Methods for the Assessment of Endobronchial Biopsies in Clinical Research

Application to Studies of Pathogenesis and the Effects of Treatment

PETER JEFFERY, STEPHEN HOLGATE, AND SALLY WENZEL ON BEHALF OF THE ENDOBRONCHIAL BIOPSY WORKSHOP AUTHORS

CONTENTS

I. Obtaining Optimal-Quality Endobronchial Biopsies
II. Processing and Methods of Visualization
III. Variability and Quantification
IV. Endobronchial Biopsy versus Bronchoalveolar Lavage versus Induced Sputum
V. Overall Conclusions

The capacity to sample directly the conducting airway tissue of subjects with asthma or chronic obstructive pulmonary disease (COPD) for comparison with tissue obtained from normal, healthy individuals has generated hypotheses and advanced our understanding of pathogenesis. Such studies in adults are now being extended, with informed consent and ethical approval, to include children. In adults, approval to obtain endobronchial biopsies on more than one occasion has also allowed us to begin to test whether or not therapeutic intervention will significantly reduce airway inflammation or restore altered tissue structure (i.e., remodeling) to normal.

Endobronchial biopsy has been considered by some the standard by which to compare the validity and interpretation of results obtained by indirect, less invasive means, such as bronchoalveolar lavage (BAL) and induced sputum.

Generating Hypotheses

Asthma and COPD are recognized as inflammatory conditions of the airways, albeit with distinctive patterns of inflammation and tissue structure (1). A variety of endobronchial biopsy methodologies are being used that may compromise valid interpretation and confirmation of previously published findings. For example, in asthma, increases in connective tissue deposition are variably reported in association with severity of disease (2, 3), and controversy remains as to whether or not there is loss of surface epithelium in asthma (4, 5).

The quality of the biopsy obtained predetermines the quality of the staining and the reliability of the data generated. Several factors contribute to quality: the number of biopsies obtained and their location, size, and morphological integrity, which may often vary from research center to research center. There are numerous protocols for tissue fixation, processing for histology, visualization by immunostaining, in situ hybridization (ISH), and electron microscopy. Variability introduced by different methods of quantification is compounded by the variability that exists naturally between subjects, patient groups, and airway generations, as well as between individual biopsies and between sections of each biopsy from any one individual. Traditionally, qualitative examination of endobronchial biopsies, often with the aid of semiquantitative scores applied in a blinded manner, has led to the generation of hypotheses. Such observations have been strengthened scientifically by more rigorous use of quantitative techniques, and there are several available.

Testing Hypotheses

Investigators have also used data obtained from biopsies to test hypotheses, and treatment-related drug effects have been published (6–14). It is not clear, however, how many such clinical studies have failed to find treatment-related effects, whether because of inadequate patient numbers or because of insufficiently or inappropriately prepared or sampled tissue. Endobronchial biopsy samples are small and, moreover, the ‘signals’ obtained from them are also small and frequently may be lost within inherently large biological variability. Thus, the analyses used need to have the capacity to discriminate subtle yet clinically important differences between treatment groups.

Study Design

Such analyses almost always benefit from the standardization of study design, avoidance of potential sources of bias, and attention to the validation of reliable endpoints, all part of well-established principles in study design (15–28). Effective a priori planning followed by strict adherence, via quality control, to the specified study design and measurement protocols is essential. Critical components include identifying clear study aims and objectives, defining a hypothesis, identifying the primary and secondary objectives, assuring experimental standardization, and selecting valid endpoints. Such information will be needed to ensure that the sample size is adequate. To avoid controversy about study interpretation, a complete protocol should be in place before the study is initiated; the data analysis plan (see APPENDIX) also should be in place before the study is initiated, or at least before the study is unblinded.

It is crucial that standard procedures (detailed in a manual) for conducting each component of the study be defined and adhered to via regular evaluation, especially in multicenter trials, to reduce variability caused by procedural differences (29). Individual study sites should be trained and certified in the proper performance of these procedures. Before initiation of the study per se, pilot biopsies outside the study should be evaluated systematically from each center. Ideally, in multicenter studies, biopsy endpoints should be measured in a single central laboratory. Endpoints whose reproducibility and responsiveness are known should be chosen and their variance, effect size, and ultimately, sample size should be estimated. Frequently, pilot or validation studies are completed before designing the final study.

Aims and Objectives

The goal of the international workshop reported here was to produce a document that would assist and guide investigators in making cross-comparisons of biopsy data. Our aim was to compare and discuss the methods currently used worldwide to obtain and analyze endobronchial biopsies in asthma and COPD and to work toward standardization of methodology.
Written documents were presented to and discussed by the entire group under the leadership of moderators and, based on the discussions, the written documents were revised. Four topics, having the following specific purposes, formed the core of the workshop:

1. **Obtaining optimal-quality endobronchial biopsies:** To discuss and recommend the best methods of safely obtaining good-quality endobronchial samples of adequate size for research in adults and also to consider the special issues related to endobronchial biopsy in children.

2. **Processing of the biopsies and methods of visualization:** To outline and record the advantages and disadvantages of different methods of processing tissue samples and of visualizing and expressing the biopsy parameters related to airway inflammation and remodeling.

3. **Variability and quantification:** To discuss the current practices of quantification to understand the limitations imposed on the interpretation of such data and to debate and clarify the use of alternative, less biased methods for the quantification of inflammatory and structural cells.

4. **Advantages and limitations of endobronchial biopsy, BAL, and induced sputum:** To highlight the distinct yet complementary data that can be gained from analyses of endobronchial biopsies, BAL, and induced sputum and to discuss how these data might relate to the clinical status of subjects.

Finally, not only can endobronchial biopsies provide information about treatment-induced changes in inflammation and remodeling, but they can also be used to isolate resident cells and to study their phenotype (30). These in vitro experimental models allow for the study of cell interactions under conditions as close as possible to their natural environment (31).

We hope that this report represents a step toward improving the reliability of testing hypotheses using airway tissue obtained by biopsy.

### I. OBTAINING OPTIMAL-QUALITY ENDOBRONCHIAL BIOPSIES

The report presented here relates primarily to research bronchoscopies aimed at obtaining endobronchial biopsies under direct vision. While we do comment briefly on transbronchial biopsies, these are not recommended routinely for research.

### Ethics

All research involving bronchoscopy must have approval by an institutional review board/ethics committee. Ethical considerations include bronchoscopy procedure, the appropriate use of samples, and the scientific validity of the study.

**Informed consent and the right to withdraw.** The subject must be fully informed about the study and of any potential risk and discomfort. This is often provided as an information sheet. In some countries, it is important to inform subjects that the samples will be archived. In both children and adults, when the bronchoscopy is being done for clinical reasons, but additional procedures are to be performed for research purposes, the ethical considerations described here also need to be followed.

**Use of biopsy samples.** Ethical concerns will vary locally and will relate both: (1) to the use of samples for purposes not originally defined in the consent form and (2) to the failure to use fully valuable archived tissue samples, a failure that would needlessly require additional volunteers to undertake bronchoscopy for the purposes of research.

### Safety

Research bronchoscopy on volunteers is an invasive investigation that usually does not improve the volunteers’ welfare directly. Thus, safety is of paramount importance (32). Although numerous reports suggest that the technique is well tolerated, even down to an FEV₁ as low as 30% of predicted (33, 34), caution should be exercised in those with a low FEV₁, particularly in subjects at risk of sudden severe bronchospasm. All research bronchoscopies should be indemnified.

### Bronchoscopic Procedures and Monitoring

Bronchoscopy needs to be performed by experienced bronchoscopists, preferably of specialist standing in appropriately equipped laboratories. Respiratory physicians in training can undertake the procedure under close supervision or after a period of training and accreditation. Research bronchoscopy is not to be used as part of the routine clinical training program (32, 35).

Continuous vigilance and monitoring throughout the procedure are essential. Subjects need to be premedicated with bronchodilators unless specifically contraindicated by the study protocol. Atropine is often used to dry secretions and decrease cough. Premedication often incorporates an opiate, such as codeine, papaveretum, or fentanyl, which may be used to suppress the cough reflex. In adults, oxygen is routinely supplemented. The use of sedatives during bronchoscopy varies. In some centers, no sedation is used; in others, benzodiazepines, such as midazolam or diazepam, are used for sedation, sometimes in conjunction with a narcotic, such as fentanyl.

