Influence of Tonsillar Size on OSA Improvement in Children Undergoing Adenotonsillectomy

Alice Tang, MD¹, James R. Benke², Aliza P. Cohen, MA³, and Stacey L. Ishman, MD, MPH¹,3,4

Abstract

Objective. To determine if pediatric obstructive sleep apnea (OSA) improves after adenotonsillectomy (AT) regardless of tonsil size.

Study Design. Case series with chart review.

Setting. Pediatric Otolaryngology Department, Johns Hopkins Hospital.

Subjects. Seventy children 1 to 18 years of age who underwent polysomnography (PSG) before and after AT.

Methods. Tonsil size was evaluated using the Brodsky grading scale.

Results. Children were stratified by tonsil size as 2⁺ (n = 20), 3⁺ (n = 36), and 4⁺ (n = 14). There was a significant improvement in obstructive apnea-hypopnea index (oAHI), apnea index (AI), and saturation nadir across all 3 groups after AT. Preoperative oAHI, AI, and hypopnea index (HI) were similar regardless of tonsil size (P > .05). Overall, oAHI improved from a median of 11.8 ± 21.7 to 2.0 ± 6.1 events/h, with 40% (28/70) of children having complete resolution. The oAHI (P < .0001-0.02), AI (P < .0001-0.017), HI (P < .0001-0.058), and saturation nadir (P < .0001-0.017) significantly improved for the 2⁺, 3⁺, and 4⁺ groups. Only the HI (P = .058) in the 2⁺ group did not. The median oAHI improvement was 3.4 ± 26.4 events/h in the 2⁺ group, 8.3 ± 16.6 events/h in the 3⁺ group, and 12.3 ± 19.5 events/h in the 4⁺ group, with 25% (5/20), 50% (18/36), and 36% (5/14), respectively, having complete resolution. There was no correlation between OSA severity and tonsil or adenoid size (P > .32).

Conclusion. Tonsil size did not correlate with OSA severity. While a larger proportion of patients classified as 3⁺ and 4⁺ had complete resolution after surgery, significant improvement was seen in AI and saturation nadir even in those classified as 2⁺.

Keywords
polysonmography, sleep study, small tonsils, obstructive sleep apnea, OSA, sleep apnea, tonsil size

Received October 23, 2014; revised February 20, 2015; accepted March 31, 2015.

Sleep-disordered breathing (SDB) occurs in 20% to 40% of children in the United States, and the most severe form of SDB is obstructive sleep apnea (OSA). Obstructive sleep apnea affects approximately 1% to 5.8% of children and is most commonly caused by adenotonsillar hypertrophy.

In view of the well-recognized neurocognitive, cardiovascular, and behavioral sequelae of untreated OSA1-3 and studies indicating that timely treatment can prevent these potentially serious sequelae, clinicians generally recommend that children with adenotonsillar hypertrophy undergo adenotonsillectomy (AT). Moreover, in many children, AT has been shown to be curative.4-7 Guidelines from both the American Academy of Otolaryngology—Head and Neck Surgery Foundation (AAO-HNSF) and American Academy of Pediatrics thus recommend AT as a first-line treatment when there is a high index of suspicion for OSA.8,9

Although the definitive diagnosis of pediatric OSA is best established by overnight polysomnography (PSG), factors such as access to sleep laboratories and long waiting times for testing preclude its widespread use prior to AT.10 Furthermore, because AT has been shown to resolve symptoms, the perceived need for preoperative PSG testing is diminished. Given these circumstances, the diagnosis of

¹Department of Otolaryngology—Head and Neck Surgery, University of Cincinnati College of Medicine, Cincinnati, Ohio, USA
²Department of Otolaryngology—Head and Neck Surgery, Johns Hopkins School of Medicine, Baltimore, Maryland, USA
³Division of Pediatric Otolaryngology—Head and Neck Surgery, Cincinnati Children’s Hospital Medical Center, Cincinnati, Ohio, USA
⁴Division of Pulmonary Medicine, Cincinnati Children’s Hospital Medical Center, Cincinnati, Ohio, USA

This article was presented at the 2014 AAO-HNSF Annual Meeting & OTO EXPO; September 21-24, 2014; Orlando, Florida.

