

## ORIGINAL ARTICLE

# Effects of Adenotonsillectomy on Parent-Reported Behavior in Children With Obstructive Sleep Apnea

Nina Hattiangadi Thomas, PhD<sup>1,2</sup>; Melissa S. Xanthopoulos, PhD<sup>1,4</sup>; Ji Young Kim, PhD<sup>3</sup>; Justine Shults, PhD<sup>3</sup>; Emma Escobar, MBA<sup>3</sup>; Bruno Giordani, PhD<sup>5</sup>; Elise Hodges, PhD<sup>5</sup>; Ronald D. Chervin, MD, MS<sup>6</sup>; Shalini Paruthi, MD<sup>7</sup>; Carol L. Rosen, MD<sup>8</sup>; Gerry H. Taylor, PhD<sup>8</sup>; Raanan Arens, MD<sup>9</sup>; Eliot S. Katz, MD<sup>10</sup>; Dean W. Beebe, PhD<sup>11</sup>; Susan Redline, MD<sup>12</sup>; Jerilynn Radcliffe, PhD<sup>2,13</sup>; Carole L. Marcus, MBBCh<sup>4,13</sup>

<sup>1</sup>Department of Child and Adolescent Psychiatry and Behavioral Sciences; <sup>2</sup>Neurobehavioral and <sup>3</sup>Biostatistical and Informatics Cores of the Clinical and Translational Research Center and <sup>4</sup>Sleep Center, Children's Hospital of Philadelphia, Philadelphia, PA; <sup>5</sup>Department of Psychiatry and Psychology, University of Michigan; <sup>6</sup>Department of Neurology and Sleep Disorders Center, University of Michigan; <sup>7</sup>Department of Pediatrics, Saint Louis University; <sup>8</sup>Department of Pediatrics, Rainbow Babies & Children's Hospital, Case Western Reserve University School of Medicine; <sup>9</sup>Department of Pediatrics, Montefiore Medical Center, Albert Einstein College of Medicine; <sup>10</sup>Division of Respiratory Diseases, Boston Children's Hospital; <sup>11</sup>Department of Pediatrics, Cincinnati Children's Hospital Medical Center; <sup>12</sup>Departments of Medicine, Brigham and Women's Hospital and Beth Israel Deaconess Medical Center, Harvard Medical School; <sup>13</sup>Children's Hospital of Philadelphia, University of Pennsylvania

**Objectives:** The childhood obstructive sleep apnea syndrome (OSAS) is associated with behavioral abnormalities. Studies on the effects of OSAS treatment on behavior are conflicting, with few studies using a randomized design. Further, studies may be confounded by the inclusion of behavioral outcome measures directly related to sleep. The objective of this study was to determine the effect of adenotonsillectomy on behavior in children with OSAS. We hypothesized that surgery would improve behavioral ratings, even when sleep symptom items were excluded from the analysis.

**Methods:** This was a secondary analysis of Child Behavior Checklist (CBCL) data, with and without exclusion of sleep-specific items, from the Childhood Adenotonsillectomy Trial (CHAT). CBCL was completed by caregivers of 380 children (7.0+1.4 [range 5–9] years) with OSAS randomized to early adenotonsillectomy (eAT) versus 7 months of watchful waiting with supportive care (WWSC).

**Results:** There was a high prevalence of behavioral problems at baseline; 16.6% of children had a Total Problems score in the clinically abnormal range. At follow-up, there were significant improvements in Total Problems ( $p < .001$ ), Internalizing Behaviors ( $p = .04$ ), Somatic Complaints ( $p = .01$ ), and Thought Problems ( $p = .01$ ) in eAT vs. WWSC participants. When specific sleep-related question items were removed from the analysis, eAT showed an overall improvement in Total ( $p = .02$ ) and Other ( $p = .01$ ) problems. Black children had less improvement in behavior following eAT than white children, but this difference attenuated when sleep-related items were excluded.

**Conclusions:** This large, randomized trial showed that adenotonsillectomy for OSAS improved parent-rated behavioral problems, even when sleep-specific behavioral issues were excluded from the analysis.

**Keywords:** obstructive sleep apnea, behavior, CBCL, CHAT.

## Statement of Significance

The effect of treatment of childhood obstructive sleep apnea syndrome on behavior remains controversial, and there have been few large studies with a randomized, controlled design. Many studies have utilized the Child Behavior Checklist parental report, but this survey includes many questions directly related to sleep. The current study analyzed behavioral data from the Childhood Adenotonsillectomy Trial (CHAT), a randomized controlled trial of surgery versus watchful waiting, in 380 school-aged children with obstructive sleep apnea. There was a high prevalence of behavioral problems at baseline, with significant improvement at 7 months in those randomized to surgery compared to watchful waiting, even when specific sleep-related questions were removed from the analysis.

## INTRODUCTION

The obstructive sleep apnea syndrome (OSAS) is common in children, with a prevalence of 1–4%.<sup>1,2</sup> Numerous studies have shown that childhood OSAS is associated with behavioral abnormalities<sup>3</sup> that may improve after treatment.<sup>4,5</sup> However, many studies have been limited by small sample size, heterogeneous study groups, or inadequate control for socioeconomic status and race, and no randomized controlled trials have been performed.

