CORRESPONDENCE

Variability of markers of airway inflammation in BAL

To the Editor:

We were pleased to see the article by WARD *et al.* [1] regarding the variability of markers of airway inflammation in bronchoalveolar lavage (BAL). In a well-designed study, the authors performed BAL twice, 1 month apart, in 20 asthmatic subjects. The subjects were symptomatic but stable during the study, as assessed by mean forced expiratory volume in one second (FEV_1), provocative dose of methacholine causing a 20% decrease in FEV_1 (PD_{20}) and the medications used. It was concluded that BAL analysis showed high variability in the indices measured, and that a sample size higher than 15 should be used to reach statistical power to detect differences.

Another way to express the results would be by the repeatability of BAL indices by the calculation of the intraclass correlation coefficient (ICC). This is appropriate since the subjects were stable. This information will be useful in accessing one of the main characteristics required from BAL, to be used as a discriminative and evaluative instrument to measure indices of airway inflammation in asthma. Variability in the dilution of airway secretions by the lavage may interfere with the repeatability of measurements. To our knowledge, this is the first study to assess this issue in stable subjects. However, to compare these results with those from other instruments, the results should be given as ICC. Can the authors give us this analysis?

The variability between indices within subjects can be observed in the figures given. As proposed by BLAND and ALTMAN [2] the differences between measures are plotted against the mean values of both. Although the figures show the variability between measurements, a better conclusion cannot be drawn since the ICC is not given. Moreover, the line which marks the absence of differences is misplaced from the zero point in most of the figures, and the 2 SD are asymmetrical. Can this be corrected? Additionally, the figure showing the repeatability of albumin has only 11 points; what happened to the other nine?

References

1. Ward C, Gardiner PV, Booth H, Walters EH. Intrasubject variability in airway inflammation sampled by bronchoalveolar lavage in stable asthmatics. *Eur Respir J* 1995; 8: 1866–1871.
2. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurements. *Lancet* 1986; i: 307–310.

E. Pizzichini, M.M.M. Pizzichini, F.E. Hargreave
Firestone Regional Chest and Allergy Unit, St. Joseph's Hospital, McMaster University, 50 Charlton Avenue East, Hamilton, Ontario L8N 4A6, Canada.

REPLY

From the authors:

The letter by Dr Pizzichini *et al.* raises some interesting issues regarding our recent publication [1]. The study basically consisted of two arms. In the first, the intrasubject variability of cellular and solute parameters was examined in repeated 180 mL bronchoalveolar lavage (BAL) procedures, in 20 clinically stable but symptomatic asthmatics of the type liable to be involved in prospective studies of the effects of asthma therapy on airway pathology. Hence, they were young, with reasonably well-preserved lung function, on little or no inhaled corticosteroid, with quite marked bronchial hyperresponsiveness and variability in lung function.

Having documented the levels of variability observed, these estimates were used to perform sample size calculations for subsequent bronchoscopic based studies. In making these estimates, an attempt was made to balance the requirements of optimizing power, while maintaining practicability, in what are demanding studies. It was felt to be necessary to document this type of information since, despite its importance for study design, it is lacking in the literature. It was suggested that a sample of around 15 subjects is consistent with a reasonable study size and this should be borne in mind at the planning stage of study protocols.

The examination of the variability between replicate measurements conducted under (as near as possible) identical conditions is a common problem in the biological sciences and Respiratory Medicine. In presenting our data, we used the approach that has consistently been advocated by BLAND and ALTMAN [2, 3] over the last 15 yrs. This approach suggests that plots of differences between measurements against their means are useful, together with approximate 95% limits of agreement (the mean difference ± two times the standard deviation of the differences). In practice, we find this type of analysis intuitive, as well as being simple to carry out. The calculation of the standard deviation of the differences also develops logically into the sample size calculations quoted in the paper, since these estimates are implicit to the calculations.