Corresponding Author:
Stacey L. Ishman, MD, MPH, Upper Airway Center, Cincinnati Children’s Hospital Medical Center, 3333 Burnet Ave, MLC 2018, Cincinnati, OH 45229, USA.
Email: Stacey.Ishman@cchmc.org
OSA has largely been based on the caregiver’s account of habitual snoring with pauses as well as the physician’s anatomic assessment, including a subjective evaluation of tonsil size.11

For children with small tonsils and significant symptoms of SDB, the AAO-HNSF guidelines suggest that clinicians should advocate preoperative PSG. This recommendation is based on the premise that there is less certainty regarding the diagnosis of OSA when tonsillar hypertrophy is absent, despite the fact that a relationship between subjective tonsil size and SDB severity has not been established.12

To our knowledge, the question of whether removing small tonsils in children will yield objective improvement in OSA has not yet been addressed. To this end, our objective was to determine if OSA improves after AT regardless of tonsil size.

Methods
Patient Selection
Consecutive children from 1 to 18 years of age who underwent AT performed by a single pediatric otolaryngologist at the Johns Hopkins Hospital and who received a pre- and postsurgical overnight PSG were included in the study. All children with signs and/or symptoms of SDB were sent for a PSG prior to AT, regardless of tonsil size. In addition, all children with an obstructive apnea-hypopnea index (oAHI, defined below) ≥5 events/h or symptoms before AT were sent for a postoperative PSG. Children were excluded if they had a syndromic disorder, craniofacial deformity, or neurologic impairment. They were also excluded if there was either incomplete data or if more than 12 months had elapsed between pre- and postsurgical PSG. Demographic data and results from pre- and postsurgical PSG were documented. Body mass index (BMI) was calculated for each patient and compared using BMI z scores as derived from the 2000 growth standards of the Centers for Disease Control and Prevention.

Tonsil size was evaluated preoperatively in the office setting using the Brodsky grading scale.13 Tonsils that did not extend beyond the tonsillar pillars received a grade of 1+, tonsils that extended just beyond the tonsillar pillars received a grade of 2+, tonsils that extended beyond the pillars but did not approach midline received a grade of 3+, and tonsils that touched or approached the midline received a grade of 4+. Adenoid size was also quantified from 1+ to 4+. When the entire vomer could be seen, adenoids were graded as 1+; when the bottom one-third of the vomer was obscured, they were graded as 2+; when two-thirds of the vomer was obscured by adenoid tissue, they were graded as 3+; and when the vomer was completely obscured by adenoidal tissue, they were graded as 4+. This study was approved by the institutional review board at Johns Hopkins School of Medicine.

PSG Data
Results of the overnight 16-channel PSG were extracted from the electronic medical records. Most studies were performed in the Johns Hopkins Pediatric Sleep Laboratory with Somnologica or REMLogic (Embla, Broomfield, Colorado). Signals included electroencephalograms (leads C3-A2, C4-A1, and O1-A2), left and right electro-oculograms, submental electromyogram, tibial electromyogram, electrocardiogram, and oxyhemoglobin saturation (Masimo, Irvine, California). End-tidal CO2 (Novametrix, Murrysville, Pennsylvania) was acquired in all participants. Airflow was acquired with a nasal cannula (Salter Labs, Arvin, California) connected to a differential pressure transducer (Pro-Tech, Mukilteo, Washington). Respiratory effort was assessed with thoracic and abdominal inductive plethysmography (Embla), and body position was monitored via infrared video camera. Several sleep studies were carried out using Alice 4.0 (Healthdyne, Marietta, Georgia), which uses a strain gauge with piezoelectric crystals to evaluate respiratory effort. Pediatric sleep medicine physicians at the Johns Hopkins Hospital interpreted the results.