Before the bronchoscope is inserted, local anesthesia with lidocaine is applied in the nares and the oral pharynx to suppress sensation, cough, and the gag reflex unless the procedure is being performed under a general anesthetic. Lidocaine is a potentially toxic drug that can cause arrhythmias and seizures. There has been only one report of death that was thought to be due to lidocaine toxicity following a research bronchoscopy (36). Lidocaine may obscure recognition of laryngomalacia; consequently, the larynx needs to be examined before topical anesthesia is applied (37). The bronchoscope is passed nasally or orally, depending on local anatomy or the size of the bronchoscope. Local anesthetic is given via the bronchoscope at the vocal cords and throughout the tracheobronchial tree.

There are no set protocols for how lidocaine is applied (38–44); thus, comparison of the various studies of lidocaine levels is difficult. It may be given in a nebulized form (1,500 mg) prior to bronchoscopy (45), or as a gel or solution applied to the larynx and bronchi (1–4%). Premedication with opiates may reduce the dose of lidocaine required. Lidocaine is metabolized in the liver with a half-life of 100 to 120 minutes (39, 43). Some manufacturers recommend that the maximum dose be less than 300 mg and 4.5 mg or less per kilogram of body weight (46). Lidocaine, however, has been given safely at 9.3 mg/kg of body weight, and only occasionally exceeded the toxic level of 5 μg/ml (40, 41). Nevertheless, it is important to use as low a dose as possible and to keep a careful record of the lidocaine use.

During all research bronchoscopies, there needs to be an adequately trained physician and nurse or other assistant able to terminate the procedure if the clinical risk becomes unacceptable or the subject is distressed. In addition to the availability of full resuscitation facilities, benzodiazepine and opiate antagonists (flumazenil and naloxone) need to be at hand to reverse respiratory depression. Noting of clotting times and platelet counts is not recommended routinely prior to bronchoscopy, but should be undertaken if there are clinical features that raise concern (these measures should always be obtained prior to transbronchial biopsies).
BAL can provide beneficial information that complements the results of biopsy. However, BAL increases the duration of the procedure and, in a small proportion of individuals, results in cough, fever, or pleuritic pain after bronchoscopy. The exact frequency of these complications will vary according to the volume of BAL and the subject group, but frequencies of fever and pleurisy of between 2 and 5% have been reported. The lavage procedure is more likely than biopsy alone to induce arterial oxygen desaturation and bronchospasm. For these reasons, if both BAL and endobronchial biopsy are contemplated, it is preferable to take the biopsy samples first and then perform the BAL procedure in a different lung or lobe to avoid hematological contamination, especially if parameters measured in biopsies are the primary outcome variable. If, however, a BAL measurement is the primary outcome variable, this procedure should be performed first.

A physician or a specialist nurse is to see all subjects before discharge, and subjects need to be given a 24-hour contact telephone number. Also prior to discharge, subjects should be warned about the effects of topical anesthesia on swallowing and of sedation on driving.

Obtaining the best endobronchial biopsy samples. Although the best results are obtained using larger instruments, this goal needs to be balanced by the subject’s comfort and safety. All aspects of the protocol and the processing of the biopsy samples are to be defined clearly before the study begins. Facilities for immediate handling and adequate storage are essential. It is important to set an order of priority in terms of sample processing to ensure that at least the primary outcome measures are obtained in case the sampling is incomplete.

Endobronchial biopsy site. Endobronchial biopsies taken under direct vision will, by definition, be from proximal airways, usually from subsegmental and segmental subcarinae and not from the lateral wall. This would usually be from the second to the fifth generation of airway branching. The choice of biopsy site needs to be defined by the protocol, especially in studies in which more than one sampling session over time is required. Because the consequences to the site of an endobronchial biopsy sample are not known, repeat samples (e.g., to assess the impact of intervention) should not be taken from the same anatomic site. Preintervention and postintervention biopsies may be randomized to each lung or come from the same airway generations in, for example, biopsies taken before and 24 hours after segmental allergen challenge.

Obtaining samples. Although no formal studies have been done, there is consensus among researchers that it is most difficult to obtain samples from normal volunteers and perhaps easiest to obtain them from subjects with asthma or COPD.

There is no systematically obtained evidence on the best way to take an endobronchial biopsy sample. Accumulated evidence suggests that reusable forceps should not be used in more than five bronchoscopies. Early experience with disposable forceps suggests a good yield. In principle, the forceps need to straddle the carina but not with excessive force, as this might increase artifactual damage. The forceps need to be closed firmly and held in place for a few seconds to allow them to grip before withdrawing. Biopsies performed in one lung usually yield eight samples. A minimum of two samples per processing procedure (e.g., embedding them in paraffin or resin or snap-freezing them) is needed to ensure at least one good sample for analysis. If multiple specimens are to be obtained, it is easiest to begin sampling in the most distal carina and subsequently to sample more proximally. This reduces the possibility of any bleeding obscuring the view. If poor samples are obtained at gross examination, consider changing the forceps. There have been no reports of excessive bleeding in otherwise healthy subjects. Therefore, use of epinephrine is not indicated. Alternatively, brush biopsy may be used to sample the endobronchial epithelium (47, 48).

Checking samples. It is important that an assistant be present at the bronchoscopy to check that the endobronchial biopsy sample is of an adequate size and quality. Rapid processing of samples and standard staining in the laboratory allow feedback as to the adequacy of the biopsy technique, which, in turn, allows for improvements to be made if needed during the course of the study.

Research Bronchoscopy in Children

All research involving bronchoscopy in children must have approval by an institutional review board/ethics committee. Ethical considerations include bronchoscopy procedure, the appropriate use of samples, and the scientific validity of the study. In young children, bronchoscopy cannot be performed solely for research purposes (49), but there is no ethical contraindication to performing an endobronchial biopsy if the child is undergoing a bronchoscopy for a clinically indicated purpose or is being intubated for another procedure, such as surgery, that permits the bronchoscope to be passed down the endotracheal tube. Indeed, research endobronchial biopsy in children has been well described and performed safely (49–53). If a research transbronchial biopsy is contemplated in a child, it is necessary to consider very carefully the increased risks of this procedure over direct endobronchial biopsy and the likelihood of obtaining worthwhile information before requesting approval for such studies (54).

A recent working group of the European Respiratory Society proposed that when general anesthesia or bronchoscopy is being performed for clinical purposes, it would be unethical not to consider taking an endobronchial biopsy and/or BAL sample after obtaining the appropriate informed consent (55). Even in children with severe asthma, bronchoscopy and airway biopsy may be performed safely under carefully controlled conditions (56, 57). Endobronchial biopsy (both forceps and brush) is a routine clinical procedure in, for example, bronchial tuberculosis in children. A series of several hundred samples taken at the time of a clinically indicated bronchoscopy in children aged 6 months to 15 years of age demonstrated that the procedure can be done without significant complication (53, 56, 58).

Infants and children may be examined under deep sedation or full general anesthesia (56, 59, 60). There are national and international guidelines for the safe administration of anesthesia and sedation (37, 61–66).

Pulmonary function should be monitored before and at 30-minute intervals after the procedure until it is within 10% of baseline (subjects are often too sleepy to give good efforts 30 minutes after the procedure). This recommendation cannot apply to children too young to have spirometry performed; in these children, pulse oximetry is usually relied upon. Intravenous atropine usually is administered to adults as part of the premedication to dry secretions and suppress laryngeal reflexes. Many anesthesiologists will choose not to give a premedication to children (61).

Procedural recommendations. During the procedure, it is essential that the child be monitored by a person independent of the bronchoscopist, usually an anesthesiologist, someone who has taken the American Academy of Pediatrics Life Support course, or an equivalently trained pediatrician. The procedure needs to be terminated immediately if any decompenbination in the child causes clinical concern to the independent subject advocate.

The largest bronchoscope that is safe should be used. This is generally a 4.9-mm endoscope (2.2-mm biopsy channel) in children at least 5 years of age and a 3.6-mm endoscope (1.2-mm biopsy channel) in younger children. The size of the broncho-
scope used should be determined by the subject’s age (bronchoscope diameter [mm] = 4 + age [years]/4) and the extent of reduction of airway caliber (airway area for ventilation = [(D – d)(D + d)/D]%), where D is the diameter of the airway and d is the diameter of the bronchoscope). Thus, a 3.6-mm bronchoscope in a 5-mm airway occludes nearly half of the airway. Other considerations are similar to those in adults.

