

## The reproducibility and sample size requirements of exercise-induced bronchoconstriction measurements

B. Dahlén\*, P.M. O'Byrne<sup>#</sup>, R.M. Watson<sup>#</sup>, A. Roquet\*, F. Larsen\*, M.D. Inman<sup>#</sup>

*The reproducibility and sample size requirements of exercise-induced bronchoconstriction measurements. B. Dahlén, P.M. O'Byrne, R.M. Watson, A. Roquet, F. Larsen, M.D. Inman. ©ERS Journals Ltd 2001.*

**ABSTRACT:** Dry air exercise challenges are frequently used to screen medications that have potential utility in the management of exercise-induced bronchoconstriction (EIB). The purpose of this study was to determine the reproducibility of three outcome measurements made using such challenges, and sample size requirements for drug evaluation studies based on these outcomes.

Forty adult subjects with asthma, who tested positively on a screening exercise challenge, were subjected to two further identical challenges, separated by 1 to >35 days. Outcome measurements included the maximum per cent fall in forced expiratory volume in one second (FEV<sub>1</sub>), after exercise (% fall<sub>max</sub>), and the area under the per cent fall in FEV<sub>1</sub>/time curve for 30 min (AUC<sub>30</sub>) and 60 min (AUC<sub>60</sub>) after exercise.

The reproducibility of these outcomes, as assessed by intraclass correlation coefficients was 0.72, 0.53 and 0.35 for % fall<sub>max</sub>, AUC<sub>30</sub> and AUC<sub>60</sub> measurements, respectively. The sample size requirements to demonstrate an attenuation of EIB equivalent to a 50% reduction in % fall<sub>max</sub> was 9, 14 and 19 subjects for the % fall<sub>max</sub>, AUC<sub>30</sub> and AUC<sub>60</sub> responses, respectively (90% power).

It is concluded that the maximum percentage fall in forced expiratory volume in one second has greater reproducibility and results in greater power in clinical trials than area under the curve measurements. Sample size calculation curves are provided which may be used in study design and interpretation of published studies.

*Eur Respir J 2001; 17: 581–588.*

Exercise-induced bronchoconstriction (EIB) occurs to some degree in 70–80% of asthmatic patients. This bronchoconstriction is thought to be a manifestation of airway hyperresponsiveness and as such, does not cause an inflammatory response or worsening of the underlying asthma. Thus, the goal in the asthmatic patient should be to encourage exercise, while providing optimal anti-inflammatory treatment of the underlying asthma and any additional treatment required to minimise symptoms associated with exercise.

While treatment with inhaled steroids can reduce the magnitude of EIB by 50% or more [1–5], treatment with additional drugs is usually required to eliminate all symptoms. Inhalation of short and long-acting  $\beta_2$ -agonists are effective in protecting against, or reversing EIB [6–8], but this effect can be reduced following periods of regular use [9–12]. More specific agents have been shown to provide partial attenuation of the response, including antileukotrienes, anticholinergics and antihistamines [13–15]. Clearly, further studies are required in this area to identify agents, or more likely combinations that can be used on a regular basis to completely prevent asthmatic symptoms associated with exercise.

In many studies where the efficacy of treatment on the magnitude of EIB has been evaluated, the outcome measurement has been the maximum per cent fall in forced expiratory volume in one second (FEV<sub>1</sub>) after exercise (% fall<sub>max</sub>) [3, 14, 16–19]. Occasionally,

additional analyses are performed on the area under the per cent fall in FEV<sub>1</sub>/time curve (AUC) [13, 15, 17, 20, 21]. Surprisingly, there is little information available concerning the reproducibility of either of these outcome measurements, or on sample size requirements for assessing and comparing agents in their ability to protect against EIB. Furthermore, to the authors knowledge, there have been no reports on the clinical relationship between these outcome variables. Thus, there is no basis for assuming that a given per cent attenuation of the per cent fall in FEV<sub>1</sub> is equivalent to the same per cent attenuation of the AUC.

The purpose of this study was to determine the reproducibility of both the % fall<sub>max</sub> as well as the AUC for 30 (AUC<sub>30</sub>) and 60 min (AUC<sub>60</sub>) after exercise. A second aim was to provide guidelines for sample size determination in exercise challenge studies of treatment efficacy, similar to those provided previously for allergen and methacholine challenges [22, 23]. Finally, the per cent fall and AUC measurements were compared in terms of clinical equivalence.

### Methods

#### Subjects

Forty asthmatic subjects (25 male, 15 female; table 1) with EIB participated voluntarily in the study, which

\*Division of Respiratory Medicine, Dept of Medicine at Karolinska Hospital, Karolinska Institute, Stockholm, Sweden. <sup>#</sup>Asthma Research Group, McMaster University, Hamilton, Ontario, Canada.

