http://dx.doi.org/10.5665/sleep.1388

The Childhood Adenotonsillectomy Trial (CHAT): Rationale, Design, and Challenges of a Randomized Controlled Trial Evaluating a Standard Surgical Procedure in a Pediatric Population

Susan Redline, MD, MPH[1]; Raouf Amin, MD[2]; Dean Beebe, PhD[2]; Ronald D. Chervin, MD, MS[3]; Susan L. Garetz, MD[4]; Bruno Giordani, PhD[5]; Carole L. Marcus, MBBCh[6]; Reneé H. Moore, PhD[7]; Carol L. Rosen, MD[8]; Raanan Arens, MD[9]; David Gozal, MD[10]; Eliot S. Katz, MD[11]; Ronald B. Mitchell, MD, PhD[12]; Hiren Muzumdar, MD[9]; H.G. Taylor, PhD[8]; Nina Thomas, PhD[13]; Susan Ellenberg, PhD[7]

[1]Division of Sleep Medicine, Harvard Medical School, Brigham and Women's Hospital and Beth Israel Medical Center, Boston, MA; [2]Children's Hospital Medical Center, Cincinnati, OH; [3]Sleep Disorders Center and Department of Neurology, University of Michigan, Ann Arbor, MI; [4]Department of Otolaryngology/Head and Neck Surgery, University of Michigan Health System, Ann Arbor, MI; [5]Neuropsychology Section, Departments of Psychiatry, Neurology, and Psychology, University of Michigan, Ann Arbor, MI; [6]Sleep Center, Children's Hospital of Philadelphia, University of Pennsylvania, Philadelphia, PA; [7]Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, PA; [8]Department of Pediatrics, Case Western Reserve University School of Medicine, University Hospitals-Case Medical Center, Rainbow Babies and Children's Hospital, Cleveland, OH; [9]Children's Hospital at Montefiore, Albert Einstein College of Medicine, Bronx, NY; [10]Department of Pediatrics, Comer Children's Hospital, Pritzker School of Medicine, The University of Chicago, Chicago, IL; [11]Division of Respiratory Diseases, Harvard Medical School, Children's Hospital, Boston, Boston, MA; [12]Pediatric Otolaryngology, Saint Louis University School of Medicine, Cardinal Glennon Children's Medical Center, St Louis, MO; [13]Neuropsychology & Assessment, Department of Child and Adolescent Psychiatry and Behavioral Science, Children's Hospital of Philadelphia, Philadelphia, PA

Each year, over 500,000 adenotonsillectomies (AT), mostly for the treatment of pediatric obstructive sleep apnea (OSA) are performed in the US in children under 15 years of age. No definitive study, however, has been yet conducted that has rigorously evaluated the effectiveness of AT for not only improving sleep disordered breathing, but also for improving clinically relevant outcomes, such as neurocognitive function, behavior, and quality of life. The Childhood Adenotonsillectomy Trial (CHAT) was designed to assess neuropsychological and health outcomes in children randomized to receive early AT (eAT) as compared to Watchful Waiting with Supportive Care (WWSC). Important secondary goals of the study are to evaluate outcomes in subgroups defined by obesity and race. This paper addresses key elements in the design and implementation of a controlled trial for a widely used "standard practice" surgical intervention in a pediatric population, that include establishment of standardized data collection procedures across sites for a wide variety of data types, establishment of equipoise, and approaches for minimizing unblinding of selected key personnel. The study framework that was established should provide a useful template for other pediatric controlled studies or other studies that evaluate surgical interventions.

Keywords: Clinical trial, adenotonsillectomy, sleep apnea, pediatrics

Citation: Redline S; Amin R; Beebe D; Chervin RD; Garetz SL; Giordani B; Marcus CL; Moore RH; Rosen CL; Arens R; Gozal D; Katz ES; Mitchell RB; Muzumdar H; Taylor HG; Thomas N; Ellenberg S. 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.

