Reliability of Apnea-Hypopnea Index Measured by a Home Bi-Level Pressure Support Ventilator Versus a Polysomnographic Assessment

Marjolaine Georges MD, Dan Adler MD, Olivier Contal PhD PT, Fabrice Espa RPSGT, Stephen Perrig MD, Jean-Louis Pépin MD, and Jean-Paul Janssens MD

BACKGROUND: Ventilators designed for home care provide clinicians with built-in software that records items such as compliance, leaks, average tidal volume, total ventilation, and indices of residual apnea and hypopnea. Recent studies have showed, however, an important variability between devices regarding reliability of data provided. In this study, we aimed to compare apnea-hypopnea indices (AHI) provided by home ventilators (AHI_NIV) versus data scored manually during polysomnography (AHI_PSG) in subjects on noninvasive ventilation (NIV) for obesity-hypoventilation syndrome. METHODS: Stable subjects with obesity-hypoventilation syndrome on NIV, all using the same device, underwent 3 consecutive polysomnographic sleep studies with different backup breathing frequencies (spontaneous mode, low and high backup breathing frequencies). During each recording, AHI_NIV was compared with AHI_PSG. RESULTS: Ten subjects (30 polysomnogram tracings) were analyzed. For each backup breathing frequency (spontaneous mode, low and high backup breathing frequencies), AHI values were 62 ± 7/h, 26 ± 7/h, and 17 ± 5/h (mean ± SD), respectively. Correlation between AHI_NIV and AHI_PSG was highly significant (r² = 0.89, P < .001). As determined by Bland-Altman analysis, mean bias was 6.5 events/h, and limits of agreement were +26.0 and −12.9 events/h. Bias increased significantly with higher AHI values. Using a threshold AHI value of 10/h to define appropriate control of respiratory events, the ventilator software had a sensitivity of 90.9%, a specificity and positive predictive value of 100%, and a negative predictive value of 71%. CONCLUSIONS: In stable subjects with obesity-hypoventilation syndrome, the home ventilator software tested was appropriate for determining if control of respiratory events was satisfactory on NIV or if further testing or adjustment of ventilator settings was required. (ClinicalTrials.gov registration NCT01130090.) Key words: apnea-hypopnea index; noninvasive ventilation; monitoring; obesity-hypoventilation syndrome; home ventilator software.

Introduction

Home noninvasive ventilation (NIV) is recognized as an effective treatment for chronic hypercapnic respiratory failure, and the number of patients using NIV devices is continuously increasing.1,2 In many western countries, obesity-hypoventilation syndrome has become the leading indication for home NIV. In ~90% of cases, obesity-hypoventilation syndrome is associated with obstructive sleep apnea syndrome, and NIV is used in these patients to...
correct nocturnal hypoventilation and apneas or hypopneas.\(^3\)\(^6\) Clinical studies have shown that stable subjects treated with NIV on a long-term basis may have pauci- or asymptomatic residual respiratory events, which may compromise the benefit of NIV.\(^7\)\(^8\) Subtle changes in ventilatory parameters may also result in unwanted respiratory events.\(^9\)\(^10\) Expert groups have emphasized the importance of appropriate monitoring of patients treated with NIV to detect and correct undesired nocturnal respiratory events.\(^11\)\(^13\) Although the American Academy of Sleep Medicine recommends monitoring NIV by polysonography, this is not technically possible in many countries because of issues such as cost and availability.\(^11\) A previous consensus suggests that monitoring of NIV should include at least medical history, daytime arterial blood gases, nocturnal pulse oximetry, and data provided by home ventilator software.\(^13\) Whenever possible, use of nocturnal capnography is also recommended.

Although home ventilators do record a considerable amount of information for clinicians, there are some doubts as to the validity and accuracy of at least part of the information provided. Compliance and pattern of use recorded by home ventilators are most probably reliable and clearly useful. In a bench study of one specific home ventilator, estimation of leaks and tidal volume (V\(_T\)) was shown to be accurate.\(^14\) However, a bench study of 7 home ventilators questioned the validity of estimated V\(_T\) (and thus total ventilation) and leaks with huge variations between brands.\(^15\) In this study, V\(_T\) was systematically underestimated, and bias increased with increasing leaks. Furthermore, leaks were reported differently according to manufacturers, some including intentional leaks and some not, whereas one ventilator averaged leaks only during expiration. There were considerable differences in estimation of leaks, and in 4 of 7 devices, estimation of leaks was clearly unreliable. These conflicting results raise some concern as to the use of other parameters provided by home ventilators, such as percentage of insufflations triggered or cycled by the device and estimations of apneas and hypopneas.

