Thorax 1984;**39**:828–832

Peak flow rate records in surveys: reproducibility of observers' reports

KATHERINE M VENABLES, P SHERWOOD BURGE, ANTHONY G DAVISON, ANTHONY J NEWMAN TAYLOR

From the Occupational Medicine Department, Brompton Hospital, London, and Solihull Hospital, Solihull, West Midlands

ABSTRACT Records of peak expiratory flow rate (PEFR), commonly used in hospital in the management of asthma, have not been evaluated as a method of identifying cases of asthma in population surveys. Four observers were asked to report on whether asthma was present or absent in 61 graphs of PEFR recorded two hourly for four weeks during surveys of working populations. Agreement within individual observers was measured using a subset of 29 graphs which had been copied and distributed at random among the set of 61; agreement was good, from 90% in one observer to 100% in two. Agreement between observers was measured on the basis of all 61 graphs. Agreement occurred between all four observers in 69% of graphs, between at least three out of four in 97%, and, when pairs of observers were examined, between 72% and 93% of graphs. Graphs assessed as showing asthma demonstrated more within day PEFR variability (expressed as the number of days in which the difference between maximum and minimum readings was at least 15%) than graphs assessed as not showing asthma. Some graphs with little within day variability were assessed as showing asthma, apparently because they demonstrated between day PEFR variability.

Asthma, defined as variable airflow limitation,[1] can be demonstrated by regular monitoring of lung function, conveniently performed by recording the peak expiratory flow rate (PEFR) over several days or weeks. Such records may be assessed by inspecting the raw readings or graphs drawn from the readings, and this is now common in hospital in the evaluation of the severity of asthma and its response to treatment.[2] PEFR recording by patients outside hospital has been encouraged by the introduction of the miniature meter,[3] which is extending the use of PEFR records to surveys of asthma in populations.

The widespread acceptance of PEFR records in clinical practice is an endorsement of the method's usefulness. Its validity in diagnosis is, however, difficult to estimate formally, there being no agreed standard test for asthma against which it could be compared. Techniques for identifying disease in epidemiological surveys should be reproducible as well as valid, and for PEFR records one important potential source of variation is differences among observers providing reports on the records. In contrast to hospital practice, records from surveys are assessed in isolation by an observer who is "blind" to other relevant information. Any abnormalities are likely to be minor and difficult to interpret. Variation in reporting might be a serious problem in surveys, as it is when physicians take a history of respiratory symptoms,[4 5] examine the chest,[5 6] or look at chest radiographs.[7 8]

We have therefore taken records made during surveys of working populations and measured observer variation in the detection of asthma from these records alone. We have also attempted to identify factors which influenced these observers in their reporting.

Address for reprint requests: Dr KM Venables, Occupational Medicine Department, Brompton Hospital, London SW3 6HP.

Accepted 21 May 1984

Methods

SUBJECTS

Recordings from 61 men formed the basis of the study. Thirty eight men were currently employed in a steel coating plant where isocyanate induced occupational asthma had occurred from 1972 to 1979[9]

and 23 were employed in an electronics factory where acid anhydride induced occupational asthma had occurred in 1979 and 1980.[10] These PEFR records formed part of follow up studies at the two workplaces in 1982 after occupational exposures had been controlled by, respectively, substitution of another chemical and improved ventilation. On clinical grounds 15 men were originally classified as having asthma related to occupation, 23 as having asthma unrelated to occupation, 20 as having respiratory symptoms not caused by asthma (but by, for example, chronic bronchitis or non-specific irritation), and three as having no respiratory symptoms (but with serum antibody against an acid anhydride-protein conjugate). Many of the subjects, including 12 of the 15 men with occupational asthma, had lost their symptoms or had partially improved by the time the PEFR records were made. We therefore confined this study to the identification of asthma rather than occupational asthma.

PEFR RECORDS

Each person was given a mini Wright peak flow meter[3] and instructed in its use. He was asked to note, on a standard form, the best of three readings taken every two hours during waking hours for four weeks. The readings were plotted as previously described[11] as graphs of the maximum, minimum, and mean of each day's PEFR, days at work being indicated by shading. The graph is plotted by computer, which also prints each day's within day PEFR variability, defined as the difference between the day's maximum and minimum readings expressed as a percentage of the maximum. For each record the number of "variable" days, with a difference of at least 15% between maximum and minimum readings, provided an index of PEFR variability.

REPORTING

Four observers (the authors) who were experienced in using these records assessed each graph independently and without knowledge of the subjects' identities, symptoms, or exposure. We used a four point scale: 4—definite asthma; 3—probable asthma; 2—probably not asthma; 1—definitely not asthma. This scale was chosen because it offered comments which resembled those made spontaneously on graphs from other surveys and it forced the observer to make a decision on whether asthma was or was not present while recognising the potential difficulties in reporting on graphs from a working population using limited information.

