Excessive variability in spirometric measurements derived from the forced expiratory volume manoeuvre, also called "test failure," is a phenomenon familiar to all those who use pulmonary function tests, whether it be in a clinical context or in the context of population-based studies. Because the manoeuvre is effort dependent, test failure is usually attributed to a failure of effort or comprehension on the part of the patient or subject (one editorialist capsulised the issue in the phrase "disability or disinclination") or to the incompetence of the technician. Adequate effort on the part of the subject tends to be particularly in question in disability assessments or when workmen's compensation is involved.

To counter this, guidelines have been developed to standardise spirometry, both in the clinical and the epidemiological context. In addition to technical recommendations on equipment, its selection and validation, these guidelines recommend that a minimum of three satisfactorily performed trials should be required from a subject undergoing tests (criteria for satisfactory performance are clearly defined), with up to eight trials if this is necessary to meet certain reproducibility criteria. For FEV₁ and FVC, the reproducibility criteria recommended by the American Thoracic Society are that the two best trials should not differ by more than 5% or 100 ml, whichever is the greater; those recommended by the European Economic Community stipulate that the difference should not exceed 300 ml for FVC.

The purpose of this editorial is to bring to attention the gathering evidence, mostly from epidemiological studies in industry, that failure to perform reproducible spirometry may itself be an indicator of respiratory ill health. For instance, in several cross-sectional studies of occupational groups test failure has been shown to relate to symptoms. In Chinese textile workers the symptoms were those of byssinosis, in Pennsylvania railroad workers chronic bronchitis, and in coal miners, wheezing and shortness of breath were implicated as well as chronic cough and sputum. Similarly, subjects who exhibit test failure have, on average, lower levels of FEV₁ and an increased annual loss of FEV₁ in two cohort studies in which this has been measured. Test failure does not, however, appear to be a significant predictor of death when function level and smoking are taken into account. A surprising and unexplained feature in two community-based studies was that test failure occurred less frequently in smokers than in non-smokers.

The physiological mechanisms underlying the phenomenon of test failure are likely to be complex. It has been known for some time that a deep inspiration increases pulmonary compliance and decreases airway resistance, both of which are important determinants of the volume flow time relations recorded by spirometry during the forced expiratory manoeuvre. The bronchodilator effect of a deep breath, however, is less consistently demonstrable or is absent in individuals with increased reactivity of the airways to non-specific challenge, usually measured as response to histamine, methacholine, or cold air inhalation, or to exercise. Airway reactivity also varies widely between individuals and is generally increased in asthmatic subjects, and often in relation to smoking. Moreover, maximal inspirations seem able to stimulate either or both bronchodilator and bronchoconstrictor responses in susceptible subjects. Thus test failure could, for instance, result from an increase in airway resistance induced by the repetition of the forced expiratory manoeuvre in individuals with hyperreactive airways that constrict in response to this manoeuvre, an increase not offset by the inspiratory manoeuvre preceding each trial; such effects could also be cumulative. Consistent with this explanation was the finding, in a study of Transvaal ceramic workers, that spirometric test failure was less frequent in post bronchodilator compared with prebronchodilator measurements, though in the author's view these findings were attributed to a learning effect.

The epidemiology of a condition encompasses its distribution and determinants—that is, who in a population has the condition and why, who does not, and why not. In the case of test failure this information is as yet incomplete, though what is currently known has important implications. For the epidemiologist planning a survey that includes spirometry, particularly in an occupational group, any attempt to improve the quality of measurements.
by excluding from analysis subjects who exhibit test failure (and who are also likely to be less healthy) will bias the study towards the null—that is, against showing the ill effects of any environmental exposure under suspicion.29 Options are to include results of all subjects, whether or not they exhibit test failure (if necessary, identifying them separately in the analysis) or to adopt more liberal criteria of reproducibility.9 11

For the general or occupational clinician, the concept of test failure as an index of ill health has some appeal, though more precision in characterising its determinants and clarifying underlying mechanisms would be welcome. For instance, the paradoxical relation to smoking described in two community based studies needs confirmation or refutation in other populations.10 11 Other pertinent questions are the relations of test failure to quantitative measurements of airway responsiveness to a non-specific challenge, whether this be to methacholine, cold air, or exercise. The important message, however, is that test failure is as likely to reflect ill health as it is to reflect poor cooperation, poor effort, or the incompetence of the technician. Clinicians should therefore be as sensitive to the possibility that the individual exhibiting test failure is disabled as they are currently to the possibility that such an individual is disinclined.

MARGARET R BECKLAKE
Pulmonary Research Laboratory,
Department of Epidemiology and Biostatistics,
McGill University,
Montreal, Quebec H3A 1A3, Canada