Mandibular advancement device therapy in obstructive sleep apnea
Aarab, G.

Citation for published version (APA):
Aarab, G. (2011). Mandibular advancement device therapy in obstructive sleep apnea
Chapter 2

Variability in the Apnea-Hypopnea Index and its Consequences for Diagnosis and Therapy Evaluation

Ghizlane Aarab, Frank Lobbezoo, Hans L. Hamburger, Machiel Naeije

Respiration 2009; 77: 32-37
Abstract

**Background:** The apnea-hypopnea index (AHI) is frequently used to recognize obstructive sleep apnea (OSA) and to evaluate therapy. **Objectives:** The aim of this study was to determine the AHI variability during a 10-week period, and to discuss its consequences for diagnosis and therapy evaluation. **Methods:** Fifteen OSA patients (50.8 ± 11.2 years) underwent four polysomnographic (PSG) recordings, with a mean interval between recordings of 3.3 weeks. **Results:** No differences were found in the average AHI values of the four PSG recordings (p = 0.985). Nevertheless, pooling all data of the 15 participants yielded a smallest detectable difference for AHI of 12.8. Linear regression between the individual means and standard deviations (SDs) of AHI showed that participants with a higher AHI tended to have a higher SD (p < 0.044). **Conclusions:** These results suggest a considerable intra-individual variability in AHI recordings. Hence, a single night recording can only recognize OSA when the AHI lies outside a cut-off band surrounding the AHI cut-off point. AHI variability should also be taken into account when evaluating OSA therapy. In this context, it should be noted that it is mainly the approach that we would like to convey to the reader and not the cut-off values per se.

**Key words:** apnea-hypopnea index · diagnosis of apnea · evaluation of obstructive sleep apnea therapy · obstructive sleep apnea · polysomnographic recordings

Introduction

Obstructive sleep apnea (OSA) is characterized by recurrent obstruction of the upper airway, often resulting in oxygen desaturation and arousal from sleep [1]. In 1999, the American Academy of Sleep Medicine Task Force reported standard definitions, criteria and severity ratings for abnormal breathing events during sleep [1]. The purpose of that report was to facilitate comparability of studies for research purposes by defining the apnea-hypopnea index (AHI). The AHI is used as the main outcome variable to recognize OSA and to evaluate therapeutic success [2]. When using the AHI for diagnostic and therapy evaluation purposes, the AHI variability has to be borne in mind.

Several reports described considerable variability in AHI values over time. In studies investigating the AHI variability during consecutive nights, the average AHI did not significantly differ between nights on a group level, but on an individual level, a considerable variability was sometimes observed [2–11]. The numerical consequences for diagnosis and therapy evaluation, however, were not elaborated in these studies. So far, the AHI variability over longer periods of time has been studied less frequently [7, 8,10]. This long-term variability, however, is clinically even more important than the short-term variability, because the time interval between diagnostic and therapeutic evaluation of polysomnographic (PSG) recordings is at least several weeks. Therefore, the aim of this study was to determine the AHI variability during a long-term follow-up study of 10 weeks, and to discuss its consequences for diagnostic and therapeutic evaluation purposes.

Methods

**Overview**

This paper is part of a series of studies that includes a controlled randomized clinical trial (RCT), in which the therapeutic efficacy of a mandibular advancement device (MAD) [12] was compared with both nasal continuous positive airway pressure (nCPAP) and a control
Mandibular Advancement Device Therapy in Obstructive Sleep Apnea

Variability in the apnea-hypopnea index

chapter 2

condition. Both MAD and nCPAP required multiple visits to our clinic and laboratory between the baseline and follow-up assessments of OSA, amongst others for titration of the devices. By introducing a placebo appliance in the control group, the three experimental conditions were made as comparable as possible, especially in terms of the personal attention that was being paid to the patients. Four extra placebo assessments were thus performed, which could be used in the present study to determine the variability in our patients, provided that the placebo appliance would turn out to be inert (i.e., no significant difference in AHI between the baseline and follow-up assessments).