In contrast, the Childhood Adenotonsillectomy Trial (CHAT) was a multicenter, randomized controlled trial evaluating a large cohort of school-aged children randomized to either early adenotonsillectomy (eAT) or 7 months of Watchful Waiting with Supportive Care (WWSC).<sup>6</sup> The CHAT study showed few improvements in psychometric-measured neurocognitive outcomes in the eAT versus WWSC group but did show improvements in global measures of behavior, quality of life, OSAS symptoms, and polysomnographic measures.<sup>6,7</sup> Of note, CHAT found improvements in parent and teacher ratings on the Conners'

Rating Scale Revised: Long Version Global Index, which primarily evaluates attention, and the Behavior Rating Inventory of Executive Function (BRIEF), which evaluates executive functioning.<sup>6</sup> The current article presents new information from CHAT on changes in dimensions of child behavior, as measured by the parent-rated Child Behavior Checklist (CBCL),<sup>8</sup> which has been commonly used in past observational and open-label studies of OSAS.<sup>9</sup> We hypothesized that Internalizing, Externalizing, and Total Problems CBCL summary scores would improve in children with OSAS randomized to eAT vs WWSC. Secondary aims were to compare dimensions of behavior at baseline in children with OSAS to normative data; assess the relationship among behavior dimensions and polysomnographic parameters; and determine the effects of sex, race, and socioeconomic status on behavior in each study arm. We also aimed to determine the robustness of changes in behavioral and emotional symptoms when sleep-related items were eliminated from the raw data and whether rating changes from baseline to follow-up on individual sleep items differed by group.

## METHODS

Methodologic details of the CHAT trial have been published.<sup>10</sup> In brief, otherwise healthy school-aged children with OSAS were randomized to eAT (within 4 weeks) or WWSC. Evaluations were conducted at baseline and 7 months postbaseline, including polysomnography and standardized assessment of cognitive and behavioral functioning. This report is limited to those participants who completed both baseline and follow-up CBCL questionnaires.

The study was approved by the institutional review board at each site. Informed consent was obtained from parents/legal guardians and assent from children 7 years or older.

### Study Group

Inclusion criteria included age 5–9 years, and an obstructive apnea–hypopnea index (AHI) of 2–30/hr or obstructive apnea index of 1–20/hr on polysomnography; children with saturation <90% for ≥2% of total sleep time were excluded. Additional exclusion criteria included recurrent tonsillitis, extreme obesity (body mass index [BMI] z-score ≥3), severe chronic illness (other than asthma), and prescribed medication for attention-deficit/hyperactivity disorder.

Overweight was defined as BMI >85th percentile, and obesity as >95th percentile.<sup>11</sup> Maternal education was chosen as a marker of socioeconomic status rather than household income due to the large number of responders who declined to provide data on income (Table 1). Analysis of the effects of race/ethnicity was restricted to blacks versus whites due to the small number and heterogeneous nature of the other categories.

### Child Behavior Checklist

The Child Behavior Checklist (CBCL) is a caregiver-completed questionnaire designed to assess multiple aspects of behavioral, emotional, and adaptive functioning in children.<sup>8</sup> It is one of the most widely used measures of child behavior and has been used in more than 4000 publications, including multiple prior studies of the emotional and behavioral sequelae of OSAS.<sup>9</sup> However, none of these studies attempted to partition out the questions directly related to sleep. The CBCL for ages 6–18 years has 112 multiple-choice items, in which the caregiver is asked to provide ratings of “not true,” “somewhat or sometimes true,” or “very or often true” regarding the child’s functioning. In addition to summary scores of Internalizing, Externalizing, and Total Problems, scales related to specific internalizing symptoms (Anxious/Depressed, Withdrawn/Depressed, Somatic Complaints), externalizing symptoms (Rule-Breaking Behavior, Aggressive Behavior), and additional concerns (Social Problems, Thought Problems, Attention Problems) are generated based on age- and sex-specific norms (Figure 1). The mean value for the CBCL T-score is 50 with 1 *SD* being 10; higher scores indicate increased severity, with a summary score >63 being considered clinically abnormal.<sup>8</sup> An additional category of “Other Problems” is not normed but is included in the Total Problems scale, including items in which the caregiver is allowed to write in concerns, which are included in the ratings.

There are several age-specific versions of the CBCL. CHAT participants ranged from 5 to 9 years at study entry. As outlined in the CBCL manual for some circumstances,<sup>8</sup> the CBCL

questionnaire for ages 6–18 was administered to all participants for consistency of data.

### Statistical Analysis

Unless otherwise specified, data are shown as median (interquartile range [IQR]) for continuously measured variables as most data were not normally distributed, or as frequency (%) for categorical variables. Data were analyzed consistent with the intention-to-treat principle. The baseline CBCL abnormalities were compared to the expected prevalence using the exact binomial test. Group differences (eAT vs. WWSC) were examined using Wilcoxon rank-sum tests. Group differences in the changes in CBCL scores/scales from baseline to follow-up were evaluated by the Wilcoxon rank-sum tests for the primary analysis, as well as by linear regression of the changes on the group indicator, adjusting for each of the subgroup factors and baseline AHI and CBCL scores, as a secondary analysis. Mixed-effects logistic regression was used to compare groups based on the changes in the percentage of participants whose scores fell in the abnormal range. The associations between polysomnographic parameters and CBCL scales were examined by Pearson correlation.