Correspondence: M.D. Inman  
Dept of Medicine,  
McMaster University  
St Joseph's Hospital  
Hamilton, Ontario  
Canada  
Fax: 19055406510

Keywords: Asthma  
exercise-induced bronchoconstriction  
intraclass correlation coefficient  
power  
study design

Received: December 31 1999  
Accepted after revision October 9 2000

Table 1. – Subject characteristics

Characteristic	Age yrs	Height cm	Mass kg	FEV <sub>1</sub> % pred
Mean±SD	26.8±6.0	175.7±8.3	73.4±9.1	88.3±7.6

was approved by the Research Ethics Committees of The Karolinska Hospital and McMaster University Health Sciences Centre. Each subject gave written informed consent before taking part. All subjects were nonsmokers and had stable asthma controlled by short-acting inhaled  $\beta_2$ -agonists alone. Three subjects used a stable dose of inhaled budesonide daily (200, 400 and 800  $\mu\text{g}$ , respectively) in addition to their  $\beta_2$ -agonist use. All subjects had FEV<sub>1</sub> >70% predicted [24]. Atopic subjects were not studied during periods associated with seasonal environmental allergen exposure. Subjects were not knowingly exposed to other environmental allergens (except house dust mite) for  $\geq 2$  weeks prior to any study visits.

### Study design

The subjects attended the laboratory for  $\geq 2$  screening sessions and then on 2 days, separated by a period of 1–21 days (33 subjects) or >35 days (7 subjects). Subjects were instructed not to use bronchodilating drugs for  $\geq 8$  h and caffeine for  $\geq 24$  h prior to all laboratory visits. Exercise challenges performed were administered at the same time of day (+/- 30 min) for each subject.

### Methods

During the initial screening visit, subject characteristics and history were documented, and an incremental cycle ergometer exercise test was performed until subjective exhaustion [25]. During the second screening visit, a dry air exercise challenge (described later) was profaned at a work rate equal to 80% of the maximum achieved during the incremental test. If this resulted in a fall in FEV<sub>1</sub> from baseline of 15–45%, then this work rate was chosen for subsequent exercise challenges. If the per cent fall in FEV<sub>1</sub> was outside this range, then further screening challenges were performed at appropriately lesser or greater work rates. At least 24 h separated all screening visits.

Subjects in whom a fall in FEV<sub>1</sub> of  $\geq 15\%$  was measured during a screening exercise challenge then performed on separate days, two further exercise challenges at the same work rate as that used during screening.

**Dry-air exercise challenge.** Subjects exercised on a stationary cycle ergometer (Karolinska site: SECA Cardiostest 100, Vogel and Halke, Hamburg, Germany. McMaster site: Ergomed 740, Siemens Mississauga, Canada) for 5 min at a constant work rate. Subjects wore nose clips and breathed dry room air (<10% relative humidity) at room temperature (21.5°C) from a Douglas bag reservoir connected *via* the inspiratory

port of a 3-way Hans-Rudolph (model 2700, Kansas City, MO, USA) valve to a mouthpiece. The FEV<sub>1</sub> was measured (Karolinska site: Vitalograph MDI Compact, Föbundsmaterial, Stockholm, Sweden. McMaster site: Collins 10 L water spirometer, Collins Inc., Braintree, MA, USA) immediately prior to and immediately following exercise, as well as at 1, 3, 5, 8, 10, 15, 20, 30, 40, 50 and 60 min postexercise. The pre-exercise (baseline) FEV<sub>1</sub> was established as the greatest of three measurements, separated by  $\geq 30$  s from each other.

### Analysis

The bronchoconstrictor response following exercise was measured as % fall<sub>max</sub>, AUC<sub>30</sub> and AUC<sub>60</sub>. Specifically, % fall<sub>max</sub> represented the lowest FEV<sub>1</sub> measured in the hour following exercise, expressed as a percentage of the pre-exercise FEV<sub>1</sub> measurement. The AUC<sub>30</sub> measurement was the area under the FEV<sub>1</sub> (expressed as a percentage of the pre-exercise value)/time curve (up to 30 min postexercise) calculated using linear trapezoidal integration. On occasions when FEV<sub>1</sub> was greater than baseline values, area was subtracted from the AUC measurement. The AUC<sub>60</sub> was calculated similarly, but for the 60 min postexercise period.

An index of the reproducibility of responses, as measured using each of these analysis techniques was obtained using the intraclass correlation coefficient [26]. This gives a number usually varying -1–1, where 1 indicates perfect agreement, 0 indicates no relationship and a negative number indicates disagreement. The magnitude of the number gives the fraction of the total variability in the measurement that can be accounted for by between-subject factors.

To obtain an indication of the utility of an exercise challenge for evaluating antiasthma treatments, the sample size required to demonstrate a given percentage attenuation of the response was calculated. Sample size was calculated based on measured SD and anticipated differences in mean responses using the following equation:

$$n = \left[ \frac{(T_{1-\alpha/2} + T_{1-\beta}) \sigma}{\mu_1 - \mu_2} \right]^2 \quad (1)$$

where  $n$  is the predicted sample size,  $T$  is the t-test t-score corresponding to the desired  $\alpha$  (probability of a type 1 error) and  $\beta$  (probability of a type 2 error),  $\sigma$  is SD and  $\mu_1 - \mu_2$  is the minimum clinically important difference that should be detected [27]. This equation is appropriate for two-tailed tests. Using an iterative process, all sample size predictions were made using t-distributions with the degrees of freedom set at a level corresponding to the resulting sample size calculation. Sample size predictions were made for % fall<sub>max</sub>, AUC<sub>30</sub> and AUC<sub>60</sub> measurements. The SD entered into the equation was of the difference between the measurement on the first and second challenges. Thus,

