Designing bronchial biopsy studies

L. W. Poulter*, C. M. Burke†, N. N. Jarjour‡ and S. D. Pyke§

*Department of Immunology, Royal Free Hospital, London, U.K.
†Department of Respiratory Medicine, James Connolly Memorial Hospital, Dublin, Ireland
‡Department of Medicine, University of Wisconsin–Madison Medical School, Madison, Wisconsin, U.S.A.
§Glaxo Wellcome plc, Stockley Park, U.K.

Bronchial biopsy provides valuable information about the inflammatory processes in lung tissue, but optimal results are only achieved if the design of intervention studies is sufficiently rigorous. The parallel-group design has merit, but the cross-over design is statistically superior, providing the wash-out period is effective. Heterogeneity of contributing pathologies in asthma patients results in large inter-patient variability which must be controlled for, for example by using strict inclusion criteria, which should ideally relate to the specific inflammatory marker being studied. The inclusion of a placebo group helps to quantify sample variability.

The study must have sufficient statistical power to detect inter-group differences for each variable; appropriate adjustments should be made when multiple tests are used. Studies with larger patient numbers are best performed using a multi-centre design, with one centre analysing all tissue samples to reduce variability. Study duration depends on the type of investigation, but should ideally be short. Longer studies are necessary to evaluate chronic changes such as tissue remodelling.

Changes in clinical status and cellular events may follow different time courses after intervention. Biopsy measurements are less reproducible than physiological tests, and diurnal variation in the number and function of inflammatory cells can further complicate measurement. The timing of clinical trial assessments needs to allow for these idiosyncrasies. Finally, a balance must be maintained between the risk, albeit small, and the benefit of performing bronchial biopsies.

Key words: asthma; biopsy; clinical trial; design; marker; variability.

Introduction

When assessing how drugs modify inflammation in the lung, the most direct form of measurement is to study samples of lung tissue. Studies using animal lung may give valuable data, but the findings need to be confirmed in the clinical setting. Although bronchial biopsies are particularly suitable for pathophysiology studies, they can also be used to provide secondary outcome measures in clinical trials, if the study objectives are kept simple.

The methodology of bronchial biopsy, i.e. how the sample is taken, preserved, sectioned and analysed, has a considerable influence on the results that are obtained and guidelines have been published for the benefit of investigators working in this field (1). However, the basic design of asthma studies that include biopsy assessments (e.g. size, duration and patient selection) may affect the validity and value of the outcome even more profoundly. This paper reviews the most important elements of biopsy study design.

Study designs

CROSS-OVER DESIGNS

The best design for biopsy studies is the cross-over, which is preferable to parallel-group, provided that steps are taken to resolve the inherent problems of the design. The main advantage of cross-over studies is that patients act as their own controls, so inter-subject variability is reduced and fewer patients are required to achieve a given statistical power than in parallel-group studies. Small-scale cross-over studies are appropriate for hypothesis generation (exploratory studies) and can be run in a single centre; this design is also particularly suitable for allergen challenge studies (2,3). A cross-over design requires more biopsies to be taken from each subject, however, this, coupled with the longer
duration of the study, may discourage patient entry. For studies with three or more treatment arms, this problem can be directly addressed by the use of an ‘incomplete block’ design.

Statistical analysis of data in cross-over studies is straightforward, although it is frequently difficult to assess how long the wash-out period should be to ensure that no drug effects carry over into the subsequent treatment period. In a long-term study, six of 18 patients with mild asthma who received placebo for 1 year after 2 year’s treatment with budesonide 1200 μg day⁻¹ maintained the lung function improvements achieved in the first 2 years, and only three of the 18 patients had to withdraw from the study during the third year because of inadequate treatment (4); this shows that participation in the trial produced clinical benefits, but also suggests that the effects of treatment with inhaled corticosteroids (ICS) may persist after treatment discontinuation.