Table 1 *Use of the asthma assessment scale by four observers, reporting on 61 graphs*

Observer		*A*	*B*	*C*	*D*
Asthma	Definite	18	20	23	11
	Probable	2	4	8	11
Not asthma	Probable	9	14	3	8
	Definite	32	23	27	31
Total		61	61	61	61

COMPARISON OF REPORTS

Twenty nine graphs were taken at random from the set. Copies were made and the copies then returned at random together with the originals. Intraobserver variation was measured by comparing each observer's reports on the 29 duplicates. All 61 graphs were used in measuring interobserver variation and, in the case of the duplicated graphs, the first of each pair was used. Each observer therefore assessed a total of 90 graphs. Complete agreement was defined as the use of the same point on the four point scale. Substantial agreement was defined as either agreement that asthma was definitely or probably present or agreement that it was definitely or probably not present. When used without qualification, "agreement" means either substantial or complete agreement.

Results

Table 1 shows the different reporting patterns of the four observers and table 2 the intraobserver variation in reporting. These reports agreed in 26–29 (90–100%) of 29 graphs and the agreement was complete in 22–27 (76–93%). Table 3 shows the interobserver variation between all four observers and between the six pairs from the four observers. All four observers agreed on the presence or absence of asthma in 42 (69%) of the 61 graphs and three out of four agreed in a further 17 (28%), so that in 59 (97%) of the graphs most or all the observers were in agreement. For pairs of observers agreement varied from 44 to 57 (72% and 93%) of the graphs.

Figure 1 shows, for the 59 graphs where observers agreed, their assessment that asthma was or was not present compared with the index of within day

Table 2 *Intraobserver variation in the assessment of 29 pairs of graphs*

Observer	*A*	*B*	*C*	*D*
Complete agreement	22	24	26	27
Substantial agreement	4	3	3	2
Disagreement	3	2	0	0
Total	29	29	29	29

Table 3 *Interobserver variation in the assessment of 61 graphs*

Observer combination	*All*	*B/D*	*B/C*	*C/D*	*A/D*	*A/B*	*A/C*
Complete agreement	29	35	43	39	41	41	37
Substantial agreement	13	22	9	13	10	10	7
Majority agreement	17						
Disagreement	2	4	9	9	10	10	17
Total	61	61	61	61	61	61	61

PEFR variability (the number of days with a difference of at least 15% between maximum and minimum readings). All 22 graphs thought by most or all of the observers to show asthma contained at least one variable day and all nine where observers were in complete agreement that asthma was present contained at least four variable days. None of the 37 graphs thought by most or all of the observers not to show asthma contained more than three variable days and none of the 20 graphs where observers were in complete agreement that asthma was not present contained any variable days. Graphs with one, two, or three variable days were classed both as asthma and as not asthma, and in none of these graphs was there complete agreement between observers.

Figure 2*a* shows a graph with 16 variable days thought by all observers to show definite asthma and figure 2*b* a graph with no variable days thought by all definitely not to show asthma. In contrast to these examples, where degree of within day variability was related to the observers' assessments, are the examples shown in figure 3. Figure 3*a* shows a graph with two variable days on which there was substantial agreement that asthma was present: it shows a consistent pattern of between day PEFR variability, suggestive of mild work related asthma, rather than within day variability. Figure 3*b* shows a graph with one variable day on which there was substantial agreement that asthma was not present: it shows no clear and consistent pattern.

In the subgroup of 23 electronics workers an assessment of respiratory symptoms was made at the start of the PEFR record. Of the eight people whose

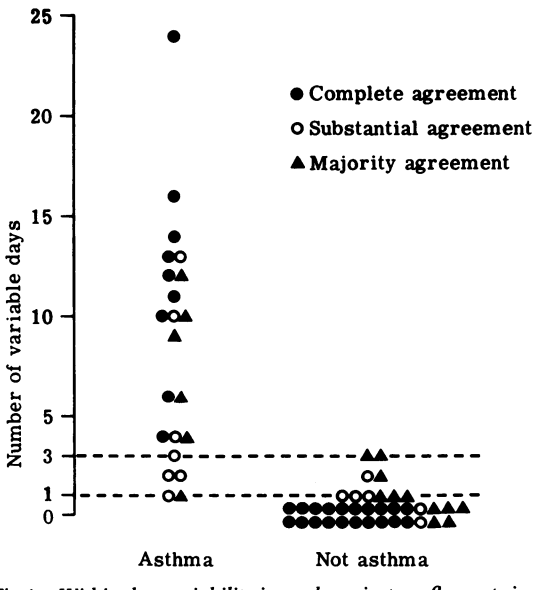

Fig 1 *Within day variability in peak expiratory flow rate in graphs assessed by the observers as either showing or not showing asthma. Within day variability is expressed as the number of "variable" days in the record (in which the difference between maximum and minimum readings was at least 15%).*