PSG Analysis
The PSGs were scored according to the American Academy of Sleep Medicine (AASM) guidelines. Apnea was defined as ≥90% reduction of airflow for at least 2 breath cycles. Apneas were identified as obstructive when associated with continued or increased inspiratory effort. A mixed apnea was identified when absence of airflow was associated with periods with and without inspiratory effort. Hypopnea was defined as a decrease in airflow of ≥50% for at least 2 breath cycles followed by a ≥3% decrease in oxygen saturation or an electrocortical arousal from sleep. Respiratory event-related arousals (RERAs) were not scored. The apnea index (AI) was calculated as the number of obstructive apneas, including mixed apneas, divided by the total sleep time. The hypopnea index (HI) was calculated as the number of obstructive hypopneas divided by the total sleep time. The apnea-hypopnea index (AHI) was calculated as the number of respiratory events (apneas and hypopneas) divided by the total sleep time. The oAHI was calculated as the number of obstructive and mixed apneas and hypopneas divided by the total sleep time. Severity of OSA was stratified by oAHI. Mild OSA was defined as 1 to <5 events per hour, moderate OSA was defined as 5 to <10 events per hour, and severe OSA was defined as ≥10 events per hour. Following AT, an oAHI of <1.0 was considered complete resolution. Saturation nadir was defined as the lowest oxygen saturation reading during an obstructive respiratory event.

Statistical Analysis
Descriptive statistics and paired analyses of PSG parameters before and after AT were performed. For comparison of demographic/baseline data in Table 1, continuous variables reported with means were compared using the Kruskal-Wallis test, and medians were compared using the nonparametric equality-of-medians test. For categorical variables, the Kruskal-Wallis test was used. Because the results were not normally distributed, the Wilcoxon sign-rank test was
used to evaluate for differences in PSG results before and after surgery. Regression was also used to evaluate the relationship between PSG parameters and tonsil and adenoid size and to examine the relationship when both tonsil and adenoid size were included in the model. Power analysis was used to detect differences in postoperative oAHI between different tonsil size groups. Age, race, BMI, and baseline OSA severity were controlled for in all multivariable regression analyses.

Values of $\leq 0.05$ were used to determine significance throughout the analysis. Data were analyzed using Stata Statistical Software: Release 10.1 (StataCorp LP, College Station, Texas).

Results

Seventy children (36 females) of 117 who underwent AT in this timeframe were included in the study. Their mean age was $5.9 \pm 4.4$ years, and 18 (26%) were younger than 3 years. Demographic information is presented in Table 1. Twenty children were classified as tonsil size 2+, 36 as 3+, and 14 as 4+. There were no significant differences among these 3 groups with regard to age ($P = 0.09$), sex ($P = 0.93$), race ($P = 0.31$), preoperative oAHI ($P = 0.62$), or BMI z score ($P = 0.62$). For the children with 2+ tonsils, 6 children had 2+ adenoids and a mean oAHI of $26.6 \pm 53.7$ events/h (median, 5.25), 12 had 3+ or 4+ adenoids with a mean oAHI of $10.4 \pm 10.7$ events/h (median, 7.05), and 2 had missing data for adenoid size with a mean of $22.5 \pm 3.5$ events/h (median, 22.5). Change in oAHI did not differ between those children with 2+ tonsils, regardless of adenoid size ($P = 0.55$).

Pre- and postoperative PSG results are presented in Figure 1 and Table 2. Overall, median oAHI improved from $11.8 \pm 21.7$ to $1.95 \pm 6.1$ events/h, with 40% (28 of 70) of the children having complete resolution. There was significant improvement in oAHI, obstructive AI, and saturation nadir across the 3 groups. Obstructive HI was