The CBCL contains 7 items clearly related to sleep (although not necessarily associated with OSAS), located in the Somatic, Thought, and Other Problems scales. It also allows caregivers to write in and rate concerns for Other Problems and factors these items into the Total Problems score (Figure 1). To determine if the changes noted in composite scales for eAT versus WWSC were robust even in the absence of specific sleep items, raw data were analyzed after exclusion of the sleep-related items and of 2 sleep-related items that individual parents had entered in the write-in “Other” category (“grinds teeth when sleeping” and “sleeps a lot”). Norm-referenced scores do not exist for this, so raw score values for each composite scale at baseline and follow-up were calculated. The groups (eAT vs. WWSC) were also compared on the sleep-specific items at baseline and at follow-up using proportional odds ordinal logistic regression to better accommodate the discrete and sparse nature of the individual item scale (0, 1, and 2). To evaluate the effect of surgery on sleep-specific items, the groups were compared on changes in raw scores for each composite scale using the Wilcoxon rank-sum test, as described in the paragraph mentioned earlier for the original scores. The effect of age and sex on the group difference was examined using regression adjusting for age and sex.

Statistical analyses were performed using R version 3.2.4 (R Foundation for Statistical Computing, [www.R-project.org](http://www.R-project.org)). The value  $p < .05$  was considered significant.

## RESULTS

### Study Group

Details of enrollment have been published.<sup>6</sup> In brief, 196 children in the eAT group and 204 in the WWSC group completed the study. Of these, 187 (95%) in eAT (97% of whom received the assigned intervention) and 193 (95%) in WWSC (94% of whom received the assigned intervention) completed the CBCL questionnaire at both baseline and follow-up visits and comprised the final study cohort.

**Table 1**—Study group demographics and polysomnography results at baseline.

	eAT	WWSC
<i>N</i>	187	193
Age, years	6.8 (5.8, 8.0)	6.8 (5.8, 8.0)
Males	84 (45%)	104 (54%)
Overweight or obese, <i>N</i> (%)	92 (49%)	89 (46%)
Obese, <i>N</i> (%)	65 (35%)	65 (34%)
Race, <i>N</i> (%)		
Black	99 (53%)	103 (53%)
White	65 (35%)	73 (38%)
Other	23 (12%)	17 (9%)
Hispanic ethnicity, <i>N</i> (%)	12 (6%)	16 (8%)
Maternal education ≤ high school, <i>N</i> (%)	59 (32%)	62 (32%)
Annual household income		
< \$20 000	46 (30%)	55 (32%)
\$20 000–40 000	47 (30%)	41 (24%)
> \$40 000	63 (40%)	75 (44%)
Polysomnographic parameters		
Total sleep time, min	465.0 (432.0, 495.5)	465.0 (426.0, 494.0)
Sleep efficiency, %	88.1 (81.5, 92.7)	88.3 (81.7, 92.0)
N3 sleep (% total sleep time)	31.1 (26.2, 36.5)	31.6 (26.5, 35.1)
Rapid eye movement sleep (% total sleep time)	18.7 (16.3, 21.6)	18.2 (15.7, 20.7)
Arousal index, N/hr	8.1 (6.4, 10.3)	7.9 (6.0, 10.1)
Obstructive apnea hypopnea index, N/hr	4.8 (2.8, 8.8)	4.4 (2.5, 9.0)
S <sub>p</sub> O <sub>2</sub> nadir, %	89.5 (86, 92)	90.0 (87, 92)
% sleep time with S <sub>p</sub> O <sub>2</sub> < 90%	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)
Peak end-tidal CO <sub>2</sub> , mm Hg <sup>a</sup>	55 (52, 58)	54 (51, 56)
% total sleep time with end-tidal CO <sub>2</sub> > 50 mm Hg <sup>a</sup>	2.0 (0.2, 14.1)	0.6 (0.1, 5.2)

Data shown as median (IQR) or *N* (%). Note there are no more than 5 missing data points for each parameter except for household income (31 missing values in eAT group and 22 missing values in WWSC group) and CO<sub>2</sub> parameters (23 missing values in eAT group and 34 missing values in WWSC group). Abbreviations: eAT, early adenotonsillectomy; WWSC, watchful waiting with supportive care.

<sup>a</sup>*p* < .01

Demographic and baseline polysomnographic variables are provided in [Table 1](#). In this primarily urban population, almost half the children were overweight/obese, a majority were black, and approximately a third of the mothers had a high school education or less. There were no significant differences between eAT and WWSC for any demographic variables. No significant differences were found in polysomnographic parameters between groups except for a statistically significant but clinically small elevation in end-tidal CO<sub>2</sub> levels in the eAT group.

#### Baseline CBCL Data

CBCL data at baseline and follow-up are shown in [Table 2](#). There were no differences in summary scores at baseline between eAT and WWSC. Group median T-scores at baseline were within the

normal range. However, the percentage of children with behavior in the abnormal range was greater than the expected 9.7% (based on the probability of scoring ≥63 with a normal distribution with a mean of 50 and *SD* of 10) for Internalizing (15.8%, *p* < .001), Externalizing (14.7%, *p* = .002), and Total Problems (16.6%, *p* < .001) scores.