All patients were selected from the multidisciplinary Center for Sleep-Wake Disorders at the Slotervaart Medical Center, which collaborates with the Academic Centre for Dentistry (ACTA). The multidisciplinary team consisted of a neurologist, ENT specialists, pulmonologists, a dentist, a psychologist, and technicians trained in sleep medicine. All patients underwent a thorough medical examination and a baseline polysomnographic (PSG) recording at the Slotervaart Medical Center. The PSG recording took place in a dark hospital room, using Siesta hardware and Pro-Fusion software (Compumedics, Abbotsford, Australia).

The criteria for inclusion of patients in the RCT were: age > 18 years and an AHI between 5 and 45 events/hour. Criteria for exclusion were: evidence of respiratory/sleep disorders other than OSA, a Body Mass Index (BMI) > 40, medication usage that could influence respiration or sleep, periodic limb movement disorder, previous treatment with continuous positive airway pressure or a mandibular advancement device, and reversible morphological upper airway abnormalities (e.g., enlarged tonsils). Additional dental exclusion criteria were: temporomandibular disorders (based on a functional examination of the masticatory system [13, 14]), untreated periodontal problems, dental pain, and a lack of retention possibilities for a mandibular advancement device or placebo appliance.

After written informed consent was obtained, the patients were randomly allocated to one of three therapies (MAD, nCPAP, or placebo). Patients were not told which one of the three therapies was the placebo therapy. The placebo group was used in the present study.

The scientific and ethical aspects of the RCT protocol were reviewed and approved by the Medical Ethics Committee of the Slotervaart Medical Center.

Patients

Initially, seventeen OSA patients were included in the placebo group, thirteen of them being men. For private reasons that were unrelated to this study, one female OSA patient dropped out of the study after the second ambulatory PSG recording (see study protocol). One male OSA patient failed one of the ambulatory PSG recordings, and was therefore excluded from the analyses of the present data. Thus, fifteen patients had a complete set of data and were included in this study. Their mean age was 50.8 years (SD, 11.2; range, 30 to 70); their mean BMI, 30.1 (SD, 4.5; range, 23 to 39). Their mean Epworth Sleepiness Score (ESS) [15] was 11.2 (SD, 4.2; range, 5 to 20). Their mean baseline AHI was 19.5 (SD, 9.6; range, 6 to 37).

Study protocol

After allocation of the patients to the placebo group, a thin palatal splint was made, using alginate impressions of both jaws, at the Clinic of the Department of Oral Function of ACTA. The splint was a thin (< 1 mm) hard acrylic resin splint, worn on the upper jaw with only a partial palatal coverage (Fig. 1). In our clinic, this splint has been used before as a placebo appliance in a study on sleep bruxism [16]. The patients were instructed to wear the splint every night during the follow-up period of ten weeks.

All patients underwent four PSG recordings (PSG1-PSG4) at home, with the splint in situ, and with an interval of 3.3 (SD, 1.7; range, 1-9) weeks between recordings. Monet hardware and Rembrandt Software (Medcare Automation B.V., Amsterdam, The Netherlands) were used for these recordings. The montage was performed at the Slotervaart Medical Center by a trained coworker. The channels of the recording consisted of two electro-encephalographic leads (C3-A2; O2-A1), two electro-oculographic leads, mental surface electromyography, nasal-oral airflow using a thermistor, oximetry, abdominal and
thoracic respiratory effort, body position, electrocardiography, leg electromyography (m. tibialis anterior), and a piezo-electric lead for the detection of snoring vibrations. Each PSG home recording was analyzed twice manually, under a blind condition, by the same examiner, who was experienced in scoring PSG recordings, using internationally accepted criteria [1, 17]. Sleep stages were scored in 30-s epochs and standard sleep and respiratory outcome variables were obtained (see left column of Table 1).

