Long-Term Follow-Up of a Randomized Controlled Trial of Oral Appliance Therapy in Obstructive Sleep Apnea

Ghizlane Aarab a Frank Lobbezoo a Martijn W. Heymans b Hans L. Hamburger c Machiel Naeije a

a Department of Oral Kinesiology, Academic Centre for Dentistry Amsterdam (ACTA), Research Institute MOVE, University of Amsterdam and VU University Amsterdam, b Department of Epidemiology and Biostatistics, VU University Medical Center Amsterdam, and c Department of Clinical Neurophysiology and Center for Sleep-Wake Disorders, Slotervaart Medical Center, Amsterdam, The Netherlands

Key Words
Obstructive sleep apnea · Long term · Mandibular advancement device · Continuous positive airway pressure · Randomized controlled trial · Treatment · Compliance

Abstract
Background: Long-term trials are needed to capture information regarding the persistence of efficacy and loss to follow-up of both mandibular advancement device (MAD) therapy and continuous positive airway pressure (CPAP) therapy. Objectives: The aim of the study was to compare these treatment aspects between MAD and nasal CPAP (nCPAP) in a 1-year follow-up. Methods: Forty-three mild/moderate obstructive sleep apnea patients (52.2 ± 9.6 years) with a mean apnea-hypopnea index (AHI) of 20.8 ± 9.9 events/h were randomly assigned to two parallel groups: MAD (n = 21) and nCPAP (n = 22). Four polysomnographic recordings were obtained: one before treatment, one for the short-term evaluation, and two recordings 6 and 12 months after the short-term evaluation. Excessive daytime sleepiness (EDS) was also evaluated at the polysomnographic recordings. Results: The initially achieved improvements in the AHI remained stable over time within both groups (p = 0.650). In the nCPAP group, the AHI improved 4.1 events/h more than in the MAD group (p = 0.000). The EDS values showed a gradual improvement over time (p = 0.000), and these improvements were similar for both groups (p = 0.367). In the nCPAP group, more patients withdrew from treatment due to side effects than in the MAD group. Conclusions: The absence of significant long-term differences in EDS improvements between the MAD and the nCPAP groups with mild/moderate obstructive sleep apnea may indicate that the larger improvements in AHI values in the nCPAP group are not clinically relevant. Moreover, nCPAP patients may show more problems in accepting their treatment modality than MAD patients.

Introduction
Obstructive sleep apnea (OSA) is characterized by recurrent obstruction of the upper airway, often resulting in oxygen desaturation and arousal from sleep [1]. Excessive daytime sleepiness (EDS), snoring and reduction in cognitive functions are among the common symptoms of this condition [2].
Although continuous positive airway pressure (CPAP) has been proposed as the most effective treatment for OSA [3], nowadays, mandibular advancement devices (MADs) play an important role in the treatment of mild/moderate OSA patients [2, 4]. These devices increase the pharyngeal space by protruding the mandible and advancing the tongue.

The short-term therapeutic efficacy of MADs has been compared with that of CPAP and was proven to be satisfactory in several randomized controlled trials [e.g. 5–12]. However, long-term parallel-group trials are needed to capture information regarding the persistence of efficacy and the loss to follow-up [3]. Therefore, the aim of the present study was to compare these treatment aspects between MAD and nasal CPAP (nCPAP) in a 1-year follow-up study.

**Materials and Methods**

**Participants**

This study is the 1-year follow-up of a short-term randomized controlled trial (RCT), in which 3 therapy groups (MAD, nCPAP and placebo) were compared [12]. OSA patients were invited for participation in the initial short-term study when they fulfilled the following inclusion criteria: age >18 years, an apnea-hypopnea index (AHI) between 5 and 45 events/h and an Epworth Sleepiness Score ≥10 [13] or at least two of the symptoms suggested by the American Academy of Sleep Medicine Task Force, for example, unrefreshing sleep and daytime fatigue [1]. The placebo group was excluded from the long-term study for ethical reasons. Moreover, OSA patients with an AHI >10 events/h and less than 50% reduction in AHI at the short-term evaluation were also excluded from the long-term study. The baseline characteristics of the patients at the time of therapy allocation are presented in Table 1. This long-term study was also approved by the Slotervaart Hospital’s Ethics Committee (No. U/1731/0326, U/2679/0326).