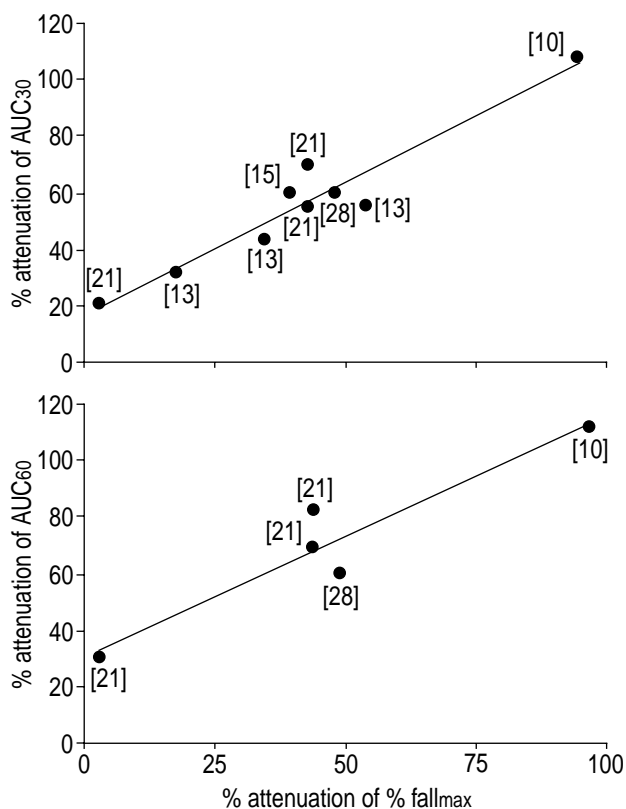


Fig. 1. – An illustration of the relationship between the degree of attenuation of maximum per cent fall in forced expiratory volume in one second (FEV<sub>1</sub>) after exercise (% fall<sub>max</sub>) and area under the curve (AUC) measurements 30 mins (AUC<sub>30</sub>) and 60 mins (AUC<sub>60</sub>) after exercise from published studies. Reference numbers are placed near each data point.

all standard deviations are based on within subject variance, which is appropriate only for repeated measures designs (either crossover or pre/post measurements). Statistical significance was set at p=0.05.

When comparing these three techniques in terms of sample size requirements for clinical trials, it is important that the degree of protection with each technique is clinically equivalent. For example, knowing that *x* subjects would be required to demonstrate a 50% attenuation in the % fall<sub>max</sub> measurement, while *y* subjects would be required to demonstrate a 50% attenuation of the AUC<sub>60</sub> measurement, would only be useful information if one assumed that these two degrees of attenuation are clinically equivalent. There is to date, no justification for this assumption. In an attempt to determine the clinical relationship between these three measurements, the present authors have reviewed the literature, identifying studies where the

degree of EIB was assessed using ≥2 of these techniques, both under placebo and treatment conditions [10, 13, 15, 21, 28]. The resulting relationships between per cent attenuation of the % fall<sub>max</sub> measurement and per cent attenuation of both the AUC<sub>30</sub> and AUC<sub>60</sub> techniques are shown in figure 1. The per cent attenuation was calculated as the percentage difference between an outcome under placebo and treatment conditions. These data sets were analysed using linear regression, resulting in the two following equations:

$$\text{attenuation (\% fall}_{\text{max}}) = 0.99 \times \text{attenuation(AUC}_{30}) - 13.33 \quad (r^2 = 0.96) \quad (2)$$

$$\text{attenuation (\% fall}_{\text{max}}) = 0.99 \times \text{attenuation(AUC}_{60}) - 25.77 \quad (r^2 = 0.95) \quad (3)$$

These equations were used to compare sample size calculations for the three analysis techniques in terms of clinically equivalent degrees of attenuation.

**Results**

Baseline FEV<sub>1</sub> was not systematically different between the first and second challenges (3.42±0.10 L versus 3.40±0.10 L, respectively; p>0.05). The mean absolute difference between the baseline FEV<sub>1</sub> values on the two challenge days was 4.25±0.70%.

The mean magnitudes of the % fall<sub>max</sub>, AUC<sub>30</sub>, and AUC<sub>60</sub> responses are included in table 2. The intra-class correlation coefficient for the % fall<sub>max</sub> measurement was 0.72, for the AUC<sub>30</sub> measurement was 0.59 and for the AUC<sub>60</sub> measurement was 0.35 (fig. 2). Thus, within-subject variability accounted for 28%, 41% and 65% of the total variance in the % fall<sub>max</sub>, AUC<sub>30</sub>, and AUC<sub>60</sub> measurements, respectively. Also in figure 2, the difference between the first and second measurements has been plotted against the mean of the two measurements for % fall<sub>max</sub>, AUC<sub>30</sub> and AUC<sub>60</sub>. These plots illustrate that the magnitude of the difference between the first and second measurements remains essentially the same over the range of measurements made from these subjects.