**Subjects with COPD**

In view of their more advanced age and their potential for comorbid conditions, such as ischemic heart disease, all COPD subjects should have a prebronchoscopy ECG and, optimally, continuous ECG recording during bronchoscopy. Subjects known to have ischemic heart disease must be excluded from research bronchoscopies. In more advanced disease, where there is risk of CO₂ retention, arterial blood gases should be measured and subjects with significant hypercapnia must be excluded (67).

**Transbronchial Biopsies**

Transbronchial biopsies should be performed only if the value of the information to be obtained outweighs the risk of the procedure as determined by the investigator and ethics committee. Because there is a lack of information about repeatability, the transbronchial technique cannot be recommended at this time for interventional (pharmacologic or challenge) studies. Transbronchial biopsies have, however, been taken for research (33, 68–70) and have provided information about inflammation in the lung parenchyma. In a series of more than 100 bronchoscopies, there was a report of one case of pneumothorax, which was treated conservatively.

The limited number of studies that have been done have used fluoroscopy to appropriately position the biopsy forceps. The routine use of fluoroscopy has not been shown to improve the safety of the procedure (71). The samples are generally obtained from the lateral segments of one lower lobe and are likely to contain both alveolar and peribronchial tissue. Samples can be taken from the standard subpleural position if one is interested in obtaining primarily parenchymal tissue or more proximally (two thirds to three quarters of the distance to the pleura) if the study is directed at smaller airways. Airway tissue is obtained in approximately 35 to 40% of samples. It is recommended that no more than four samples be taken for research purposes. A small-bore chest tube needs to be immediately available in case of pneumothorax. A postprocedure chest X-ray needs to be obtained. Subjects should be instructed to avoid the use of aspirin for 2 weeks prior to the procedure. Epinephrine should also be available in case of significant bleeding. In the opinion of the authors, this procedure can be performed for research purposes but only in centers with considerable research bronchoscopy experience. The safety of this procedure needs to be reviewed regularly. Pooling of data from different institutions would help to clarify the role of this procedure in research.

**Audit and Recommendations for Research Bronchoscopies**

Considerable progress has been made in defining the best way to sample the airways, and this has revolutionized the concept of airway inflammation. However, continued audits, preferably using pooled data, will further the development of safer and more effective procedures. In the interests of safety and the ethical practice of this procedure, researchers are encouraged to report both positive and negative experiences with endobronchial biopsies. It is recommended that these reports be audited by panels of experts made up of representatives of relevant national and international societies.

**II. PROCESSING AND METHODS OF VISUALIZATION**

Both the bronchoscopist and laboratory representatives need to participate directly in the initial planning. Prior to the study, pilot biopsies (i.e., samples not part of the study per se) should be taken to control for the adequacy of biopsy size and quality. The histology laboratory should ideally report back as soon as possible to the bronchoscopist on biopsy integrity and its viability for immunohistologic, molecular, and ultrastructural studies as appropriate to the study protocol. There may be logistical issues when there is a single central reporting laboratory in a study of multicenter/multicountry design.

**Biopsy Quality**

For the purposes of research, an evaluable sample in adults should ideally have a subepithelial area, excluding crush artifact, cartilage, or blood clot, of at least 0.3 to 0.5 mm (72, 73). Samples of blood and mucus, scrapings of epithelium, and fragments of glands are unacceptable. If there are to be three distinct processing techniques, it is recommended that a minimum of six samples be taken from each subject, two samples for each processing technique. Endobronchial biopsy samples should not be placed in sterile phosphate-buffered saline as an interim holding procedure but must be fixed or snap-frozen immediately following their removal from the patient.

**Tissue Fixation and Embedding**

Methods used for tissue fixation and processing vary among researchers depending on study endpoints, local expertise, and experience (74–76). Whatever the method, there needs to be a standard operating protocol adhered to throughout the study. This is particularly important when comparing samples taken before and after challenge or therapeutic intervention. Fixation time needs to be long enough to preserve tissue morphology but not so long as to mask antigens, reduce accessibility, or destroy messenger RNA (mRNA)/DNA. Typically, fixation times vary between 4 and 24 hours, and for ISH, fixatives need to be free of RNAses and DNAses (77, 78).

Snap-freezing tissue. For immunohistology, immediate snap-freezing of freshly obtained samples in melting isopentane (i.e., previously cooled in liquid nitrogen) and cryostat sectioning has the advantage of a rapid processing method (74–76, 79–81). An alternative is embedding the samples in optimal cutting temperature compound prior to freezing. Freezing enables all antigens of interest to be detected by immunohistochemistry (IHC). If ISH is to be performed on such snap-frozen samples, prefixation using freshly prepared 4% paraformaldehyde for 4 hours at room temperature and cryoprotection in 15% sucrose in phosphate-buffered saline is recommended. Fixation prior to freezing enhances morphology but can compromise IHC; for example, cell surface adhesion molecules and activation markers such as CD25 may be lost. Snap-frozen samples are usually cut at a thickness of 5 to 7 μm.

In comparison with samples embedded in paraffin or glycol methacrylate (GMA), frozen tissue provides relatively poor morphology because of its greater section thickness and has the potential disadvantage of ice crystal artifact. In the absence of prior fixation, there may be unwanted diffusion of antigens and proteins, and there is a need for low-temperature storage (–80°C in a freezer or storage in liquid nitrogen). In cut and stored sections, some antigens are lost quite rapidly. Also, because of the thickness of these sections, a smaller number can be prepared from each sample when compared with paraffin and GMA, both of which use thinner sections. Collaboration with centers world-wide requires the use of special thermal packaging for shipping.
TABLE 1. FEATURES OF DIFFERENT METHODS OF ENDOBRONCHIAL BIOPSY FIXATION AND EMBEDDING GIVEN OPTIMUM TISSUE PROCESSING

<table>
<thead>
<tr>
<th>Embedding Medium</th>
<th>Recommended Fixation</th>
<th>Morphologic Preservation (Graded)</th>
<th>Ease of Use</th>
<th>Antigen Preservation: IHC</th>
<th>Cytoplasmic and Nuclear Markers</th>
<th>Surface Markers</th>
<th>Cytokines and Chemokines</th>
<th>Adhesion Molecules</th>
<th>ISH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Snap-frozen</td>
<td>Acetone or methanol</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Fixed, then snap-frozen</td>
<td>4% para-formaldehyde</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Paraffin</td>
<td>10% neutral buffer formalin or 4% para-formaldehyde</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>GMA</td>
<td>Acetone</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>0</td>
</tr>
</tbody>
</table>

Definition of abbreviations: GMA = glycol methacrylate; IHC = immunohistochemistry; ISH = in situ hybridization.

Scale is from 0 to 5, where 0 = not detectable, 1 = poor, and 5 = excellent.

of frozen biopsies, but samples cannot be guaranteed to remain frozen during transit and customs inspection.

Paraffin wax embedding. Processing samples into paraffin wax (77), usually following fixation in 10% neutral-buffered formalin for IHC (77) or 4% paraformaldehyde or 10% neutral-buffered formalin for ISH (78), offers some advantages over snap-freezing (74–76, 82, 83). First, tissue architecture is better preserved. In addition, there is little or no diffusion of antigens, specialized storage facilities are not required, and facilities for processing are universally available. Finally, there are no difficulties of tissue preservation during transport of samples between laboratories, and both IHC and ISH procedures can be performed using paraffin-embedded material.

There are also some disadvantages to paraffin embedding. Antigen retrieval techniques may be required due to use of cross-linking fixatives that may mask selected antigens from detection by IHC; some antigens may thus not be retrievable. Tissue shrinkage occurs and is often ignored; this is usually of the order of 10%, but can be as much as 50%, depending on the temperature and the organic solvents used during the processing and embedding procedure.