**Randomization and Interventions**

At the start of the short-term RCT, using block randomization, consenting patients were allocated to the interventions. The allocation sequence was automatically generated and concealed by an independent co-worker. The two interventions studied in this parallel-group follow-up study were: MAD [14, 15] and nCPAP (REMstar Pro; Respironics, Herrsching, Germany).

Both MAD and nCPAP were titrated before the start of the treatment [12]. The titration of nCPAP was performed during a polysomnographic (PSG) recording. The pressure was increased in incremental steps of 1 cm H₂O/h, until respiratory disturbances and respiration-related arousals were reduced to ≤5/h, and snoring was minimized. The average value of the pressure was 7.3 cm H₂O (SD 1.9, range 4–11). For the titration of the MAD, 4 ambulatory PSG recordings were obtained at regular intervals [15]. The most effective protrusion position of the MAD (that is, the mandibular position that yielded the lowest AHI value) was chosen from among 4 randomly offered positions (0, 25, 50 and 75% of the maximum protrusion). The MAD was set at 25% of the maximum protrusion in 1 patient, at 50% in 7 patients and at 75% in 12 patients.

Analyst blinding was ascertained by assigning codes to data sets and by analyzing these sets in random blocks. For more details, see Aarab et al. [12].

**Procedure**

From all patients, 4 PSG recordings were obtained in the sleep laboratory of the Slotervaart Medical Center: 1 before treatment, 1 for the short-term evaluation (approximately 6 months after therapy assignment) and 2 for the long-term evaluation (approximately 6 and 12 months after the short-term evaluation). The montage was performed at the Slotervaart Medical Center by a trained co-worker. Each PSG recording was analyzed manually, under blind conditions, by the same examiner, who was experienced in scoring PSG recordings, using internationally accepted criteria [1, 16]. Sleep stages were scored in 30-second epochs and standard sleep and respiratory outcome variables were obtained. The mounting and procedure of the PSG recordings are described in detail in Aarab et al. [17]. The primary and secondary outcome measures were obtained at the time of the PSG recordings.

The therapy evaluation PSG recordings were followed by a visit at ACTA, during which the participants were interviewed about their compliance (percent of nights per week of usage) and possible side effects (nature and number).

**Outcome Measures**

The change in the AHI between baseline and therapy evaluation (ΔAHI) was the primary outcome variable. Secondary outcome variables were the changes in sleep variables and in EDS between baseline and therapy evaluation. Other secondary outcome variables were self-reported compliance and side effects.

**Statistical Analyses**

Differences in patient characteristics at baseline between the two therapy groups were analyzed using independent t tests and χ² tests. Outcome variables that showed significant between-groups differences at baseline were used as covariate in the subsequent analyses (see below).

### Table 1. Patient characteristics (mean ± SD) at baseline of the MAD and nCPAP groups

<table>
<thead>
<tr>
<th></th>
<th>MAD (n = 21)</th>
<th>nCPAP (n = 22)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, years</td>
<td>50.4 ± 8.9</td>
<td>54.9 ± 10.1</td>
</tr>
<tr>
<td>Number of men/women</td>
<td>17/4</td>
<td>15/7</td>
</tr>
<tr>
<td>Apnea-hypopnea index, events/h</td>
<td>21.4 ± 11.0</td>
<td>20.1 ± 9.0</td>
</tr>
<tr>
<td>Body mass index*</td>
<td>27.1 ± 3.1</td>
<td>30.5 ± 3.4</td>
</tr>
<tr>
<td>Neck circumference, cm</td>
<td>41.7 ± 2.9</td>
<td>43.2 ± 3.8</td>
</tr>
<tr>
<td>Excessive daytime sleepiness</td>
<td>12.0 ± 5.7</td>
<td>11.0 ± 4.4</td>
</tr>
</tbody>
</table>

* Statistically significant at the 0.05 probability level.
The associations between one or more predictors and missing values in AHI at the therapy evaluations were studied using logistic regression analyses. Several variables were found to be related to the missing values. These predictors of missing values were included in an imputation model to estimate the missing values by applying multiple imputation (MI) [18]. MI was based on the multivariate imputation by chained equations procedure [19]. This procedure allows one to specify the multivariate structure in the data as a series of conditional regression models, based on the information of other variables included in the imputation model. Ten separate imputation samples were generated, for both treatment groups separately.