INTRODUCTION

Pediatric obstructive sleep apnea (OSA) affects between 2% to 3% of children, with rates two- to four-fold higher in certain subgroups, such as African American children and children from families of low socioeconomic status.[1,2] The disorder is characterized by increased upper airway resistance, associated with narrowing and intermittent pharyngeal collapse leading to snoring and periods of apnea and hypopnea. Periodic upper airway obstruction often results in intermittent hypoxemia, hypercapnia, and sleep disruption. A wide range of adverse health outcomes has been associated with untreated OSA, including cognitive deficits, behavioral problems (inattention, hyperactivity, aggression, conduct problems), mood impairments, excessive daytime sleepiness, impaired school performance, and poor quality of life. Compared to unaffected children or those who have snoring without OSA, children with OSA have been reported to have higher levels of blood pressure, C-reactive protein (CRP), insulin resistance, and left ventricular mass, and thus may be at increased risk for chronic cardiovascular and metabolic morbidity. OSA also has been associated with failure to thrive in young children, enuresis, and with overall increased health-related costs.

A number of risk factors likely influence airway patency during sleep, and thus the propensity for OSA. These include anatomic characteristics of the nose and throat and neuromuscular factors that modulate the tone and responsiveness of the airway muscles. Adenotonsillar enlargement is the most commonly recognized anatomic risk factor for pediatric OSA. In the US, standard practice usually involves adenotonsillectomy (AT) as the primary treatment for childhood OSA. Over 500,000 ATs are performed each year for OSA in the US,[3] resulting in substantial health care expenditures and exposure to the risks of surgery and anesthesia for large numbers of children. Despite the high frequency of AT and its use as first-line treatment for pediatric OSA, information on its effectiveness is limited. To date, evidence of the utility of AT for OSA has been evaluated only in small to

Submitted for publication February, 2011
Submitted in final revised form April, 2011
Accepted for publication April, 2011

Address correspondence to: Susan Redline, MD, MPH, Harvard Medical School, Department of Medicine and Division of Sleep Medicine, Brigham and Women's Hospital and Beth Israel Deaconess Medical Center, 221 Longwood Avenue, Room 225, Boston, MA 02115; Tel: (617) 732-4013; Fax: (617) 732-4015; E-mail: sredline1@partners.org

modestly sized and uncontrolled or not randomized studies.[4-9] Although the existing data do suggest that OSA improves after AT in the majority of children, residual OSA may occur in 20% to 40% of children from various clinical settings.[8,10] Further, it is not clear how much of the observed improvements are due to extraneous factors such as regression to the mean, growth of the child, or other confounders. Data indicate that a large proportion of children recently referred for AT is overweight or obese.[11] Some, although not all,[8] studies have indicated that obesity is associated with poorer response to surgery, with residual OSA occurring in as many as 75% of obese children.[9,10] Thus, rigorous assessment is needed of the role for AT in contemporary practice where surgery is often performed on overweight or obese children.

Data regarding changes in function and health status following surgery are even more limited. Few studies have evaluated changes in perceptions of sleep quality or daytime functioning after AT. Although several studies have shown improved sleep quality, snoring, and daytime fatigue in approximately 70% to 80% of children one year following AT,[12,13] the studies were uncontrolled. Results of uncontrolled studies also suggest that treatment of childhood OSA with AT may result in improvements in learning, aggression, and hyperactivity.[14-16] A US non-randomized study showed marked improvements in academic performance after AT among children who initially scored academically in the 10th percentile for academic performance,[17] suggesting a potentially important role of OSA in influencing cognitive function and the possibility that AT may improve academic performance. However, these results were subject to potential confounding due to regression to the mean (children selected for low performance will improve as a group on follow-up testing regardless of intervention) and the impacts of enhanced expectations and increased maturity, as well as a selection bias because families who pursued more aggressive treatment may differ in important ways from families who did not seek treatment for their children. A non-randomized study that compared children who had clinically indicated AT to those who had unrelated surgical or medical care demonstrated substantially improved objective measures of cognition, parent-rated behavior, objectively assessed sleepiness, and psychiatrist-assessed mental health one year after AT.[18-20] However, in this study of only moderate size, no outcomes except for improved sleepiness appeared to be predicted by documented OSA presence or severity prior to AT.

The lack of evidence-based data on the role of AT for treatment of pediatric OSA likely contributes to the large geographic variation in the use of this procedure.[21] A similar paucity of data regarding AT for treatment of children with chronic, mild throat infections has been implicated in the large heterogeneity of AT use for treatment of infections. In response, a multi-center controlled trial of AT was conducted between 2000 and 2003 in the Netherlands.[22] This study showed little clinical benefit of AT, which was associated with a large cost.[23] Children diagnosed with OSA were excluded from this study, so the findings do not address the role of AT as a treatment for pediatric OSA.

