Monaldi Arch Chest Dis
2001; 56: 3, 281–288

OASYS-3: improved analysis of serial peak expiratory flow in suspected occupational asthma

P. Bright*, D.T. Newton#, P.F.G. Gannon#, C.F.A. Pantin#, P.S. Burge*

ABSTRACT: *OASYS-3: improved analysis of serial peak expiratory flow in suspected occupational asthma. P. Bright, D.T. Newton, P.F.G. Gannon, C.F.A. Pantin, P.S. Burge.*

Serial peak expiratory flow (PEF) records are useful for the screening and diagnosis of patients with occupational asthma. We have presented work on a method based on linear discriminant analysis (OASYS-2) and have now tried to improve on this technique and investigate the repeatabilities of expert interpretation.

268 serial PEF records made by workers with possible occupational asthma were divided into four sets. The first two were development sets; development set 1 was used to develop a discriminant analysis to match the human expert, who scored each complex (a sequence of work days, then rest days, then work days, or its counterpart: rest-work-rest) for its likelihood of having a work-related effect. This was modelled using 729 new measurements from each complex, together with the 50 measurements originally used in the development of OASYS-2. Linear discriminant analysis is a statistical technique which can refine from a large number of indices a narrower range that best predict a known outcome. Combinations of these indices are weighted and used to assign outcomes for novel cases. Models were produced containing a combination of measurements that generated scores best matching the expert. Development set 2 was used to test the model. Sets 3 and 4 were "gold standard" sets where the diagnosis had been made independently of the PEF record. Set 3 was used to set the cut-off for an occupational effect, the sensitivity and specificity for the combined model was determined from the fourth (gold standard) set. The fourth set was also used to determine the sensitivity and specificity of the human expert.

The repeatability of the human expert re-scoring the same complexes had a weighted kappa score of 0.71. OASYS-3 was 100% repeatable. Comparing the scores awarded to whole records by the new (OASYS-3) and OASYS-2 analysis methods with the scores awarded by the human expert revealed mean (95% CI) differences of –0.28 (0.30, –0.26) and –0.34 (–0.37, –0.31) respectively. Hence both OASYS-3 and OASYS-2 tended to score records less positively for work-related changes in PEF than the expert. OASYS-3 scored complexes marginally better than OASYS-2. The sensitivity of OASYS-3 was better than OASYS-2 (82% and 75% respectively) for an equivalent specificity (94%). The sensitivity of the human expert was 100% with a specificity of 93%.

OASYS-3 provides an objective method of interpreting serial peak flow records with a sensitivity and specificity approaching that of a human expert and is a modest improvement on OASYS-2.
Monaldi Arch Chest Dis 2001; 56: 3, 281–288.

***Keywords:** Peak expiratory flow, serial measurement, occupational asthma, discriminant analysis, repeatability.*

**Occupational Lung Disease Unit Centre, Birmingham Heartlands Hospital, Birmingham, #Industrial and Community Health Research Centre, Keele University School of Postgraduate Medicine, North Staffordshire Hospital, Stoke-on-Trent, U.K.*

Correspondence: Dr. P. Bright, Occupational Lung Disease Unit, Birmingham Heartlands Hospital, Birmingham, B9 5SS; e-mail: Phil.bright@btinternet.com

Received: December 5, 1997; accepted after revision: April 2, 2001.

Introduction

A reliable, objective and readily available method for the interpretation of serial peak expiratory flow (PEF) records is needed for the screening and diagnosis of patients with occupational asthma. Until recently the most specific and sensitive method for the interpretation of these records was by expert visual inspection. This requires expertise that is not widely available, is subjective and not necessarily reproducible between individuals. BURGE *et al.* [1] described the results of a study on colophony workers using the maximum, mean and minimum plots. They required 75% of the working week to show specific (but undefined) patterns. The sensitivity was 77%, specificity 100%. Côté [2] 'visually' analysed records on workers with cedar asthma and found sensitivity of 86% and specificity of 89%. Agreement was needed by 2 out of 3 readers for 2 out of three weeks at work. Lack of expert availability may be a factor in the under-diagnosis of occupational asthma. Several attempts have been made to produce objective statistical based methods for the interpretation of serial PEF records but these have failed to approach the sensitivity and specificity of expert interpretation. We previously presented work [3] on an interpretation method based on linear discriminant analysis – OASYS-2. OASYS-2 plots the serial PEF record as the daily maximum, mean and minimum. The record is divided into complexes of days at work followed by rest days followed by a further period of work days (Work-Rest-Work, WRW complex) or rest days then work days then rest days (Rest-Work-Rest, RWR complex). From each complex, measurements are made from two adjacent periods and entered into the analysis program. For whole PEF records, this method has a specificity of 94%, approaching that of human experts but a sensitivity lower than human experts of 75%. The score for whole records is calculated from the scores of the component complexes. OASYS-2 tends to underscore those parts of a record judged to have a definite work-related effect by the expert [3]; this may lead to under-interpretation of the whole record. The problem with OASYS-2 may lie with the method used to determine the measurements made from the PEF record or with the technique of discriminant analysis used to derive the analysis model.