Following the MI procedure, generalized estimating equation analyses were performed to study differences between both groups (MAD and nCPAP) for the primary and secondary outcome variables. For each variable, its baseline value was used as a covariate to protect against potential regression to the mean effects. Interactions of treatment groups with time were used to study if differences in treatment effects increased or decreased over time. Generalized estimating equation analyses were done in each imputed dataset, and the results were summarized using Rubin’s rules [21].

All statistical tests were performed with the SPSS 17.0 (SPSS Inc., Chicago, Ill., USA) and R (R Foundation for Statistical Computing, Vienna, Austria) software packages.

Results

A total of 64 patients were enrolled in the initial short-term study, and were randomized at the start of the RCT as shown in figure 1. The placebo group was excluded from the long-term part of the study. Table 1 illustrates that the MAD group had a significantly lower BMI than the nCPAP group (T = 3.921; p = 0.001). This difference was constant over time (F = 1.456, p = 0.242).

The mean (±SD) baseline values of the respiratory, subjective, and sleep variables as well as the changes in these variables from baseline to therapy evaluation are shown in table 2.

Loss to Follow-Up

At the short-term evaluation, three patients in the MAD group were instructed to stop with the therapy, because it was not effective (AHI >10 events/h and less than 50% reduction in AHI). These 3 patients were offered a treatment with nCPAP instead. After the short-term evaluation, a total of 35 patients in the MAD and nCPAP...
groups started with a 1-year follow-up. In the MAD group, 2 patients dropped out, because they experienced more side effects than benefits from the treatment. These 2 MAD patients reported the following side effects: discomfort in wearing (n = 2), tenderness in the masseter muscle region upon awakening (n = 1) and feeling of a changed occlusion upon awakening (n = 1). In the nCPAP group, 2 patients dropped out because of private reasons that were unrelated to the study and 3 patients dropped out because they experienced more side effects than benefits from the treatment. The following side effects were reported by these 3 nCPAP patients: problems with expiration against the positive pressure (n = 3), pain due to pressure of the mask (n = 1) and difficulty in changing sleep position (n = 2). Finally, a total of 28 patients completed the entire study protocol (see fig. 1).

At the end of the present study, patients were advised to continue with their treatment and were monitored with polysomnography on a yearly basis.

Primary Outcome Variables
Analyses of the imputed data sets showed that the MAD group had a significantly smaller ΔAHI value than the nCPAP group (p = 0.000; table 3). The mean difference over time within groups was not statistically significant in the nCPAP group (p = 0.650). However, in the MAD group, the ΔAHI value was statistically significant (p = 0.000; table 3).

Table 2. The mean (± SD) baseline and Δ values (that is, difference between baseline and therapy evaluation) of the respiratory, subjective and sleep outcome variables of the MAD group and the nCPAP group

<table>
<thead>
<tr>
<th></th>
<th>MAD</th>
<th>nCPAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline value</td>
<td>(n = 21)</td>
<td>(n = 22)</td>
</tr>
<tr>
<td>Δ value at short-term therapy evaluation</td>
<td>21.4 ± 11.0 to 16.3 ± 10.3</td>
<td>20.1 ± 9.0 to 19.5 ± 8.7</td>
</tr>
<tr>
<td>Δ value at 6 months after short-term therapy evaluation</td>
<td>15.0 ± 10.5 to 4.8 ± 5.1</td>
<td>11.0 ± 4.3 to 1.2 ± 4.9</td>
</tr>
<tr>
<td>Δ value at 12 months after short-term therapy evaluation</td>
<td>15.0 ± 10.5 to 4.8 ± 5.1</td>
<td>11.0 ± 4.3 to 1.2 ± 4.9</td>
</tr>
<tr>
<td>Δ value at 12 months after short-term therapy evaluation</td>
<td>15.0 ± 10.5 to 4.8 ± 5.1</td>
<td>11.0 ± 4.3 to 1.2 ± 4.9</td>
</tr>
</tbody>
</table>

