Editorial

Decline of FEV$_1$

The one second forced expiratory volume, the FEV$_1$, remains one of the most useful tests of lung function in epidemiological studies. Despite the introduction of more subtle and more specific tests, none has supplanted it as a means of detecting and quantifying environmental effects on respiratory health. Recently convincing evidence has appeared[1] that the FEV$_1$ not only usefully describes current levels of respiratory fitness but also has considerable prognostic significance for eventual respiratory death.

One characteristic feature of the FEV$_1$ is that older people have on average lower values than those who are young. The FEV$_1$ is said to decline with age. What do we know about the decline of FEV$_1$—what is as it were received knowledge? It is of some interest to examine the origin of this received knowledge, how it was acquired, and how strongly based it is. So far as possible, statistical matters and indeed numerical values will be avoided in this review.

The FEV$_1$, after increasing throughout childhood to a maximum at an age somewhere from 18 to 25, declines in men at a rate, according to the best authorities,[2] of 31 ml a year—that is, 1 litre in 32¼ years. This is a well established fact, but there is some small print associated with it. First, the result was based on men with no symptoms of respiratory disease and no history of chest illness, though no notice was taken of whether a man was a smoker or not if he had no symptoms or history. In practice the rate of loss will be higher than this: nearer to 40 ml a year in nonsmokers and approaching 50 ml a year in smokers. Second, allowance had to be made for height in establishing the norm of 31 ml a year because large men have larger FEV$_1$s, though they are not necessarily any better for it.

Here we come to the most important reservation about this established norm. It was not based on the observation of individual men throughout life. It was based on the comparison of different men of different ages at the same moment of time or epoch. Men of 60 had FEVs that averaged 1 litre less than men of 27 years 9 months. No one has yet observed a group of individuals throughout life and observed the average of their declines in FEV$_1$. We do not know at all if a steady, smooth decline in FEV$_1$ is indeed characteristic, or if individuals show sustained plateaux separated by step like falls. The nearest we have to such

Address for reprint requests: Professor PD Oldham, Department of Mathematical Statistics and Operational Research, University College Cardiff, PO Box 78, Cardiff CF1 1XL.

data are perhaps those obtained in the classic study of bronchitis by Fletcher and his associates,[3] in which the FEV$_1$ was measured in 792 men twice a year for eight years. Unfortunately multiple observers had to be used, and there seemed to be systematic differences between them and from one occasion to another. Consequently the data had to be adjusted, and the criterion for adjustment was to achieve a smooth, constant rate of decline. Thus they cannot tell us what actual patterns of decline occurred.

We must apparently now modify our received knowledge about the decline of FEV$_1$. It is not a true decline at all, but merely a statistical phenomenon, expressing the average difference in FEV$_1$ seen in men of different ages at the same moment of time. As in all such cases the average behaviour of a statistical aggregate in no way implies similar behaviour of the individuals in the aggregate. Indeed, there is an obvious example where it does not. At present human stature is found, by exactly the same method of observation as is used for FEV$_1$, to decline by 0·14 cm a year or about 1 inch (2·5 cm) in 20 years. Such a decline is not seen in healthy individuals; a loss of about ½ inch in 40 years may not be untypical. We see at once how complicated is the basis for adjusting FEV$_1$ for stature. To what extent does lung size decline with stature—in proportion to the actual loss of height of the individual, or in proportion to the observed statistical decline (which we would probably attribute to secular changes in nutrition)?

The decline of FEV$_1$ is not, of course, an absolute fact of physiology but a convenient standard against which we can measure declines seen in persons whose life history can be described as in some way special—marked by chronic respiratory disease, for instance, or subject to a constant or occasional exposure to some pollutant of the environment. It should be used only in this comparative way, as a convenient but arbitrary index, not unlike a standardised mortality rate, which in no way measures an observable rate of dying.

The use of the decline of FEV$_1$ in epidemiological studies should be by regarding a year of life as having potentially a distinctive effect when it has been lived in one environment rather than another. To begin with it would naturally be assumed that the effect is proportional to the number of such years of life, but there is no difficulty in exploring more complicated ideas; that there is a diminishing or accelerating effect of the number of years, or that the interval of time

since the environment was first experienced, or last experienced, should be allowed for.

In the simplest case we have to find out the average level of FEV_1 that corresponds to a specified number of years in particular environments. This, of course, is what is achieved by calculating the multiple regression of FEV_1 on the years lived in these environments. The regression coefficients are then precisely the conventional, arbitrary measures of decline in FEV_1, in so many litres or millilitres a year. The calculation is a standard one, but whether it is meaningful or not depends on the information available.