To determine whether the degree of reproducibility was affected by the time interval allowed between the challenges, the absolute value of the difference between the % fall<sub>max</sub> measurement on the first and second challenges was plotted against the interval (fig. 3). The difference in this measurement between the first and second challenge did not increase as the time interval

Table 2. – Data for maximum per cent fall in forced expiratory volume in one second (FEV<sub>1</sub>) after exercise (% fall<sub>max</sub>), and area under the per cent fall in FEV<sub>1</sub>/time curve for 30 (AUC<sub>30</sub>) and 60 mins (AUC<sub>60</sub>), during two challenges, and the difference between the two (Δ)

Challenge	% fall <sub>max</sub>			AUC <sub>30</sub>			AUC <sub>60</sub>		
	1	2	Δ	1	2	Δ	1	2	Δ
Mean±SD	23.3±111.9	21.4±12.5	-1.9±9.0	384±257	354±281	-30±244	561±453	557±477	-4±531

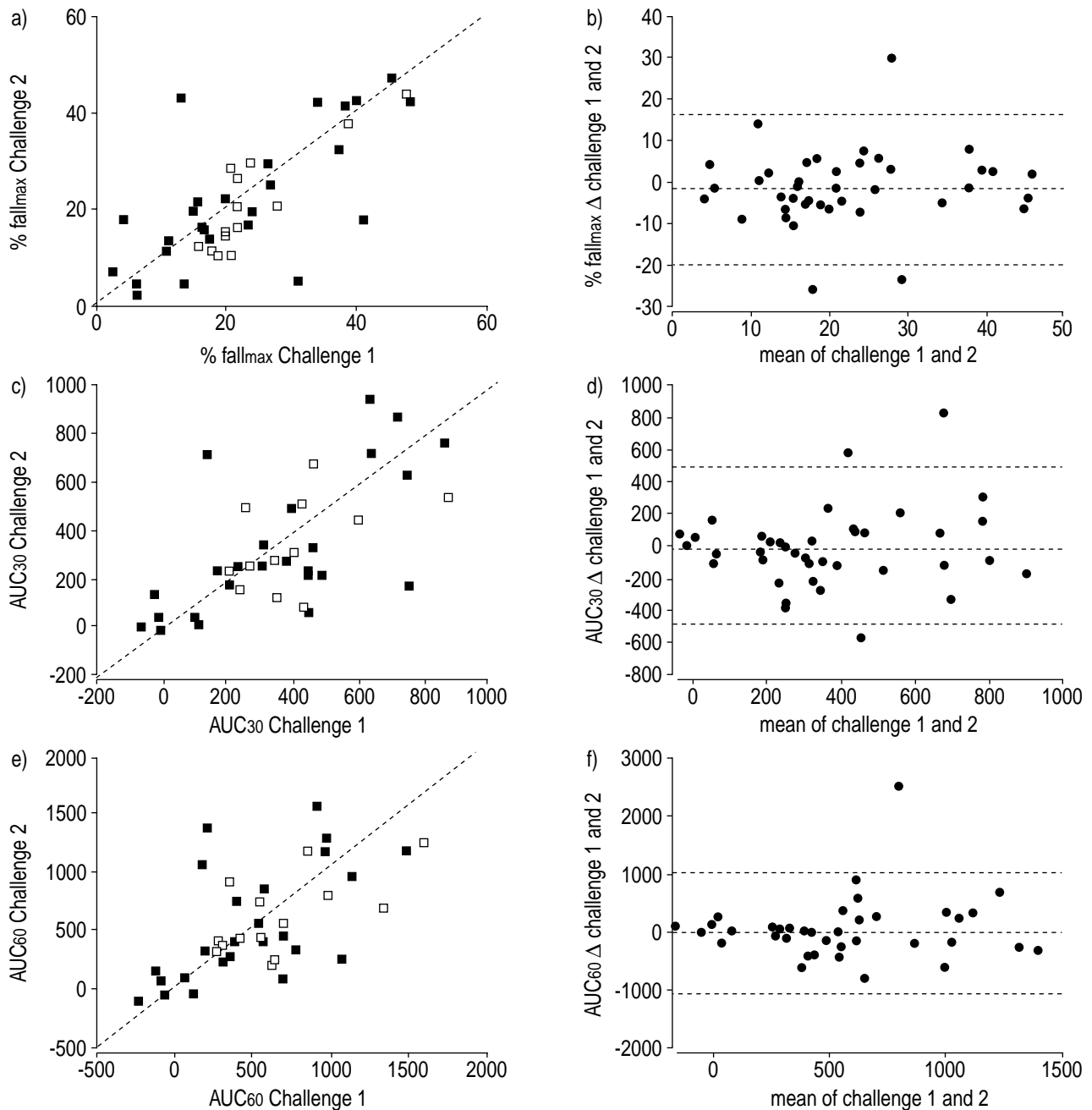


Fig. 2. – Relationship between 1st and 2nd challenges for maximum per cent fall in forced expiratory volume in one second (FEV<sub>1</sub>) after exercise (% fall<sub>max</sub>): Intraclass correlation coefficient (ICC)=0.72; b) The difference between ( $\Delta$ ) the first and second measurement, plotted against the mean of the two measurements for % fall<sub>max</sub>; c) Relationship between 1st and 2nd challenges for area under the per cent fall in FEV<sub>1</sub>/time curve (AUC), 30 mins after exercise (AUC<sub>30</sub>): ICC=0.59; d) The difference between, plotted against the mean of, the two measurements for AUC<sub>30</sub>; e) Relationship between 1st and 2nd challenges for AUC, 60 mins after exercise (AUC<sub>60</sub>): ICC=0.35; and f). The difference between, plotted against the mean of, the two measurements for AUC<sub>60</sub>.

increased from 1–>35 days; in fact it appeared to decrease.

The estimated sample size requirements to illustrate statistically significant attenuation of the % fall<sub>max</sub>, AUC<sub>30</sub> and AUC<sub>60</sub> responses, based on the mean of the two responses and the SD of the delta values in table 2 are illustrated in figure 4. Thus, the curves in figure 4a are calculated, based on attenuation of a 22.3% fall<sub>max</sub> with SD of the difference between the treated and untreated responses of 9.02% fall<sub>max</sub> units.

