Interpretation of occupational peak flow records: level of agreement between expert clinicians and Oasys-2


Background: Oasys-2 is a validated diagnostic aid for occupational asthma that interprets peak expiratory flow (PEF) records as well as generating summary plots. The system removes inconsistency in interpretation, which is important if there is limited agreement between experts. A study was undertaken to assess the level of agreement between expert clinicians interpreting serial PEF measurements in relation to work exposure and to compare the responses given by Oasys-2.

Method: 35 PEF records from workers under investigation for suspected occupational asthma were available for review. Records included details of nature of work, intercurrent illness, drug therapy, predicted PEF, rest periods, and holidays. Simple plots of PEF and the Oasys-2 generated plots were available. Experts were advised that approximately 1 hour was available to review the records. They were asked to score each work-rest-work (WRW) period and each rest-work-rest (RWR) period for evidence of occupational effect. At the end of each record scores of 0–100% were given for evidence of "asthma" and "occupational effect" for the whole record. Kappa values were calculated for each scored period and for the overall score on the whole PEF record. The scores were converted into four groups (0–25%, 26–50%, 51–75%, 76–100%) and two groups (0–50% and 51–100%) for analysis. This is relevant to scores produced by Oasys-2. Agreement between Oasys-2 scores and each expert was calculated.

Results: 24 of 35 records were analysed by seven experts in the allotted time. For whole record occupational effect, median kappa values were 0.83 (range 0.56–0.94) for two groups and 0.62 (0.11–0.83) for four groups. For asthma, median kappa values were 0.58 (0.0–0.67) and 0.42 (0.15–0.70) for two and four groups respectively. For all WRW and RWR periods kappa values were 0.84 (0.42–0.94) and 0.70 (0.46–0.87) respectively. Agreement between Oasys-2 and individual experts showed median kappa values of 0.75 (0.50–0.92) for two groups and 0.50 (0.39–0.70) for four groups. Kappa values for the median expert score v Oasys-2 were 0.75 for two groups and 0.67 for four groups. Agreement was poor for records with intermediate probability, as defined by Oasys-2.

Conclusion: Considerable variation in agreement was seen in expert interpretation of occupational PEF records which may lead to inconsistencies in diagnosis of occupational asthma. There is a need for an objective scoring system which removes human variability, such as that provided by Oasys-2.

METHOD
Clinicians expert in the diagnosis of occupational asthma were asked to spend 1 hour reviewing 35 original peak flow records from patients under investigation for suspected occupational asthma. The records were selected at random and were not selected “difficult cases”. They included details of nature of work, intercurrent illness, drug therapy, predicted peak flow, rest periods, and holidays. Simple plots of peak flow (fig 1) and Oasys-2 generated summary plots were available (fig 2). The latter included plots of daily minimum, mean and maximum PEF; number of recordings each day; and measures of diurnal variation. Each expert was asked to score every work-rest-work (WRW) complex (two periods of days at work, separated by a period of days away from work) and every rest-work-rest (RWR) complex (two periods of days away from work, separated by a period at work) for likelihood of occupational effect (scale 0–100). At the end of each record they were asked to give two separate overall scores for “occupational effect” and “asthma” (scale 0–100). Experts were not asked to score in any other way other than 0–100 and they were not aware how the analysis would proceed. They were not aware that they would be compared with Oasys-2.

Each expert was also asked to comment if they felt the record was invalidated by any of the confounding factors they recognised.

Statistical analysis
Kappa values were calculated for each scored period and for the opinion on the whole record. Each expert was compared

Thorax 2002;57:860–864

See end of article for authors' affiliations

Correspondence to: Dr D R Baldwin, Department of Respiratory Medicine, City Hospital, Nottingham, NG5 1PB, UK, david.baldwin@nottingham.ac.uk

Revised version received 29 May 2002
Accepted for publication 29 May 2002
with others and median values calculated. Oasys-2 gives scores to two decimal places from 1 (no occupational effect) to 4. To allow comparison with the experts' scores they were also converted into percentages using the following formula:

\[
\text{(OASYS-2 score - 1)}/3 \times 100
\]

Two analyses were performed, one with two and one with four groups. Firstly, expert scores were split into two groups (0–50% and 51–100%) which correspond to predictions of occupational effect given by Oasys-2 which equate to a cut off score of 2.5. At this score the positive predictive value of Oasys-2 is 92% and negative predictive value is 80%.\(^1\) Expert scores were also converted into four groups (0–25%, 26–50%, 51–75%, and 76–100%). These scores represent negative, possible, probable, and positive, respectively, and also correspond to different levels of sensitivity and specificity of Oasys-2.\(^4\) Median expert scores were generated for whole record occupational effect and these were tested for agreement with the Oasys-2 score by calculating kappa values.

RESULTS

Agreement between experts

Only two experts were able to score all 35 records within an hour. Eight completed 13 records and seven completed 24. The results are based on the seven experts who completed 24 records.