In the ideal cross-over design, which should be double blind, four bronchoscopy procedures with biopsy sampling should be performed, before and after each treatment period (Fig. 1). The pretreatment sample allows intra-subject variability to be better controlled for. If the treatment and wash-out periods are short, four samples may not be practicable, as patients cannot tolerate frequent bronchoscopies; however, omitting the pretreatment sample need not be a problem if placebo is included as one of the treatments. Some variability between subjects may still remain in cross-over studies, resulting from the differences in underlying pathology between study participants. Analysis of the study results requires treatment effects to be averaged across subjects; therefore, the patient sample needs to be made as homogeneous as possible. This can be achieved by using strict inclusion criteria to select specific patient subsets.

### PARALLEL GROUP

Strict inclusion criteria can also be employed to reduce variability in parallel-group studies; another common approach is to recruit large numbers of patients. In a heterogeneous patient population such as asthmatics, the recruitment of larger treatment groups allows for the increased variability, and increases the probability of the groups being well balanced. In large studies, it may also be possible to create better-balanced groups by stratifying patients into sub-groups; differences in the pathophysiology can be expected according to a patient’s atopic or smoking status, the co-existence of COPD, and disease severity (5-8).

For a given treatment protocol, the treatment period in a parallel-group study is shorter than in a cross-over study. When the treatment period is longer than 6 months, patient withdrawals may become a serious problem, and parallel-group studies may therefore be the more feasible option. However, performing biopsies on large numbers of patients may increase the complexity and cost of the study.

If the levels of the chosen inflammatory marker vary with time in the normal disease state, inclusion of a placebo group is essential (irrespective of study design). A placebo group can also help to quantify the population variability in studies where this is expected to be large. A beneficial placebo effect is often observed (9), which must be accounted for when interpreting treatment effects. It may be unethical, however, to withhold active treatment in long-term studies unless patients have very mild disease.

### Controlling variability

#### PATIENT NUMBERS AND STATISTICAL CONSIDERATIONS

The statistical plan of a clinical study which includes bronchial biopsies must address two common problems—multiple comparisons and large inter-patient variability. It is quite common for 10 or more markers to be measured in biopsy studies, but as the number of tests increases, so does the chance of finding a significant difference where one does not exist (Type I error) (10). For example, if eight tests are performed, the probability of making a Type I error (α) may be as great as one in three, depending on the degree of correlation between measures, and so adjustment methods must be used (11). The use of a Bonferroni correction to reduce α is not appropriate when the measurements being made are highly correlated (12). For a given inter-treatment difference, the only way to increase the power of the study without affecting the probability of a Type I error is to increase the sample size. The study protocol should include a calculation of sample size and study power (on the basis of the primary efficacy parameter), and it scrutiny of the study data shows that these are not achieved, the differences should be noted in the study report.

An additional source of variability in bronchial biopsy studies is the change in the disease process over time. In one group of 12 patients with stable asthma, the intra-subject differences in inflammatory cell counts (T-cell and eosinophil markers) between two biopsies taken 1 month apart was much greater than the intra-subject difference between upper and lower lobe samples, despite no change being observed in treatment, forced expiratory volume in 1 second (FEV₁), symptoms or bronchial hyper-reactivity (13). By analysing the observed variation in cell counts in this study, the authors concluded that 15 patients would be...
FIG. 2. Relationship between the number of patients enrolled in a biopsy study and the detectable difference between pre- and post-treatment biopsy cell counts (for three lymphocyte markers and an eosinophil marker), assuming a power of 80% to detect a difference significant at the $P < 0.05$ level. [Reproduced from Richmond et al. (1996) with permission (13)].

needed per treatment group in a parallel-group biopsy study to obtain 80% statistical power to detect a significant difference in inflammatory cell counts, and eight per group in a cross-over design (13); however, the differences detected would need to be large (Fig. 2).

The large variability is often evident in the form of large differences in baseline values of inflammatory cell count or concentration of markers; these may be greater than the difference that results from the study treatment (14). On such occasions, the most useful measure for comparison may be change from baseline (effectively normalising the baseline readings), but caution is still needed with interpretation. The variation in baseline readings may indicate that patients have different underlying pathologies.