Included in these figures is an illustration of the estimated sample size requirements to demonstrate, with 90% power, a 50% attenuation of each response; this being 9, 21 and 40 subjects for the % fall<sub>max</sub>, AUC<sub>30</sub> and AUC<sub>60</sub> measurements, respectively. In figure 5, separate x axes have been included for % fall<sub>max</sub>, AUC<sub>30</sub>, and AUC<sub>60</sub> curves and these have been aligned relative to each other, according to the relationships in equations (2) and (3), so that there is a clinically equivalent degree of attenuation on all three curves at

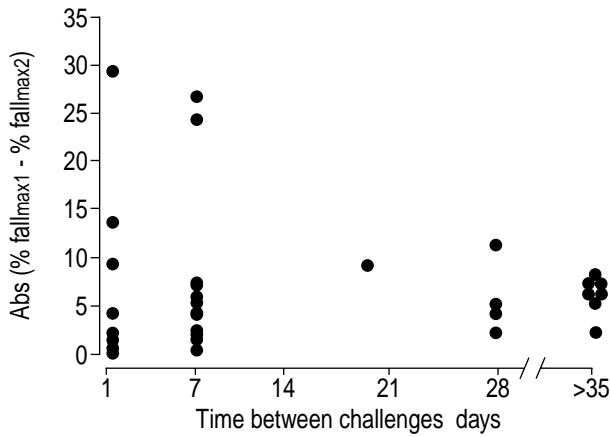


Fig. 3. – The absolute value of the difference between the two maximum per cent fall in forced expiratory volume in one second (FEV<sub>1</sub>) after exercise (% fall<sub>max</sub>) measurements, plotted against the time interval between the two challenges.

any point where a vertical line is placed. After making this correction, the number of subjects required to demonstrate an attenuation equivalent to a 50% reduction in % fall<sub>max</sub> was 9, 14 and 19 subjects using the % fall<sub>max</sub>, AUC<sub>30</sub>, and AUC<sub>60</sub>, measurements, respectively.

**Discussion**

In this study, the reproducibility and power of three techniques for quantifying the degree of exercise-induced bronchoconstriction has been compared. The greatest reproducibility has been observed for the maximum % fall in FEV<sub>1</sub> and the least reproducibility for the area under the 60 min FEV<sub>1</sub>/time curve. Furthermore, the power of these techniques to detect the attenuating effects of treatment are such that small degrees of attenuation are more likely to be detected using the % fall<sub>max</sub> technique with the same sample size.

The intraclass correlation coefficient (ICC) for the

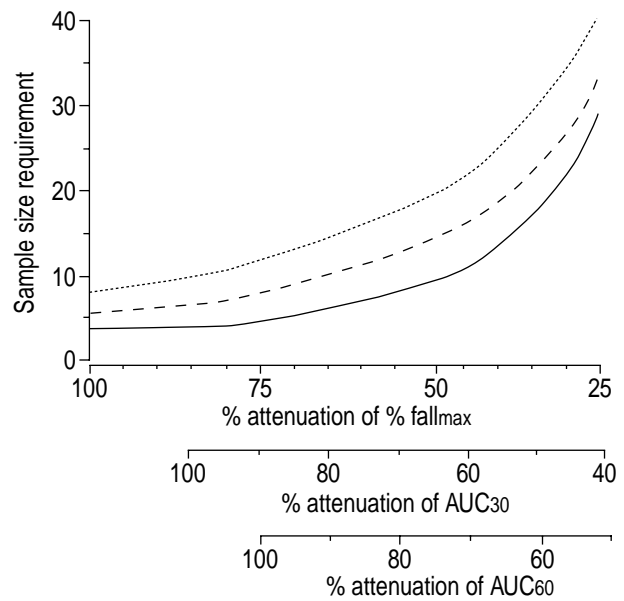


Fig. 5. – Sample size requirements to demonstrate a range of expected attenuations of maximum per cent fall in forced expiratory volume in one second (FEV<sub>1</sub>) after exercise (% fall<sub>max</sub>; —); area under the per cent fall in FEV<sub>1</sub>/time curve (AUC) for 30 min after exercise (AUC<sub>30</sub>; - - -); and for 60 min after exercise (AUC<sub>60</sub>; · · · ·) response with 90% power. The x axes are calibrated and positioned based on equations (2) and (3), so that the degree of attenuation of each response is clinically equivalent at any point.

AUC<sub>30</sub> technique measured in this study was 0.59, and was similar to the value of 0.67 observed by HOFSTRA *et al.* [29]. The observation of a higher ICC for the % fall<sub>max</sub> technique (0.72) differs from HOFSTRA *et al.* [29], who observed a lower ICC for this outcome (0.57). Generally, however, the two studies are in agreement that less than half of the variability of the EIB response is within-subject or error variance. The present observations also suggest that this reproducibility extends to time intervals between challenges of greater than 35 days.