Plastic resin. A hydrophilic resin, GMA has recently been employed for IHC (84) and applied to numerous studies (70, 85, 86). When coupled with acetone fixation (with protease inhibitors), GMA offers some advantages over frozen and paraffin preparations. Acetone is a non-cross-linking fixative that, together with processing at a low temperature, ensures antigen preservation; all antigens demonstrable in frozen samples can also be seen in GMA but, compared with frozen samples, morphologic preservation is greatly improved. This is because sections are only 1 to 2 μm thick and provide essentially a single focal plane when viewed through the light microscope. Due to their thinness, adjacent sections can show more than one cut through a single cell, enabling the detection and “co-localization” of more than one epitope to a cell. Relatively large numbers of sections can be cut from a single sample, allowing for comparatively more markers to be investigated than with the other embedding techniques. As with paraffin embedding, GMA-embedded samples can be stored safely for long periods. However, there are disadvantages: GMA embedding is more time-consuming and less widely used than either frozen or paraffin preparations and is not yet suitable for studies of RNA/DNA and their visualization by ISH.

The methods used for sample fixation and embedding are summarized in Table 1 with ranked scores of preference for their use when visualizing particular cellular characteristics.

Immunohistochemistry

A dedicated laboratory and the support of an experienced immunohistochemist are needed. For batch staining of biopsy samples in clinical trials, an automatic staining machine is favored to ensure consistency and good quality of immunostaining over time. A high-quality microscope with photographic capability and access to integrated, interactive image analysis equipment is recommended. Various IHC methods have been reviewed and described in detail (77, 87, 88). Both monoclonal and polyclonal antibodies can be applied. The initial use of pilot tissue and positive controls known to contain the antigens of interest is recommended so that each antibody can be titrated by limiting dilution to provide an optimal working concentration (i.e., to obtain the best signal-to-noise ratio). Positive and negative controls should always be included with each subsequent run of tissue samples. Negative controls include omission of the primary antibody, using an isotype-matched Ig, or preabsorption of antigen prior to the IHC procedure. Although the methods of visualization for IHC are many and varied, most groups use horseradish peroxidase, alkaline phosphatase, or fluorescent two–three stage detection systems. If an enzymatic detection system is used, endogenous tissue enzymes must be blocked (89, 90). The following techniques are commonly available: streptavidin–biotin conjugates (91) (full horseradish peroxidase or alkaline phosphatase), alkaline phosphatase antialkaline phosphatase (92) (for monoclonal antibodies only), EnVision (horseradish peroxidase or alkaline phosphatase) (93), and fluorescence labeling. New techniques, such as immunoperoxidase chain reaction (94) are now being validated for the detection of weakly expressed antigens. Detailed written standard operating protocols and standardized criteria for identifying and measuring the immunostained parameter of interest should be agreed upon and followed consistently, with appropriate quality control checks during the study.

In Situ Hybridization

ISH requires an investigator with extensive experience in the field and dedicated facilities. There is an absolute requirement to minimize RNAse contamination, which is ubiquitous, if riboprobes are to be used (95). If using radiolabeled probes, appropriate safety equipment and procedures are essential, including approval for radioactive work and a well-equipped darkroom. A microscope fitted for observation of fluorescence as well as dark- and bright-field viewing is recommended (90, 95, 96).

In order to visualize and localize mRNA, different types of probes are available, including DNA oligonucleotide probes and complementary RNA probes. The most commonly used RNA probes (97) offer the advantage of being more sensitive and specific for the detection of a particular mRNA. The probes can be labeled with radioisotope or with biotin, digoxigenin, or fluorescein isothiocyanate. Radiolabeled probes provide great sensitivity and should be used to detect mRNA encoding of actively synthesized proteins that are rapidly turned over and
Electron Microscopy

Because it is limited by the wavelength of light, bright-field microscopy has a resolution of up to 0.25 μm and a maximum useful magnification of 2,000 ×. In contrast, transmission electron microscopy (TEM), which uses electrons rather than light, allows much greater resolution of biological samples, of the order of 1 to 2 nm, and magnifications greatly in excess of 200,000 ×. However, TEM requires an investment in expensive equipment and a dedicated laboratory with appropriately trained staff to process samples and to operate and maintain the equipment.

An advantage of TEM over immunologic methods is that tissue components and cells are all simultaneously visualized and may be quantified in one and the same section. In particular, they require skill and a maximum investment in expensive equipment, particularly of epithelia. The major advantage of nonradioactive probes is that the immunolabeled mRNA hybrids can be visualized by bright field microscopy without the need for autoradiography. Also, unlike the autoradiography of radioisotopes, which take a week or more to develop, nonradioactive probes can be viewed immediately after the IHC detection procedure. Double-labeling techniques in which the ISH procedure is followed by IHC can be used to identify the cell expressing a gene of particular interest. Appropriate controls are essential and need to include the complementary sense, an irrelevant sense, or an alternative irrelevant mRNA probe, preferably of similar length to the antisense probe for the mRNA of interest. RNAse pretreatment (prior to hybridization) can be used in addition as a negative control. Positive controls must also be included when optimizing each probe.

An experienced investigator with knowledge of endobronchial biopsy, morphology, and ISH should interpret the results. A number of inflammatory cells, including eosinophils and macrophages, can bind probes nonspecifically, especially if radioactive probes are used. In the case of radioactive probes, caution is to be exercised in interpreting the autoradiographic signal at the edge of the section and at the inner surface of blood vessels or airways (because of the “edge effect”) (102). There have been correlations measured between clinical pulmonary function and mRNA expression of inflammatory mediators (95, 103–105). The absence of a hybridization signal does not necessarily indicate the inability of cells to produce a specific protein. In the case of a transient protein product, mRNA can be present in low copy number and may be beyond the sensitivity of the ISH technique (106). Moreover, posttranscriptional and posttranslational events may affect gene expression. Small differences in the number of ISH-positive cells may have important functional and immunoregulatory implications.

III. VARIABILITY AND QUANTIFICATION

Depending on the nature of the research question, semiquantitative or quantitative methods may be used. Indeed, there is evidence that a semiquantitative approach can provide results and interpretation similar to those derived from morphometric evaluation in, for example, assessments of lung fibrosis (123). Having said that, the basic equipment required for the objective application of quantitative methods is not extensive and has already been used in pulmonary medicine to quantify both epithelial and interstitial cells (123–129).

Counts of Inflammatory Cells: Area Profiles

The standard practice of counting the number of cut cell profiles of interest in a tissue section using light microscopy and normalizing these counts to submucosal area or to length of the epithelial reticular basement membrane continues to be a popular quantitative approach. The methodologies involved in some of these procedures and their application to assess inflammation in bronchial biopsies have been summarized and discussed previously (130, 131).
The probability of visible cells being counted in a two-dimensional section is proportional to their size, density, and orientation, thus introducing a bias in favor of the larger cut profile (Figure 1) (132). However, assuming that nuclear size varies less than overall cell size, opting to count only cells whose nucleus appears in the plane of the tissue section theoretically reduces this bias. Application of stereology and assessment of the cell’s “volume density” is a simple and reproducible alternative that has already been applied successfully to endobronchial biopsies (18, 19). There are as yet, however, no reports of the use of stereology in the assessment of treatment effects (i.e., hypothesis testing).

**Stereology**

The present authors recommend investigators consider the application of stereology, a less biased sampling strategy than that used thus far. Those interested are recommended to Howard and Reed’s *Unbiased Stereology: Three Dimensional Measurement in Microscopy* (133), a well written, easy-to-read book that contains an excellent discussion of sampling and stereology-based morphometric techniques (see Chapter 5, pp 69–106). Stereology is the study of the three-dimensional properties of objects usually seen in two dimensions. Thus, by using stereology-based techniques, one can obtain quantitative three-dimensional information from measurements in the two dimensions of histology.

**Key Concepts**

**Randomness.** Because inflammation and remodeling in the airway in asthma and COPD are likely to be unevenly distributed in the airway mucosa throughout the bronchial tree, it is essential that biopsy sampling and the selection of biopsy sections and microscopic areas (fields) of tissue within biopsy sections be random. Random sampling eliminates bias and maximizes accuracy in the data generated. Randomness cannot be assumed; sampling strategies to ensure randomness need to be built into the measurement protocol.

**Variability.** Variability is introduced at the level of the subject, the biopsy (block), the biopsy section, the microscopic field of tissue, and the specific measurement used. Using the stereologic technique of “point counting,” the relative contributions of variance at these five levels have been estimated (134) and include 70% variability between subjects, 20% between blocks, 5% between sections, 3% between fields, and 2% between measurements.