In this study we present an attempt to improve the accuracy and robustness of the discriminant analysis approach. We have altered the method used to produce the measurements entered into the program to more closely resemble that used during expert interpretation. The expert's sensitivity and specificity were assessed against gold standard set 4, which was also used to determine that of the OASYS-3 analysis. We have also investigated the reproducibility of the expert, and the influence of the expert seeing a complex in the context of the whole record rather than as isolated complexes on the scoring of the individual complex. It may be, for instance, that the score awarded by the expert to an individual complex is adjusted according to an overall impression of the record.

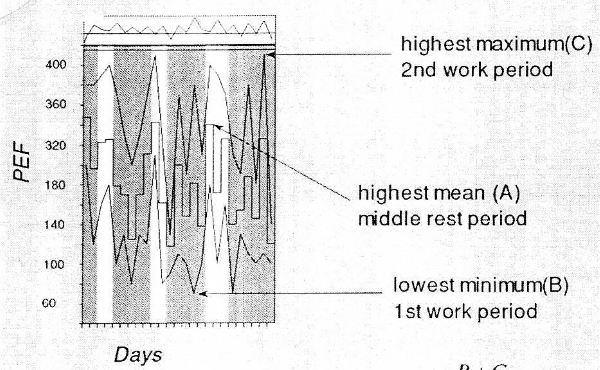
Method

Workers attending a specialist occupational lung disease clinic for investigation for occupational asthma were asked to record serial PEF records two-hourly from waking to sleeping, as previously described [3] using a mini-Wright meter with a Wright scale. Serial PEF records were plotted as the daily maximum, mean and minimum as described by Burge [1, 4].

Two hundred and sixty eight collected records were divided into four sets; development sets 1 & 2, for various stages of the development and testing of the discriminant analysis models, and gold standard sets 3 & 4, for evaluation against an external standard. Development sets 1 and 2 were identical to those used in the previous study [3]. Each record was divided into a series of overlapping complexes, either a period of days at work either side of a period away from work (a rest-work-rest complex), or its counterpart, a work-rest-work complex. Each complex was scored subjectively by an expert for the presence of a work-related effect on a scale of 0-100%, and translated into a score of 1 to 4. A zero probability of a work effect was given a score of 1, probabilities between 1 and 49% were given a score of 2, between 50 and 99% a score of 3 and complexes showing a definite work effect a score of 4. Any record showing evidence of a respiratory tract infection or a progressive increase or decline in the PEF over several days (learn and laze effects [5]) was excluded. The presence of respiratory tract infections was determined either because the worker reported consistent symptoms or because a typical pattern suggestive of a respiratory tract infection was observed in the peak flow record [6].

Development set 1 was used to develop the discriminant analysis. For a Work-Rest-Work complex measurements between the initial work, the central rest and the final work period for combinations of the highest, average and lowest readings of the daily maximum, mean and minimum were made. A total of 729 such measurements are possible. An equivalent combination of measurements were made for Rest-Work-Rest complexes. An illustration of the calculation of one of the measurements for a Work-Rest-Work complex is given in figure 1. In this example the average of the PEF values of the lowest minimum from the first work period and the highest maximum from the second work period is subtracted from the highest mean from the rest period. In the terminology used in the Appendix, the average of the measurements in columns '1' and '3' are subtracted from the measurement in column '2', to give a PEF measurement that is entered into the discriminant analysis. In addition the original fifty measurements used in the development of the 2-way analysis were included. Two models were produced by entering measurements made on development set 1 separately for Work-Rest-Work and Rest-Work-Rest complexes as the independent variables into a linear discriminant analysis programme [7]. For each model forward selection of variables was employed. The possible explanatory variables were added separately to initial models containing only a constant term and the significant variables determined. Returning to the initial model containing only the constant term, the most significant variable was included and the improvement in the model tested. The process was repeated until none of the remaining variables made a significant contribution to the model. For discriminant analysis the criterion used to assess variables for inclusion is the increment in the model sum of squares. An F-value corresponding to the increment given by each variable is calculated and the largest F-value chosen by comparison with F-values from tables corresponding to the chosen significance level (5% in this study). If the derived F-value is smaller than the value from tables the procedure stops without adding the variable under consideration to the current model [8]. These models were used to generate scores for development set 2 and the gold-standard sets.