VARIATION IN CONTRIBUTING PATHOLOGY

Although different asthmatic patients may have similar clinical signs and symptoms, the immunological, biochemical and structural abnormalities that contribute to the pathology of each patient may differ. Various alterations in the structure or function of cells or tissues may be present at any one time, yet the sum of all the abnormalities gives rise to similarity in the clinical expression of asthma. A patient only presents with asthma if the aggregate effect of these alterations is large enough to reach the threshold for clinical symptoms. This is illustrated in Fig. 3, where each contributing factor is represented by a letter. In patient 1, a combination of abnormalities in factors A and B is present, but does not reach a level sufficient to cause a clinically detectable change; however, patient 2, whose major abnormality is factor A, presents with asthma because the aggregate of the abnormal factors exceeds the threshold. Although the clinical presentation of patient 3 is identical to patient 2, the major underlying abnormality is B. Patient 4, despite no abnormality in A or B, also presents with asthma because of abnormalities in factors C, D, E and F. Despite their contrasting underlying pathological profiles, patients 2, 3 and 4 cannot be distinguished from each other simply on the basis of clinical measurements, e.g. FEV₁, airway hyper-reactivity or symptom measures.

When patients are recruited for clinical trials in which only clinical outcomes are measured, the use of clinical tests as inclusion criteria is appropriate. However, when a trial is designed to assess the capacity of a drug to correct one abnormal factor, recruitment according to results of clinical tests is inadequate, and is certain to increase the variability of responses to the agent under study. For example, a drug which was designed to reduce numbers or function of T-cells would be more likely to be effective in a patient with an abnormality in T-cell reactivity than a patient with normal cell reactivity. By adding an assessment of the specific mechanistic marker to the inclusion criteria, either directly or via a surrogate, those patients who are more likely to respond (at both the cellular and clinical level) can be distinguished, making the trial more productive. In order to achieve greater homogeneity of patient phenotype in clinical trials, validated surrogate markers in sputum, blood, urine or exhaled air are required for a variety of inflammatory cells and processes. These markers will allow appropriate subjects to be selected for specific investigations, and make research outcomes more reliable and reproducible.

PATIENT SELECTION

Inter-patient variability may also be reduced by stratifying patient groups in biopsy studies as described previously. An alternative approach to stratification is to perform a ‘run-in’ biopsy; only those patients whose levels of a target inflammatory marker exceed a set threshold are randomized to treatment. However, performing a separate biopsy
analysis in this way runs counter to best biopsy practice, i.e. preparing, sectioning and staining all biopsies from one study in a single operation. This approach will also increase regression to the mean, since some of the patients who meet the test criterion may have unusually high levels of the target marker compared to their normal range, and these are likely to fall subsequently. This will not cause a treatment group imbalance if patients are randomized after the result of the run-in test is known. As previously stated, it is easier to assess the scale of inter-patient variability when large numbers of patients are enrolled; subgroup analysis is also more practicable. The fastest way to recruit large patient numbers is to perform a multicentre study. Although there may be an increase in variability because of inter-centre differences in biopsy preparation and bronchoscopy method, the variability of methodology is small in comparison with that between patients. Indeed, inter-centre variability can be obviated by nominating one centre to process and analyse all biopsies, enabling the methodology to be better standardized. It is vital that study centre staff are properly trained to take and prepare biopsies, and that there is sufficient time and funding to produce high-quality results.

Large-scale clinical trials that include biopsy assessment are costly, but their value can be increased by 'banking' the biopsy samples after completion of the study. Samples taken from healthy volunteers or from asthma patients treated with placebo may be a source of historical control data for future work. In addition, banked samples can be re-analysed at a later date, either to assay markers that are subsequently characterized or discovered, or to make use of improved assay methods that may become available in the future.