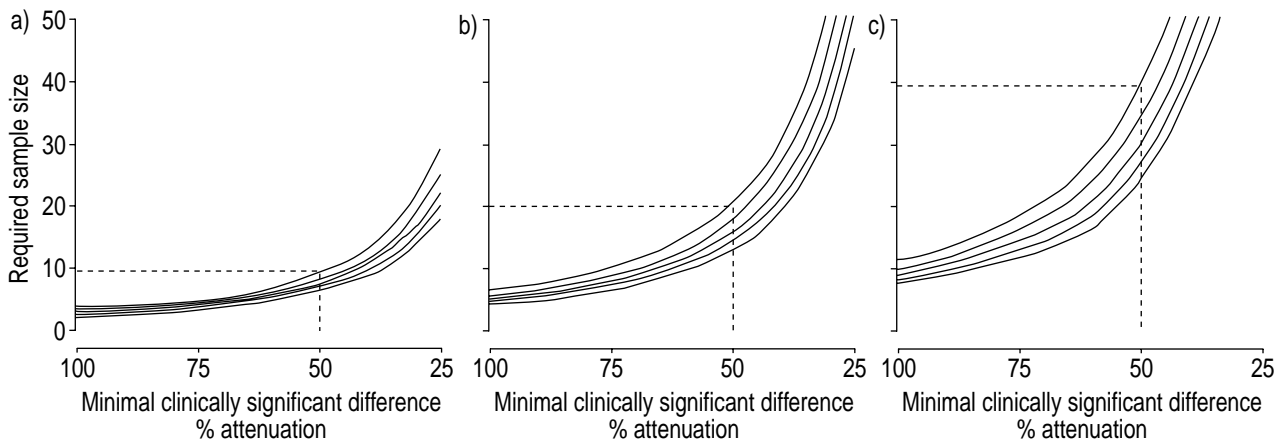


Fig. 4. – Sample size requirements to demonstrate a range of expected attenuations of a) maximum per cent fall in forced expiratory volume in one second (FEV<sub>1</sub>) after exercise (% fall<sub>max</sub>); b) area under the per cent fall in FEV<sub>1</sub>/time curve (AUC) for 30 min after exercise; c) AUC 60 min after exercise and responses. The five curves (top to bottom) represent 90%, 85%, 80%, 75% and 70% power, respectively. The dashed lines illustrate how to calculate the sample size requirements in order to demonstrate 50% attenuation of each response with 90% power.

Although it has been observed that there is good reproducibility of the % fall<sub>max</sub> measurement, the authors argue that this is not sufficient to infer a high degree of utility in all research applications. In crossover studies where the change in a measurement is the important outcome variable, it is the sample size, variability in the change, and clinically important magnitude of the change that determine the power of the study. For this reason the SD of the change in % fall<sub>max</sub>, AUC<sub>30</sub>, and AUC<sub>60</sub> between the first and second challenges have been measured and used, along with a range of degrees of attenuation and desired power levels, to predict sample sizes that should be used in clinical trials. The authors feel that this information gives much more information on the utility of each technique than the ICC does.

It has been estimated that a sample size of  $\geq 9$  subjects should be sufficient to demonstrate, with 90% power, a 50% attenuation in the % fall<sub>max</sub> measurement in response to treatment. It must be stressed that if a degree of attenuation of  $>$  or  $<50\%$  was thought to be the minimal clinically important difference, then correspondingly less or more subjects would be required in the study design, as indicated in figure 4. The present estimation of nine subjects being required to demonstrate 50% attenuation of the response is greater than the sample size of six estimated to be required to demonstrate the same effect in children as reported by HOFSTRA *et al.* [29]. In that study, sample sizes were estimated using a Z distribution, while in the present study, a t-distribution has been used, given that SDs were estimated from a relatively small sample of the population ( $n=40$ ). If a Z-distribution had been used, then a sample size of 7 would have been estimated, similar to that of HOFSTRA *et al.* [29]. Unlike HOFSTRA *et al.* [29], the present results suggest that greater sample sizes would be required to demonstrate a similar attenuation of the AUC<sub>30</sub> and AUC<sub>60</sub> outcomes. Reasons for the disagreement between the present findings and those of HOFSTRA *et al.* [29] are not clear, but may be due to their calculations being based on paediatric subjects and a treadmill based protocol.

By reviewing studies where both % fall<sub>max</sub> and AUC<sub>30</sub> or AUC<sub>60</sub> measurements were made, estimates of equations 1 and 2 were made, with which an attempt has been made to convert attenuation of % fall<sub>max</sub> measurements to clinically equivalent attenuation of AUC<sub>30</sub> and AUC<sub>60</sub> measurements. For both equations, it is clear that a given degree of attenuation of % fall<sub>max</sub> (when the treated response is expressed as a percentage of untreated response) is equivalent to greater attenuation of AUC measurements. Even after applying these equations, it was observed that clinically equivalent degrees of attenuation would be detected with smaller sample sizes using the % fall<sub>max</sub> measurement.

While sample size requirements for crossover designed studies have been presented, calculations can be made to determine sample sizes for parallel group studies. Equation 1 would still apply, but  $\sigma$  should be the SD of the measurement itself, rather than the SD of the difference between two repeated measurements. This equation could then be used to calculate the subjects required in each treatment arm. Performing

this calculation to determine the number of subjects required to demonstrate a 50% attenuation of % fall<sub>max</sub> these equations calculated that 15 would be required in each treatment arm (assuming an untreated response of 22.32% fall and an SD of the treated and untreated responses of 12.2% fall units). This number is similar to the recommendation of 12 subjects in each treatment arm made by HOFSTRA *et al.* [29] in their study of asthmatic children. This calculation assumes that the outcome measurement in the parallel group design would be the response to a single exercise challenge. If in fact, the outcome was the change in the response from pre- to post-treatment conditions, then the SD of the difference between two challenges (9.02 % fall<sub>max</sub> units as measured here) should be used to estimate sample size. In this case, the curves in figure 3 could be used to estimate the sample size requirements for each arm of the study.