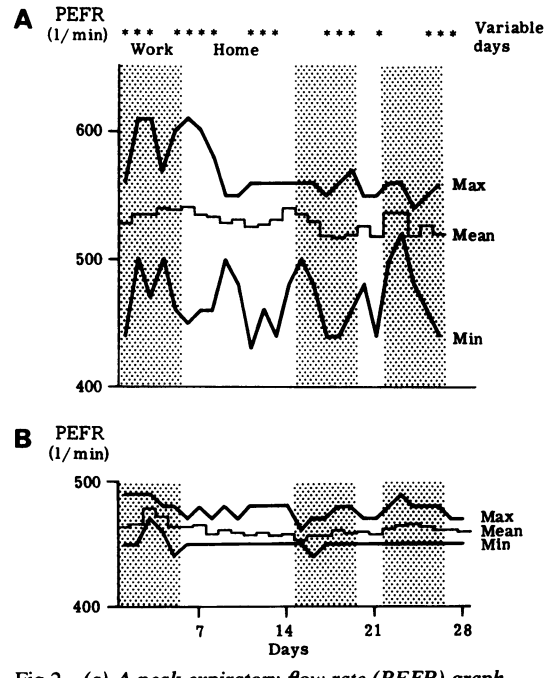

Fig 2 *(a) A peak expiratory flow rate (PEFR) graph thought by all observers to show definite asthma; (b) a graph thought by all observers definitely not to show asthma.*

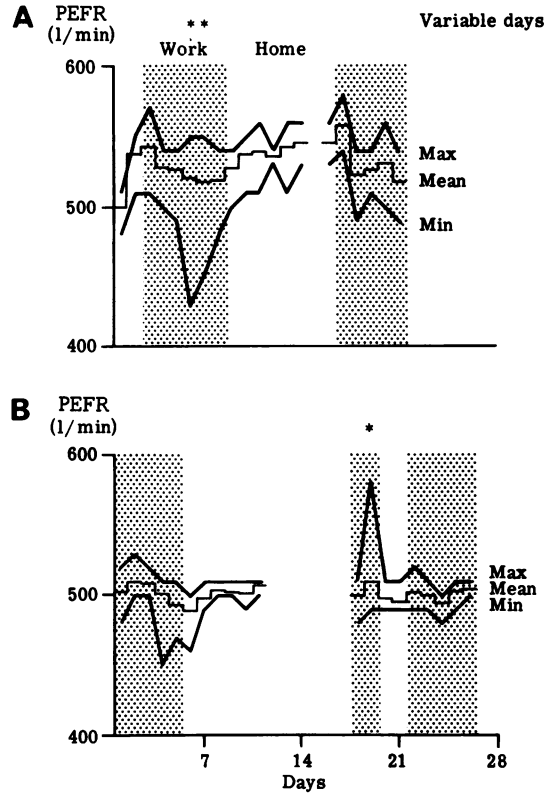

Fig 3 *(a) A peak expiratory flow rate (PEFR) graph with two variable days thought by all observers to show either definite or probable asthma; (b) a graph with one variable day thought by all observers either definitely or probably not to show asthma.*

graphs were reported by either three or four observers to show asthma, seven experienced current or intermittent wheeze, breathlessness, chest tightness, or unproductive cough and one was symptomless but had serum antibody against an acid anhydride-protein conjugate. Of the 13 thought by either three or four observers not to show asthma, four were symptomless, four complained only of chronic productive cough, four complained of either unproductive cough or intermittent wheeze or breathlessness (none experiencing these symptoms at the time of the record), and one complained of current wheeze and breathlessness but had inhaled corticosteroid and bronchodilator treatment regularly throughout the period of the record.

Discussion

In clinical practice patients referred to hospital with asthma usually have PEFR records showing clear variability within and between days. In contrast, the subjects of this study were working people; the asthmatic subjects had mild asthma and few had consulted a doctor about their symptoms; the others had no symptoms or trivial ones. The fact that physiological changes in this working population were, in general, minor was probably a factor in producing the observer variation we describe.