<table>
<thead>
<tr>
<th>Table 1. Patient Demographics.\textsuperscript{a}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total No. of patients</td>
</tr>
<tr>
<td>Age, mean $\pm$ SD, y</td>
</tr>
<tr>
<td>Age, median, y</td>
</tr>
<tr>
<td>Sex, No. (%)</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Race, No. (%)</td>
</tr>
<tr>
<td>White</td>
</tr>
<tr>
<td>Black</td>
</tr>
<tr>
<td>Hispanic</td>
</tr>
<tr>
<td>Other</td>
</tr>
<tr>
<td>BMI z score, mean $\pm$ SD</td>
</tr>
<tr>
<td>BMI z score, median</td>
</tr>
<tr>
<td>BMI z score, %, mean $\pm$ SD</td>
</tr>
<tr>
<td>BMI z score, %, median</td>
</tr>
<tr>
<td>Adenoid size</td>
</tr>
<tr>
<td>1+</td>
</tr>
<tr>
<td>2+</td>
</tr>
<tr>
<td>3+</td>
</tr>
<tr>
<td>4+</td>
</tr>
<tr>
<td>Not recorded</td>
</tr>
</tbody>
</table>

Abbreviation: BMI, body mass index.

\textsuperscript{a}For continuous variables, means for the 2+, 3+, and 4+ tonsil groups were compared using the Kruskal-Wallis test, and medians were compared using the nonparametric equality-of-medians test. For categorical variables, the Kruskal-Wallis test was used.

Figure 1. Median obstructive apnea-hypopnea index (oAHI) before and after adenotonsillectomy by tonsil size. All groups had a significant improvement in median oAHI as indicated by the asterisk.
also significantly improved in the 3+ and 4+ tonsil groups but did not reach statistical significance in the 2+ group ($P = .058$). Complete resolution occurred in 25% (5/20) of children in the 2+ tonsil group, 50% (18/36) in the 3+ tonsil group, and 36% (5/14) in the 4+ tonsil group. There was no correlation between OSA severity and either tonsil or adenoid size ($P > .32$). Regression analysis also revealed no significant relationship between tonsil and adenoid size when controlling for age, race, BMI z score, and baseline OSA severity.

**Power Analysis**

Power analysis to detect differences in the postoperative oAHI between groups found that with 20 children in the 2+ tonsil group and 50 in the combined 3+/4+ tonsil group, a difference in means of 2.91, and a standard deviation of 2.5, the power was greater than 0.95.

**Discussion**

In the present study, we found that oAHI, obstructive AI, and oxygen saturation nadir improved significantly, regardless of tonsil size. Despite improvement in these specific respiratory parameters, however, complete resolution (oAHI <1) was seen in only 25% to 50% of children, even in those with 4+ tonsils. We also found that tonsil size did not correlate with baseline OSA severity, even when adenoid hypertrophy was accounted for in the analysis.

Our results are consistent with previous literature evaluating the relationship between subjective tonsil size and OSA severity. A systematic review of this topic reported the results of 16 case series and 4 case-control studies with a mean sample size of 161 patients. The authors note that the 4 studies with the highest level of evidence reported no association between tonsil size and OSA severity. Overall, 9 of 20 studies found no association between tonsil size and OSA severity, whereas 11 studies (all case series) reported a positive relationship between the two.

The lack of a clear association between subjective tonsil size and OSA disease severity may be attributed to difficulty in accurately estimating tonsil size due to the position of the tonsil in the fossa (eg, endophytic tonsils) or difficulty in accurately estimating the anterior-posterior dimension of the tonsil. Patient-related factors such as poor cooperation may also play a role. Howard and Brietzke studied the relationship between objective tonsil size and AHI in 34 children, determining objective tonsil size by weight and volume measurements after surgical removal of the tonsils. They reported that although there was a strong correlation between subjective and objective tonsil size, only objective tonsil size was predictive of preoperative AHI. While the intent of this study was to further evaluate the resolution rate of OSA after AT, only 50% (17/34) of children underwent postoperative PSG. The authors noted concern for significant selection bias toward patients with persistent OSA, as those with residual symptoms were most likely to undergo repeat PSG. Of children who had postoperative PSG, those with larger tonsils were more likely to have a lower AHI; however, the study was underpowered to determine significance.

