



OPINION

Are reference equations for spirometry an appropriate criterion for diagnosing disease and predicting prognosis?

Guy B Marks

Correspondence to

Dr Guy B Marks, Department of Respiratory and Environmental Epidemiology, Woolcock Institute of Medical Research Sydney, PO Box M77, Missenden Road PO, Sydney, NSW 2050, Australia; guy.marks@sydney.edu.au

Received 8 June 2011
Accepted 12 July 2011
Published Online First
8 August 2011

ABSTRACT

In the last few years, there has been considerable debate on the use of threshold criteria for the diagnosis of obstructive lung disease based on FEV₁ and FEV₁/FVC ratio. It has been argued that a fixed ratio and fixed percentage criterion result in misclassification. The author argues that this critique is based on a false presumption about the validity of reference equations as a criterion for normality. The flaw lies in the methods used to derive reference equations, which involve arbitrary and circular criteria for exclusion of some members of the population, use potentially non-representative reference populations and include predictive variables that are really risk factors for disease or for adverse outcomes of disease. The author argues for a new interpretative approach for the use of lung function data in clinical practice based on prognostic equations analogous to the Framingham cardiovascular risk factor equations. These interpretative equations should be based on data from cohort studies and randomised controlled trials, rather than cross-sectional studies, and if properly formulated, will prove to be valuable aids to clinical decision making.

In the last few years, there has been considerable debate on the use of threshold criteria for the diagnosis of obstructive lung disease based on the spirometric ratio (forced expiratory volume in one second/forced vital capacity (FEV₁/FVC)) and also spirometric volumes (specifically FEV₁). The 'M' word (misclassification) has been invoked to criticise one criterion, or set of criteria, with respect to another.¹ I argue that this critique is based on a false presumption about the truth (validity) of the proposed criterion for normality. It is underpinned by a limited and limiting framework for understanding and using the information that measurement of spirometric function provides. Unfortunately, the current Global Lungs Initiative,² which aims to provide the world with predictive equations for lung function, is operating within this limited framework and missing the opportunity for extending the value of this measurement.

Measurement of spirometric function has a long history beginning with the first report on a device for measuring vital capacity in 1846³ and was greatly advanced 101 years later by Tiffeneau and Pinelli's description of the timed forced expiratory manoeuvre and derivation of the FEV₁/FVC ratio as an indicator of airflow obstruction.⁴ The prognostic significance of FEV₁ for chronic obstructive

pulmonary disease^{5 6} and its importance in the diagnosis of asthma⁷ have been central to our understanding of these diseases for decades. Its broader relevance as a prognostic indicator for cardiovascular outcomes is also well established.⁸ Spirometric function is a key indicator of health status.

The attempt to interpret the spirometric function of individual patients or subjects with respect to reference values also has a long history. Its beginnings are in the identification of 'normal' variation attributable to gender differences demonstrated by Hutchinson himself and extend to racial differences in soldiers of the Union Army during the American Civil War (reviewed by Braun⁹). Modern reference equations incorporating age, sex, height and race or ethnic origin have been used in North America^{10 11} and Europe^{12–14} for several decades. However, there has been little critical analysis of the basis of these reference equations and the claim that they represent the criterion against which the presence or absence of disease, specifically obstructive lung disease, can be ascertained.

The current model for detecting abnormality or disease based on lung function measurements is to compare an individual's observed values with a reference range. These reference ranges are universally derived from spirometric surveys conducted in apparently representative populations of apparently normal individuals. Regression equations are used to estimate the expected or average value based on selected predictors, usually functions of age, height, sex and, sometimes, race. The expected value and residual variance are then used to define the range of values within which 95% of the reference normal population values would be expected to lie. The lower limit of normal is the lower end of this range. This method has been applied to all spirometric variables including FEV₁, FVC and FEV₁/FVC ratio, as well as peak expiratory flow rate. This method is simple, based on sound statistical principles, and appealing. Some would argue that values within this reference range are defined as normal or not diseased and values outside this range are defined as abnormal and hence diseased. Any diagnostic criteria that result in a different classification of disease or non-disease are regarded as a misclassification.¹ This claim requires some analysis.