Addressing the concern about the observed non-zero mean difference between replicate measurements: in repeatability studies, the assumption is that the observed

cell count for a patient is their "true" count, measured with random error. Thus although the underlying population mean difference between replicate measurements is hypothesized to be zero, it is unlikely ever to be exactly zero, due to random variation. Only if the mean difference is large relative to the observed standard deviation of the differences is there evidence of bias between the two repeat measurements. For example, in our paper ([1], fig. 1 and table 2) we observed that between BAL 1 and BAL 2 there was an overall difference in percentage lymphocyte count of 3%, with a mean percentage lymphocyte count at BAL 1 of 29% and at BAL 2 of 26%. The standard deviation of the differences was approximately 13. The 95% limits of agreement were, therefore, $3\pm(2\times13)\%$. The fact that these limits, and that of the other plots in the paper (for the key cell data), were not symmetrical about zero, was therefore consistent with plotting biological data and not an error, as suggested by Dr Pizzichini *et al*.

Dr Pizzichini *et al*. suggest that the quotation of the intraclass correlation coefficient (ICC) may be useful in interpreting our data further. The ICC is the ratio of the between-patient variation to total variation in the cell count data and is interpreted as the correlation that can be expected between replicate measurements. The ICC is a relative measure of variability ranging 0–1.

To calculate the ICC, it is necessary to carry out a one-way analysis of variance, with patients as the group variable. This analysis separates the sources of variation in the data into between-patient variation and within-patient variation, the components needed for the calculation of the ICC. Clearly, as the within-patient variation becomes smaller (*i.e*, the measurement error becomes smaller), the ICC increases. Unfortunately, this also occurs if the between-patient variation increases. Therefore, just as for the familiar Pearson correlation coefficient, the ICC is affected by the range of measurements in the sample.

In contrast, the within-patient standard deviation, quoted in the same units as the data, should be independent of the between-patient variation. For these reasons, we did not advocate the use of the ICC in our paper and we would point out that there is an increasing body of statistical literature expressing very strong reservations about the use of ICCs [3, 4]. We recognize, however, that functionally the ICC is used by researchers and we provide the values for key cell data in table 1. In table 1, we also present our preferred measure of variability, two times the standard deviation of the differences between repeat measures, known as the coefficient of repeatability (CR).

Table 1. – Measures of variability

Parameter	Paired n	CR	ICC
% macrophages	18	32	0.33
% lymphocytes	18	26	0.60
% eosinophils	18	2	0.91
% mast cells	12	0.9	0.37

CR: coefficient of repeatability; ICC: intraclass correlation coefficient.

Dr Pizzichini was quite right to point out that repeated albumin assays on archived supernatants were only available in 11 subjects from the total of 20 asthmatics recruited.This does, nevertheless, give an estimate about the variability of measurement for this solute, and we therefore felt it was worth quoting the data in our paper.

In summary, we would like to point out that the purpose of our study, together with a recently published paper relating to biopsy work [5], was not to didactically construct absolute numbers for study design, but was rather an attempt to contribute to a development of methodology in an increasingly important area of clinical research. We would like to thank Dr Pizzichini and colleagues for their interest and constructive questioning, and the European Respiratory Journal for publishing our work.

References

1. Ward C, Gardiner PV, Booth H, Walters EH. Intrasubject variability in airway inflammation sampled by bronchoalveolar lavage in clinically stable mild-to-moderate asthmatics. *Eur Respir J* 1995; 8: 1866–1871.
2. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; i: 307–310.
3. Bland JM, Altman DG. Comparing two methods of clinical measurement: a personal history. *Int J Epidemiol* 1995; 24: S7–S14.
4. Muller R, Buttner P. A critical discussion of intraclass correlation coefficients. *Statistics in Medicine* 1994; 13: 2465–2476.
5. Richmond I, Booth H, Ward C, Walters EH. Intrasubject variability in airway inflammation in biopsies in mild-to-moderate stable asthma. *Am J Respir Crit Care Med* 1996; 153: 899–903.

C. Ward*, V. Ryan, E.H. Walters***
*Dept of Respiratory Medicine, Alfred Hospital and Monash University Medical School, Prahran, Victoria Australia, and **Statistical Consulting Centre, The University of Melbourne, Parkville, Victoria, Australia.