The records in development set 2 were scored four times by the expert at least two weeks apart and without knowledge of the previous scores. So



$$measurement = A - (\frac{B+C}{2})$$

Fig. 1. – Illustration of the calculation of a measurement from a Work-Rest-Work complex for OASYS-3.

WORKED EXAMPLE

3-way Measurement

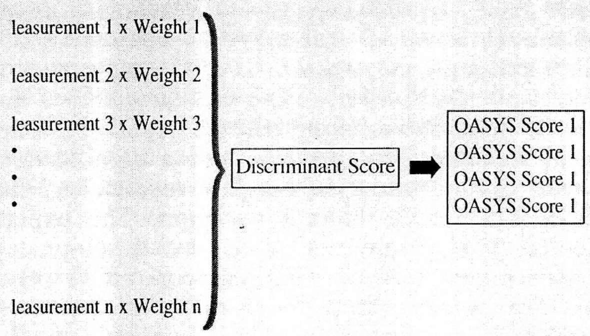
See Figure 1

Lowest Minimum	Highest Mean	Highest Maximum
74 L/min	340 L/min	418 L/min

Number entered into Discriminant Analysis =

$$340 - \frac{74 + 418}{2}$$

For a 2-Way Measurement the number entered into the discriminant analysis is simply the difference between the two measurements.
The Measurements are then combined to produce a discriminant analysis score which determined allocation of the complex to a particular OASYS complex score.



that each record only contributed one complex, a complex from each record was randomly chosen and used to compare the repeatability of the expert in scoring complexes. Repeatability was calculated by comparing the weighted kappa scores for combinations of pairs of scoring sessions.

Two sets of gold standard records used in the previous study were also utilised. Each gold standard set consisted of groups of records with and without occupational asthma. The sets of 'gold-standard positive' records were collected on subjects in whom the diagnosis of occupational asthma had been substantiated by methods other than the serial PEF records. To be included subjects had to have a history consistent with occupational asthma and either a positive specific bronchial challenge test, a positive RAST to the suspect substance or a change of four fold in the histamine reactivity (Yan's method [9]) after a week at work compared to measurement after at least a week away from work. The serial PEF records on these subjects had to be recorded during a time when they were known to be working with the suspect substance. The sets of gold-standard negative records were collected by combining records from two sources; (a) from workers who had been completely removed from exposure to an agent known to have induced occupational asthma in them, (b) this group was supplemented by asymptomatic post-office workers who had participated in a cross-sectional survey of respiratory symptoms. Various indices that described the nature of the sets of records used were calculated. All diurnal variations (DV) are stated as % predicted after linearisation [10]. That is, the PEF measurements (all made on mini-Wright peak flow meters) were corrected for the non-linearity of the meter [11] and then the diurnal variation calculated as the daily maximum minus the minimum expressed as a percentage of the predicted PEF for that patient.

Evaluating the Models

Using development set 2, the scores awarded for each Work-Rest-Work or Rest-Work-Rest complex were compared with those awarded by the human expert by Cohen's weighted Kappa. Whole record scores were calculated as the average of the scores for a record's component complexes with a weighting of x2 given to scores of 1 and 4 so as to add emphasis to definitely negative and definitely positive complexes. Whole record scores were compared using the method of Bland and Altman [12]. The whole record scores for gold standard set 1 were used to determine a cut-off score between records with and without a work-related effect. This cut-off score was then applied to gold standard set 2 to derive the sensitivity and specificity of the models.

The human expert sensitivity and specificity were determined by blinded evaluation of whole records from gold standard set 4 for the presence or absence of a work related effect.