To further investigate the relationship between tonsil size and OSA resolution rates, we evaluated factors associated with persistent OSA in children, including age, race, and obesity. We found no significant differences in these factors between the 3 tonsil groups. Ulualp and Szmk also studied children with persistent OSA and small tonsils to investigate the likelihood that multilevel upper airway collapse could explain differences in resolution rates. Thus far, however, their work has not substantiated this hypothesis. These authors evaluated the role of lateral wall collapse in children with OSA, reporting that those with 3+ and 4+ tonsils were more likely to have lateral wall collapse than those with 1+ and 2+ tonsils.

Because our study was retrospective, it has the inherent limitations of all such studies. The first is likely selection bias; this is illustrated by the fact that none of the children undergoing AT in this study had tonsils classified as 1+. It would be ideal for future studies to include these children. We also did not do volumetric analysis of the tonsil or adenoid tissue since this is not part of our routine pathologic evaluation of tonsils; this may be a useful additional measure since objective tonsil size has been shown to have a

**Table 2. Polysomnographic Parameters before and after Adenotonsillectomy by Tonsil Size.**

<table>
<thead>
<tr>
<th></th>
<th>Over All</th>
<th>2+ Tonsil</th>
<th>3+ Tonsil</th>
<th>4+ Tonsil</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Preop</td>
<td>Postop</td>
<td>$P$ Value</td>
<td>Preop</td>
</tr>
<tr>
<td>oAHI$^b$</td>
<td>11.8 (21.7)</td>
<td>1.95 (6.1)</td>
<td>&lt;.001</td>
<td>6.2 (29.8)</td>
</tr>
<tr>
<td>Hi$^b$</td>
<td>7.8 (18.5)</td>
<td>1.5 (5.6)</td>
<td>&lt;.001</td>
<td>5.3 (26.3)</td>
</tr>
<tr>
<td>AI$^b$</td>
<td>1.1 (7.0)</td>
<td>0 (1.0)</td>
<td>&lt;.001</td>
<td>1.0 (6.8)</td>
</tr>
<tr>
<td>Sat Nadir$^c$</td>
<td>85 (8.2)</td>
<td>91 (5.0)</td>
<td>&lt;.001</td>
<td>86 (8.0)</td>
</tr>
</tbody>
</table>

Abbreviations: AI, apnea index; HI, hypopnea index; oAHI, obstructive apnea-hypopnea index; Preop, preoperative; Postop, postoperative; Sat Nadir, oxygen saturation nadir; SD, standard deviation.

$^a$All reported values are median (SD) values. The $P$ values represent a comparison between preoperative and postoperative results using the Wilcoxon rank-sum test.

$^b$Values are events/hour.

$^c$Values are percent saturation.
better correlation with AHI than subjective tonsil size. However, it was the intent of this study to understand the utility of the subjective tonsil grading system to predict resolution of OSA. In addition, only children with PSG-diagnosed OSA were included in this study. While we routinely obtained PSG for all children with SDB during the period of this study, 90% of children in the United States who undergo AT for SDB do so without the benefit of a formal diagnosis of OSA. Therefore, the true effect of tonsil excision on SDB requires further exploration. In addition, we used regression analysis to determine if adenoidectomy size might determine resolution of the PSG parameters, but our sample size of children with 2+ tonsils and 2+ adenoids was too small to do a subgroup analysis. This is an area that definitely deserves further study. Moreover, most children included in this study were black, which may limit the generalizability of our results. Last, children with an oAHI <5 and no symptoms did not routinely undergo a postoperative PSG and were not included in this study.

**Conclusion**

We found that children with small tonsils had significant clinical improvement in PSG respiratory parameters and overall low complete resolution rates. While a larger proportion of patients classified as 3+ and 4+ had complete resolution of OSA after surgery, significant improvement was seen in AI and saturation nadir even in those classified as 2+. Moreover, tonsil size did not correlate with OSA severity.

**Author Contributions**

Alice Tang, acquisition of data, analysis, draft manuscript, find approval; James R. Benke, acquisition of data, analysis, draft manuscript, find approval; Aliza P. Cohen, analysis, draft manuscript, find approval; Stacey L. Ishman, conception design, acquisition of data, analysis, draft manuscript, find approval.

**Disclosures**

**Competing interests:** None.

**Sponsorships:** None.

**Funding source:** None.

**References**