Value of nocturnal oxygen saturation as a screening test for sleep apnoea

B G Cooper, D Veale, C J Griffiths, G J Gibson

Abstract
The sensitivity and specificity of overnight recording of arterial oxygen saturation (SaO₂) in routine clinical practice was evaluated in 41 subjects who were being investigated for possible sleep apnoea-hypopnoea syndrome. SaO₂ was measured with an ear probe oximeter (Biox IIa) and chart recorder during an “acclimatisation” night immediately before a detailed polysomnographic study. The recordings were classified by two observers as positive, negative, or uninterpretable. Twelve of the 41 patients had the obstructive sleep apnoea syndrome when defined in terms of an apnoea-hypopnoea index greater than 15 events an hour on the second night. The sensitivity of nocturnal SaO₂ on the acclimatisation night when the diagnostic criterion was an apnoea-hypopnoea index of >5, >15, and >25/h was 60%, 75%, and 100% respectively. Corresponding values for specificity were 95%, 86%, and 80%. Oximetry alone therefore allowed recognition of a moderate or severe sleep apnoea syndrome. In routine practice an appreciable number of equivocal results is likely and repeat oximetry or more detailed polysomnography will then be required if clinical suspicion is high.

STUDY DESIGN
All patients were admitted to hospital for two consecutive nights. Oxygen saturation was recorded with a Biox IIa ear oximeter with the output signal connected to a Rikadenki three channel chart recorder. The recordings were made at 12 cm/hour and the start time was recorded by the subject. The fitting of the ear probe was explained to the subject and this was attached with the assistance of a nurse before the subject went to sleep. Nursing staff checked whether the probe was in place on occasion during the night but the subject was not under constant supervision. Polysomnography on the second night included measurement of chest wall movement (by magnetometers), airflow at the nose and mouth (thermocouples), sleep stage (electroencephalogram (EEG) and electro-oculogram (EOG)), and ear oximetry (Biox IIa). These were recorded on to a multichannel chart recorder and FM tape or on to synchronised video tape.

ANALYSIS
Oximetry records from the acclimatisation night were coded and analysed “blind” by two experienced observers and classified in one of three categories: positive—sleep breathing disorder present; negative—sleep breathing disorder not present; uninterpretable—technically unsatisfactory or “don’t know.”

The term sleep breathing disorder was used to imply the presence of either the sleep apnoea or the sleep hypopnoea syndrome. Records regarded as “technically unsatisfactory” included those where the chart recorder had failed or where it was suspected from the pattern of the recording that the oximeter probe had come adrift for a large proportion of the night. A recording was classified as positive on the basis not of exact diagnostic criteria but of “pattern recognition” of repetitive dips in oxygen saturation of more than 5%. The observers were aware of the aim of the study and the two nights’ traces were read separately, without knowledge of the recordings on the other night. Disturbance of breathing during sleep was classified from the polysomnographic recordings on the subsequent night by an independent observer on the basis of the frequency of episodes of apnoea and hypopnoea (sleep apnoea-hypopnoea index). Samples of the record were also checked by a second observer. Because different values of the apnoea-hypopnoea index are used in the definition of the syndrome, patients were diagnosed on the basis of an

Methods
SUBJECTS
Forty one subjects (15 female) referred for possible sleep apnoea were included in the survey, which looked at the SaO₂ recordings made during the “acclimatisation” night immediately before a more detailed nocturnal study in relation to the polysomnographic recordings on the second night.
Agreement between observers on results of oximetry. Both agreed that the result was positive on 13 occasions. Where one or both reported a negative or uninterpretable record the result was taken as negative (28 occasions in all).

Both reported record oximetry. Both on Agreement between result was taken as positive in 13 occasions. Where one or both observers agreed that the result was positive or negative, or were agreed that the result was positive, the result was considered positive. Where one or both observers disagreed, the result was considered uninterpretable. Where one or both observers disagreed, the result was considered negative.

A simple screening test for sleep apnoea is desirable because polisomnography is expensive and time consuming. Several screening investigations have been proposed, including laryngeal sounds, electrocardiography, flow detection, and oximetry; but formal evaluation has rarely been performed. The recording of oxygen saturation on the night before polysomnography in this study is likely to be equivalent to the clinical screening method used in many centres.

There is still controversy in published papers about the defining criterion for the sleep apnoea-hypopnoea syndrome. The originally proposed criterion of more than 5 episodes of apnoea an hour is now generally regarded as too strict and an apnoea-hypopnoea index above 15 is considered more realistic. An ideal screening test should have a high sensitivity with a reasonable specificity. Clearly, however, the sensitivity and specificity of screening oximetry are dependent on the criterion used for a positive diagnosis: the lower the value of the apnoea-hypopnoea index chosen for defining the lower limit the higher the specificity of oximetry but the poorer the sensitivity, whereas higher values lead to higher sensitivity but reduced specificity. Our results suggest that oximetry alone allows confident recognition of moderate and severe cases of the sleep apnoea-hypopnoea syndrome but it is inadequate for exclusion of milder cases.