Results

Details of the records are shown in table 1. Both development sets' records were approximately the same length and had the same number of readings per day, consistent with the instruction to perform PEF measurement every 2 hours. Approximately half of each set had a mean diurnal variation of <15% but very few did not have at least one day with a diurnal variation >15%. Workers providing records had a wide range of occupational exposures representative of those seen in a spe-

Table 1. – Details of the data sets used in the development and testing of OASYS-3

Data Set	N	Length in days mean (95% CI)	Readings /day mean (95% CI)	% pred diurnal variation mean (95% CI)	% of record with diurnal variation >15% mean (95% CI)	Number of records with mean diurnal <15% variation	Number of records having no day with >15% diurnal variation
Development set 1	81	32.2 (29.1, 35.4)	7.9 (7.5, 8.4)	16.8 (14.6, 19.0)	44.2 (36.9, 51.5)	44	3
Development set 2	60	36.7 (32.0, 41.4)	7.7 (7.3, 8.2)	16.3 (14.0, 18.7)	43.1 (34.0, 52.2)	30	3
Gold Standard Sets							
Set 1 +	27	29.8 (26.8, 32.9)	7.6 (7.0, 8.1)	21.1 (16.7, 25.5)	59.2 (45.8, 72.5)	10	0
Set 1 –	33	22.5 (20.0, 25.0)	8.0 (7.3, 8.7)	12.5 (10.4, 14.7)	29.9 (19.0, 40.7)	20	0
Set 2 +	32	33.1 (29.1, 37.2)	7.4 (6.8, 8.0)	27.0 (21.9, 32.0)	71.6 (61.5, 81.8)	6	2
Set 2 –	35	20.5 (19.2, 21.9)	8.4 (7.8, 9.0)	11.7 (9.3, 14.1)	21.2 (10.6, 31.7)	17	1

+ Evidence of occupational asthma diagnosed by methods other than serial peak flow records. – Workers who previously had occupational asthma but who have been removed from exposure, or asymptomatic post-office workers.

cialist occupational lung disease clinic [3]. The gold standard negative sets had a significantly lower diurnal variation but few records had no days with >15% diurnal variation. The expert scored most complexes in the two development sets as either showing a definite work effect or no definite work effect; those with intermediate scores of 2 or 3 were equally less common (table 2). The mean weighted kappa for the expert's repeatability was 0.71 (range 0.49 to 0.82).

Variables included in the discriminant analysis models to predict the scores for Work-Rest-Work and Rest-Work-Rest complexes were the same for each model and are given in the appendix. There were fourteen 2-way measurements and twelve 3-way measurements in each model. The most common component of the measurements was the Highest Maximum which formed part of approximately 65% of all measurements included in the OASYS-3 model and 57% of all measurements included in the OASYS-2 model; all other components occurring relatively infrequently (<10% for each).

Table 2. – Details of the Work-Rest-Work (WRW) and Rest-Work-Rest (RWR) scores awarded by the expert to the complexes in the development sets

Data Set	Complex Type	Score 1	2	3	4
Development set 1	WRW	108	39	39	51
	RWR	91	45	29	52
Development set 2	WRW	75	37	24	69
	RWR	74	44	29	66

Score 1 = definitely no work effect (0% probability)
Score 2 = probably no work effect (1-49% probability)
Score 3 = probable work effect (50-99% probability)
Score 4 = definite work effect (100% probability).

The weighted kappa for the comparison between complexes from development set 2 scored by the expert and OASYS-2 analysis was 0.55, for the expert and the OASYS-3 analysis, 0.51 and for OASYS-2 compared to OASYS-3, 0.65. The differences in record scores between scores awarded by OASYS-2, OASYS-3 and the expert showed a normal distribution (Kolmogorov-Smirnov Test [7]). The mean difference (limits of agreement) for the whole record scores for OASYS-2 minus the expert was –0.34 (–0.57 to –0.12), for OASYS-3 minus the expert was –0.28 (–0.44 to –0.13) and for OASYS-2 minus OASYS-3 was –0.06 (–0.3 to 0.18). Both the OASYS-2 and OASYS-3 methods tended to award scores to whole records significantly lower than the expert. OASYS-3 scored whole records marginally higher than OASYS-2's analysis. In all three comparisons the limits of agreement were narrow and within acceptable limits.

The details of the agreement for development set 2 between the expert, OASYS-2 and OASYS-3 for both Work-Rest-Work and Rest-Work-Rest complexes combined are given in table 3. Agreement with the expert for the important scores of 1 and 4 was relatively poor. Agreement with the expert for scores of 4 was better for OASYS-3 than OASYS-2.