There was no significant correlation between polysomnographic markers of OSAS severity (including AHI, SpO<sub>2</sub> nadir, percentage of sleep time with SpO<sub>2</sub> < 90%, peak end-tidal CO<sub>2</sub> and percentage of sleep time with CO<sub>2</sub> > 50 mmHg) and Internalizing, Externalizing, or Total Problems T-scores.

#### Changes at 7 Months

At follow-up, Total Problems (*p* < .001) and Internalizing (*p* = .04) CBCL T-scores decreased more in eAT than WWSC

CBCL/6-18 - Syndrome Scale Scores for Girls 6-11 Scored Using T Scores for United States

T S C O R E	Internalizing				Externalizing				C L I N I C A L  N O R M A L
	Anxious/ Depressed	Withdrawn/ Depressed	Somatic Complaints	Social Problems	Thought Problems	Attention Problems	Rule-Breaking Behavior	Aggressive Behavior	
100									
95									
90									
85									
80									
75									
70									
65									
60									
55									
≤ 50									
<b>Total Score</b>	<b>2</b>	<b>10</b>	<b>1</b>	<b>5</b>	<b>5</b>	<b>7</b>	<b>1</b>	<b>11</b>	
<b>T Score</b>	<b>51</b>	<b>80-C</b>	<b>53</b>	<b>59</b>	<b>66-B</b>	<b>62</b>	<b>52</b>	<b>63</b>	
<b>Percentile</b>	<b>54</b>	<b>&gt;97</b>	<b>62</b>	<b>81</b>	<b>95</b>	<b>89</b>	<b>58</b>	<b>90</b>	
0	14.Cries	5.EnjoysLittle	0 47.Nightmares	1 11.Dependent	0 9.MindOff	1 1.ActsYoung	0 2.Alcohol	0 3.Argues	
0	29.Fears	1 42.PreferAlone	0 49.Constipate	1 12.Lonely	0 18.HarmSelf C	2 4.FailsToFinish	0 26.NoGuilt	0 16.Mean	
0	30.FearSchool	1 65.WonTalk	0 51.Dizzy	0 25.NotGetAlong	0 40.HearsThings C	1 8.Concentrate	0 28.BreaksRules	1 19.DemAtten	
0	31.FearDoBad	0 69.Secretive	1 54.Tired	1 27.Jealous	0 46.Twitch	0 10.SitStill	0 39.BadFriends	1 20.DestroyOwn	
0	32.Perfect	2 75.Shy	0 56a.Aches	0 34.OurToGet	0 58.PicksSkin	0 13.Confused	0 43.LieCheat	0 21.DestroyOther	
0	33.Unloved	1 102.LacksEnergy	0 56b.Headaches	0 36.Accidents	0 59.SexPartsP C	1 17.Daydream	1 63.PreferOlder	0 22.DishHome	
0	35.Worthless	2 103.Sad	0 56c.Nausea	0 38.Teased	0 60.SexPartsM	0 41.Impulsive	0 67.RunAway c	0 23.DishSchool	
0	45.Nervous	2 111.Withdrawn	0 56d.EyeProb	0 48.NotLiked	0 66.RepeatsActs	1 61.PoorSchool	0 72.SetsFires c	0 37.Fights	
0	50.Fearful		0 56e.SkinProb	1 62.Clumsy	0 70.SeesThings C	0 78.Inattentive	0 73.SexProbs	0 57.Attacks C	
0	52.Guilty		0 56f.Stomach	1 64.PreferYoung	2 76.SleepsLess	1 80.Stares	0 81.StealsHome	0 68.Screams	
1	71.SelfConsc		0 56g.Vomit	0 79.SpeechProb	0 83.StoresUp	0 82.StealsOut	0 82.StealsOut	2 86.Stubborn	
0	91.TalkSuicide C				0 84.StrangeBehv	0 90.Swears	0 96.ThinksSex	2 87.MoodChange	
1	112.Worries				0 85.StrangeIdeas	0 99.Tobacco	0 99.Tobacco	1 88.Sulks	
					1 92.SleepWalk	0 101.Truant	0 105.UsesDrugs c	1 89.Suspicious	
					2 100.SleepProblem	0 106.Vandalism	0 106.Vandalism	1 94.Teases	
								2 95.Temper	
								0 97.Threaten	
								0 104.Loud	

Copyright 2007 T.M. Achenbach

B = Borderline clinical range; C = Clinical range

Broken lines = Borderline clinical range

Other Problems	
0	6.BMOOut C
0	7.Brags
1	15.CruelAnimal C
0	24.NotEat
0	44.BiteNail
0	53.Overeat
0	55.Overweight
0	56h.OtherPhys
0	74.ShowOff
0	77.SleepsMore
0	93.TalkMuch
1	98.ThumbSuck
0	107.WetsSelf C
0	108.WetsBed
0	109.Whining
0	110.WishOppSex
2	113.OtherProb

c = critical item

Figure 1—The Child Behavior Checklist (CBCL) scales with sleep questions highlighted. Copyright T.M. Achenbach. Reproduced by permission.