Multiple regression works by discovering how, among persons with the same amount of exposure to all other relevant agents, some index, in our cases the FEV_1, varies as the degree of exposure to a particular agent varies. This it does simultaneously for each relevant agent. The arithmetical calculation can almost always be made, but whether the answers are truly what they seem to be depends on what Fisher[4] called the configuration of the set of exposure data. Consider an industry in which everyone works either in the office or on the shop floor, and there is no interchanging between these two environments. Multiple regression will give the effect on FEV_1 of years of work in each environment perfectly correctly, but clearly the interpretation of the results depends on our being able to make the assumption that the two distinct groups of people, office workers and shop floor workers, were otherwise strictly similar; the results were not obtained from persons who had experienced both environments. If, at the other extreme, in this industry everyone does 25 years on the shop floor followed by 25 years in the office, the effect on the FEV_1 of these environments would be correctly estimated, but we might feel some anxiety because one environment was experienced only up to age 40 and the other only after age 40. One might be tempted to add age as another variable, but this would produce disaster, because age is no more than a constant plus the sum of the two durations of exposure. If the employment rule had been exactly enforced the multiple regression calculation would fail; if, on the other hand, there had been some deviations from the rule, so that some employees had done 23 years and some 27 on the shop floor before moving to the office, the calculation might proceed, but the results could well be entirely misleading since the effects of the two environments and of age would have been distinguished solely from what evidence these small discrepancies provided. In this as in other circumstances, the mere carrying out of a statistical calculation is an objective process, but the interpretation of the result of the calculation depends on every aspect of the raw data, and on the tacit assumptions that are made as well as on the explicit ones. In looking for factors related to the decline of FEV_1 the possible hidden links between these factors must be considered. In the example where there is interest in possible differential effects of years of work in several different environments several such links may occur.

First, men who have long durations of employment in one such environment will inevitably tend to have short durations of employment in others. Fortunately such negative links are seldom so great as to lead to difficulties, but clearly the data must be inspected to ensure that nothing remarkable in this way is present. The limiting case here is seen in the first example, where each man works only in one of the environments; the interpretive problem is then only that of establishing reasonable comparability of the groups.

Second, the more comprehensive the set of environments that are covered the more the total of the number of years spent in them will approach the age of the individual, or his age less the length of his schooldays. There is then a real risk that the individual results will become quite meaningless. The solution here, of course, is not to include age at all. In this case age in relation to decline of FEV_1 is only a pseudonym for a "variety of environments," the "constitutional tendency for the lungs to degenerate," or other imagined or unimagined processes linked to the passage of time. We might as well dispense with it, including "years in a neutral environment" to make up the occupational history to the full length of adult life. The only disadvantage of this would be that, if we wished to give measures of relative hazard to the various suspect environments we would need to subtract the contribution of "years in a neutral environment" from each before calculating their ratios.

In all such investigations the possible effects of smoking have to be allowed for. There is one totally unsatisfactory way of attempting it; this is to correct for smoking by the addition or subtraction of a constant appropriate to the smoking class. Evidently any effect of smoking would be expected to be related to the duration of the habit, being greater in older subjects. A device that achieves this is the introduction of a so called interaction between age and smoking, so that the age effect is modified when the dummy variable representing smoking is not zero. This leads to a rather complicated looking formulation, which is still somewhat too restrictive in that it assumes that the inherent variability of the FEV_1 is the same in all smoking classes. It is by no means certain that this is so. It is possible that smokers may be a little more variable than non-smokers and that ex-smokers may be even more variable (unpublished observations). The simplest and best solution is to make entirely separate analyses for the smoking classes, which are now not far from equality in numbers—4:3:3 for smokers:ex-smokers:non-smokers. This enables other

suspected or unsuspected differential effects of the various environments to be detected.

Clearly the investigation of factors relevant to the decline of FEV_1 is not and must not be taken as something that computers do without the intervention of human intelligence. Recently it has been strongly suggested that the sort of investigation discussed here can give entirely misleading results, and that the remedy lies in more complicated algebraic manipulation within the computer.[5] The first part of this suggestion is true, but the second part is false.

We have already seen the sort of background to the data, the sort of configuration of the data, that can lead to a misleading conclusion at the end of an investigation. In essence, this occurs when there are close but not exact relationships between several factors whose role in affecting decline of FEV_1 is to be investigated. An example was when the total number of years spent in certain environments was nearly equal to the subjects' ages, or at least working lifetimes. In these circumstances a computer program is given, say, three sets of data when in fact nearly all the available information is contained in any two of them. Precisely what effects are attributed to the individual sets of data will then depend solely on the pattern of deviations from the exact equivalence of any two to the third. Usually, in such circumstances, the results will be so remote from anything reasonable that it will at once be appreciated that something has gone wrong. The results will also be highly variable if the ingredients of the analysis are changed slightly—say, by omitting one of the environments or including a new one. Generally it should not be necessary to establish by some formal analysis that the results are unreliable, though such an analysis can readily be made.