It is important for investigators to acquire data in their own laboratories on the within- and between-observer variability for key outcomes and for the between-subject, between-biopsy, between-section, and between-field variability for these outcomes. Information about variability is essential for estimating required sample size for clinical studies involving bronchoscopy. For example, when Carroll and colleagues (135) examined airway tissues obtained postmortem from asthmatic persons, they demonstrated within- and between-observer coefficients of variation of approximately 1% for measurements of the dimensions of tissue structures and as high as 60% for counts of inflammatory cells. Additionally, there are examples in the pulmonary literature that demonstrate approximately 10 to 25% error (defined bias) in developing and injured airways using appropriate stereologic methods (125, 127). The variation found for counts of inflammatory cells and measures of reticular basement membrane thickness in endobronchial biopsies in subjects with asthma or COPD are on the order of 1 to 6%, and the variations between airway generations and subjects have been reported (72, 80, 136–139). High variability may preclude meaningful interpretation of the data in a study designed to test the effect of a treatment. This has led recently to a change in the current practice of analysis of one or in some cases two biopsy samples from which several sections and fields are quantified, to at least two biopsies per fixative, more subjects (as many as is practically and ethically possible), and relatively fewer sections and fields per biopsy. This change in approach to the quantitative analysis is captured in the saying “do more less well.”

**Tissue orientation.** There is relatively little bias in volumetric tissue analysis when “point counting” is used to establish volume fractions and the “disector” method is used to determine cell number (133) (see *Number Estimation by the Disector Method*, below). While not applicable to measurements of tissue volume or cell number, unbiased measurement of certain features that are orientation-sensitive, such as thickness and surface area, requires that the sections be random in their three-dimensional orientation (140, 141). As endobronchial biopsy tissue tends to curl on sampling, because of inherent tissue elasticity, it favors a random orientation of tissue structures included during embedding and in subsequent sectioning. This topic, referred to as isotropic randomization, can be complex and is reviewed clearly and simply in *Unbiased Stereology* (133).

**Application of Concepts to Practice**

**Obtaining and processing endobronchial biopsy samples.** To obtain a random sample, the authors recommend that a list of 10 to 15 potential biopsy sites located in the second, third, and fourth airway generations of the right or the left lung be generated before initiating the procedure. During bronchoscopy and moving in a distal (caudal) to proximal (rostral) direction, six to eight sites selected at random from this list need to be considered. Ideally, both lungs would be sampled equally, but potential complications secondary to endobronchial biopsy (bleeding and, rarely, pneumothorax) limit this possibility, and thus, only one lung should be biopsied during a bronchoscopy. If, for example, two fixation procedures are to be used and six samples are to be obtained, one sample from each of three generations would be placed in each fixative.

**Sectioning samples.** Several sections can be cut initially to determine orientation, followed by a new random start. In general, when there are layered structures (such as the mucosal layer of the airways) and the tissue sample is relatively large (not the case with endobronchial biopsies), unbiased estimates can be obtained by vertically sectioning (perpendicular to the surface epithelium and mucosa) the embedded mucosa and then randomly selecting the sections to be analyzed (133). Serial sections are cut and three to four unpaired sections are chosen randomly on a slide for each particular marker.

Given the small amount of tissue obtained by endobronchial biopsy, it is important to select the priority markers to be evaluated before sectioning the tissue. The distance between the sections needs to be greater than the largest element to be counted. With the newly recommended approach “do more less well,”
an “adequate” biopsy section need have only one evaluable field of interest (a distinct difference from the technique used most widely at present, in which several fields are considered necessary for evaluation). Of course, if a biopsy has only one evaluable field, the remaining biopsies to be counted must be larger in order to yield a sufficient number of fields to complete the analysis for the subject.

Measurements in fields of interest require an invariant reference compartment that could be either the reticular basement membrane or the volume of the sample. Normalization to surface area of basement membrane (i.e., number of cells per surface area) is one recommended approach that has been used to document changes in epithelial and inflammatory cells in biopsies of airways (142).

**Selecting fields for analysis.** The first step in selecting high-power fields for analysis is to decide on the tissue compartment of interest. Once an area has been chosen (e.g., epithelium or submucosa), the next step is to choose fields randomly. This can be done by dividing the area into quadrants and choosing fields randomly or by employing computerized sampling systems or random number or field generation.

**Quantification. Volume density.** Measures of density of cells, blood vessels, collagen, smooth muscle, or other tissue components can be determined by the simple technique of point counting. This determines the volume density (mm³/mm³) of objects in a reference volume (140), using the formula: \( V_{V}(O,R) = PO/PR \), where \( V_{V}(O,R) \) denotes the volume of objects \( O \) contained in a reference volume \( R \), \( PO \) is the number of points from a grid falling onto the structure of interest, and \( PR \) is the number of points falling into the reference space. \( V_{V} \) is the volume density and corresponds to the volume proportion that the structure of interest occupies in the reference space. The beauty of this ratio of points is that it is proportional to volume (140). Therefore, the ratio of three-dimensional measurements can be obtained by assessing tissue elements in a two-dimensional plane, without taking any more (or even less) time than do currently used methods of counting cell area profiles.

**Surface area.** A simple modification is used to determine surface area (e.g., of basement membrane). A grid with lines (as distinct from points) is superimposed on the tissue section, and the number of intersections of the test lines with the surface of interest (e.g., the reticular basement membrane) is recorded. The test grid can be provided in the form of an eyepiece reticule inserted into the microscope or by using an attached computerized system that can also facilitate the mathematical transformation of the intersection count to a surface area measurement (143). The subsequent determination of surface density (mm²/mm³) of objects in a reference volume is estimated therefore from both intersect and point counting (140) according to the formula: \( S_{V}(O,R) = 2 \times I(O)/L(R) \), where \( S_{V}(O,R) \) denotes the surface area of an object \( O \) contained in a reference volume \( R \), \( I(O) \) is the number of intersections of the test line of length \( L \) in reference space \( R \) with the trace of the object (see Unbiased Stereology, Chapter 6, pp 107–123 in [133]) (140).

Determination of volume density is an elegant and simple way to avoid the criticisms of counts of cut cell profiles but does not directly provide an estimate of cell number. One simple approach to obtaining an estimate of cell number is to divide the calculated volume/density for any given cell type by an estimate of its overall cell size (144). Alternatively, cell number per volume can be more accurately calculated using the disector method of analysis, but this is more labor intensive and consumptive of tissue (see subsequent section).

**Length.** Determination of the length of a tissue component may be required by the researcher in endobronchial biopsy tissue (e.g., measurement of the average length of a blood vessel in the submucosal compartment). The length density of a blood vessel can also be estimated using stereology and given as its length per unit volume of the reference space. In a section, it will be obtained by counting the number of times the vessel transects the section.

**Number estimation by the disector method.** The number of cells can also be estimated and expressed per unit volume of tissue (i.e., the “reference space”) using serial sections and the disector method, which can only be outlined briefly here. The disector method relies on the principle that all cells will be counted with equal probability if the method is applied in one of two ways: use of either the physical disector or the optical disector. The physical disector is the most practical of the two for analysis of bronchial biopsies and consists of a pair of serial sections separated by a known distance; the distance chosen depends on the size of the smallest cell of interest. The principle is that a cell is counted only if it is seen in one section and not in the next (16, 133).

In conclusion, biopsy tissue can yield valuable quantitative information about both airway inflammation and remodeling. For interpretation of the data, we must assume that endobronchial biopsy of large airway subcarinae provides unbiased and representative samples of the entire bronchial tree, which currently is unproven and needs to be determined in larger lung samples obtained at surgery or postmortem. Animal studies, at least, indicate that the airway wall cellular characteristics are significantly different from those obtained from subcarinae (17). Stereology requires a common denominator reference space for counts of cells or measures of tissue structures to be expressed meaningfully. The surface area of the reticular basement membrane is one potentially useful reference, and simultaneous measurements of subepithelial and epithelial volume may also be useful as a way of normalizing count data (18, 19). In animal experiments, lung volume has been used as the reference space (145). Data from computer-assisted tomography has been used to provide a useful estimate of lung volume in humans (146).

Although area profile counts still remain acceptable, we recommend that investigators consider and try to apply less biased, stereology-based sampling methods, which are likely to be simpler, more discriminatory, and more powerful for hypothesis testing. In addition, the general incorporation of these stereologic methods will facilitate greater opportunities for meta-analyses of data generated from similarly designed studies worldwide (see Table 2 for a list of advantages and disadvantages to several methods of quantitation.) Ultimately, comparisons of data obtained from both stereologic and the currently used “area profile” techniques applied to bronchial biopsies need to be made and reported.