Problems in using oximetry for recognising the sleep apnoea syndrome may be technical or physiological. An example of the former is poor contact of the probe with the ear, which occasionally produces signals resembling multiple falls in oxygen saturation; an example of the latter is periodic nocturnal desaturation in patients with chronic airways disease. Such errors increase the false positive results and decrease the specificity. Our criteria were relatively strict in that for a record to be classified as positive both observers had to agree; records where one or both observers considered the saturation record as uninterpretable were taken as negative. Such a policy

Results

The two observers agreed on 27/41 (66%) occasions (figure). Of these records, six were classified by both as uninterpretable, four because of chart recorder or oximeter failure. The observers disagreed on 14 occasions: in 12 of these one or other observer regarded the record as uninterpretable and in two cases there was complete disagreement (figure). The table shows values for the sensitivity and specificity of oximetry as a screening test with the three

"diagnostic" thresholds of the apnoea-hypopnoea index. For indices exceeding 5, 15, and 25 the sensitivity was 60%, 75%, and 100% respectively, with corresponding specificity of 95%, 86%, and 80% (table).

Examination of the results classified as false negatives showed that in each case one or other observer had classified the Sa2 record as "uninterpretable."

Discussion

A simple screening test for sleep apnoea is desirable because polysomnography is expensive and time consuming. Several screening investigations have been proposed, including laryngeal sounds, electrocardiography, flow detection, and oximetry; but formal evaluation has rarely been performed. The recording of oxygen saturation on the night before polysomnography in this study is likely to be equivalent to the clinical screening method used in many centres.

There is still controversy in published papers about the defining criterion for the sleep apnoea-hypopnoea syndrome. The originally proposed criterion of more than 5 episodes of apnoea an hour is now generally regarded as too strict and an apnoea-hypopnoea index above 15 is considered more realistic. An ideal screening test should have a high sensitivity with a reasonable specificity. Clearly, however, the sensitivity and specificity of screening oximetry are dependent on the criterion used for a positive diagnosis: the lower the value of the apnoea-hypopnoea index chosen for defining the lower limit the higher the specificity of oximetry but the poorer the sensitivity, whereas higher values lead to higher sensitivity but reduced specificity. Our results suggest that oximetry alone allows confident recognition of moderate and severe cases of the sleep apnoea-hypopnoea syndrome but it is inadequate for exclusion of milder cases.

Problems in using oximetry for recognising the sleep apnoea syndrome may be technical or physiological. An example of the former is poor contact of the probe with the ear, which occasionally produces signals resembling multiple falls in oxygen saturation; an example of the latter is periodic nocturnal desaturation in patients with chronic airways disease. Such errors increase the false positive results and decrease the specificity. Our criteria were relatively strict in that for a record to be classified as positive both observers had to agree; records where one or both observers considered the saturation record as uninterpretable were taken as negative. Such a policy

Results of polysomnography and classification of oximetry recordings

<table>
<thead>
<tr>
<th>Criterion (events per hour)</th>
<th>Polysomnography positive</th>
<th>Oximetry records</th>
<th>Sensitivity of oximetry (TP/(TP + FN))%</th>
<th>Specificity of oximetry (TN/(TN + FP))%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apnoea-hypopnoea index</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 5</td>
<td>20</td>
<td>12 8 20 1</td>
<td>60</td>
<td>95</td>
</tr>
<tr>
<td>&gt; 15</td>
<td>12</td>
<td>9 3 25 4</td>
<td>75</td>
<td>86</td>
</tr>
<tr>
<td>&gt; 25</td>
<td>6</td>
<td>6 0 28 7</td>
<td>100</td>
<td>80</td>
</tr>
</tbody>
</table>

TP—true positive; FN—false negative; TN—true negative; FP—false positive.
inevitably reduces the sensitivity of a diagnostic investigation. In the clinical context, however, technical errors or an equivocal result would be regarded as a “negative” outcome and the study would be repeated. One observer produced three more false negative results than the other; in each instance the apnoea-hypopnoea index was 5–15/h.

It might be argued that our group of subjects had already been screened by clinical assessment and the referral procedure, but this again reflects the usual clinical circumstances in which oximetry is used as a screening test. Furthermore the oximetry records may be influenced by the “first night” effect of sleep study measurements. This may mean that, although the diagnosis of the sleep apnoea-hypopnoea syndrome is likely to be reliable in severe cases, it will be much less so in patients with milder disease.

In conclusion, nocturnal oxygen saturation alone allowed confident recognition of moderate and severe cases of obstructive sleep apnoea, but it is likely to be inadequate for excluding milder cases in clinical practice. Repeat oximetry or more detailed polysomnography is then required if clinical suspicion is high.