**Table 2**—Changes in CBCL in eAT compared to WWSC.

CBCL summary scores	eAT		WWSC		<i>p</i> value for comparison of change between eAT and WWSC
	Baseline	Change from baseline to 7 months	Baseline	Change from baseline to 7 months	
Full CBCL (T-scores)					
<i>N</i>	187	187	193	193	
<i>Total Problems</i>	52 (46, 59)	-4 (-8, 1)	53 (46, 62)	-1 (-6, 4)	<0.001
<i>Internalizing</i>	50 (43, 59)	-3 (-9, 2)	52 (45, 61)	-1 (-6, 6)	0.04
<i>Externalizing</i>	51 (44, 58)	-2 (-7, 2)	51 (44, 60)	0 (-6, 3)	0.19
Scale Scores					
Anxious/Depressed	51 (50, 54)	0 (-1, 0)	51 (50, 57)	0 (-2, 1)	0.39
Withdrawn/Depressed	52 (50, 56)	0 (-4, 0)	52 (50, 58)	0 (-4, 0)	0.76
<i>Somatic Complaints</i>	53 (50, 64)	0 (-4, 0)	57 (50, 64)	0 (-3, 3)	0.01
Social Problems	53 (51, 58)	0 (-3, 1)	53 (51, 58)	0 (-3, 1)	0.67
<i>Thought Problems</i>	54 (51, 61)	-1 (-7, 0)	54 (51, 62)	0 (-4, 1)	0.01
Attention Problems	53 (51, 59)	0 (-3, 0)	53 (51, 61)	0 (-3, 2)	0.08
Rule-Breaking Behavior	52 (50, 58)	0 (-3, 0)	53 (51, 59)	0 (-3, 1)	0.49
Aggressive Behavior				0 (-3, 1)	0.75
CBCL with sleep-specific items excluded (raw scores) <sup>b</sup>					
<i>N</i>	184	184	192	192	
<i>Total</i>	21 (12, 34)	-4 (-12, 2)	22 (12, 39)	-1 (-9, 6)	0.02
Internalizing	3 (2, 8)	-1 (-3, 1)	4 (2, 9)	0 (-3, 2)	0.22
Externalizing	5 (2, 10)	-1 (-3, 1)	5 (2, 12)	0 (-3, 2)	0.24
Scale Scores					
Anxious/depressed	2 (0, 4)	0 (-1, 1)	2 (0, 4)	0 (-1, 1)	0.46
Withdrawn/depressed	1 (0, 2)	0 (-1, 0)	1 (0, 2)	0 (-1, 0)	0.48
Somatic complaints	1 (0, 3)	0 (-1, 0)	1 (0, 3)	0 (-1, 1)	0.13
Social problems	2 (1, 5)	0 (-2, 1)	2 (1, 4)	0 (-1, 1)	0.37
Thought problems	1 (0, 2)	0 (-1, 0)	1 (0, 2)	0 (-1, 1)	0.33
<i>Attention problems</i>	4 (1, 6)	-1 (-2, 0)	4 (1, 7)	0 (-2, 2)	0.04
Rule-breaking behavior	1 (0, 3)	0 (-1, 0)	2 (1, 3)	0 (-1, 1)	0.76
Aggressive behavior	4 (2, 8)	-1 (-3, 1)	4 (1, 8)	0 (-2, 1)	0.32
<i>Other Problems</i> <sup>c</sup>	4 (2, 6)	-1 (-2, 1)	4 (2, 6)	0 (-1, 1)	0.01

Abbreviations: eAT, early adenotonsillectomy; WWSC, watchful waiting with supportive care.

<sup>a</sup>Data shown as median (IQR). *p* values are from Wilcoxon rank-sum tests for the group. The *p* values refer to the difference in the changes from the baseline to 7 months, with significant parameters shown in italics. There were no significant differences in summary scores at baseline between the eAT and WWSC groups.

<sup>b</sup>Sleep-specific items were excluded from the Total, Internalizing, Somatic, Thought and Other scores; however, raw scores from all scales are provided for the sake of completeness. Raw data were unavailable for 4 participants (see text).

<sup>c</sup>Data for Other Problems are not provided for the summary scores as norm-referenced scores are unavailable (see text).

(Table 2). The eAT group also showed a significantly greater decrease in Somatic Complaints and Thought Problems.

A sensitivity analysis was performed for the primary analysis using a per protocol analysis (i.e., reassigning participants) based on the actual treatment that participants received, rather than the arm to which they were assigned. There were 6

children randomized to eAT who did not undergo surgery and 12 children randomized to WWSC who underwent eAT. This analysis showed qualitatively similar results to the intent-to-treat analyses but with more significant *p* values, with a larger decrease in the Internalizing (*p* = .004) and Total Problems (*p* < .001) CBCL T-scores in eAT compared to WWSC group

as well as a newly apparent decrease in the Externalizing score ( $p = .01$ ).

The only significant group difference in change across follow-up in the percentage of participants scoring in the abnormal range was in the Somatic Complaints scale. The odds of falling in the abnormal range for somatic complaints decreased 98% for the eAT group between baseline and follow-up, whereas the odds decreased 68% for the WWSC group ( $p = .01$  for the group difference).