Study duration

Deterioration in compliance with treatment and the loss of patients to follow-up impose a practical limitation on the duration of asthma studies; however, valuable information on the natural history of asthma may be obtained by performing repeat biopsies on patients for many years. In order to assess the long-term effects of treatment on inflammation and airway remodelling, a study should last for a minimum of 1 year, with biopsies taken at baseline, 3, 6 and 12 months. However, the study design should include adequate controls to account for the natural variability of the target inflammatory marker over this time period. Indeed, reproducibility of inflammatory cell counts from biopsies shows a pronounced decline as the interval between consecutive biopsies is prolonged (15).

Studies should not be so short that treatments have insufficient time to achieve clinical benefit. For example, 2-3 months of ICS therapy may be required before effects on some lung function and immunological parameters are observed (16,17). An observed effect of drug treatment on the number of cells in a biopsy sample may be made more definitive by measuring whether the cell count returns to normal after withdrawal of treatment. The interval between stopping treatment and taking the final biopsy should be carefully judged, to avoid provoking ethical objections if patients have severe disease.

Integrating biopsy assessments into clinical trials

The time points selected for biopsy assessments during a clinical study may introduce additional sources of variability. The diurnal variation of lung function measures such as peak expiratory flow is well recognized (18) and is routinely taken into account in the timing of clinical assessments. However, the time of day at which bronchoscopy is performed also has an influence on the test result; inflammatory cell counts in the alveoli of patients with nocturnal asthma were shown to be significantly higher at 04.00 hours than at 16.00 hours in biopsies taken 1 week apart (19).

Results from a controlled parallel-group study that featured clinical and biopsy assessments suggest that the timescale of changes at the clinical and cellular level are different. In that study, biopsies were taken at baseline, 2 and 8 weeks from 27 asthmatic patients treated with either fluticasone propionate (FP) 2 mg/day or placebo (17). Significant improvements from baseline were recorded in spirometry variables after 2 weeks in the FP group compared with placebo, but there was no further improvement at 8 weeks. However, reductions in the number of inflammatory cells followed different time courses; after 2 weeks, the number of T-cells and primed T-cells (expressed as a ratio to the baseline count) was significantly lower in the FP group than the placebo group, but the corresponding change in activated eosinophil (EG2\textsuperscript{+}) numbers took 8 weeks to reach significance.

The exact timing of clinical assessments such as the 

Safety

Bronchoscopy is a potentially hazardous procedure: the most common problem is bronchospasm and associated dyspnoea, but there is a small risk of fever, bleeding and pneumothorax (21). Exposing the patient to these risks must always be justified by the potential value of the research results. If the target variable can be satisfactorily measured by using another safer assessment method, bronchoscopy should not be performed.
This risk analysis is particularly important for patients who will receive placebo.

Consideration should also be given to the patient’s comfort during and after bronchoscopy. The use of the short-acting general anaesthetic propofol to sedate patients during bronchoscopy is associated with faster recovery and a lower incidence of drowsiness after the procedure than sedation with midazolam (C. Burke, personal communication), although the product licence for propofol stipulates the attendance of an experienced anaesthetist.

Study centres should have established procedures to ensure adherence to appropriate guidelines for asthma studies and for bronchoscopy. For example, clinicians who perform bronchoscopies for investigating centres should participate in an ‘on call’ system to ensure that expert help is permanently available by telephone to patients or volunteers enrolled in the study, their family doctors and the hospital’s emergency department. If the patients are not well acquainted with the medical or paramedical staff, the study team should develop a system for follow-up calls to the patients at 12 and 24 h after bronchoscopy.

Conclusions

Bronchial biopsy studies need to be carefully designed and executed at the clinical and physiological level. It is not ideal for biopsy assessments to be simply added on to an existing clinical trial protocol.

When planning a study, the main elements of the protocol, e.g. design, scale, patient selection, duration and assessment methods should be considered with particular reference to the special requirements of biopsy. It is important that a primary biopsy variable is defined for confirmatory studies, as well as the size of effect that is sought. Measures should be taken to control variability, and when assessment time points are selected, allowance must be made for differences in the time course of changes at the cellular and physiological levels. By paying sufficient attention in advance to these elements of the design process, the quality of the study will inevitably be improved.

References