Prior to each PSG recording, the patient received sleep hygiene advises, e.g., avoiding copious meals and alcoholic beverages 3 hours preceding bedtime, and creating a good sleeping environment [18]. Each PSG recording was followed by a visit to ACTA. During this visit, the BMI (kg/m²) and ESS were determined. The patients were also asked about their compliance (i.e., frequency of wearing the placebo splint) and about possible side effects of the appliance.

After the four recordings at home, all patients had a follow-up PSG recording in the hospital, under the same conditions as the baseline PSG recording (see overview), with the placebo appliance in situ.

Fig. 1. The placebo splint in situ; a thin (< 1mm) hard acrylic splint, worn on the upper jaw with only a partial palatal coverage.

Statistics

The BMI, ESS, AHI, and standard sleep variables were analyzed for the within-subject effects (i.e., the four home recordings: Night effect), using a General Linear Model (GLM) with repeated measures, which was preceded by Mauchly’s Test of Sphericity as to asses normality of the variables measured.

The standard error of measurement (SEM; the average SD among observations from the same subject) of the AHI, a measure of the within-subjects variability, was calculated according to Fleiss and Kingman [19]. The smallest detectable difference (SDD) in the AHI (the SDD between two AHI outcomes of an individual, which is significant at a 95% probability level [19-21]) was calculated according to the following formula: SDD= 1.96*√2*SEM.

The intra class correlation coefficient (ICC) was used to determine the intra-examiner reliability of the AHI scoring of the home recordings. Further, paired samples t-tests were used to test the null hypothesis that there were no within-subject differences in the AHI between the baseline and follow-up PSGs that were both obtained in the hospital. Finally, the linear regression was calculated between the individual’s mean values and standard deviations from the four AHI recordings. All statistical tests were performed with SPSS 12.0 software package (SPSS Inc., Chicago, IL). Probability levels of p < 0.05 were considered statistically significant.

Results

For the duration of the study, none of the patients showed any adverse effects due to wearing the placebo appliance. All patients wore their appliance during at least 90% of the nights of the 10-week study period. Table 1 shows the average (± SD) values of the BMI, ESS, AHI, and standard sleep variables for each of the four home PSG recordings. The BMI and ESS did not differ significantly between the four nights, nor did the standard sleep variables (F = 0.167 – 1.078; p = 0.918 – 0.369).
The examiner’s intra-observer reliability in the AH1 scoring was excellent, with an ICC of 0.96; that of the sleep scoring could be qualified as at least fair-to-good, with ICC values ranging from 0.63 to 0.94. The AH1 values of the baseline hospital recordings without the placebo appliance did not differ significantly from the AH1 values of the hospital recordings with the placebo appliance in situ (T = 1.881; p = 0.079).

Figure 2 shows the course of the AH1 of the fifteen patients for the four home recordings (PSG1-4). No differences were found between the average AH1 values of the four nights (F = 0.051; p = 0.985). Based upon the data of the four nights, the standard error of measurement SEM of the AH1 was 4.6 events/hour and the smallest detectable difference was 12.8 events/hour. For each participant, the mean value and standard deviation of the four AH1 outcomes were also calculated. A significant linear regression was found between the mean value and the standard deviation (R^2 = 0.276; p < 0.044), with a tendency of higher standard deviations for participants with higher mean values (Fig. 3).

Table 1. Descriptive statistics (mean ± SD) of the patient characteristics, the apnea-hypopnea index (AH1), and sleep variables of the fifteen OSA patients as derived from four polysomnographic recordings (PSG1-4)