### Analysis of the Contribution of Sleep-Specific Items

Raw data for both baseline and follow-up were available for all but four children in the original sample (184/187 eAT and 192/193 WWSC). The effect of age and sex including main effect terms and interaction terms were not significant after removing sleep items and were therefore removed from the final model. The only difference between groups at baseline was overtiredness ( $p = .032$ ), which was higher in eAT. At follow-up, however, four of the other symptoms (sleeps less [ $p = .042$ ], trouble sleeping [ $p = .004$ ], sleeps more [ $p = .012$ ], and talks/walks in sleep [ $P = .006$ ]) were significantly more improved in eAT versus WWSC (Table 3). “Overtired” remained significantly more improved in eAT versus WWSC after adjusting for baseline differences ( $p = .010$ ).

Even when sleep-specific items were omitted from the composite scores, children in eAT had significantly greater reductions in Total Problems from baseline to follow-up than WWSC (Table 2). Changes in Internalizing score, and Somatic

Complaints and Thought Problems scales, previously found to be significant when sleep items were included, were no longer significant, suggesting that sleep concerns calculated into these scales may have been driving this finding in the initial analysis. The composite of Other Problems, even with sleep items removed, was significantly improved in eAT versus WWSC at follow-up (this scale was not examined in the prior analysis because norm-referenced scores are unavailable).

### Subgroup Analyses

Models were fitted with interaction terms to evaluate the effects of sex, obesity, maternal education status, and race on CBCL summary T-score changes at 7 months. Analyses were adjusted for differences in baseline CBCL scores and AHI. There was no effect of sex, obesity or maternal education on changes in T-scores for Total Problems, Internalizing or Externalizing, with and without the sleep-specific items. Although there were no significant differences in CBCL scores between black and white children at baseline, black children randomized to eAT were significantly less likely, relative to WWSC to show an improvement in T-scores for Internalizing ( $p = .033$ ), Externalizing ( $p = .006$ ) or Total Problems ( $p < .001$ ) compared to white children. Specifically, the mean reduction in the Total Problems score for the eAT relative to WWSC group was estimated to be 6.70 points for white children but only 0.44 points for black children. Similarly, the mean reduction in the Internalizing score for eAT participants relative to WWSC participants was estimated to be 4.91 for white children compared to 1.07 points

Table 3—Sleep item frequencies<sup>a</sup>.

Item	Response	Baseline			Follow-up			p value for change
		Not true	Somewhat/sometimes true	Very/often true	Not true	Somewhat/sometimes true	Very/often true	
54. Overtired	eAT	129 (70.1%)	41 (22.3%)	14 (7.6%)	162 (88.0%)	20 (10.9%)	2 (1.1%)	.01 <sup>b</sup>
	WWSC	153 (79.7%)	30 (15.6%)	9 (4.7%)	156 (81.3%)	29 (15.1%)	7 (3.7%)	
76. Sleeps less	eAT	139 (75.5%)	34 (18.5%)	11 (6.0%)	160 (87.0%)	22 (12.0%)	2 (1.1%)	.41
	WWSC	137 (71.4%)	37 (19.3%)	18 (9.4%)	153 (79.7%)	26 (13.5%)	13 (6.8%)	
100. Trouble sleeping	eAT	108 (58.7%)	48 (26.1%)	28 (15.2%)	152 (82.6%)	25 (13.6%)	7 (3.8%)	.06
	WWSC	112 (58.3%)	43 (22.4%)	37 (19.3%)	135 (70.3%)	37 (19.3%)	20 (10.4%)	
77. Sleeps more	eAT	144 (78.3%)	35 (19.0%)	5 (2.7%)	166 (90.2%)	15 (8.2%)	3 (1.6%)	.19
	WWSC	147 (76.6%)	32 (16.7%)	13 (6.8%)	156 (81.3%)	24 (12.5%)	12 (6.3%)	
108. Wets the bed	eAT	130 (70.7%)	30 (16.3%)	24 (13.0%)	148 (80.4%)	20 (10.9%)	16 (8.7%)	.63
	WWSC	129 (67.2%)	39 (20.3%)	24 (12.5%)	140 (72.9%)	38 (19.8%)	14 (7.3%)	
47. Nightmares	eAT	118 (64.1%)	60 (32.6%)	6 (3.3%)	133 (72.3%)	46 (25.0%)	5 (2.7%)	.57
	WWSC	116 (60.4%)	64 (33.3%)	12 (6.3%)	121 (63.0%)	66 (34.4%)	5 (2.6%)	
92. Talks/walks in sleep	eAT	121 (65.8%)	54 (29.4%)	9 (4.9%)	146 (79.4%)	35 (19.0%)	3 (1.6%)	.01
	WWSC	128 (66.7%)	51 (26.56%)	13 (6.77%)	129 (67.19%)	52 (27.08%)	11 (5.73%)	

Abbreviations: eAT, early adenotonsillectomy; WWSC, watchful waiting with supportive care.

<sup>a</sup>Data shown as N (%). Significant parameters are shown in italics.

<sup>b</sup>Adjusted for baseline.

for black children. For the Externalizing score, the mean reduction for eAT participants relative to the WWSC group was estimated to be 3.90 points for white children, compared to a mean increase of 0.60 points for black children. Because black children were less likely than white children to have normalization of their polysomnogram at 7 months, we further evaluated the effect of race on the summary T-scores by adjusting for the change in AHI. This did not substantially change the results, with Black children randomized to eAT versus WWSC still significantly less likely to show an improvement in Total Problems ( $p < .001$ ), Internalizing ( $p = .034$ ), and Externalizing ( $p = .005$ ) T-scores compared to white children. However, these race effects were no longer significant after omitting the sleep-specific items.