A plot of the sensitivities and specificities from gold standard set 1 scored by OASYS-3, for a range of cut-off scores between 1 and 4, is given in figure 2. To maximise specificity a cut-off score of ≥2.88 was chosen to differentiate positive from negative records. Using this cut-off score the sensitivity was 82% and specificity was 94% for gold standard set 2. There were six gold standard positive records and two gold standard negative records classified incorrectly. The false negative records all had areas where there appeared to be a possible change in exposure, although none was

Table 3. – Scores awarded by the expert for Work-Rest-Work (WRW) and Rest-Work-Rest (RWR) complexes from development set 2 compared to the scores awarded by OASYS-2 and OASYS-3

		Expert							
		1		2		3		4	
		Oasys-2	Oasys-3	Oasys-2	Oasys-3	Oasys-2	Oasys-3	Oasys-2	Oasys-3
	1	92	86	20	26	6	7	5	5
OASYS	2	47	42	45	34	23	22	19	20
	3	8	18	16	21	18	19	50	38
	4	0	1	2	2	6	5	61	72

recorded by the subject. A number of complexes were underscored by OASYS-3 when the record was assessed by the expert. The two false positive records were of asymptomatic post-office workers. These were the same records reported by Gannon [3] as false positive when scored by OASYS-2. Figure 3 gives the plot of one such worker who worked a night shift and had definite work related declines in PEF as assessed by the expert.

The human expert's sensitivity was 100% with a specificity of 93%. Of the three records where the expert disagreed with the gold standard classification, all were Gold Standard positive records, scored as negative for occupational asthma by the expert. The expert commented in one that there was a high degree of peak flow variation across the whole record making interpretation difficult. In the second, that there was a long decline in the peak flow transcending work and rest periods. While in the third there was only one obvious night shift with a marked deterioration in the peak flow, other changes being relatively minor.

Discussion

Several studies have demonstrated the value of PEF in the diagnosis of occupational asthma [1, 4, 13-15]. To date OASYS-2 is the best computer based method for analysing serial PEF records in workers suspected of having occupational asthma. It is however not as sensitive as an expert [1, 4, 16, 17], and relies on measurements which form only a small part of the record [3]. OASYS-3 alters the method for measurement between work and rest periods, incorporating information from all three parts of a complex and improves the sensitivity for an equal specificity. When scoring the same set of records OASYS-3 still did not match the sensitivity of the human expert. OASYS-3 was better at scoring complexes showing a definite work related effect than OASYS-2. Both appeared to be sensitive to different aspects of the record, as agreement between the two systems had a weighted kappa of 0.65.

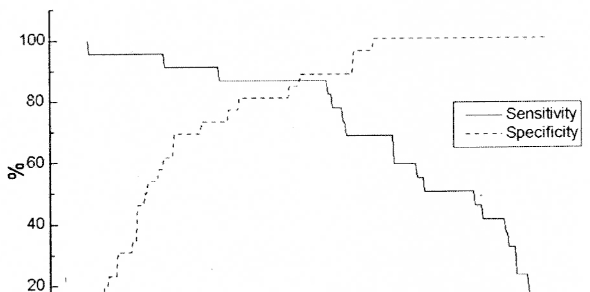


Fig. 2. – Plot of the sensitivities and specificities generated by moving the cut-off score for gold standard sets GS1⁺ and GS1⁻.