Apneas and hypopneas have been provided by CPAP devices for many years. However, a previous consensus statement noted that, in obstructive sleep apnea syndrome, CPAP devices estimated apneas and hypopneas using different algorithms according to each manufacturer, with all definitions being different from American Academy of Sleep Medicine consensus definitions.\(^16\) Available studies show good agreement between apnea indices (AI) measured by CPAP devices and polysomnography, but much lower correlations for hypopnea indices (HI).\(^17\)\(^22\)

To our knowledge, the validity and accuracy of apnea-hypopnea indices (AHI) provided by home bi-level pressure support ventilators have not been independently studied. The complexity of bi-level versus continuous positive pressure may add to the discrepancy between estimated and measured AHI in patients on NIV. Furthermore, apnea episodes are rare, and hypopnea episodes are more frequent, and this may add to a potential lack of agreement between AHI provided by home ventilators and values recorded by polygraphy or polysomnography.

In this study, stable subjects on NIV for obesity-hypoventilation syndrome underwent 3 consecutive polysomnogram recordings with different ventilator settings. Changes in backup breathing frequency led to marked changes in central and obstructive hypopneas and apneas, providing an excellent database for comparing data provided by home ventilators and those simultaneously measured by polysomnography.\(^9\) This comparison was among the end points prespecified in the study design (ClinicalTrials.gov registration NCT01130090). All subjects were equipped with the same NIV device for this study.

**Methods**

The flow chart, methods, and inclusion criteria for this study have been described in detail previously.\(^9\)\(^23\) Briefly, 10 subjects treated with home NIV using the same device for obesity-hypoventilation syndrome were recruited by the Division of Pulmonary Diseases of Geneva University Hospitals. Obesity-hypoventilation syndrome was defined as the association of morbid obesity (body mass index of > 30 kg/m\(^2\)) and daytime hypercapnia (P\(_{\text{ACO}}\) > 45 mm Hg) without any other obstructive or restrictive pulmonary pathology.\(^4\)\(^6\)\(^24\) All subjects had begun home NIV after at
least one episode of acute hypercapnic respiratory failure. All subjects were in stable condition at inclusion. Exclusion criteria were: < 18 y of age, FEV₁/FVC < 0.70, poor compliance, and hospitalization for an acute episode of cardiac and/or respiratory failure within the previous 3 months. The 10 subjects underwent 3 consecutive polysomnographic sleep studies on NIV with 3 different backup breathing frequencies in random order: (1) spontaneous mode without backup breathing frequency, (2) spontaneous-timed mode with low backup breathing frequency (backup rate set at 2 points below the average nocturnal breathing frequency recorded by the built-in software of the subject’s home ventilator during the previous 2 weeks), and (3) spontaneous-timed mode with high backup breathing frequency (set at the 95th percentile of the average nocturnal breathing frequency). No other ventilatory parameters were modified during the 3 nights. All subjects used a VPAP III ST-A ventilator (ResMed, San Diego, California). The study protocol was approved by the Geneva University Hospitals ethics committee for medical research, and written informed consent was obtained for all subjects.

Polysomnography, Ventilator Software, and Respiratory Events

Polysomnogram recordings, all performed on NIV, included standard electroencephalography (7 electrodes: F4, F3, C3, C4, Cz, O1, O2), left and right electrooculography, and submental electromyography. Left and right anterior tibialis electromyography was used to detect leg movements, and a bipolar electrocardiograph was used for cardiac monitoring. Air flow (pneumotachograph), thoracic and abdominal movements (XactTrace respiratory inductive plethysmograph, Embla, Buffalo, New York), SpO₂, body position, and video were simultaneously and continuously monitored (RemLogic 1.1, Embla).

Sleep and microarousals were scored manually according to American Academy of Sleep Medicine criteria. The following parameters were quantified: total sleep time, sleep efficiency (total sleep time/total recording time) × 100, expressed as a percentage), duration of each sleep stage, percentage of total sleep time, wake-after-sleep onset, microarousal index (microarousals/h), and sleep latency. Microarousals were defined as an abrupt shift in electroencephalography frequency, including alpha waves, theta waves, and/or frequencies > 16 Hz lasting at least 3 s.

