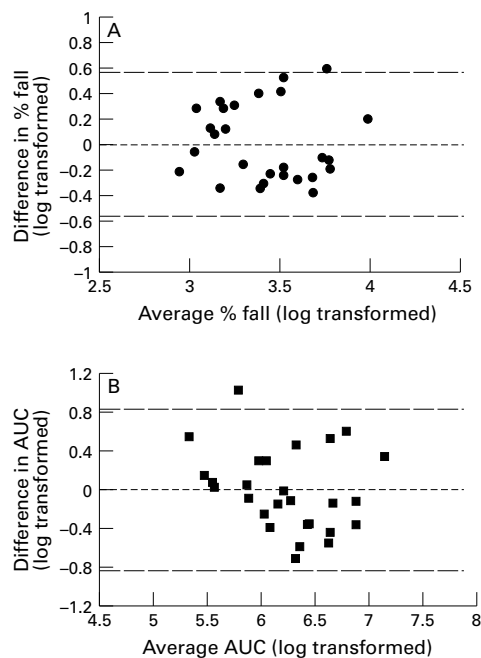

Figure 1 Repeated measures of log_e-transformed % fall in FEV_1 (A) and area under the curve (AUC) (B), the difference in repeated measures plotted against their means and shown with line of no difference (dotted lines) and 95% agreement limits (dashed lines).

values for the same α and β, but based on n – 1 degrees of freedom (with n = estimated sample size).)

Results

All children completed the study according to the protocol. Between children, the duration of the exercise tests varied from 3.5 to 6 minutes. Mean (SD) heart rate during the last minute of the test did not differ significantly between the test days (188 (9)/min and 189 (8)/min, respectively, p = 0.44). The mean (SD) % fall for the first test was 32.1 (10.5)%, and for the repeat test 31.7 (10.3)%. Mean AUC was 582 (307)%min for test 1, and 596 (297)%min for test 2. Severity of EIB did not differ significantly between the two tests (% fall, p = 0.87; AUC: p = 0.80). Because the differences in % fall and AUC between the two tests were proportional to their mean, both parameters were log-transformed, thereby achieving normality of the differences (fig 1). The intra-class correlation coefficients (based on these log-transformed data) were 0.57 for % fall and 0.67 for AUC. Curves for estimation of sample sizes at given power levels were constructed for parallel group studies (fig 2) and for crossover studies (not shown). The standard deviation of difference (Sdd) in EIB for the two tests used in the formulae to calculate sample size was 0.296 for the log_e-transformed data of % fall, and 0.416 for the log_e-transformed data of AUC. The estimated sample sizes needed at different power levels to discriminate statistically significant reductions in EIB of 40% and 50%, respectively, are presented in table 1, the required number of subjects being dependent on
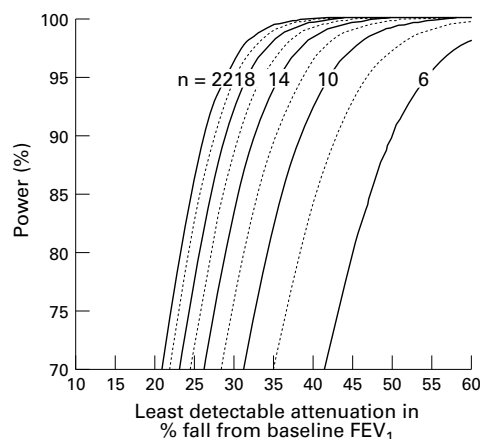
Figure 2   Curves based on log-transformed data for estimation of sample size needed in a parallel group study to show variable attenuations in exercise-induced bronchoconstriction at certain power levels for % fall in FEV₁ with n = the required number of subjects per treatment arm (p = 0.05).

Table 1   Sample size estimations (n) at given power levels required in crossover and parallel group studies to discriminate the expected protective effect of a drug and placebo against exercise-induced bronchoconstriction with statistical significance (α = 0.05, two-sided)

PI (%)	Power (%)	Crossover		Parallel*	
		% fall	AUC	% fall	AUC
40	80	5	8	16	26
	90	6	10	20	32
	95	7	11	24	40
50	80	4	6	12	16
	90	4	6	12	20
	95	5	8	16	24

PI = protection index = $\{(EIB_{post\text{-}placebo} - EIB_{post\text{-}drug})/EIB_{post\text{-}placebo}\}$ $\times 100\%$; % fall = % fall in FEV₁ from baseline; AUC = area under time-response curve.
* n for drug and placebo group together, hence 0.5 n per treatment arm.

the choice of design (crossover or parallel) and outcome variable (% fall or AUC). For example, if a drug-induced reduction of 50% is considered to be clinically effective, only five (% fall) or six (AUC) patients are needed to achieve statistical significance (α = 0.05, two-sided, β = 0.90, one-sided) using a crossover design, while a parallel group design requires a total of 12 (% fall) to 20 (AUC) patients.