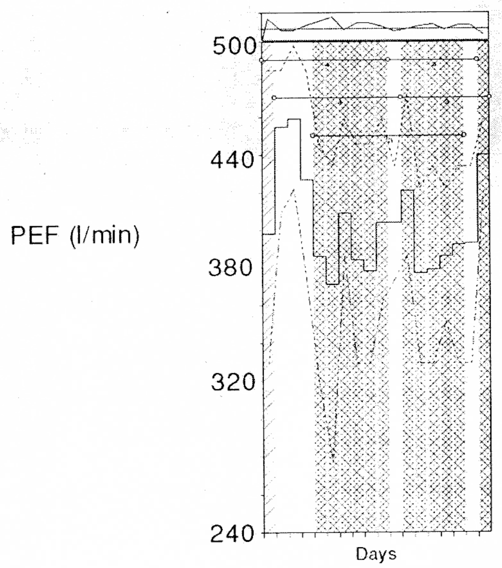


Fig. 3. – OASYS-3 plot one of the two Gold standard negative (asymptomatic) workers classified as having an occupational effect. The top panel shows the daily diurnal variation (percent predicted), showing that although asymptomatic the worker had a diurnal variation in the level usually classified as asthma. The middle panel shows the daily maximum (top line), mean (middle bar), and minimum (lower line) PEF. Days away from work have a clear background, days at work have a cross-hatched background (denoting night shift work), or a dashed background (afternoon shift work). The small numbers overlying the record are the scores allotted by OASYS-3. Visually there is improvement in the first weekend off work, deterioration in the first period of night shifts, with improvement on the last Sunday and the one day off work. There is deterioration in three of the next four night shifts, and improvement over the next weekend off work. There is no deterioration in the last work-period. The OASYS-3 scores are similar to the subjective expert analysis. The bottom panel shows the date, and the number of readings making up each "day". There are two days with three or less readings which may account for the changes on these days.

Most records in the development sets had a majority of days with the diurnal variation (expressed as percent predicted) above 15% but only half had a diurnal variation (measured as record average diurnal variation) of greater than 15%. Even so they form a group of records with a high degree of variability in the PEF. This is useful as the majority of patients presenting for investigation will usually have asthma and will need a determination made as to whether there is an occupational cause. Of the Gold Standard negative sets the diurnal variation tended to be lower than the Gold Standard positive sets. This raises the possibility that OASYS-3 is simply selecting out as positive those records with a high degree of variation, i.e. 'asthmatic' records. However, although the mean diurnal variations in the Gold Standard negative sets is lower, 13 of 33 for $GS1^-$ and 18 of 35 for $GS2^-$ had mean diurnal variations of greater than 15%. In addition very few records had no day with a diurnal variation of less than 15%. One strength of the study is the diversity of exposures present in the patients providing records for all the data sets. This avoids the possibility of training the analysis to recognise one stereotyped pattern of change produced by a particular agent.

The repeatability of the expert reader was good to excellent [18]. Agreement between readers and the repeatability of individual readers is important but not always appreciated. In studies to date, looking at relatively clear cut cases, Liss [16], for example, quoted a kappa score of 0.62-0.83 for inter-reader agreement, while MALO *et al.* [19] quoted intra-reader agreement as 83-100%. The repeatability of a single reader and the agreement between readers is likely to be a function of the subtlety of changes in the record and the method of plotting of the data used.

This study can be criticised because the analysis is modelled on only one expert. However, the expert has considerable experience in examining PEF records and has a known repeatability. Using several experts would inevitably lead to disagreements especially in the more difficult cases. Failure to include such cases at the development stage would be detrimental to the robustness of the analysis method. The final evaluation of the system is completely independent of the expert's opinion, being based on external standards. The gold standard positive cases, particularly those diagnosed by specific bronchial provocation testing, usually have the more obvious changes on the serial PEF record, such that the quoted sensitivity and specificity may not be matched with other less definite records. The group with positive specific IgE to an occupational allergen are particularly important in this regard, as no particular physiological change is needed for their inclusion. Studies have found that the relationship between non-specific bronchial hyper-responsiveness and work-related asthmatic symptoms alone was useful but of limited value [20-22]. In this study non-specific bronchial hyper-responsiveness was used in conjunction with other investigations as one of the definitions of a Gold Standard positive record. In contrast specific bronchial challenge testing is an accepted method of defining Gold Standard positive cases although can only be applied to a limited range of substances.

The 'subjective' assessment of PEF record by visual inspection often involves some objective rules. For instance, in the study by Burge [1] 75% of the working weeks or weekends had to show specific patterns; Côté [17] defined a positive record as one in which two out of three physicians agreed that two out of three weeks showed a work related change. In the Liss study [16] several criteria were employed. In this study we did not ask the expert to produce a subjective whole record score but split the record into complexes which were then subjectively scored. The method used to reassemble the scores for individual complexes was based on a set of arbitrary rules: the weighting of definitely negative and definitely positive complexes. This approach was chosen as it would have been impossible to develop a computer based method that interpreted the record as a whole. Expert interpretation of whole records may well be more sensitive than the OASYS models and certainly more sensitive than simple statistical methods especially when the work-related effect is intermittent. In such circumstances the expert will be able to focus on relatively small areas of the record showing a work-related effect, while statistical techniques would dilute such isolated work-related effects, especially in longer records. It is difficult to mimic the expert 'whole view' of the record using any kind of systematic approach suitable for computer analysis. Little is known about how an expert's mind actually works in interpreting the record; which part of the record has greatest influence, for example.