In summary, despite the high prevalence of pediatric OSA and its associated comorbidities, no definitive study has documented the effectiveness of AT, the standard treatment modality for pediatric OSA. Such data could also contribute critical evidence that OSA directly contributes to the adverse outcomes that are known to be associated with the sleep disorder. The **Ch**ildhood **A**denotonsillectomy **T**rial (CHAT) was therefore designed to assess neuropsychological and health outcomes in children with OSA randomized to receive early AT (eAT) as compared to Watchful Waiting with Supportive Care (WWSC). Important secondary goals of the study are to evaluate outcomes in subgroups defined by obesity and race. The design and implementation of a controlled trial for a widely used "standard practice" surgical intervention in a pediatric population raises a number of methodological challenges.

METHODS

Study Aims

The primary objective of CHAT is to test whether after a 7-month observation period, children with mild to moderate OSA randomized to eAT will show greater levels of neurocognitive functioning, specifically in the attention-executive functioning domain, than children randomized to WWSC. We also will evaluate whether children randomized to eAT will show greater improvements in behavior, other indices of neurocognitive functioning (learning and memory, information processing, etc.), physical growth, blood pressure, metabolic profile, and quality of life. We will also assess whether direct measurements of OSA, including the number of breathing disturbances recorded on overnight polysomnography and level of oxygenation during sleep, improve more in the eAT than the WWSC arm, and will explore the extent to which improvement in sleep and breathing indices correlate with improvement in neuropsychological and health indices. Finally, we will investigate subgroup differences in response, specifically in regard to obese compared to non-obese children and ethnic/racial minorities compared to other children.

Study Organization

The study is supported by a Data Coordination Center (University of Pennsylvania; Philadelphia, PA), charged with development of the study's statistical design and monitoring plans, construction and management of the study database and study materials, generation of statistical reports to investigators and the CHAT Data and Safety Monitoring Board (DSMB), and quality assurance through surgical and neuropsychology cores that operate at the University of Michigan, Ann Arbor, MI. The study is also supported by a Scientific Coordinating Center/Sleep Reading Center (Brigham and Women's Hospital, Boston, MA) charged with oversight of the general scientific objectives, centralized polysomnographic scoring, and generation of standardized polysomnographic variables. Clinical sites are each headed by an experienced pediatric sleep specialist or otolaryngologist and are responsible for recruitment and follow-up of participants. Initially, 4 clinical sites (Children's Hospital of Pennsylvania, Philadelphia, PA; Cincinnati Children's Medical Center, Cincinnati, OH; Kosair Children's Hospital, Louisville, KY; Rainbow Babies and Children's Hospital, Cleveland, OH) were identified to participate in the study. One of these sites (Kosair Children's Hospital) was removed after its Principal Investigator relocated, and 3 new sites (Children's Hospital, Boston, Boston, MA; Cardinal Glennon Children's Hospital, St. Louis, MO; Montefiore Medical Center, Bronx, NY) were added to improve subject accrual, resulting in 6 sites that contribute participants.

Modality and operational committees are organized to address the multiple quality control and monitoring needs of the study: Surgical Quality Control, Neuropsychology Quality Control, Polysomnography Quality, Recruitment and Operations, and Publications and Presentations. Study governance is through a Steering Committee with representation from each participating site, key quality control cores, and NHLBI program staff. An Executive Committee, consisting of the Study Chair, the DCC Director and project manager, and the NHLBI project officer, meets twice monthly by telephone to address emerging issues. An independent DSMB, with expertise in pediatric ethics, surgery, sleep apnea, clinical trials, and biostatistics, appointed by and reporting directly to the National Heart, Lung and Blood Institute, meets regularly to assess the emerging data and make recommendations. An independent board-certified sleep specialist, or his back-up, is continuously available as a medical monitor (MM).