Table 3. Outcomes of the GEE analyses for the primary and secondary outcome variables

<table>
<thead>
<tr>
<th></th>
<th>Treatment effect(^1) (MAD vs. nCPAP)</th>
<th>Time effect(^1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean difference between groups p(^2)</td>
<td>mean difference over time within groups p(^2)</td>
</tr>
<tr>
<td>Primary outcome variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ΔAHI, events/h</td>
<td>-4.1 (-5.7, -2.5) 0.000(^*)</td>
<td>0.2 (-0.7, 1.1) 0.650</td>
</tr>
<tr>
<td>ΔEDS</td>
<td>-0.9 (-2.8, 1.0) 0.367</td>
<td>1.9 (1.3, 2.5) 0.000(^*)</td>
</tr>
<tr>
<td>Δ Respiratory arousals, events/h</td>
<td>-3.2 (-4.5, -1.8) 0.000(^*)</td>
<td>0.1 (-0.6, 0.7) 0.816</td>
</tr>
<tr>
<td>Δ Stage NREM, %</td>
<td>0.2 (-1.8, 2.2) 0.854</td>
<td>0.2 (-0.7, 1.1) 0.692</td>
</tr>
<tr>
<td>Δ Stage REM, %</td>
<td>-0.1 (-2.2, 1.9) 0.897</td>
<td>-0.2 (-1.1, 0.7) 0.703</td>
</tr>
</tbody>
</table>

\(^1\) For all variables, there was no significant interaction of the groups with time. Therefore, only the between-groups effect and the time effect within groups are reported.

\(^2\) p value as result of the GEE-analyses, controlled for the effects of the baseline value and of BMI.

Δ Values show the difference between baseline and therapy evaluation value. Figures in parentheses are 95% confidence intervals. REM = Rapid-eye-movement; NREM = non-rapid-eye-movement. * Statistically significant at the 0.001 probability level.

Long-Term Effects of Oral Appliances
Respiration 2011;82:162–168
165
ence between both groups in ΔAHI was 4.1 events/h (table 3). The change in AHI was stable over time as indicated by the non-significant time effect (p = 0.650; table 3). For the ΔAHI, there was no significant interaction of the groups with time.

Figure 2 shows the AHIs over time for each patient who completed the trial. nCPAP treatment was offered to the patient who had an AHI of 20 events/h at the 6-month evaluation, because the therapy was considered ineffective. This patient, however, wanted to continue the MAD treatment for 6 months, because he experienced subjective benefits of the treatment (improvement in EDS and a decrease in snoring sound).

Secondary Outcome Variables

The MAD group had a significantly smaller change in respiratory arousal index than the nCPAP group (p = 0.000; table 3). The mean difference between both groups in the change of the respiratory arousal index was 3.2 events/h (table 3). There was no significant difference between both groups in the change of EDS (ΔEDS). The ΔEDS increased over time, as indicated by the time effect (p = 0.000; table 3).

The MAD patients who completed the trial used their appliance 85.8% (SD 18.8) of the nights, the nCPAP patients 84.8% (SD 20.6) of the nights. There was no significant difference between both groups in compliance.

The nature and number of side-effects at the first evaluation are described in detail in Aarab et al. [12]. In most cases, the side effects in the MAD group had a dental nature (for example, sensitive teeth upon awakening, tenderness in maseter muscle region and feeling of changes in occlusion upon awakening). The side effects in the nCPAP group were in most cases related to the mask and the cumbersome nature of the CPAP device (for example, pain due to pressure of the mask and problems with expiration against the positive pressure). The number of side effects decreased over time within both groups (p = 0.000). In the MAD group, the number of side effects reduced from 1.5 at the short-term evaluation to 0.7 at the 12-month evaluation. In the nCPAP group, the number of side effects was reduced from 2.2 at the short-term evaluation to 1.0 at the 12-month evaluation. For all secondary outcome variables, there was no significant interaction of the groups with time.

Discussion

The short-term improvement in AHI was maintained in both the MAD group and the nCPAP group in this 1-year follow-up. The EDS further improved over time for both treatment modalities.

