Validation of automatic analysis of respiratory events

Sigríður Sigurðardóttir, Kristín Anna Óláfsdóttir, Thorarinn Gíslason, Olfra Asa Johannsdóttir, Omar Hilmarsson, Erna Sif Arnardóttir

1Department of Respiratory Medicine and Sleep, Landspitali
2The National University Hospital of Iceland, Reykjavik, Iceland.
3Faculty of Medicine, The University of Iceland, Reykjavik, Iceland.
4Nox Medical, Reykjavik, Iceland.

Introduction
A reliable automatic analysis of respiratory and oxygen desaturation events in sleep studies can decrease the time needed for manual scoring. This study aimed to assess the validity of automatic vs. manual analysis and verify obstructive sleep apnea (OSA) categorization.

Methods
A general population cohort (n = 27) was assessed twice (2006/2012). The first study was performed with Embla A10 (Natus Medical Inc) but the latter with Nox T3 device (Nox Medical). All studies were scored in Nocturnal software with manually set start and stop times. Recordings with no cannula flow signal were scored using respiratory inductive plethysmography (RIP) flow.

Results
The apnea hypopnea index (AHI) ranged from 0 - 20 events/hour in the Embla recordings and 0 - 23 in the Nox by manual scoring. AHI ranged from 0 - 22 in the Embla recordings and 0 - 34 in the Nox by automatic analysis. The correlation between automatic and manual scoring is very high (Table 1). Sensitivity (reflects true positive events) and specificity (reflects true negative events) of automatic analysis are shown in Table 2. Figure 1 shows examples of false positive events and false negative events by automatic analysis.

When looking at how subjects were classified by OSA category (non OSA (AHI<5), mild OSA (AHI 5-15), moderate to severe OSA (AHI>15)) using automatic analysis only, we found that 19% of individuals were categorized wrongly when diagnosed with Embla and 41% when diagnosed with Nox. AHI was overestimated in 93% of wrongly classified cases. See Figure 2.

Discussion
- The automatic scoring was acceptable regarding sensitivity and specificity of apnea-hypopnea events and oxygen desaturation events.
- However, its tendency to overestimate respiratory events makes OSA categorizing inaccurate.
- The automatic analysis did worst when flow signal quality was bad. One way to improve the automatic analysis would be to let the software change between signals i.e. from cannula flow to RIP flow like the manual scorer does when needed.
- The difference between the automatic and manual analysis would most likely have been greater if the start and stop times had not been added manually prior to analysis.
- Manual overview is still recommended for all sleep studies.

Table 1. Correlation between automatic and manual scoring

<table>
<thead>
<tr>
<th></th>
<th>Embra recordings</th>
<th>Nox recordings</th>
</tr>
</thead>
<tbody>
<tr>
<td>AHI</td>
<td>ODI</td>
<td>AHI</td>
</tr>
<tr>
<td>r</td>
<td>r</td>
<td>r</td>
</tr>
<tr>
<td>p</td>
<td>p</td>
<td>p</td>
</tr>
<tr>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Table 2. Sensitivity and specificity of the automatic scoring

<table>
<thead>
<tr>
<th></th>
<th>Embra recordings</th>
<th>Nox recordings</th>
</tr>
</thead>
<tbody>
<tr>
<td>AH events</td>
<td>OD events</td>
<td>AH events</td>
</tr>
<tr>
<td>sensitivity</td>
<td>specificity</td>
<td>sensitivity</td>
</tr>
<tr>
<td>0.89</td>
<td>0.91</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Figure 1.
- a) Shows two false positive events.
- b) Shows the same frame scored manually.
- c) Shows a false negative event.
- d) Shows the same frame scored manually.

Figure 2. Difference in OSA classification between automatic and manual scoring.
The lower bar represents the manual scoring and the upper bar the automatic analysis, except in one case (*).
- a) Nox.
- b) Embla.

Conflict of interest: Erna Sif Arnardottir is a part-time consultant for Nox Medical, Reykjavik, Iceland.
Grants: The presenting author received a travel grant from the Icelandic Sleep Research Society.