**IV. ENDOBRONCHIAL BIOPSY VERSUS BRONCHOALVEOLAR LAVAGE VERSUS INDUCED SPUTUM**

Airway inflammation includes the presence of activated inflammatory cells, the release of cellular mediators, and the functional and structural consequences of such inflammation. Response to tissue injury, abnormal tissue repair, and restructuring, inappropriate to the maintenance of normal function, are referred to as “airway wall remodeling” (147, 148). Inflammation and remodeling may result from parallel or sequential pathways, and the techniques of endobronchial biopsy, BAL, and induced sputum may be used to sample complementary components of these pathways.

**Advantages and Limitations of Endobronchial Biopsy, Bronchoalveolar Lavage, and Induced Sputum**

The potential for the use of less invasive alternatives to endobronchial biopsy for measuring inflammation and remodeling in
**TABLE 2. COMPARISON OF COMMONLY USED MORPHOMETRIC METHODS IN ASTHMA RESEARCH**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>2-Dimensional: One Section per Biopsy or &gt; Two-Step Section</th>
<th>3-Dimensional: Multiple Biopsies and Multiple Sections per Biopsy</th>
<th>Semiquantitative Methods</th>
<th>Automated Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advantages</td>
<td>No. per area No. per length</td>
<td>Volume density (mm²/mm³) Volume to surface</td>
<td>Scoring system (1+, 2+, 3+)</td>
<td>No. per area</td>
</tr>
<tr>
<td></td>
<td>Uses established stereologic principles</td>
<td>Volume-oriented measurement</td>
<td>Useful for pilot studies</td>
<td>Fast</td>
</tr>
<tr>
<td></td>
<td>Can use technique to quantify cells and structures</td>
<td>Easily mastered</td>
<td>Consistency of staining</td>
<td></td>
</tr>
<tr>
<td>Disadvantages</td>
<td>May invite bias because larger cells are overestimated, smaller cells underestimated</td>
<td>&quot;Do more less well&quot;—requires more subjects and more biopsies</td>
<td>Not quantitative</td>
<td>Systems may have difficulty discerning cells; differences based on color, not cell type</td>
</tr>
</tbody>
</table>

Endobronchial biopsy. **Advantages.** Endobronchial biopsy, including endobronchial brush biopsy, directly samples the resident cells. Biopsies alone maintain the spatial relationships of structural components that may be important to functional change. The airway wall is composed of many interrelated structural components such as epithelium, connective tissue, vessels, muscle, and mucus-secreting gland acini. Endobronchial biopsy can provide an assessment of the status of the airway mucosa, whereas BAL and induced sputum cannot. A biopsy preserves structural relationships that may be visualized by microscopy. Individual structural components can be designated for study and, if required, removed and studied in isolation using laser capture dissection (154). Inflammatory cell subtypes can be identified by immunostaining, and biomarkers of epithelial damage, such as mitosis and apoptosis, can be measured. Immunologic cascades can be studied using immunohistochemistry, ISH, and microarray analysis (for assessment of gene expression) (see III. PROCESSING AND METHODS OF VISUALIZATION). Tissue explants and cells derived from biopsies can also be preserved in culture, and resident cells, such as airway smooth muscle myocytes, have successfully been dissected from endobronchial biopsies from subjects with asthma, allowing measurements of the proliferative responses of these cells (155).

**Limitations.** Endobronchial biopsy samples are limited in overall size to between 1 and 2 mm in diameter; they may not be representative of the entire conducting airways because they come mainly from relatively proximal airways and from subcarinae rather than the lateral walls. Thus, there is a bias inherent in the sampling procedure. Moreover, peribronchial tissue (deep to the mucosa) is not included in endobronchial samples. Depending on the operator’s experience, the procedure itself may induce artifacts, especially due to forceps damage apparent as crush artifact and mechanical loss of epithelium (5). In general, the more normal the airway, the more likely is tissue damage. In repetitive studies, the biopsy procedure itself creates scar tissue and can initiate cellular processes that can proceed during the interval (seconds) between sampling and fixation. It should be remembered that biopsy provides only a snapshot in time of an ongoing disease process. Finally, most endobronchial biopsies have been performed in adults and not children, although pediatric data are now forthcoming (50, 57).

Bronchoalveolar lavage. **Advantages.** Samples obtained by BAL include luminal cells (predominantly macrophages) and soluble factors coming from a relatively large airway surface area, including the peripheral airways. With sequential lavage, it is possible to observe the results of allergen and placebo challenge in the same individual over time. BAL includes both mediators and cells released into the fluid phase that may be studied individually or in specific combinations in suspension or in culture. International guidelines and monographs on BAL are already available (156, 157).

**Limitations.** There is no spatial, structural, or histologic information in BAL samples, and the cellular signals may be affected by the method of collection. The source of the cell sample is imprecise and may include cells that have migrated into the lumen or that were flushed from the airway wall. The volume of sample recovered may relate to disease severity, and the sources and the dilution factors of the solutes obtained are imprecise. The procedure itself or the subsequent handling and processing may introduce artifacts (e.g., lidocaine and bronchodilators used during bronchoscopy may affect cell activity). The removal, washing, centrifuging, and resuspension of mucus can activate cells and confound the cell and solute information. Finally, the BAL process may lead to transient fever within 24 hours after the lavage, occasionally associated with lung infiltrates (134).

**Induced sputum.** **Advantages.** Hypertonic saline-induced sputum is increasingly used as a minimally invasive procedure to obtain markers of airway inflammation, such as differential cell counts and solutes in asthma and COPD (158). Recently, detailed international recommendations on safety precautions and standardized induction, processing, and analysis of sputum have become available (158). Sputum can even be obtained during an acute exacerbation of the disease, and it can also be obtained (albeit less successfully) from children (159). The precision, accuracy, and responsiveness to challenges and therapy are well validated in asthma. Sputum can be repeatedly sampled, after sufficient time (i.e., 48 hours) has elapsed between inductions to allow cellular events caused by the procedure itself to subside. It can be studied using immunocytochemistry or ISH (158) and by a limited range of cell culture procedures. Many cytokines/mediators can now be measured by bio-, enzyme-, or immunosassays and real-time polymerase chain reaction in the fluid phase of the sputum sampled. Interestingly, induced sputum appears to be useful clinically in the management of asthma because therapy, guided by sputum eosinophil counts, has been shown to result in a considerable improvement of disease outcome (160).

**Limitations.** Airway structural information is not available from sputum. The exact site of origin of the material and the dilution factor are largely unknown. Hypertonic saline inhalation is not without risk in subjects with poorly controlled asthma: isotonic saline can be a good alternative (161). About 10 to 20%
of subjects are unable to produce an adequate sputum specimen
(up to 50% in children), especially those who are clinically stable
and well controlled with therapy. Sputum requires the use of
reducing agents to process the samples, which may affect the
results of subsequent procedures. Cytokines, growth factors, and
mediators may be encrypted in mucus, and the cellular profile
and staining patterns are more difficult to interpret than they
are with BAL preparations.

**Issues Still To Be Investigated**

We are just starting to appreciate the pros and cons of the
available biopsy techniques. There is little doubt that bronchial
biopsies, BAL, induced sputum, and also markers in exhaled
breath (162) provide complementary biological information on
the complex and dynamic features of inflammation. The distribu-
tion of such abnormalities within the airway wall and along the
tracheobronchial tree determines the severity of airway nar-
rrowing (163, 164). Each method samples specific parts of the
pathologic process. Eventually, the choice among the available
methods for monitoring airway inflammation critically depends
on the specific hypotheses and primary objectives of each individ-
ual study in conjunction with the burden that such measurements
impose on the subject.