Respiratory events were scored when associated with a drop in SpO₂ of at least 4% and/or a microarousal and classified in 3 previously defined groups: central, mixed, and obstructive events. Patient-ventilator asynchrony was quantified as reported by Guo et al as a percentage of total sleep time. Data recorded by ventilator software were downloaded after each polysomnogram recording (ResScan 4.1.0.0, ResMed), including mean and 95th percentile values for leaks, which have been shown to be reliable with the VPAP III ST-A device.

Statistical Analysis

According to distribution of variables, data are expressed as mean ± SD or median (interquartile range) when appropriate. Data analysis was performed using SPSS 20.0 (IBM, Armonk, New York). Correlation between polysomnographic and ventilator software variables was performed with the Pearson test. Bias and limits of agreement between AHI values obtained from polysomnography (AHI PSG) and ventilator software (AHI NIV) were calculated according to Bland and Altman. Using an arbitrary threshold level of 10/h for AHI, sensitivity, specificity, and negative and positive predictive values of data provided by ventilator software were calculated. Association between AHI PSG and AHI NIV was tested by the Fisher exact test. The Friedman test was used to determine whether results between the 3 ventilator modes tested were significantly different. The Pearson test was used to compute correlation between the oxygen desaturation index and indices provided by ventilator software (AHI, AI, and HI). For all tests, P < .05 was considered significant.

Results

Ten subjects with a median age of 52 (51–61) y and a median body mass index of 46.8 (45.6–54.4) kg/m² on NIV for 19.5 (16–67) months were prospectively included between October 2008 and December 2010. Subjects had normal arterial blood gas values at inclusion: P aO₂, 72.7 ± 10.9 mm Hg; P aCO₂, 41.3 ± 5.6 mm Hg; and pH, 7.44 ± 0.03. Pulmonary function test results were: FEV₁, 87 ± 23% of predicted; FVC, 86 ± 22% of predicted; and FEV₁/FVC, 100 ± 1% of predicted. Baseline ventilator settings were: inspiratory positive airway pressure, 20.5 ± 1.2 cm H₂O (range of 16–28); expiratory positive airway pressure, 9.2 ± 0.6 cm H₂O (range of 7–12); and backup breathing frequency, 14 ± 1.5 cycles/min. All subjects but one were equipped with a full face mask. Changes in backup breathing frequencies were 11 (10–12) for low backup breathing frequency and 21 (18–22) for high backup breathing frequency.

Twenty-seven paired tracings were compared (3 recordings could not be completely downloaded from home ventilators). Leaks were low on average (19.0 ± 3.1 L/min), as were periods with either patient-ventilator asynchrony or autotriggering. As shown in Table 1, residual respiratory events were mainly central in spontaneous mode and predominantly obstructive with both low and high backup breathing frequencies.
Table 1. Results of Pulse Oximetry, Polysomnography, and Capnography and Data Provided by Ventilator Software During Recordings Performed in Spontaneous Mode (No Backup Breathing Frequency) and in Spontaneous-Timed Mode With Low and High Backup Breathing Frequencies