The basis for the concern about the use of reference equations as the criterion for diagnosing

disease lies in the representativeness of the populations on which it is based, the criterion for defining normal and the selection of predictive factors. All studies used to derive reference equations seek to limit their study populations to 'normal' individuals. This is usually achieved by a questionnaire-based selection criteria related to respiratory symptoms, diagnosed respiratory disease and smoking (a risk factor for respiratory disease). This raises several questions. If reported diagnoses and symptoms are an adequate basis for distinguishing normal from not normal, why do we need to measure lung function at all? Surely, the main rationale for using a spirometric criterion for disease is that it is independent of subjective criteria such as symptomatic status or reported diagnosis. The inclusion of this criterion in the spirometric definition of 'normal' means that the definition is not independent of subjective factors. In developed nations, smoking is the main risk factor for obstructive lung disease. However, is this an adequate justification for excluding smokers from the reference populations? If so, what about people with other risk factors for low lung function such as airway hyper-responsiveness, exposure to biomass fuel smoke or occupational exposure to dust and fumes? Is there a definitive reason for the priority given to smoking status as the only risk factor justifying exclusion from the reference populations? The definition of 'normal' used in selecting reference populations is convenient but it is arbitrary and, in relation to the exclusion of people with symptoms and diagnosed disease, circular.

In order to make generalisable statements based on a sample survey, it is important that the sample is representative of the population for whom the statements are to apply. Representativeness of the study populations used to derive spirometric reference equations is a problem. Even when the original study population is selected using sound sampling principles, as in the Third National Health and Nutrition Examination Survey, the final population used for the derivation of reference equations is potentially severely biased. After exclusion of people with respiratory symptoms or diagnosed respiratory disease, smokers and people who could not perform reproducible spirometry, the original sample of 16 484 individuals aged 17 years and older was reduced to 4634 individuals (28.1% of the original sample) for derivation of the reference equations.¹¹ There is a substantial risk that this population will not be representative of the general population and that this lack of representativeness will not be confined to the absence of disease and the presence of smoking. In some countries, the prevalence of smoking in men approaches 70%. Those who do not smoke, and hence are eligible for inclusion as reference 'normals', are a small, and probably non-representative, segment of the population. It is questionable whether the reference equations derived from highly selected subgroups of the population can be considered generalisable to the population as a whole.

The final problem with the use of reference equations to define normality, and hence the presence or absence of disease, is the choice of predictors used in the reference equations. If one simply wants to describe lung function in a population, then it is appropriate to include all potentially predictive factors. This includes the conventional predictors such as age, height, gender and race but should also include other potential explanatory variables such as environmental exposures and other constituent factors such as atopy and genetic factors. However, spirometric reference equations are not used for this purpose. They are used to define normality and, hence by exclusion, for the diagnosis of disease, independent of the presence of risk factors. The inclusion of potential risk factors for disease in the reference equations reduces the likelihood that people with those risk factors

will be diagnosed with disease. The question of appropriate selection of covariates for regression models has been widely canvassed in the epidemiological literature, where these potential risk factors for the outcome are referred to as 'intervening variables'.¹⁵ Which of the conventional lung function predictors could be considered a risk factor for obstructive lung disease? Probably height is not a risk factor for disease. All the other predictors are potentially risk factors. Age is a strong risk factor for disease and for mortality. In many societies, race is strongly correlated with risk factors for disease including environmental exposures and nutrition. Sex may also be a risk factor for disease both due to correlation with environmental exposures¹⁶ and due to constitutional factors. Inclusion of each of these covariates within the prediction equations tends to reduce the likelihood that members of high risk groups defined by these factors will be diagnosed with disease. The inclusion or exclusion of these factors as predictors in the reference equations comes down to answering difficult questions such as: is the presence of lower lung function in older people, some racial groups and women normal or an indicator of higher prevalence of disease? Or alternatively, does the prevalence of disease increase with age, in certain racial groups and in women? The obvious dilemma posed by these questions points to the problem in choosing covariates for inclusion in reference equations and in defining normality based on these equations.

These problems with the definition of normal, representativeness of the study populations and the selection of predictors for the spirometric reference equations have two important consequences. First, they call into question the implicit assumption that reference equations represent an absolute truth and hence that they can be used as a criterion or gold standard for classification of normality or disease. It is not valid to claim, as some do, that another criterion for defining disease, such as $FEV_1/FVC < 0.7$, must be wrong because it misclassifies subjects compared with the lower limit of the normal derived from reference equations. The second implication is that we need a new approach to the interpretation of spirometry for informing decisions in clinical practice.

The best place to start in designing a new approach is to ask how we use spirometry or other tests in clinical decision making. Ultimately, we are seeking to make a diagnosis and, with this, enable advice about prognosis, risk factor modification and the likely benefit of alternative treatment regimens. This information is available from cohort studies and from randomised controlled trials, but not from cross-sectional studies. The important dimension is time. We perform tests to give information about the future, something we cannot already know. Cross-sectional studies tell us only about the present, which we can already know.

We also need an approach that allows the incorporation of other information into the interpretation of the results of spirometry. In clinical practice, advice and decisions are not made on the basis of a single test. This Bayesian approach was elucidated by Sackett and his colleagues two decades ago,¹⁷ but uptake into respiratory medicine has been slow.