In this study, it has been recommended that fewer than 10 subjects are required to demonstrate a 50% attenuation of % fall<sub>max</sub>. Clearly, more subjects are required to illustrate a smaller degree of attenuation. Furthermore, more subjects may be required to demonstrate that the protective effects of two treatments are different. For example, if one drug blocked 75% of the % fall<sub>max</sub> and another drug blocked 50% of the % fall<sub>max</sub>, both of these effects should be detected at least 90% of the time with nine subjects. However, the difference in the degree of attenuation between the two drugs is only 25% (75%-50%), a difference that would require 30 subjects to demonstrate using the % fall<sub>max</sub> outcome (fig. 3). Thus, when designing studies to compare two drugs, it is important to decide what difference between the two drugs is thought to be clinically relevant. This difference should then be used in equation 1 or plotted on figure 3a to calculate the appropriate sample size.

A further use of these sample size requirements is to evaluate the power of published studies. Usually, studies of the efficacy of various drugs in the prevention of EIB have  $\geq 10$  subjects [10, 14, 16, 20, 30–33] and are thus adequately powered, assuming that a degree of protection equivalent to  $<40$ – $50\%$  attenuation of the % fall<sub>max</sub> is considered not to be clinically important. Often, studies are performed to compare two drugs or doses with respect to their ability to attenuate EIB. Frequently these studies contained no more subjects than studies evaluating single compounds [13, 18, 19, 21, 34] and were therefore probably underpowered to detect clinically important differences between drugs or doses.

The mean % fall<sub>max</sub> in this study was  $\sim 22\%$ . In some studies a greater fall in FEV<sub>1</sub> is achieved under placebo treatment conditions. It is likely that the sample sizes estimated in this study will overestimate those required in studies where the untreated % fall<sub>max</sub> is larger than that reached by subjects in this study.

It is interesting to note that the % fall<sub>max</sub> value did not exceed 15 in seven of the subjects in response to the first exercise challenge, and 13 of the subjects in response to the second exercise challenge. There were six subjects for whom the % fall<sub>max</sub> value was  $<15$  on both exercise challenges. While it could be argued that these subjects represented screening failures, the

authors elected to include them in the analysis, as these subjects, having met screening criteria, would be included in clinical trials for which the sample size estimates will be used.

In summary, it has been shown that the maximum per cent fall of forced expiratory volume in one second after exercise has greater reproducibility than area under the curve measurements. More importantly, information on sample size requirements has been provided that can be used both in the design of well powered studies and in the evaluation of already published studies. These results show that for drug efficacy and drug comparison studies, maximum per cent fall in forced expiratory volume in one second is more powerful than area under the curve measurements.

**Acknowledgements.** Statistical advice from C. Goldsmith was greatly appreciated in the preparation of this manuscript.