<table>
<thead>
<tr>
<th></th>
<th>No Backup Breathing Frequency</th>
<th>Low Backup Breathing Frequency</th>
<th>High Backup Breathing Frequency</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean nocturnal $S_{pO_2}$, %</td>
<td>92.4 ± 0.4</td>
<td>92.1 ± 0.4</td>
<td>92.2 ± 0.4</td>
<td>.58</td>
</tr>
<tr>
<td>Minimal nocturnal $S_{pO_2}$, %</td>
<td>77.9 ± 2.5</td>
<td>80.4 ± 2.3</td>
<td>81.0 ± 2.2</td>
<td>.59</td>
</tr>
<tr>
<td>Time spent with $S_{pO_2}$ at &lt; 90%, %</td>
<td>14.7 ± 3.5</td>
<td>14.9 ± 3.7</td>
<td>11.0 ± 3.7</td>
<td>.67</td>
</tr>
<tr>
<td>Mean nocturnal $P_{teCO_2}$, mm Hg</td>
<td>46.7 ± 4.0</td>
<td>45.1 ± 2.1</td>
<td>45.6 ± 2.5</td>
<td>.12</td>
</tr>
<tr>
<td>Total sleep time, min</td>
<td>403.2 ± 77.7</td>
<td>460.2 ± 57.9</td>
<td>391.3 ± 97.1</td>
<td>.27</td>
</tr>
<tr>
<td>Sleep efficiency, %</td>
<td>71.1 ± 13.6</td>
<td>80.2 ± 8.5</td>
<td>66.5 ± 18.3</td>
<td>.27</td>
</tr>
<tr>
<td>Wake-after-sleep onset, min</td>
<td>142.7 ± 77.4</td>
<td>106.3 ± 53.6</td>
<td>148.0 ± 97.1</td>
<td>.91</td>
</tr>
<tr>
<td>N1 sleep, % total sleep time</td>
<td>17.0 ± 7.3</td>
<td>13.4 ± 5.9</td>
<td>14.4 ± 3.7</td>
<td>.50</td>
</tr>
<tr>
<td>N2 sleep, % total sleep time</td>
<td>55.4 ± 9.6</td>
<td>54.8 ± 11.1</td>
<td>59.8 ± 10.9</td>
<td>.27</td>
</tr>
<tr>
<td>Slow-wave sleep, % total sleep time</td>
<td>9.3 ± 6.9</td>
<td>13.9 ± 9.5</td>
<td>12.6 ± 8.8</td>
<td>.12</td>
</tr>
<tr>
<td>Rapid-eye-movement sleep, % total sleep time</td>
<td>18.8 ± 3.8</td>
<td>17.9 ± 6.2</td>
<td>13.1 ± 6.7</td>
<td>.067</td>
</tr>
<tr>
<td>Patient-ventilator asynchrony, % total sleep time</td>
<td>11 (7–20)</td>
<td>8 (2–32)</td>
<td>6 (0–31)</td>
<td>.73</td>
</tr>
<tr>
<td>Leaks, L/min</td>
<td>3.6 (1.8–7.8)</td>
<td>6 (1.2–9)</td>
<td>6 (2.7–15.2)</td>
<td>.60</td>
</tr>
<tr>
<td>Oxygen desaturation index, n/h</td>
<td>59.3 ± 6.2</td>
<td>29.9 ± 6.1</td>
<td>21.1 ± 4.7</td>
<td>.002</td>
</tr>
<tr>
<td>Microarousal index, n/h</td>
<td>9.4 ± 2.5</td>
<td>12.9 ± 2.1</td>
<td>17.3 ± 1.8</td>
<td>.15</td>
</tr>
<tr>
<td>Central event index, n/h</td>
<td>26.9 ± 6.8</td>
<td>2.6 ± 1.2</td>
<td>2.9 ± 1.7</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Mixed event index, n/h</td>
<td>9.7 ± 1.9</td>
<td>0.6 ± 0.3</td>
<td>0.4 ± 0.2</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Obstructive event index, n/h</td>
<td>25.6 ± 4.7</td>
<td>22.5 ± 7.1</td>
<td>13.4 ± 5.1</td>
<td>.067</td>
</tr>
<tr>
<td>AHI_{PSG}, n/h</td>
<td>62.3 ± 7.2</td>
<td>25.7 ± 7.0</td>
<td>16.7 ± 5.2</td>
<td>.006</td>
</tr>
<tr>
<td>AHI_{NV}, n/h</td>
<td>51.9 ± 4.1</td>
<td>23.0 ± 7.1</td>
<td>13.4 ± 3.5</td>
<td>.050</td>
</tr>
</tbody>
</table>

Values are expressed as mean ± SD or median (interquartile range). Leaks reported are unintentional. P values were obtained by the Friedman test. Significant P values (<.05) are shown in boldface.

$P_{teCO_2}$ = transcutaneously measured partial pressure of carbon dioxide
AHI_{PSG} = apnea-hypopnea index measured by polysomnography
AHI_{NV} = apnea-hypopnea index reported by ventilator software

Figure 1 shows the correlation between AHI_{PSG} and AHI_{NV}. Values for $r^2$ were: 0.89 ($P < .001$) for all tracings, 0.74 ($P = .001$) for tracings in spontaneous mode, 0.90 ($P < .001$) for tracings with low backup breathing frequency, and 0.83 ($P < .001$) for tracings with high backup breathing frequency. Correlations between the oxygen desaturation index obtained during polysomnography and indices obtained from ventilator software were: $r^2 = 0.85$ ($P < .001$) for AHI_{NV}, $r^2 = 0.62$ ($P < .001$) for AHI_{NV}, and $r^2 = 0.226$ ($P = .12$) for HINIV.