We do not need to look too far to find evidence for the value of an alternative approach to the use of test data for informing prognosis. The Framingham risk factor equations are widely used for predicting risk for a range of cardiovascular outcomes, based on the results of a range of tests and observations.¹⁸ Our cardiologist colleagues had the advantage of the Framingham cohort to derive these equations.

We do have the data from a range of cohort studies and randomised controlled trials that allow us to examine the

prognostic consequences of various levels of spirometric function for a range of clinically important outcomes including the onset of respiratory symptoms, accelerated decline in lung function, disability, hospitalisations and death. Appropriate analysis of these data, together with the incorporation of data on other risk factors, should allow the estimation of new risk equations for respiratory outcomes incorporating spirometry. Quantitative estimates of prognostic risk can be obtained, providing a strong basis for advice and for clinical intervention.

In conclusion, I argue that the current approach to the interpretation of spirometry is flawed. Reference equations have no special status as the repository of the truth about normal lung function. The debate about the lower limit of the normal versus a fixed ratio and the attempt to provide 'world' reference equations are distractions from the real task at hand: to develop respiratory risk equations based on spirometric measurements but incorporating other relevant risk factors and biomarkers of prognostic significance. Once developed, these equations can be readily translated into clinically useful and usable practice tools.

Competing interests None.

Contributors The author conceived the idea for this paper and drafted the manuscript alone.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. **Miller MR**, Quanjer PH, Swanney MP, *et al.* Interpreting lung function data using 80% predicted and fixed thresholds misclassifies more than 20% of patients. *Chest* 2010;**139**:52–9.
2. www.lungfunction.org. *Lung Function in Growth and Aging*. <http://www.lungfunction.org> (accessed 8 Jun 2011).
3. **Hutchinson J**. On the capacity of the lungs, and on the respiratory functions, with a view of establishing a precise and easy method of detecting disease by the spirometer. *Med Chir Trans* 1846;**29**:137–61.
4. **Tiffeneau R**, Pinelli A. Air circulant et air captif dans l'exploration de la fonction ventilatrice pulmonaire. *Paris Med* 1947;**133**:624–8.
5. **Fletcher C**, Peto R, Tinker C, *et al.* *The Natural History of Chronic Bronchitis and Emphysema*. Oxford: Oxford University Press, 1976.
6. **Peto R**, Speizer FE, Cochrane AL, *et al.* The relevance in adults of air-flow obstruction, but not of mucus hypersecretion, to mortality from chronic lung disease. Results from 20 years of prospective observation. *Am Rev Respir Dis* 1983;**128**:491–500.
7. **Lowell FC**. Spirometry in asthma. *N Engl J Med* 1973;**288**:262–3.
8. **Kannel WB**, Lew EA, Hubert HB, *et al.* The value of measuring vital capacity for prognostic purposes. *Trans Assoc Life Insur Med Dir Am* 1980;**64**:66–83.
9. **Braun L**. Spirometry, measurement, and race in the nineteenth century. *J Hist Med Allied Sci* 2005;**60**:135–69.
10. **Morris J**, Koski A, Johnson L. Spirometric standards for healthy non-smoking adults. *Am Rev Respir Dis* 1971;**103**:57–67.
11. **Hankinson JL**, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. *Am J Respir Crit Care Med* 1999;**159**:179–87.
12. **Cotes JE**, Rossiter CE, Higgins IT, *et al.* Average normal values for the forced expiratory volume in white Caucasian males. *Br Med J* 1966;**1**:1016–19.
13. **Polgar G**. The influence of pulmonary growth and development on paediatric respiratory diseases. *Acta Paediatr Acad Scient Hungaricae* 1974;**15**:183–91.
14. **Roca J**, Burgos F, Sunyer J, *et al.* Reference values for forced spirometry. *Eur Respir J* 1998;**11**:1354–62.
15. **Hernan MA**, Hernandez-Diaz S, Werler MM, *et al.* Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am J Epidemiol* 2002;**155**:176–84.
16. **Jarvis D**, Chinn S, Luczynska C, *et al.* Association of respiratory symptoms and lung function in young adults with use of domestic gas appliances. *Lancet* 1996;**347**:426–31.
17. **Sackett D**, Haynes R, Guyatt G, *et al.* *Clinical Epidemiology. A Basic Science for Clinical Medicine*. 2nd edn. Boston: Little, Brown and Company, 1991.
18. **Anderson KM**, Odell PM, Wilson PWF, *et al.* Cardiovascular disease risk profiles. *Am Heart J* 1991;**121**:293–8.