## DISCUSSION

To our knowledge, the CHAT study was the first to rigorously evaluate changes in behavior in response to treatment of childhood OSAS using a randomized controlled study design and a large study sample. The CBCL analysis confirmed an elevated prevalence of behavioral problems in children with OSAS at baseline, across multiple dimensions. At follow-up, there was a highly significant improvement in Total Problems, Internalizing behaviors, Somatic Complaints, and Thought Problems in children randomized to surgery as compared to WWSC. When specific sleep-related questions were removed from the analysis, the eAT group continued to show overall improvement in Total and Other Problems compared to WWSC.

Behavioral abnormalities were described in the earliest reports of childhood OSAS.<sup>12</sup> However, systematic studies have not always demonstrated behavioral abnormalities. In a comprehensive review, Kohler *et al* found that less than half of 22 studies reported an increased prevalence of behavioral problems in children with OSAS.<sup>13</sup> For example, O'Brien found no difference in behavior, assessed with the CBCL, between 35 children with OSAS compared to controls.<sup>14</sup> Similarly, although many studies have shown improvements in behavior after adenotonsillectomy,<sup>13</sup> others have not.<sup>15,16</sup> Variable types of behavioral improvements have been reported after surgery even among those studies documenting positive changes.<sup>13</sup> Limitations of these studies include small sample sizes and lack of randomization.

Although the CBCL is commonly used to assess behavior in children with OSAS,<sup>9,13</sup> it includes several questions directly related to sleep (Figure 1). In the current study, the Somatic and Thought scores improved the most after surgery. Reanalysis of these scores with sleep items removed indicated that sleep questions were likely contributing to group differences. However, the finding in the eAT group of greater improvement in Total Problems at follow-up remained robust, and additional significant differences in Other Problems emerged when raw score change was examined with sleep items excluded. In addition, sensitivity analyses using the treatment that participants actually received showed a significant improvement in the Externalizing score, which does not include any sleep questions. Finally, the improvements in behavior in the eAT group as shown by the CBCL mirror the previously reported changes in other dimensions of neurobehavioral functioning in the eAT group, such as

the Conners and BRIEF.<sup>6</sup> The Conners includes only one question related to sleep ("seems tired or slowed down all the time") and the BRIEF has none.

There were no significant correlations between key polysomnographic parameters and behavioral outcomes. Although counterintuitive, this is not surprising as previous studies in both adults<sup>17</sup> and children<sup>18–20</sup> have shown a poor relation between polysomnographic parameters and neurocognitive/behavioral outcomes. This may be because of the complex interplay of the many polysomnographic variables and multiple other factors that affect behavior, or it may be that the effect of OSAS on behavior reflects a threshold rather than a linear dosing effect.

Black children in the eAT group showed less improvement in behavior than white children. Once sleep-specific questions were excluded, there were no longer differences among racial groups. This suggests that black children were either less likely to have improvements in sleep following surgery (even when adjusted for differing levels of AHI) or that the effects of surgery on sleep were obscured by race differences in other factors such as sleep schedules and practices.<sup>21</sup>

Strengths of this study include the large cohort, rigorous polysomnographic methods, and a randomized controlled trial design. A limitation is the subjective and unblinded nature of any caregiver-administered survey such as the CBCL, although the CBCL has been well validated and several findings persisted even after removal of the sleep items most likely to be biased by lack of blinding. Moreover, the subjective experience of families with regard to behavioral outcomes may reflect more clearly than objective testing the issues of prime importance to parents who confront decisions about adenotonsillectomy. Teacher ratings were not obtained in order to alleviate teacher burden, as the teachers were requested to fill out other study surveys. CHAT was limited to school-aged children with OSAS without prolonged desaturation; thus, the study findings should not be extrapolated to younger children or those with more severe disease.

In summary, this large randomized trial among of school-aged children with OSAS has shown an elevated prevalence of behavioral problems across multiple dimensions. Children randomized to adenotonsillectomy showed greater improvement in both sleep-specific and more general behavioral problems than those randomized to watchful waiting. Further randomized controlled studies in younger children are warranted, as these children may be at increased risk for behavioral problems resulting from OSAS.

## REFERENCES

1. Lumeng JC, Chervin RD. Epidemiology of pediatric obstructive sleep apnea. *Proc Am Thorac Soc.* 2008; 5(2): 242–252.
2. Spilsbury JC, Storfer-Isser A, Rosen CL, Redline S. Remission and incidence of obstructive sleep apnea from middle childhood to late adolescence. *Sleep.* 2015; 38(1): 23–29.
3. Marcus CL, Brooks LJ, Draper KA, et al.; American Academy of Pediatrics. Diagnosis and management of childhood obstructive sleep apnea syndrome. *Pediatrics.* 2012; 130(3): e714–e755.
4. Marcus CL, Radcliffe J, Konstantinopoulou S, et al. Effects of positive airway pressure therapy on neurobehavioral outcomes in children with obstructive sleep apnea. *Am J Respir Crit Care Med.* 2012; 185(9): 998–1003.