Figure 2 shows agreement between AHI_{PSG} and AHI_{NV} according to Bland and Altman.26 Mean bias (d) was 6.5 events/h; limits of agreement (d ± 2 SD) were +26.0 and −12.9 events/h. As shown, bias increased significantly as AHI values increased.

We considered arbitrarily that subjects with a residual AHI of < 10/h had an appropriate control of respiratory events. Based on this threshold, Table 2 shows sensitivity, specificity, and negative and positive predictive values for detection of an abnormally high residual AHI using AHI_{NV}.
AH hypertension, interest, and positive and negative predictive values of AHI obtained by ventilator software were high. This suggests that AHI values obtained by ventilator software with the device tested allow the clinician to determine whether further testing is warranted (polysomnography, polygraphy) or whether central and/or obstructive respiratory events are adequately controlled by NIV.

During these sleep studies, unintentional leaks and patient-ventilator asynchrony were very low (see Table 1). This may be relevant because prolonged leaks and/or patient-ventilator asynchrony either induces or reflects a dissociation between the patient’s spontaneous respiratory pattern and detection by the ventilator of the patient’s efforts. Patient-ventilator asynchrony and leaks may thus affect the reliability of AHI provided by ventilator software, as well as that of other variables such as VT, total ventilation, and percentage of cycles triggered and cycled by the patient. The clinician must thus be aware that the reliability of AHI obtained by ventilator software can change from one device to another and that these discrepancies require regular independent testing.

For this particular device, however, the present data substantiate the algorithm proposed by the SomnOx group, which states that medical history, daytime arterial blood gases, nocturnal pulse oximetry, and data downloaded using ventilator software suffice for regular monitoring of patients on long-term NIV.

This study has a few limitations. First, the number of recordings is modest. In our country and in many countries of Western Europe, polysomnography for patients on NIV is not routine, and the availability of polysomnography is limited, which explains the limited number of inclusions. Second, observations are limited to one type of ventilator. Similar studies should theoretically be performed either independently or provided by manufacturers to inform clinicians. Third, it is presently impossible for clinicians to analyze the agreement between events recorded by polygraphy or polysomnography and those recorded by ventilator software on an event-by-event basis, as ventilator software provides only an overall estimation of events detected by polygraphy or polysomnography and those recorded by ventilator software were high. This suggests that AHI values obtained by ventilator software with the device tested allow the clinician to determine whether further testing is warranted (polysomnography, polygraphy) or whether central and/or obstructive respiratory events are adequately controlled by NIV.

The computation of AHI remains somewhat mysterious for clinicians, and little information is provided by manufacturers. As for CPAP devices, it probably changes from one manufacturer to another and evolves with different generations of devices from a given manufacturer. Previous bench tests have shown to what extent performances and reliability of ventilator software can change from one device to another, and these discrepancies require regular independent testing.

For this particular device, however, the present data substantiate the algorithm proposed by the SomnOx group, which states that medical history, daytime arterial blood gases, nocturnal pulse oximetry, and data downloaded using ventilator software suffice for regular monitoring of patients on long-term NIV.

This study has a few limitations. First, the number of recordings is modest. In our country and in many countries of Western Europe, polysomnography for patients on NIV is not routine, and the availability of polysomnography is limited, which explains the limited number of inclusions. Second, observations are limited to one type of ventilator. Similar studies should theoretically be performed either independently or provided by manufacturers to inform clinicians. Third, it is presently impossible for clinicians to analyze the agreement between events recorded by polygraphy or polysomnography and those recorded by ventilator software on an event-by-event basis, as ventilator software provides only an overall estimation of events for the entire night. The results presented thus reflect an approximation of the agreement between AHI and AHI. An analysis of agreement event by event (ie, AIPSG vs AINIV and HIPSG vs HINIV) has yet to be per-
formed. Finally, we lack information as to how ventilators report patient-ventilator asynchrony (whether unrewarded efforts or autotriggering affect $AHI_{NIV}$) and to what extent leaks affect $AHI_{NIV}$.

Conclusions

In summary, in spite of the abovementioned limitations, this study suggests that the reliability of AHI provided by ventilator software of the device tested is sufficient for monitoring subjects on long-term NIV for obesity-hypoventilation syndrome and determining whether control of residual obstructive or central respiratory events during NIV is sufficient or if further testing (polygraphy, polysomnography) is required. Further studies are necessary to explore other devices and subjects with other indications for NIV.

REFERENCES