### References

- Henriksen JM, Dahl R. Effects of inhaled budesonide alone and in combination with low dose terbutaline in children with exercise-induced asthma. *Am Rev Respir Dis* 1983; 128: 993-997.
- Henriksen JM. Effect of inhalation of corticosteroids on exercise-induced asthma: randomized double blind crossover study of budesonide in asthmatic children. *BMJ* 1985; 291: 248-249.
- Vathenen AS, Knox AJ, Wisniewski A, Tattersfield AK. Effect of inhaled budesonide on bronchial reactivity to histamine, exercise, and eucapnic dry air hyperventilation in patients with asthma. *Thorax* 1991; 46: 811-816.
- Molema J, van Herwaarden CL, Folgering HT. Effects of long-term treatment with inhaled cromoglycate and budesonide on bronchial hyperresponsiveness in patients with allergic asthma. *Eur Respir J* 1989; 2: 308-316.
- Farrero E, Llundell A, Canete C, Rodriguez B, Mauresa F, Romero PV. Inhaled steroids treatment and withdrawal in asthmatic children. *Allergologia et Immunopathologia* 1995; 23: 182-188.
- Anderson SD, Seale JP, Rozea P, Bandler L, Theobald G, Lindsay DA. Inhaled and oral salbutamol in exercise-induced asthma. *Am Rev Respir Dis* 1976; 114: 493-500.
- Bundgaard A, Rasmussen FV, Madsen L. Pretreatment of exercise-induced asthma in adults with aerosols and pulverized tablets. *Allergy* 1980; 35: 639-645.
- Konig P, Hordvik NL, Serby CW. Fenoterol in exercise-induced asthma. Effect of dose on efficacy and duration of action. *Chest* 1984; 85: 462-464.
- Gibson GJ, Greenacre JK, Konig P, Conolly ME, Pride NB. Use of exercise challenge to investigate possible tolerance to beta-adrenoceptor stimulation in asthma. *Br J Dis Chest* 1978; 72: 199-206.
- Inman MD, O'Byrne PM. The effect of regular inhaled albuterol on exercise-induced bronchoconstriction. *Am J Respir Crit Care Med* 1996; 153: 65-69.
- Ramage L, Lipworth BJ, Ingram CG, Cree IA, Dhillon DP. Reduced protection against exercise induced bronchoconstriction after chronic closing with salmeterol. *Respir Med* 1994; 88: 363-368.
- Simons FE, Gerstner TV, Cheang MS. Tolerance to the bronchoprotective effect of salmeterol in adolescents with exercise-induced asthma using concurrent inhaled glucocorticoid treatment. *Pediatrics* 1997; 99: 655-659.
- Finnerty JP, Holgate ST. The contribution of histamine release and vagal reflexes, alone and in combination, to exercise-induced asthma. *Eur Respir J* 1993; 6: 1132-1137.
- Manning PJ, Watson RM, Margolskee DJ, Williams VC, Schwartz JI, O'Byrne PM. Inhibition of exercise-induced bronchoconstriction by MK-571, a potent leukotriene D4 receptor antagonist. *N Engl J Med* 1990; 323: 1736-1739.
- Finnerty JP, Wood-Baker R, Thomson H, Holgate ST. Role of leukotrienes in exercise-induced asthma. Inhibitory effect of ICI204219, a potent leukotriene D4 receptor antagonist. *Am Rev Respir Dis* 1992; 145: 746-749.
- Meltzer SS, Hasday JD, Cohn J, Bleecker ER. Inhibition of exercise-induced bronchospasm by zileuton: a 5-lipoxygenase inhibitor. *Am J Respir Crit Care Med* 1996; 153: 931-935.
- Van Schoor J, Joos GF, Kips JC, et al. The effect of ABT-761, a novel 5-lipoxygenase inhibitor, on exercise- and adenosine-induced bronchoconstriction in asthmatic subjects. *Am J Respir Crit Care Med* 1997; 155: 875-880.
- Woolley M, Anderson SD, Quigley BM. Duration of protective effect of terbutaline sulfate and cromolyn sodium alone and in combination on exercise-induced asthma. *Chest* 1990; 97: 39-45.
- Henriksen JM, Agertoft L, Pedersen S. Protective effect and duration of action of inhaled formoterol and salbutamol on exercise-induced asthma in children. *J Allergy Clin Immunol* 1992; 89: 1176-1182.
- Twentyman OP, Finnerty JP, Holgate ST. The inhibitory effect of nebulized albuterol on the early and late asthmatic reaction and increase in airway responsiveness provided by inhaled allergen in asthma. *Am Rev Respir Dis* 1991; 144: 782-787.
- Adelroth E, Inman MD, Summers E, Pace D, Modi M, O'Byrne PM. Prolonged protection against exercise-induced bronchoconstriction by the leukotriene D4-receptor antagonist cinalukast. *J Allergy Clin Immunol* 1997; 99: 210-215.
- Inman MD, Watson R, Cockcroft DW, Wong BJ, Hargreave FE, O'Byrne PM. Reproducibility of allergen-induced early and late asthmatic responses. *J Allergy Clin Immunol* 1995; 95: 1191-1195.
- Inman MD, Hamilton AL, Kerstens HAM, Watson RM, O'Byrne PM. The utility of methacholine airway responsiveness measurements in evaluating anti-asthma drugs. *J Allergy Clin Immunol* 1998; 101: 342-348.
- Morris JF, Koski A, Johnson LC. Spirometric standards for healthy nonsmoking adults. *Am Rev Respir Dis* 1971; 103: 57-67.
- Jones NL. *Clinical Exercise Testing*. 3 ed. Philadelphia, W.B. Saunders Co, 1988.
- Streiner DL, Norman GR. Reliability. In: Streiner DL, Norman GR, eds. *Health Measurement Scales*. Oxford, Oxford University Press, 1989; 79-96.
- Kupper LL, Hafner KB. How appropriate are popular sample size formulas? *Am Statistician* 1989; 43: 101-105.

28. Makker HK, Lau LC, Thomson HW, Binks SM, Holgate ST. The protective effect of inhaled leukotriene D<sub>4</sub> receptor antagonist ICI204,219 against exercise-induced asthma. *Am Rev Respir Dis* 1993; 147: 1413–1418.
29. Hofstra WB, Sont JK, Sterk PJ, Neijens HJ, Kuethe MC, Duiverman EJ. Sample size estimation in studies monitoring exercise-induced bronchoconstriction in asthmatic children. *Thorax* 1997; 52: 739–741.
30. Anderson SD, Rodwell LT, Du Toit J, Young IH. Duration of protection by inhaled salmeterol in exercise-induced asthma. *Chest* 1991; 100: 1254–1260.
31. Robuschi M, Riva E, Fuccella LM, *et al.* Prevention of exercise-induced bronchoconstriction by a new leukotriene antagonist (SK&F 104353). A double-blind study *versus* disodium cromoglycate and placebo. *Am Rev Respir Dis* 1992; 145: 1285–1288.
32. Newnham DM, Ingram CG, Earnshaw J, Palmer JB, Dhillon DP. Salmeterol provides prolonged protection against exercise-induced bronchoconstriction in a majority of subjects with mild, stable asthma. *Respir Med* 1993; 87: 439–444.
33. Waalkens HT, van Essen-Zandvliet EE, Gerritsen J, Duiverman EJXKK, Knol K. The effect of an inhaled corticosteroid (budesonide) on exercise-induced asthma in children. Dutch CNSLD Study Group. *Eur Respir J* 1993; 6: 652–656.
34. Sichletidis L, Daskalopoulou E, Kyriazis G, *et al.* Comparative efficacy of salbutamol and salmeterol in exercise-induced asthma. *J Internat Med Res* 1993; 21: 81–88.