The concept that airway inflammation plays a central role in the pathogenesis and propagation of asthma has resulted in an interest in methods of evaluating and following the progression of airway inflammation which are rapid, reproducible, relevant and, ideally, noninvasive. The methods used to study airway inflammation in the past have involved collecting samples (airway biopsies, bronchoalveolar lavage or bronchial washings) through a bronchoscope (usually fiberoptic) or measuring changes in peripheral blood inflammatory cells. These methods have been very useful and have greatly advanced our understanding of asthma pathogenesis. They are, however, either invasive, as in the case of bronchoscopy, and therefore not amenable to frequent events occurring in the airways.

Evidence that elevated numbers of eosinophils can be measured in sputum from asthmatics has been available since the end of the 19th century [1]. However, induced sputum has only been used clinically since the 1940's as a method of collecting samples from the airways, initially to test for the presence of tuberculous bacilli [2], and more recently for Pneumocystis carinii infection in patients with acquired immunodeficiency syndrome (AIDS) [3]. In 1992, PIN et al. [4] described a method of inducing sputum using hypertonic saline to evaluate airway inflammation in asthmatic subjects. Subsequently, several research groups have developed their own, slightly different methods [5–9]. These methods have all used hypertonic saline, inhaled in different concentrations and for varying durations, to induce sputum. Some methods select sputum from the mixture of saliva and sputum that the subject produces during the induction [4, 7], while others process both the sputum and the saliva [5, 6, 8] or ask subjects to collect saliva in one container and sputum in another [9]. A comparison of the last two methods, published in this issue of the Journal [10], suggests that the separate collection of sputum and saliva is better than collecting both together, as it reduces the contamination with buccal squamous cells and results in significantly higher levels of eosinophil cationic protein (ECP). Also, a previous report of a comparison between the selected sputum and the residual portion (saliva) has demonstrated that when sputum is selected, buccal squamous cell contamination is almost eliminated and the levels of ECP are higher [11]. Also, the selected sputum method gives results for the cell differentials and fluid phase measurements that are qualitatively similar to spontaneous sputum [12].

There are two publications reporting on the reproducibility of both the whole sample [8] and the selected sputum [13] techniques, one of which [8] is reported in this issue of the Journal. In these studies, measures of reproducibility, also termed reliability or repeatability, are concerned with the agreement between repeated measurements made on the same subjects, with a short time between the two measurements. Close agreement between these measurements is obviously necessary for the method to have a high degree of utility (or usefulness). In both studies, the authors have measured the reproducibility of several cellular and soluble markers of inflammation measured in induced sputum. On the whole, these measurements proved reproducible, with intraclass correlation coefficients as high as 0.94. Based on these measurements, IN'T VEEN and colleagues [8] have concluded that induced sputum is a potentially useful in clinical asthma studies to monitor inflammatory processes in the lungs. We agree with the authors that high measurement reproducibility is an important contributor to a test's utility in clinical or research studies. However, in order to document the usefulness of a test, it is equally important to show that the measurement has validity, which may be demonstrated by comparing it to other established "gold standard" measurements (criterion validity) or by demonstrating that it responds as expected in different disease states or following intervention (content validity).

The criterion validity of several indices measured from induced sputum has been investigated using comparisons with bronchoalveolar lavage or bronchial washings in asthmatic subjects [14–16]. In these studies most correlations between cell differentials made from sputum and the bronchoscopy samples were poor. Furthermore, soluble inflammatory mediators were more concentrated in sputum samples and correlated poorly with the bronchoscopy samples [14–16]. It is important to point out that the lack of agreement between sputum and the bronchoscopy samples may differ for reasons other than a lack of validity of the sputum technique. For example, the two techniques may be sampling from different sites [16] or be subject to different concentrating/diluting effects. As this is likely to be the case, the lack of a gold standard measurement of airway inflammatory indices at the sites sampled by induced sputum makes attempts at criterion validation (gold standard comparison) difficult.

Without a gold standard test for comparison, the validity of induced sputum measurements must be demonstrated by providing evidence of content validity. Content validation is arguably more valuable than criterion validation, as it allows for inference based on measurements made with the validated tool. There are multiple sources of evidence of content validity of induced sputum measurements. Studies performed to date indicate that asthmatic subjects have higher numbers of inflammatory cells, such as eosinophils and metachromatic cells, in induced sputum, when compared to nonasthmatic subjects and
Subjects with chronic bronchitis [4, 5, 17, 18]. The methods are also responsive to changes in the state of asthma control, with increases in eosinophil and neutrophil numbers during asthma exacerbations [19, 20] or in eosinophil numbers after allergen [21, 22] or isocyanate [7] inhalation, and eosinophil decreases associated with symptomatic improvement after treatment with oral or inhaled corticosteroids [22–24], or allergen avoidance [25]. Thus, the methods of induced sputum which have been evaluated in this way, have been shown to have content validity in that they are able to detect differences between nonasthmatics and asthmatics and between asthmatics at different stages of disease as well as their response to treatment. The other valuable information that can be obtained when the reproducibility of these methods is reported is to use it to estimate sample size for future studies.

These studies reported to date demonstrate that, while different methods have been developed to collect and process induced sputum, the high degree of reproducibility and content validity of the methods that have been studied strongly supports a role for their use in both clinical and research investigations of airflow inflammation in asthma. However, care must be taken that the results of these carefully performed studies on reproducibility and validity are not extrapolated to all methods that have been, or might be, developed. Each method should undergo the same rigorous evaluation and ideally the methods should be directly compared in the same subject population to allow interpretation of results when different methods are employed.

References