The reproducibility of the observers' reports was good, particularly the reproducibility of individual observers, who agreed with their own assessments in the great majority of the 29 graphs included in the study of intraobserver variation. Interobserver variation was greater and some may have been semantic, observers differing on the meaning of the assessment scale rather than the meaning of the graphs. This may explain, for example, the pattern of observers B and D, who agreed on the presence or absence of asthma in almost all the 61 graphs but differed on whether it was "definitely" or "probably" present or absent in many. On the other hand, observers A and C disagreed on the presence or absence of asthma in over a quarter of the graphs. Semantic differences may have played a part, or possibly the two used slightly differing criteria for the detection of asthma; or their difference may be explained by differences in experience or personality impossible to evaluate in this study. One of the important functions of studies of observer variation is the stimulation of debate and the development of consensus on criteria for disease detection. We would expect that a second study, similar to this one, with the same observers would show less variation.

As expected, both within day and between day variability of PEFR were criteria used in the detection of asthma. An index of within day variability, the number of variable days, correlated well with the observers' combined assessments, although not necessarily with those of individual observers. In general, records with four or more variable days were regarded as showing asthma and those containing none were not thought to show asthma. Those with one, two, or three variable days could be classed either way, and there was never perfect agreement on the classification of this group of graphs. The use of between day variability as a criterion was most easy to identify in this group, which is exemplified by figure 3*a*; and this criterion was presumably also used in the interpretation of the other graphs. The pattern of a PEFR graph should be consistent with our knowledge of PEFR variability in health and disease; isolated high readings, such as that in figure 3*b*, are not and may result from errors in technique or transcription during these unsupervised records. We have noted this previously[12] and

have also noted that PEFR variability in the first days of a record, while the subject is learning to use the meter, may be spurious. In clinical practice doubts about graphs like these can be resolved by questioning the patient or repeating the record, but in surveys this is not possible.

PEFR records should be susceptible to a mathematical analysis and our preliminary work[13] (also KM Venables, unpublished cusum analyses) suggests that the complex and various patterns of asthma will require a sophisticated analysis for computer based reporting to be reliable. The human observer will, however, be necessary, at the least to test the validity of new mathematical models and to supplement computer reporting.

We were able to compare symptoms and PEFR reports in the subgroup of 23 electronics workers and found a close relationship between the presence or absence of asthmatic symptoms at the time the record was made and the PEFR reports made without knowledge of these symptoms.

The epidemiology of asthma is a developing subject and the PEFR record appears to be a useful method of identifying asthma. The technique is non-invasive, inexpensive, and in our experience acceptable to subjects. We have demonstrated that observers report reproducibly on the records. We hope that this report will encourage the use of PEFR records in surveys in occupational and non-occupational contexts.

We wish to acknowledge the help given by Rosemarie Hawkins, who transferred the readings into a computer file; by David Hughes, who designed a program for plotting them in a graphical form; and by Jeanie Thomson, who typed the manuscript.

References

[1] Scadding JG. Meaning of diagnostic terms in bronchopulmonary disease. *Br Med J* 1963;ii:1425–30.
[2] Turner-Warwick M. On observing patterns of airflow obstruction in chronic asthma. *Br J Dis Chest* 1977;**71**:73–86.
[3] Wright BM. A miniature Wright peak-flow meter. *Br Med J* 1978;ii:1627–8.
[4] Cochrane AL, Chapman PJ, Oldham PD. Observers' errors in taking medical histories. *Lancet* 1951; i:1007–9.
[5] Schilling RSF, Hughes JPW, Dingwall-Fordyce I. Disagreement between observers in an epidemiological study of respiratory disease. *Br Med J* 1955;i:65–8.
[6] Fletcher CM. The clinical diagnosis of pulmonary emphysema—an experimental study. *Proc R Soc Med* 1952;**45**:577–86.
[7] Fletcher CM, Oldham PD. The problem of consistent radiological diagnosis in coalminers' pneumoconiosis. *Br J Ind Med* 1949;**6**:168–83.
[8] Cochrane AL, Garland LH. Observer error in the interpretation of chest films. *Lancet* 1952;ii:505–9.
[9] Dally MB, Burge PS, Davis G, *et al*. Occupational asthma in a steel coating plant: population study leading to identification of a cause. *Thorax* 1982;**37**:229.
[10] Howe W, Venables KM, Topping MD, *et al*. Tetrachlorophthalic anhydride asthma: evidence for specific IgE antibody. *J Allergy Clin Immunol* 1983;**71**:5–11.
[11] Burge PS, O'Brien IM, Harries MG. Peak flow rate records in the diagnosis of occupational asthma due to colophony. *Thorax* 1979;**34**:308–16.
[12] Venables KM. Two objective methods for the diagnosis of occupational asthma in epidemiological surveys. MSc dissertation, University of London, 1980.
[13] Burge PS. Prolonged and frequent recording of peak expiratory flow in workers with suspected occupational asthma due to colophony or isocyanate fumes. MSc dissertation, University of London, 1978.