5. Goldstein NA, Fatima M, Campbell TF, Rosenfeld RM. Child behavior and quality of life before and after tonsillectomy and adenoidectomy. *Arch Otolaryngol Head Neck Surg.* 2002; 128(7): 770–775.
6. Marcus CL, Moore RH, Rosen CL, et al.; Childhood Adenotonsillectomy Trial (CHAT). A randomized trial of adenotonsillectomy for childhood sleep apnea. *N Engl J Med.* 2013; 368(25): 2366–2376.
7. Taylor HG, Bowen SR, Beebe DW, et al. Cognitive effects of adenotonsillectomy for obstructive sleep apnea. *Pediatrics.* 2016; 138(2): pii: e20154458.
8. Achenbach TM, Rescorla LA. *Manual for the ASEBA school-age forms & profiles.* Burlington, VT: University of Vermont, Research Center for Children, Youth & Families; 2001.
9. Beebe DW. Neurobehavioral morbidity associated with disordered breathing during sleep in children: a comprehensive review. *Sleep.* 2006; 29(9): 1115–1134.
10. Redline S, Amin R, Beebe D, et al. The Childhood Adenotonsillectomy Trial (CHAT): rationale, design, and challenges of a randomized controlled trial evaluating a standard surgical procedure in a pediatric population. *Sleep.* 2011; 34(11): 1509–1517.
11. Himes JH, Dietz WH. Guidelines for overweight in adolescent preventive services: recommendations from an expert committee. The Expert Committee on Clinical Guidelines for Overweight in Adolescent Preventive Services. *Am J Clin Nutr.* 1994; 59(2): 307–316.
12. Guilleminault C, Eldridge FL, Simmons FB, Dement WC. Sleep apnea in eight children. *Pediatrics.* 1976; 58(1): 23–30.
13. Kohler MJ, Lushington K, Kennedy JD. Neurocognitive performance and behavior before and after treatment for sleep-disordered breathing in children. *Nat Sci Sleep.* 2010; 2: 159–185.
14. O'Brien LM, Mervis CB, Holbrook CR, et al. Neurobehavioral correlates of sleep-disordered breathing in children. *J Sleep Res.* 2004; 13(2): 165–172.
15. Constantin E, Kermack A, Nixon GM, Tidmarsh L, Ducharme FM, Brouillette RT. Adenotonsillectomy improves sleep, breathing, and quality of life but not behavior. *J Pediatr.* 2007; 150(5): 540–546, 546.e1.
16. Biggs SN, Vlahandonis A, Anderson V, et al. Long-term changes in neurocognition and behavior following treatment of sleep disordered breathing in school-aged children. *Sleep.* 2014; 37(1): 77–84.
17. Kingshott RN, Engleman HM, Deary IJ, Douglas NJ. Does arousal frequency predict daytime function? *Eur Respir J.* 1998; 12(6): 1264–1270.
18. Chervin RD, Ruzicka DL, Giordani BJ, et al. Sleep-disordered breathing, behavior, and cognition in children before and after adenotonsillectomy. *Pediatrics.* 2006; 117(4): e769–e778.
19. Melendres MC, Lutz JM, Rubin ED, Marcus CL. Daytime sleepiness and hyperactivity in children with suspected sleep-disordered breathing. *Pediatrics.* 2004; 114(3): 768–775.
20. O'Brien LM, Holbrook CR, Mervis CB, et al. Sleep and neurobehavioral characteristics of 5- to 7-year-old children with parentally reported symptoms of attention-deficit/hyperactivity disorder. *Pediatrics.* 2003; 111(3): 554–563.
21. Crosby B, LeBourgeois MK, Harsh J. Racial differences in reported napping and nocturnal sleep in 2- to 8-year-old children. *Pediatrics.* 2005; 115(1 Suppl): 225–232.

## FUNDING

Funded by National Institutes of Health grants HL083075, HL083129 and UL1 TR000003.

## ACKNOWLEDGMENT

We thank the children and their families for participating in the study. We thank all the members of the CHAT research team for their assistance, and Nicholas Ambrulavage for assistance with data entry.

## SUBMISSION & CORRESPONDENCE INFORMATION

Submitted for publication November, 2016

Submitted in final revised form December, 2016

Accepted for publication January, 2017

Address correspondence to: Carole L. Marcus, MBCh, Children's Hospital of Philadelphia Sleep Center, Suite 9NW50, Main Hospital 34<sup>th</sup> Street & Civic Center Boulevard Philadelphia, PA 19104, USA. Telephone: 267-426-5842; Fax: 267-426-9234; Email: [marcus@email.chop.edu](mailto:marcus@email.chop.edu)

## TRIAL REGISTRATION

ClinicalTrials.gov number NCT00560859.

## DISCLOSURE STATEMENT

Drs. Thomas, Xanthopoulos, Kim, Shults, Giordani, Hodges, Paruthi, Taylor, Arens, Katz, Beebe, Redline, Marcus and Ms. Escobar have no conflicts of interest. Dr. Chervin is president of the American Academy of Sleep Medicine, has intellectual property rights and royalties in Zansors and royalties in Uptodate and Cambridge University Press; not related to the current article. Dr. Rosen is involved with investigational drugs/devices with Jazz Pharmaceutical and Flamel and is a consultant for Jazz Pharmaceutical and Advance-Medical; not related to the current article. Dr. Radcliffe is on the Board of Directors for the Pennsylvania Association for Infant Mental Health; not related to the current article.