**V. OVERALL CONCLUSIONS**

Given that the safety of the volunteer subject is of paramount
importance, bronchoscopy has an invaluable place in airway
research, even in children. The conduct of clinical trials should be
based on well established principles of study design and include
effective a priori planning followed by strict adherence to the
specified study design and measurement concepts. Laboratory
research personnel or senior representatives concerned with tis-
seau preparation and analysis need to be involved during both
study design and execution. The procedures of tissue sampling
and processing of biopsies require meticulous attention to the
high premium placed on each biopsy: this is the only technique
that directly samples the resident cells and maintains the spatial
relationships with structural components—the information is ad-
ditional and complementary to that obtained by lavage and sput-
num. Standard operating protocols need to be agreed upon and
adhered to and quality controls are needed. Minimizing the bias
and variability introduced by the sampling procedure and the
biology will improve accuracy and the capacity to test hypothe-
ses. Counting area profiles of inflammatory cells, where the nu-
cleus is seen in the plane of the section, remains an acceptable
procedure. However, the application of stereology for quantifying
cells and tissue structures minimizes bias. The last consider-
ation may improve statistical power, better enabling investiga-
tors to discriminate and determine the potential for current and
new treatments aimed at attenuating the inflammation and re-
modeling of diseases such as asthma and COPD.

**Workshop Authors**

**I. Obtaining Optimal-Quality Endobronchial Biopsies**

Neil Barnes, London, United Kingdom
Andrew Bush, London United Kingdom
Ratko Djukanovic, Southampton, United Kingdom
Peter Howarth, Southampton, United Kingdom
Douglas Robinson, London, United Kingdom
John Warner, Southampton, United Kingdom

**II. Processing and Methods of Visualization**

Qutayba Hamid, Montreal, Quebec, Canada
Peter Jeffery, London, United Kingdom

**ANNIKA LAITINEN, Helsinki, Finland**
MAURIZIO VIGNOLA, Palermo, Italy
SUSAN J. WILSON, Southampton, United Kingdom

**III. Variability and Quantification**

Peter H. Burri, Bern, Switzerland
Neil Carroll, Nedlands, Australia
John Faiy, San Francisco, California
Peter Gehr, Bern, Switzerland
James Hogg, Vancouver, British Columbia, Canada
Dallas Hyde, Davis, California
Peter Jeffery, London, United Kingdom
Monica Kraft, Denver, Colorado

**IV. Endobronchial Biopsy versus Bronchoalveolar Lavage**

Louis-Philippe Boulet, Quebec City, Quebec, Canada
Pascal Chanez, Montpellier, France
Kathleen Haley, Boston, Massachusetts
Stephen Holgate, Southampton, United Kingdom
Makoto Hoshino, Tokyo, Japan
Yutaka Nakamura, Madison, Wisconsin
Theodore Reiss, Rahway, New Jersey
Peter J. Sterk, Leiden, The Netherlands
E. Hayden Walters, Tasmania, Australia
Sally Wenzel, Denver, Colorado

**References**

2. Chu HW, Halliday JL, Martin RJ, Leung DY, Szefler SJ, Wenzel SE.
Collagen deposition in large airways may not differentiate severe asthma from milder forms of the disease. Am J Respir Crit Care Med 1998;158:1936–1944.


10. O’Sullivan S, Akvold M, Burke CM, Poulter LW. Effect of the addition of montelukast to inhaled fluticasone propionate on airway inflamma-


85. Nakamura Y, Christodoulopoulos P, Hamid Q. Localization of IL-4 and IL-5 mRNA by in situ hybridization in bronchial biopsies. In:


Variable Characteristics

Variables can be either categoric or numeric. Categoric variables fit into discrete groups or categories. Typically, they are characterized by displaying the frequency with which each group occurs and its proportionate representation of the total. Categoric variables can be further classified as “nominal” or “ordinal.” Nominal variables are those with only two categories (e.g., treatment and age groups in Table A1) or those with categories that cannot be ranked. Ordinal variables are those with three or more categories that are capable of being ranked (e.g., FEV<sub>1</sub> percentage groups in Table A1).

Numeric variables are those that take values within a given (valid) range. They can be subclassified as discrete if they can take only integer values (e.g., raw cell counts) or continuous if they can take any value (e.g., cell density). Numeric values are characterized by measures of central location, variability, and skew. For example, the left side of Figure A1 depicts the frequency histogram for cell density at baseline in the example study. Discrete cell density ranges were obtained by dividing the overall cell-density distribution into 10 equal intervals with bar heights used to depict the frequency of observations within each of these ranges. This histogram clearly indicates that the frequency distribution is skewed to the left (positive skew). For skewed distributions, the median is a better estimate of central location than the mean, with the degree of skewing determined by the location of the median within the 95% confidence interval (CI). In addition, normality of the distribution is assessed by superimposing the expected bell-shaped curve for a normal distribution on the histogram, using the mean and standard deviation calculated from the study data.

As depicted in the Figure A1, baseline cell density is not normally distributed. Normally distributed data are preferred in statistical analyses because they permit easier identification of outlying data points, clearer differentiation between observations, and valid inferences when applying statistical methods that assume a normal distribution. For these reasons, nonnormally distributed data are usually mathematically transformed in an attempt to normalize them. The right side of Figure A1 depicts the log transformation of the cell density values. These log-transformed values are more evenly distributed and approximate a normal distribution.

The disadvantage in mathematically transforming data is that it alters the interpretation of a unit change in the transformed parameter. Trivial transformations of a variable (e.g., X to X<sup>2</sup> or X<sup>3</sup>) do not adversely alter the meaning of the variable. However, more complex transformations, especially those that change the shape of the distribution, may leave the variable difficult to interpret. The ideal transformation is one that achieves the desired distribution while ensuring that estimates based on the transformed data remain interpretable.

For example, although 1/√X and log(X) are both valid normalizations, the log transformation is preferable because it has multiplicative meaning (e.g., 0.69 increment in log(X) always corresponds to a doubling of values on the original X scale). Log transformation of variables with zero values require a “trivial” modification (e.g., log(X + 0.5)). If the proportion of zero values is large, the transformed data may look bimodal (two peaks) and will not be distributed normally. Results of analyses using transformed data should be back-transformed and presented in the original units (e.g., geometric mean cell density in cells/0.1 mm<sup>2</sup>, not mean log cell density [Figure A1, right panel]).

Association between Variables

The next step in data analysis is assessing possible associations between two or more variables of interest. In the example study,
association between variables is used to assess baseline characteristics, summarize study design, and estimate and characterize treatment effects.

Descriptive methods can be used to summarize the association between each treatment category and baseline variables, such as age, %FEV₁, and cell density (Table A1). The difference in cell density between treatment groups can be estimated as either an absolute difference or a proportional difference (ratio). However, the estimated difference in median values is not calculated as the difference between medians. Hence, it does not correspond to 46.4 minus 80.0. Instead, it and its 95% CI are calculated based on an extension of the Mann–Whitney test, which does not require a normal distribution (2). With respect to the log-transformed data, the back-transformed difference between log cell densities is equivalent to the ratio of the geometric means (i.e., 50.3/55.6 = 0.90, with the 95% CI calculated using a t-test [Table A1]) (3).

Association between two numeric variables (e.g., pre- and posttreatment cell densities) are often summarized with a Pearson product-moment (if both variables are normally distributed) or a Spearman rank correlation coefficient (if both variables are not normally distributed) with the 95% CI calculated via the z-transform method (2). Values vary between −1 and 1, with 0 indicating no association. In the example study, the Spearman rank correlation between baseline and posttreatment mast cell densities separately for placebo and active groups are $r_s = 0.38$ (95% CI, 0.03 to 0.65) and $r_s = −0.14$ (−0.47 to 0.24), respectively.

Longitudinal plots, by linking data from the same subject, can be used to depict the relationship between baseline and posttreatment values (Figure A2). In the example study, the y-axis is plotted using the log-transformed data. Because these data have been normalized, the mean log cell density and its 95% CI were chosen for the summary statistic, although the median would have been equally appropriate. Qualitative evalu-

---

**Table A1. Baseline Characteristics of Study Population**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Placebo ($n = 14$)</th>
<th>Active ($n = 16$)</th>
<th>Difference</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\geq 65$ years</td>
<td>6 (43)</td>
<td>8 (50)</td>
<td>–</td>
<td>0.70†</td>
</tr>
<tr>
<td>$&lt; 65$ years</td>
<td>8 (57)</td>
<td>8 (50)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>FEV₁ %pred, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>0.64**</td>
</tr>
<tr>
<td>$&lt; 30$</td>
<td>3 (21)</td>
<td>2 (13)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>30–49</td>
<td>8 (57)</td>
<td>12 (75)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>$\geq 50$</td>
<td>3 (22)</td>
<td>2 (12)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Mast cell density, cells/0.1 mm²</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median (95% CI)*</td>
<td>80.0 (22.9 to 158.3)</td>
<td>46.4 (24.4 to 77.8)</td>
<td>−5.2† (−80.0 to 25)§</td>
<td>0.71‖</td>
</tr>
<tr>
<td>Geometric mean (95% CI)†</td>
<td>55.6 (24.5 to 126.3)</td>
<td>50.3 (29.3 to 86.3)</td>
<td>0.90 (0.36 to 2.25)††</td>
<td>0.82‖</td>
</tr>
</tbody>
</table>

Definition of abbreviation: CI = confidence interval.
*Non-parametric binomial-based 95% CI.
† t-Test.
‡ Median difference between groups (cells/0.1 mm²).
§ Ratio of geometric means (no units).
†† Mann–Whitney test-based 95% CI.
‖ χ² Test.
** Fisher’s exact test.
†† Mann–Whitney U test.