<table>
<thead>
<tr>
<th></th>
<th>PSG 1</th>
<th>PSG 2</th>
<th>PSG 3</th>
<th>PSG 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patient characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMI</td>
<td>30.1 ± 4.5</td>
<td>30.2 ± 4.6</td>
<td>30.2 ± 4.6</td>
<td>30.1 ± 4.8</td>
</tr>
<tr>
<td>ESS</td>
<td>9.9 ± 4.2</td>
<td>9.4 ± 5.2</td>
<td>9.1 ± 5.5</td>
<td>9.4 ± 5.2</td>
</tr>
<tr>
<td><strong>Respiratory</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AH1 (events/hour)</td>
<td>10.7 ± 8.2</td>
<td>10.1 ± 8.3</td>
<td>9.9 ± 6.2</td>
<td>9.2 ± 6.7</td>
</tr>
<tr>
<td><strong>Sleep</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total sleep time (min)</td>
<td>432.7 ± 61.9</td>
<td>449.2 ± 85.4</td>
<td>415.4 ± 73.2</td>
<td>421.9 ± 64.6</td>
</tr>
<tr>
<td>Stage 1 and 2 (%)</td>
<td>65.5 ± 10.2</td>
<td>64.3 ± 13.3</td>
<td>66.9 ± 12.4</td>
<td>66.3 ± 15.2</td>
</tr>
<tr>
<td>Stage 3 and 4 (%)</td>
<td>12.8 ± 8.9</td>
<td>14.4 ± 11.9</td>
<td>12.8 ± 12.6</td>
<td>11.9 ± 12.6</td>
</tr>
<tr>
<td>Stage REM (%)</td>
<td>21.3 ± 6.7</td>
<td>21.0 ± 6.1</td>
<td>20.1 ± 9.2</td>
<td>21.4 ± 6.5</td>
</tr>
<tr>
<td>Total sleep supine (%)</td>
<td>39.4 ± 21.3</td>
<td>34.9 ± 30.7</td>
<td>36.8 ± 25.9</td>
<td>43.5 ± 28.1</td>
</tr>
</tbody>
</table>

Fig. 2. Individual values of the apnea-hypopnea index (AH1) in the 15 OSA patients, as derived during the four polysomnographic recordings (PSG1-4).

Fig. 3. Linear regression between the mean values and the standard deviations of the AH1 of the 15 OSA patients. The broken vertical line represents the cut-off criterion of 5 for the AH1.
Discussion

In this study, a considerable intra-individual variability was found between the four AHI recordings obtained during a follow-up period of ten weeks. It is of importance to take this variability into account when using this index in clinical practice and research.

A so-called first-night effect may contribute to the AHI variability over time. To avoid this first-night effect, a first PSG recording, used solely for habituation to sleeping under experimental conditions, often precedes the recording that is used for diagnosis and research purposes [22]. In this study, a first-night effect could not be detected in the four home PSG recordings. The standard sleep variables and the AHI recorded during the first night were not significantly different from those recorded during the following nights. This may be due to the fact that the first home PSG recording was preceded by a baseline PSG recording at the hospital (as part of the inclusion procedure), thus giving the patients the opportunity of habituation to the equipment.

In the present study, the AHI variability was studied in a group of patients with a placebo appliance in situ. One could wonder whether this placebo appliance could have influenced the respiratory variables. However, the AHI did not differ significantly between the baseline PSG (without the placebo appliance) and the follow-up PSG recording (with placebo appliance in situ) that were both obtained at the hospital. This indicates that the placebo appliance used in this study may be considered inert. Hence, this study provides proper insight into the natural variability in the AHI in OSA patients.

In this study, no differences in the average AHI values from one night to the other were found. Similar results were found in other studies [3-5, 7-10]. However, in a prospective study, Bittencourt et al. [2] already pointed out that a small AHI variability between consecutive nights at a group level does not necessarily mean also a small variability at an intra-individual level: thirteen out of their twenty patients presented a difference between nights in the AHI value equal to or higher than 10 events/hour. A similar result could be found in six of the fifteen patients in the present study which had a much longer period between the four recording nights. The excellent intra-examiner reliability found in this study indicates that the variability of the examiner only played a minor role in the AHI variability. A more plausible explanation for AHI variability would be a relation to the intrinsic variability of the OSA phenomenon under study, which in turn can be influenced by variables like behavioural factors and differences in body position and sleep architecture between recordings. Thus, despite the small variability in the AHI observed at a group level, this long-term study revealed that there may be a considerable variability in AHI present at an intra-individual level.