Randomized clinical trials are a powerful tool for investigating treatment effects, but in human trials there are often problems of noncompliance, where the patient does not adhere to the treatment assigned. A common approach to the analysis of data with missing values is to...
exclude the patients with missing values. Typically, this leads to a reduction in statistical power and to estimates that can potentially be biased when the probability of a missing value is related to the characteristics of the patients [22]. To overcome this problem, imputation methods for missing data have been developed [23]. There is increasing evidence of the superiority of multiple imputation methods to replace missing values, suggesting that these methods should be preferred over other imputation methods [18, 22]. Therefore, in this study, the multiple imputation method was used to replace the missing values.

In the short-term evaluation, no difference in the ΔAHI was found between the MAD and nCPAP therapies. Only in the worst case scenario, with the failure and success patterns set at their extreme values in favor of nCPAP, the difference between the two treatment modalities was significant [12]. In the present long-term evaluation, with more measurement points, a significant difference in ΔAHI of 4.1 events/h was found between the two therapies. However, this small difference may not be clinically relevant, because there was no significant difference between the two groups in the improvement of EDS.

Similar to the present study, a 4-year follow-up study of CPAP therapy reported that AHI values were stable over time [24]. Surprisingly, there are no other studies to determine whether the efficacy of CPAP is still adequate more than 3 months after the start of the treatment [25]. In line with our findings, MADs were still effective in the long term in other studies [26, 27]. Others also found that MADs were effective, but they both observed a tendency for the efficacy to reduce over time [28, 29]; a tendency which was not found in the present study. Since OSA is usually a lifelong condition [30], it is of importance that therapy is effective in the long term. Although all above-mentioned studies followed patients in the long term, studies including both MAD therapy and nCPAP therapy are lacking [3]. Differences between these treatment modalities can only be found by including both modalities in a single study. This is the first study in which this treatment aspect is compared between MAD and nCPAP in a single study.

Interestingly, the improvement in EDS, which was already seen in the short-term evaluation [12], further improved in this 1-year follow-up. This was surprising, because the AHI value and the respiratory arousal index value did not reduce anymore. It indicates that EDS in OSA patients may need time to show further improvement in mild/moderate OSA patients, which was also found in another study [24]. As hypothesized by Meurice et al. [24], a slow progressive reversibility of abnormal cerebral functions under long-term treatment may be possible. On the other hand, also a deterioration or no change in the initially achieved improvement in EDS has been reported in the long term [26, 27]. Future studies are needed to confirm and explain a possible delayed effect on EDS.

The side effects reported by both groups were comparable with those in previous studies [31–34]. Side effects can lead to the discontinuation of the treatment [34, 35], which was also found in the present study. From the start of the short-term RCT until the end of the long-term RCT, 6 patients in the nCPAP group and 2 patients in the MAD group withdrew from treatment due to the occurrence of side effects, suggesting that nCPAP patients show more problems in accepting their treatment modality than MAD patients. Further, it should be noted that in the MAD group, 3 patients withdrew from treatment after the short-term evaluation, because the therapy was not effective.

In conclusion, the absence of significant long-term differences in EDS improvements between the MAD and the nCPAP groups with mild/moderate OSA may indicate that the larger improvements in AHI values in the nCPAP group are not clinically relevant. In the nCPAP group, more patients withdrew from treatment due to the occurrence of side effects, suggesting that nCPAP patients show more problems in accepting their treatment modality than MAD patients.

Acknowledgements

The authors thank the staff of the Center for Sleep-Wake Disorders of Slotervaart Medical Center in Amsterdam, The Netherlands, for their assistance with this work, Dr. Corine Visscher and Dr. Irene Aartman for their assistance with the statistical analyses of this study, and the Dental Laboratory Excent, Amsterdam, The Netherlands, for the assistance in the development of the MAD used in this study. The Netherlands Institute for Dental Sciences (IOT) supported this work.

Financial Disclosure and Conflicts of Interest

No actual or potential conflicts of interest exist for any of the authors, nor is there any personal or financial support and author involvement with organizations with financial interest in the subject matter of the paper to be disclosed for any of the authors.
References