---

**Figure A1.** Distribution of cell density and log cell density at baseline ($n = 30$). Frequency histogram (left panel) shows cell density at baseline in the example study, indicating that the frequency distribution is skewed to the left (positive skew). Nonnormally distributed data are usually mathematically transformed in an attempt to normalize them. The log-transformed values (right panel) are more evenly distributed and approximate a normal distribution. 1-Dimensional scatter plot (top), frequency histogram (middle), summary statistics (bottom). 95% confidence interval (CI) calculated by *nonparametric binomial-based or **parametric t-distribution.
Assessment of the plotted data reveals similar baseline means for the placebo and active treatment groups, similar between-subject variability for the active group compared with the placebo group at both baseline and posttreatment, an increase in mean cell density comparing baseline with placebo, and a decrease in mean cell density comparing baseline with active treatment.

**Hypothesis Testing**

Up to now, the analyses in this hypothetical example have focused on estimation (e.g., the study sample median is used as an estimate of the unknown true population median, and the 95% CI is used to quantify the uncertainty of this estimation). Assessing the truth concerning associations requires hypothesis testing. Hypothesis testing starts by specifying two mutually exclusive hypotheses referred to as the null hypothesis and the alternative hypothesis. The probability of obtaining either the observed or a more extreme result in the study is then calculated, assuming the null hypothesis to be true. This probability, which falls between 0 and 1, is referred to as the p value. By convention, probabilities of 0.05 or less (or in some cases 0.01 or less) are assumed to be statistically significant, i.e., so unlikely to have occurred by chance as to cause rejection of the null hypothesis and acceptance of the alternative hypothesis. Of note, the cutoff chosen for this statistical significance also defines the probability of a Type I error, i.e., rejecting the null hypothesis when it is in fact true. In contrast, failure to reject a false null hypothesis constitutes a Type II error. The probability of a Type II error is related to the power of the analysis (power equals 1 minus the probability of a Type II error) and is a function of sample size. Usually, a study should be designed to provide a power of at least 80%.

Sample size calculations can be approached from either a hypothesis testing or an estimation point of view. For hypothesis testing, power calculations are done to control the probability of a Type II error, assuming the alternative hypothesis to be true. With increasing sample size, the probability of a Type II error decreases, but the probability of a Type I error does not vary from 5% (when the significance level is 5%) regardless of sample size. The estimation approach uses precision calculations to control the maximum width of 95% CI. Large studies have smaller intervals and thus a good chance of estimating the true value of the parameter. With increasing sample size, the probability of a Type II error, assuming the alternative hypothesis to be true, decreases, but the probability of a Type I error does not vary from 5% (when the significance level is 5%) regardless of sample size. The estimation approach uses precision calculations to control the maximum width of 95% CI. Large studies have smaller intervals and thus a good chance of estimating the true value of the parameter.

A common statistical problem in the analysis of biopsy data is the inflation of Type I error brought about by assessing outcomes for multiple cell types. Ideally, the Type I error for the whole analysis, termed the ‘experiment-wise error rate,’ should not exceed 5%. However, when hypothesis tests are performed for three separate outcomes, each with a 5% Type I error rate, the experiment-wise error rate is inflated to 15%. The Bonferroni correction (4, 5) is a traditional method for capping the experiment-wise error rate at 5%. However, it has been shown to be overly conservative (experiment-wise error rate < 5%) as the correlation between the various outcomes being assessed increases (6). Holm’s sequential testing procedure (7) and O’Brien’s global test (8) are alternative methods with better overall validity.

The χ² test (9) (see Age, Table A1) is frequently used to test hypotheses involving associations between categoric variables. However, this test makes a large sample approximation that requires the expected frequency in each cell of the tabulation to be greater than 5 to be valid. When this is not satisfied, the usual alternative is the Fisher’s exact test (see %FEV₁, Table A1) (9).

Differences in mast cell density at baseline can be assessed using a t-test and assuming equal variances on the log-transformed median data, because it is plausible that all subjects were sampled from the same population at baseline (Table A1). Alternatively, differences in the nonnormally distributed median data can be assessed using a Mann–Whitney test (9). Although this test is distribution-free, meaning that it does not assume any particular shape for the distribution of the data, it does assume that the distributions of the two groups being compared are similar.

Two additional issues must be considered prior to hypothesis testing of the primary objective. Namely, what is the comparative measure of treatment effect, and is the study treatment the only (important) source of variation in mast cell density following treatment?

Analysis of cell data usually assumes a multiplicative treatment effect—meaning that, on average, the posttreatment cell density for a subject on active therapy is X% lower (or higher) than on placebo—and is estimated as a ratio between means (or medians). Additionally, in the example study, baseline cell density, %FEV₁, and age were a priori confounding variables that could bias the point estimate of treatment effect (and the position of the 95% CI, which centers around this point estimate) if imbalances between groups were not adequately adjusted for. In the example study, the cell density imbalance between groups at baseline was small (Table A1) and the point estimates for the treatment effect are similar with (ratio = 0.56) or without (ratio = 0.50) adjusting for this baseline imbalance (Table A2).
### TABLE A2. TREATMENT EFFECT ESTIMATION—LOG MAST CELL DENSITY

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Placebo Mean (SE)</th>
<th>Active Mean (SE)</th>
<th>Effect Estimate Mean (SE)</th>
<th>Ratio (95% CI)</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline log ((p_0))</td>
<td>4.01 (0.38)</td>
<td>(p_0 )</td>
<td>(a_0 )</td>
<td>(a_0 )</td>
<td>0.50 (0.21 to 1.19)</td>
</tr>
<tr>
<td>Posttreatment log ((p_1))</td>
<td>4.27 (0.35)</td>
<td>(a_1 )</td>
<td>(0.69 ) (0.42)</td>
<td>0.59 (0.57)</td>
<td>0.56 (0.17 to 1.80)</td>
</tr>
<tr>
<td>Postbase change log ((p_1/p_0))</td>
<td>0.25 (0.44)</td>
<td>(a_1/a_0 )</td>
<td>(-0.34 ) (0.38)</td>
<td>(0.59 ) (0.57)</td>
<td>0.59 (0.57)</td>
</tr>
</tbody>
</table>

Definition of abbreviations: CI = confidence interval; SE = standard error.
Values presented are log cells/0.1 mm².
* t-test.
† Mann–Whitney U test.

In general, less biased point estimates are obtained with adjusted analyses even when imbalances are small. However, this adjustment also makes the baseline-adjusted comparison less precise (wider 95% CI) and less powerful (larger p value). Therefore, the extent of adjustment should be weighted by the true correlation between outcome and covariate in order to obtain unbiased estimates with greater precision (analysis of covariance) (10).

In the example study, the estimated correlation between baseline and posttreatment is low (see previously discussed Spearman rank correlations). Therefore, the value of \(k\) is approximately 0 and the unadjusted analysis is more powerful.

Because the example study was randomized with important confounders well balanced between treatment groups, it is valid to maximize statistical power (and precision) by using the posttreatment-only analysis. This analysis demonstrates that, on average, active drug therapy reduced mast cell density by 50% relative to placebo (95% CI of ratio, 0.21 to 1.19), but that this reduction was not statistically significant (\(t\)-test, \(p = 0.11\)).

### Conclusions

Data analysis should consist of both descriptive and statistical components. Thorough descriptions of the data are necessary before formal statistical tests are performed to understand the effects at work within the study data and to ensure validity of statistical comparisons.

### Appendix References