![Figure 1](image1.png)

Figure 1 Simple plots of peak flow (half of record). Shaded areas represent time at work.

![Figure 2](image2.png)

Figure 2 Oasys-2 plot of the same patient as in fig 1 (complete record). The upper panel shows daily diurnal variation, expressed as percentage predicted. The central panel shows daily maximum, mean and minimum PEF, days "interpreted" to start with the first reading at work and stop with the last reading before work on the next day. Days at work have a shaded background, days away from work a clear background. The bottom panel shows the date and the number of readings each day.

Table 1 shows that agreement between experts for an occupational effect, as shown by median kappa value, was good. However, the range of values is very wide, indicating how inconsistent that agreement was. Kappa values were lower for asthma.

Reason for disagreement

Seven records were identified where there were clear disagreements within experts and between the Oasys-2 score. Four of the records were felt to be similar in the pattern of peak flow (low diurnal variation) and it was noted that each expert was consistent in their scoring for these records, even though there was disagreement. For one record where one expert differed from the rest, it was felt that this was a genuine matter of interpretation of the changes in peak flow which were not typical, but could represent occupational effect. A further record was extremely difficult to interpret because of possible confounders with low number of records, variable treatment, and a respiratory tract infection. One disagreement was felt to be a genuine error by the expert. Examples of peak flow records where experts disagreed are shown in figs 3 and 4. The reasons for the poor agreement between experts on occupational effect were found to be consistent with experts either missing changes...
regarded as significant by others, or finding other reasons to increase the likelihood of an occupational effect.

**Agreement between Oasys-2 and individual experts**

The comparison of Oasys-2 with each expert for whole record occupational effect revealed that experts were able to achieve median kappa values of 0.75 for two groups and 0.50 for four groups (table 2). For four groups, any one of all experts disagreed with Oasys-2 in 19 of 24 cases and, for two groups, any one expert disagreed with Oasys-2 in 10 of 24. Kappa values for median expert scores v Oasys-2 were 0.75 for two groups and 0.67 for four groups. There was general underscoring of

<table>
<thead>
<tr>
<th></th>
<th>Kappa score (2 groups)</th>
<th>Kappa score (4 groups)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All work-rest work and rest work-rest periods (n=159)</td>
<td>0.84 (0.48 to 0.87)</td>
<td>0.70 (0.48 to 0.74)</td>
</tr>
<tr>
<td>Whole record: occupational effect (n=24)</td>
<td>0.83 (0.76 to 0.88)</td>
<td>0.62 (0.46 to 0.71)</td>
</tr>
<tr>
<td>Whole record: “asthma” (n=24)</td>
<td>0.58 (0.42 to 0.65)</td>
<td>0.42 (0.24 to 0.30)</td>
</tr>
</tbody>
</table>

Values are median (interquartile range).

Kappa values of agreement are shown for two groups (negative and positive) and for four groups corresponding to clinical probabilities of negative, possible, probable, and positive.

---

**Table 1** Kappa values of agreement between the seven experts who completed 24 records

---

**Figure 3** Peak flow recording (Oasys plot) showing low diurnal variation. One expert scored this record as positive for occupational effect, another gave it a score of 50%, and the rest less than 30%.

---

**Figure 4** Oasys plot showing an asthmatic patient with confounding factors including a respiratory tract infection and variable treatment. There are also low numbers of daily measurements. Experts were very divided on their interpretation of this record.
the records compared with Oasys-2 (fig 5) with consistent differences between experts, some generally scoring higher than others. This explains why experts seemed to agree better between themselves than with Oasys-2.

No records were considered invalidated by the experts, but some were correctly noted to have missing values (not by all experts). Change of therapy, variable exposure, and suspected respiratory tract infection were all noted, but not by all experts.

DISCUSSION

This study has shown that, when experts recognised in the field of occupational medicine are asked to interpret PEF records which they use regularly in clinical practice, the level of agreement about what they mean is relatively low. Correct interpretation of PEF records is essential if they are to be an accurate measure of occupational airways disease. The interpretation of PEF records is therefore a potentially important function to prevent incorrect conclusions. This study has identified differences in the interpretation of occupational PEF records by expert clinicians. Observer variation alone may account for difficulties which some have with the use of this method in the diagnosis of occupational airways disease. The solution to this problem is to standardise the interpretation of the records with computer systems such as Oasys-2.

REFERENCES

3 Bright P, Burge PS. The diagnosis of occupational asthma from serial measurements of lung function at and away from work. Thorax 1996;51:857–63.
Interpretation of occupational peak flow records: level of agreement between expert clinicians and Oasys-2


Thorax 2002 57: 860-864
doi: 10.1136/thorax.57.10.860

Updated information and services can be found at:
http://thorax.bmj.com/content/57/10/860

References

This article cites 10 articles, 4 of which you can access for free at:
http://thorax.bmj.com/content/57/10/860#BIBL

Email alerting service

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Topic Collections

Articles on similar topics can be found in the following collections

- Asthma (1782)
- Airway biology (1100)
- Lung function (773)

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/