According to the American Academy of Sleep Medicine Task Force [1], OSA is defined as the combination of symptoms (e.g., excessive daytime sleepiness) with 5 or more obstructive breathing events per hour of sleep. The number of obstructed breathing events is usually quantified by the number of apneas and hypopneas per hour of sleep (AHI). The cut-off point of 5 for the AHI is widely used to recognize OSA, but it should be noted that the clinical value of any particular cut-off point has not been adequately determined yet and is mainly based on consensus [1, 23]. Moreover, in a systematic review of Franklin et al. [24] the establishment of specific AHI cut-off criteria for both genders and for different age groups is recommended. To that end, large study samples should be included in future studies to this topic. However, the time-variant nature of the disorder complicates the use of any cut-off point of the AHI. If the recorded values of the AHI lie close to this point, the value may for one night be above and for another night below this point. This suggests that conclusions about the presence or absence of OSA can only be drawn when the single-night AHI recordings lie far enough away from the cut-off point; that is outside the limits of a cut-off band surrounding that point. Supposing that the chances of a single-night PSG recording of a non-OSA patient to indicate the presence of OSA should be smaller than 5% (false positive) and that the chances of a PSG recording of an OSA patient to deny the presence of OSA should also be smaller than 5% (false negative), then, under the assumption that the single-night recordings are normally distributed, the cut-off band is the 90% probability interval around the cut-off point, with its upper and lower limits at a distance of 1.645 times the standard deviation from
that point. In the present study, the linear regression found between the standard deviations and mean values of the AHI variable (Fig. 3) suggests that the best estimate for the standard deviation of single-night AHI recordings for patients with an AHI value close to the cut-off point of 5 is 2.9. Based upon this estimate and within the limitations of the present study, the cut-off band for the AHI variable will range from 0.2 to 9.8 (5 ± 1.645*2.9). For single-night AHI recordings lying above or below this cut-off band, the PSG recordings indicate that OSA is present or absent. For single-night AHI recordings within the cut-off band, the PSG recordings do not allow conclusions to be drawn about the presence or absence of OSA. When decisions are not based upon single-night AHI recordings but upon the average of n recordings, the cut-off band will become narrower, since the standard deviation of the average of n recordings is that of a single recording divided by √n.

An observed difference between pre- and post-therapeutic outcome variables does not necessarily reflect a possible therapeutic effect, but may also be related to the natural course of the disorder, the biological fluctuation of the variables measured, and possible inconsistencies in measurement. The smallest detectable difference (SDD) is an appropriate measure for assessing these possible sources of variability within individual patients [19-21]. In this study, a SDD of the AHI of 12.8 events/ hour was found when the follow-up data of all 15 participants were used in its calculation. This suggests that a change in AHI should be > 12.8 events/ hour before it can be considered statistically significant at a 95% probability level. However, Fig. 3 indicates that participants with a lower mean AHI value have a tendency to have a smaller SDD than those with a higher mean value. In the calculation of the SDD, this tendency was ignored because taking it into account would have required a much larger group of patients than the 15 participating in the present study. In this context, it should be mentioned that the variability in AHI is not expected to decrease with a larger sample size; only its estimate will be more accurate. Further, it should be noted that the calculated SDD may be dependent on factors, which are specific for the study setting (e.g., sample characteristics). Therefore, we recommend establishing SDDs for individual clinics and sleep laboratories when using the AHI for research purposes. Moreover, it should be kept in mind that a statistically significant change in the AHI value is not necessarily also a clinically relevant change; the latter remains to be determined by the responsible clinician and must also be based upon medical grounds.

In conclusion, our results suggest a considerable intra-individual variability between AHI recordings. Hence, single-night recordings can only confirm or deny the presence of OSA when the recordings lie outside a cut-off band surrounding the AHI cut-off point. The large AHI variability should also be taken into account when evaluating therapy success. In this context, it should be noted that it is mainly the approach that we would like to convey to the reader; not the cut-off values per se.
References