The question then arises of what can be done to salvage some solutions. It was because of the configuration of the data that the problem arose; the information contained in the data was insufficient to enable the questions posed—"What roles do a set of environments play in determining the decline of FEV_1?"—to be answered. Clearly therefore further manipulations of these data cannot save the situation.

Sometimes it may be practicable to simplify the questions asked and still achieve useful answers. If it is possible to recognise which of the factors being studied is so closely linked to others that its separate effect cannot be distinguished, its omission may solve the problem—leaving, however, the answer in the ambiguous form 'Factor A or perhaps factor B, which is closely linked to it, or perhaps both cause a considerable increase in the rate of decline of FEV_1."

Sometimes, in an overambitious desire to extract as much as possible from the data, interactions between factors may be looked for in the analysis. Many computer packages enable this to be done at the touch of a button. Unfortunately the manuals for such packages do not usually include warnings about what exactly is implied by the term interaction. In this context it means that the dependence of FEV_1 on the years spent in a particular environment is thought perhaps to depend on the size of some other factor, such as age: for example, does the suspect environment have more effect on FEV_1 in an old man than in a young one?

The calculations necessary to estimate whether such an interaction exists require the multiplication of the values of the two possibly interacting variables, and multiplication requires the existence of meaningful zero points for each variable. With "years in a particular environment" the zero point is sensible and interpretable—there might have been no years in the environment, and this is reasonably counted as zero. But what about age? No one in the study will have zero age, though zero age at least means something. But in terms of level of FEV_1 number of years above age 20 or 25 is really a more appropriate measure, for this will indicate the time during which decline of FEV_1 has been occurring. On the other hand, in relation to the original concept of a differential effect in the old and the young the deviation of age from a zero value at 45 or 50 might be most appropriate.

The point of this is that the results obtained depend critically on the zero point chosen. The interaction term will not change, but the simple effect of the exposure variable will. In a recent study the level of FEV_1 in non-smokers depended not only on age (declining by 37 ml per year of age) but also on the amount of radiological simple pneumoconiosis, declining by 218 ml for each category of the ILO classification—(PD Oldham, unpublished observations). This is a case where it might be asked whether the effect of pneumoconiosis was the same at all ages, and so, for the purposes of this review the results of estimating an interaction were examined. First age itself was put in, multiplied by the pneumoconiosis score. The interaction term was insignificant but negative, suggesting that the older the subject the less effect pneumoconiosis would have on his FEV_1. The contribution of age itself was almost unchanged, the loss being 34 ml a year, but the contribution of pneumoconiosis changed from a significant loss of 218 ml per category to an insignificant *gain* of 18 ml. This could be a very deceptive result, because it is the artificial effect of measuring age from zero. Measuring age positively or negatively from age 40 instead of zero restored the effect of pneumoconiosis to a loss of 160 ml per category, the age and interaction effects being unchanged. This, of course, was because around the age of 40, near the mean age of the group, the interaction con-

tributes nothing, and the simple additive effects apply.

It may therefore be overambitious to try to look for complicated effects by incorporating them uncritically in an analysis. There have been cases where workers have deceived themselves badly by doing so. A little ingenuity in using the graphical facilities of modern computers, to look for the goodness of fit at the more extreme levels of the variables, may be a better and safer way of showing one's sophistication.

In summary, statistical calculations can be carried out without much thought, but the results may not then mean what they appear to mean. The decline of FEV_1 remains a most powerful measure of the possible adverse effects of an environment, but understanding its meaning may not be achieved without a fair degree of careful thought and the due application of common sense.

PD OLDHAM
University College
Cardiff CF1 1XL

References

1 Peto R, Speizer FE, Cochrane AL, *et al*. The relevance in adults of airflow obstruction, but not of mucus hypersecretion, to mortality from chronic lung disease. *Am Rev Respir Dis* 1983;**128**:491–500.
2 Cotes JE. *Lung function: assessment and application in medicine*. 4th ed. Oxford: Blackwell Scientific Publications, 1979.
3 Fletcher C, Peto R, Tinker C, Speizer FE. *The natural history of chronic bronchitis and emphsema*. Oxford: Oxford University Press, 1976.
4 Fisher RA. Uncertain inference. *Proceedings of the American Academy of Arts and Science* 1936;**71**:245–58.
5 Perry J. Account of 145th ordinary meeting of the British Region of the Biometric Society. *Biometric Bulletin* 1986;**3**(2):5.