Studies also differ in the method used to assess the success of a particular analysis. In both OASYS-2 and OASYS-3 the sensitivities and specificities were derived from gold standard sets in which the diagnosis of occupational asthma was established independently of the PEF record. Methods for establishing a positive record are not themselves absolute and false-positive records may arise. Gold standard positive cases may not all have occupational asthma, as false positive challenge tests can occur, as well as changes in reactivity unrelated to work exposure. All however had a compatible history. Likewise, gold standard negative records may include workers with occupational asthma. Subjects thought no longer exposed may still be experiencing some exposure to substances known to cause their occupational asthma, or may be working with new provoking agents. Workers may have changes in their serial PEF record highly suggestive of occupational asthma but be included among gold standard negative cases because they remain asymptomatic and work in low risk occupations. Two asymptomatic workers in the gold standard negative set had features on their PEF records suggestive of work-related changes; both received high whole record scores from OASYS-3. The one illustrated in figure 3 commenced work at 2 am; the record could be confounded by changes in diurnal variation between work and rest days. Two other workers from the same sorting office did however have symptomatic occupational asthma, which was the reason for the original study.

The data entered into the discriminant analysis is only a small proportion of that available to the expert when interpreting a complex. This may account for the agreement with the expert for complexes measured by the Kappa scores, being lower than the repeatability of the expert. For example, in a standard 5 day work, 2 day off, 5 day work (Work-Rest-Work) complex there are a total of 36 (3x5 + 3x2 + 3x5) possible data points to enter. The number of possible measurements that could be made between all the data points (each day has a maximum, mean and minimum) is 1350 (15^2x6). In contrast there are eight data points used from the first period of a complex, nine from the second and seven from the third (Appendix), combined to form twenty six separate measurements. Entering all possible measurements into the discriminant analysis would be difficult, as the analysis would require a much larger amount of training data. Indeed, this study can already be criticised, as the number of indices entered into the analysis is already large in comparison to the amount of data used to train the analysis. Secondly, statistical methods like discriminant analysis competitively enter indices into the model. If a particular measurement does not significantly improve the model it will be excluded from the final model. It is possible therefore that if a particular pattern of change in a complex, for example, a three day recovery, occurs infrequently in the training data set, indices capable of recognising it will not be included in the final model. Finally, as with OASYS-2, OASYS-3 tends to mis-classify work-rest differences where the pattern of change is definite but the absolute difference is small. This occurs because the model uses absolute measurements in PEF rather than the pattern of change.

OASYS-3 has attempted to improve the sensitivity and specificity of computer assisted assessment of serial peak flow records compared with a human expert. The major change from OASYS-2 has been the inclusion of '3-way' measurements and a small improvement in overall sensitivity has been achieved. The excellent concordance with Gold Standard records is likely to deteriorate when the system is used on less selected data. It is likely that further improvement in the objective classification of occupational PEF records will require different techniques of analysis.

An improved method of defining gold standard negative cases is also needed.

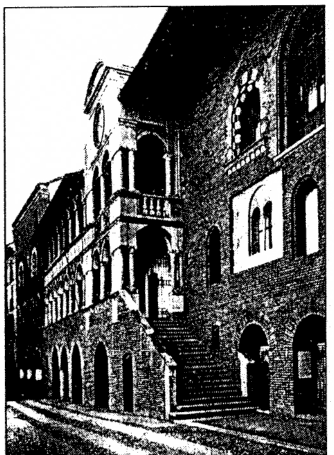
APPENDIX

Two-way measurements used in the OASYS-3 analysis model	*1*	2	*3*
	Highest Maximum	Highest Maximum	
	Highest Maximum	Lowest Minimum	
	Highest Maximum	Average Mean	
	Highest Maximum	Average Maximum	
	Highest Maximum	Average Minimum	
		Highest Maximum	Highest Maximum
	Average Mean	Highest Maximum	
	Highest Mean	Highest Maximum	
	Lowest Mean	Highest Maximum	
		Highest Maximum	Average Mean
		Highest Maximum	Highest Mean
		Highest Maximum	Minimum Mean
	Lowest Minimum	Highest Maximum	
		Highest Maximum	Lowest Minimum