Sample Population and Enrollment

A total of 460 children from 6 clinical sites will be randomized to one of the two treatment arms. The targeted study population is children between 5.0-9.99 years of age with mild to moderate OSA, as defined by parental report of the child's snoring and a standardized and centrally scored polysomnogram showing an obstructive apnea index (OAI, number of obstructive apneas per hour of sleep) ≥ 1 or apnea hypopnea index (AHI) ≥ 2. Participants must also have an OAI < 20 and AHI < 30. The complete eligibility criteria are shown in Table 1. These criteria reflect slightly more inclusive criteria than when the study was originally designed. Specifically, soon after initiating the study, the upper range for body mass index (BMI) z-score was raised from 2.5 to 2.99 (since a BMI z-score < 3.0 would not routinely require specialized perioperative management), and the criteria for tonsillar hypertrophy were changed from ≥ 2 to ≥ 1 and considered to be an appropriate surgical candidate by an otolaryngologist. The age range chosen for this study reflects a need to measure key endpoints using comparable methods with reliability across the age range in the study population. Although OSA may be more severe in younger children, we restricted the sample to children older than 5 years of age due to the lack of normative data for our primary attention/executive outcome for younger children. Above age 10 years, pubertal changes could affect study outcomes, such as sleep patterns, growth, and hormones in a nonlinear fashion that could be difficult to dissect from treatment influences. Eligibility criteria are assessed through a series of screening and evaluation procedures, including chart review and parent interview (see Figure 1). Additionally, prior to randomization, the child is evaluated by an otolaryngologist to ensure that the child would be an appropriate surgical candidate should she or he be randomized to eAT.

The study was initially designed with the intention to utilize pediatric Sleep Centers/Sleep Laboratories and pediatric ENT clinics as primary recruitment sources. However, due to lagging recruitment in several sites, recruitment was broadened to general pediatric clinics and to the general community through the use of public advertising. It was anticipated that 40% to 50% of children referred for sleep studies would meet the AI/AHI eligibility criteria, and of these, 40% would agree to participate. Thus, initial projections were that approximately 3000 families would need to be approached to meet the enrollment targets.

Table 1—CHAT eligibility criteria

Inclusion Criteria

1. Ages 5.0 to 9.99 years at time of screening.
2. Diagnosed with obstructive sleep apnea defined as:
   - OAI ≥ 1 or AHI ≥ 2, confirmed on nocturnal, laboratory-based PSG *and*
   - Parental report of habitual snoring (on average occurring > 3 nights per week).
3. Tonsillar hypertrophy ≥ 1 based on a standardized scale of 0-4
4. Deemed to be a surgical candidate for AT by otolaryngologist (ENT) evaluation.

Exclusion Criteria

1. Recurrent tonsillitis that meets published clinical practice guidelines for surgery defined as: ≥ 3 episodes in each of 3 years, 5 episodes in each of 2 years, or 7 episodes in one year.
2. Craniofacial anomalies, including cleft lip and palate or submucosal cleft palate or any anatomic or systemic condition which would interfere with general anesthesia or removal of tonsils and adenoid tissue in the standard fashion.
3. Obstructive breathing while awake that merits prompt AT in the opinion of the child's physician.
4. Severe OSA or significant hypoxemia requiring immediate AT as defined by: OAI > 20 or AHI > 30; desaturation defined as SpO_2 < 90% for > 2% sleep time
5. AHI in the normal range (OAI < 1 and AHI < 2)
6. Evidence of clinically significant cardiac arrhythmia on PSG: non-sustained ventricular tachycardia, atrial fibrillation, second degree AV block, sustained bradycardia < 40 bpm (> 2 min), sustained tachycardia > 140 bpm (> 2 min)
7. Extremely overweight defined as: body mass index > 2.99 for age group and sex z-score
8. Severe health problems that could be exacerbated by delayed treatment for OSA, including: severe cardiopulmonary disorders (e.g., cystic fibrosis, congenital heart disease); sickle cell disease; poorly controlled asthma (with > 1 hospitalization in last year); epilepsy requiring medication; diabetes (type 1 or type 2) requiring medication; doctor-diagnosed heart disease or cor pulmonale; History of stage II hypertension (HTN) defined as > 99% percentile and/or requiring medication; mental retardation; chronic infection; or HIV.
9. Psychiatric or behavioral disorders requiring or likely to require initiation of new medication, therapy, or other specific treatment during the 7-month trial period.
10. Known genetic, craniofacial, neurological, or psychiatric conditions likely to affect the airway, cognition, or behavior.
11. Current use of: ADHD medications, psychotropic medication, hypnotics, hypoglycemic agents or insulin, antihypertensives, growth hormone, anticonvulsants, anticoagulants, daily oral corticosteroids.

Based on the characteristics of the recruitment sites, it is anticipated that 50% to 60% of the sample will be African American or Hispanic and 50% will be overweight or obese.