Discussion

This study has shown that the response to standardised exercise testing in asthmatic children is adequately repeatable using inspiration of dry air and monitoring of heart rate. Sample sizes needed to detect significant differences in EIB are relatively small, supporting the feasibility of research studies monitoring EIB in childhood asthma.

Two methodological points need to be addressed when comparing our results with those in the literature.[4–7] Firstly, the repeatability is influenced by the degree of EIB of the selected patients. This was illustrated in an epidemiological study by Haby *et al*[7] in which the reproducibility of the % fall to a standardised free range running exercise test was assessed by the calculation of the single determination 95% range. Taking into account all children in

that study, a 95% range of ±12% was calculated, meaning that there is a 95% chance of the true value for a subject to be found within the range of 12% fall in FEV₁ around the single measurement value. However, when re-analysing the published data in children with at least 20% fall in FEV₁ post-exercise (comparable to our study population), we have estimated the single determination 95% range to be ±18.5%.[7] Secondly, in the studies published different indices of repeatability are used such as coefficient of variation (CV = standard deviation divided by mean),[4 5] intra-class correlation coefficient,[6] or 95% CI of a single measurement.[7] The CV is only to be used when the standard deviation is proportional to the mean,[10] otherwise the CV will vary with the mean value of EIB, as was elegantly shown for the % fall by Eggleston *et al*.[4] In our study we chose to use the intra-class correlation coefficient as an index of repeatability as has recently been advised because, in all circumstances, the intra-class correlation coefficient relates the size of the error variation to the size of the variation of interest.[10] However, a low intra-class correlation coefficient does not necessarily implicate a larger sample size, as the sample size estimation is dependent on the standard deviation of the difference between two tests, and not the intra-class correlation coefficient itself.[6]

What are the implications of these data? The short term repeatability of the response to standardised dry air exercise testing is adequate enough to allow drug evaluation for EIB in limited numbers of children, with the required study sample sizes being influenced by the choice of design (crossover or parallel) and the outcome variable (% fall or AUC). Thus, standardised dry air exercise testing can be an important tool in the management of childhood asthma. Whether repeatability can be improved by measuring ventilation instead of heart rate during testing remains to be investigated.

1 Cockcroft DW. Nonallergic airway hyperresponsiveness. *J Allergy Clin Immunol* 1988;**81**:111–9.
2 The British Thoracic Society, the British Paediatric Association, the Research Unit of the Royal College of Physicians of London, the King's Fund Centre, the National Asthma Campaign, the Royal College of General Practitioners, *et al*. Guidelines on the management of asthma. *Thorax* 1993;**48**:S1–24.
3 Sterk PJ, Fabbri LM, Quanjer PhH, Cockcroft DW, O'Byrne PM, Anderson SD, *et al*. Airway responsiveness. Standardized challenge testing with pharmacological, physical and sensitizing stimuli in adults. *Eur Respir J* 1993;**6**(Suppl 16):53–83.
4 Eggleston PA, Guerrant JL. A standardized method of evaluating exercise-induced asthma. *J Allergy Clin Immunol* 1976;**58**:414–25.
5 Henriksen JM. Reproducibility of exercise-induced asthma in children. *Allergy* 1986;**41**:225–31.
6 Inman MD, Watson R, Wooley KL, Manning PJ, O'Byrne PM. Reproducibility of bronchoconstriction following dry air exercise challenge. *Am J Respir Crit Care Med* 1994;**149**:A1048.
7 Haby MM, Peat JK, Mellis CM, Anderson SD, Woolcock AJ. An exercise challenge for epidemiological studies of childhood asthma: validity and repeatability. *Eur Respir J* 1995;**8**:729–36.
8 Snedecor GW, Cochran WG. *Statistical methods*. 7th ed. Ames: Iowa State University Press, 1980: 2825.
9 Thomson NC, Roberts RS. Measurement of effect of drugs on airway responsiveness. In: *Airway responsiveness: measurement and interpretation*. Mississauga: Astra Pharmaceuticals, 1985:105–18.
10 Chinn S. Statistics in respiratory medicine. Scale, parametric methods, and transformations. *Thorax* 1991;**46**:536–8.