Three-way measurements used in the OASYS-3 analysis model	*1*	2	*3*
	Highest Maximum	Highest Mean	Highest Maximum
	Highest Maximum	Highest Minimum	Highest Maximum
	Highest Maximum	Lowest Maximum	Highest Maximum
	Highest Maximum	Minimum Mean	Highest Maximum
	Highest Maximum	Highest Maximum	Highest Mean
	Highest Maximum	Highest Maximum	Highest Minimum
	Highest Maximum	Highest Maximum	Lowest Maximum
	Highest Maximum	Highest Maximum	Minimum Mean
	Highest Mean	Highest Maximum	Highest Maximum
	Highest Minimum	Highest Maximum	Highest Maximum
	Lowest Maximum	Highest Maximum	Highest Maximum
	Minimum Mean	Highest Maximum	Highest Maximum

References

1. Burge PS, O'Brien IM, Harries MG. – Peak flow rates in the diagnosis of occupational asthma due to colophony. *Thorax* 1979; 34: 308–16.
2. Cote J, Kennedy S, and Chan-Yeung M. – Sensitivity and specificity of PC_{20} and peak expiratory flow rate in cedar asthma. *Journal of Allergy and Clinical Immunology* 1990; 85: 592–598.
3. Gannon PFG, Newton DT, Belcher J, Pantin CFA, Burge PS. – The development of OASYS-2, a system for the analysis of measurements of peak expiratory flow in workers with suspected occupational asthma. *Thorax* 1996; 51: 484–9.
4. Burge PS, O'Brien IM, Harries MG. – Peak flow rate records in the diagnosis of occupational asthma due to isocyanates. *Thorax* 1979; 34: 317-23.
5. Bright P, Burge PS. – Characterisation of "Laze" and "Learning" effects on plots of serial peak flow records. *Eur Respir J* 1995; 8: 271s.
6. O'Brien C, Bright P, Nicholson C, Burge PS. – Patterns of peak expiratory flow response to upper respiratory infections in asthmatics. *Eur Respir J* 1995; 8: 272s.
7. SPSS for Windows. – SPSS for Windows (Version 6.1). SPSS Inc, Chicago, Illinois, USA, 1994.
8. Armitage P, Berry G. – Statistical Methods in Medical Research. Blackwell Scientific Publications, Oxford, UK, 1987.
9. Yan K, Salome C, Woolcock AJ. – Rapid method for measurement of bronchial responsiveness. *Thorax* 1983; 38: 760–5.
10. Gannon PFG, Newton DT, Pantin CFA, and Burge PS. – Different methods for calculating diurnal variation in peak expiratory flow (PEF). *Eur Respir J* 1992; 5: 200s.
11. Miller MR, Dickinson SA, Hitchings DJ. – The accuracy of portable peak flow meters. *Thorax* 1992; 47: 904–9.
12. Bland JM, Altman DG. – Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1: 307–10.
13. Burge PS, Perks W, O'Brien IM, Hawkins R, Green M. – Occupational asthma in an electronics factory. *Thorax* 1979; 34: 13–8.
14. Burge PS, Perks WH, O'Brien IM, Burge A, Hawkins R, Brown D *et al.* – Occupational asthma in an electronics factory; a case control study to evaluate aetiological factors. *Thorax* 1979; 34: 300–7.
15. Cartier A, Malo J-L, Forest F, Lafrance M, Pineau L, St-Aubin J-J *et al.* – Occupational asthma in snow crab processing workers. *J Allergy Clin Immunol* 1984; 74: 261–9.
16. Liss GM, Tarlo SM. – Peak expiratory flow rates in possible occupational asthma. *Chest* 1991; 100: 63–9.
17. Cote J, Kennedy S, Chan-Yeung M. – Quantitative versus qualitative analysis of peak expiratory flow in occupational asthma. *Thorax* 1993; 48: 48–51.
18. Landis JR, Koch GG. – The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159–74.
19. Malo J-L, Cote J, Cartier A, Boulet L-P, L'Archeveque J, Chan-Yeung M. – How many times per day should peak flow rates be assessed when investigating occupational asthma? *Thorax* 1993; 48: 1211–7.
20. Kongerud J, Soyseth V. – Methacholine responsiveness, respiratory symptoms and pulmonary function in aluminium potroom workers. *Eur Respir J* 1991; 4: 159–66.
21. Hargreave FE, Ramsdale EH, Pugsley SO. – Occupational asthma without bronchial hyper-responsiveness. *Am Rev Respir Dis* 1984; 130: 513–5.



Pavia - Palazzo Broletto