Table 2—Primary and key secondary endpoints

Primary Outcome: NEPSY A/E Subscore

Secondary and Mediator Outcomes

Sleep Apnea: AHI; % of total sleep time with SpO_2 < 92%

Sleep Symptoms and QoL: Pediatric Sleep Questionnaire (total score, SRBDS); OSAS-18 (total score); Modified Epworth Sleepiness Scale

Cognition: GCA total score from DAS-2

Behavior: Regulation Total Score from BRIEF; ADHD Index from the Connors Rating Scale

Metabolic: *CRP, HOMA-IR*

Anthropometry: Change in height, weight, and BMI percentiles

Blood Pressure: Mean arterial pressure

Generic Quality of Life: PedsQL (total score)

Study Interventions

The intervention period is 7 months. Equal numbers of participants will be randomized to eAT or WWSC.

Watchful Waiting with Supportive Care (WWSC) refers to conservative medical management, with treatment or referral for treatment of comorbidities (e.g., asthma, allergic rhinitis), education regarding general sleep hygiene and healthy behaviors, and use of nasal saline spray as needed for nasal mucosal crusting or dryness. Children randomized to WWSC are to be reevaluated by an otolaryngologist after the 7-month observational period and could be recommended for surgery at that time.

For children randomized to eAT, surgery will be performed by participating ENT surgeons within 4 weeks of study enrollment, with treatment or referral for treatment of comorbidities (e.g., asthma, allergic rhinitis), and education regarding general sleep hygiene and healthy behaviors, and use of nasal saline spray as needed for nasal mucosal crusting or dryness. Since AT is considered a routine clinical procedure, its cost is covered by medical insurance and not by the research project.

Endpoints

A summary of the primary and secondary endpoints is shown in Table 2. The primary outcome is a measure of attention/executive function assessed with the Attention/Executive (A/E) Functioning Domain Index from the Developmental Neuropsychological Assessment (NEPSY).[24] Although we are interested in measuring the range of comorbidity associated with pediatric OSA, we chose to highlight attention/executive function based on several areas of investigation: (1) animal and other experimental data showing adverse neuropsychological effects of physiological stresses common in OSA such as intermittent hypoxemia, arousal, and sleep deprivation; (2) cross-sectional studies showing neuropsychological differences in children with OSA compared to controls; (3) uncontrolled studies showing improved neuropsychological outcomes in children with OSA studied before and after AT. The specific choice of the NEPSY Attention/Executive domain composite as the primary outcome was similarly made on several grounds. First, growing empirical and theoretical evidence suggests that the domains of attention and executive functioning may be differentially sensitive to the adverse effects of OSA.[25-27] Second, the NEPSY A/E composite has excellent technical qualities. Unlike most tests of attention and executive functioning, it is applicable to children throughout the age range being investigated. Moreover, it summarizes multiple aspects of attention and executive functioning by combining relevant subtest scores, providing a comprehensive range of executive and attention functions, including inhibition, selective attention, planning, maintaining and changing set, and motor persistence. The composite A/E Index has strong internal reliability, reasonable error of measurement, and good stability measures. Third, and most germane, the NEPSY A/E composite has been reported to be sensitive to untreated OSA (effect size: > 0.3).[28]

Other cognitive and behavioral measures for secondary analyses were chosen to assess additional domains sensitive to sleepiness or intermittent hypoxemia. Additional tests evaluate overall cognitive ability, memory, language abilities, psychomotor skills, behavior, and mood. Several measurements of generic and disease-specific quality of life are measured. Each test is standardized, has normative values across the study age range, and acceptable psychometric properties.Tests are administered by centrally trained psychometricians blinded to treatment group. A licensed psychologist supervises each psychometrician. Reports from teachers are obtained to the extent possible to get additional perspectives on the child's performance and behavior.

Several other secondary outcomes, including blood pressure, fasting insulin levels, and CRP levels, were identified. OSA has been associated with elevations in inflammatory cytokines,[29] potentially related to systemic responses to recurrent upper airway obstruction and hypoxia in altering critical metabolic pathways implicated in atherosclerosis. Thus, secondary aims of the study were to address whether markers of cardiovascular risk, such as C-reactive protein (CRP) or fasting insulin, improve with OSA treatment. Changes in weight, body mass index, and growth velocity are of interest, but the clinical interpretation of these outcomes is recognized to be dependent on the baseline characteristics of the child. On the one hand, among growth-delayed children, an increased growth velocity after AT would be interpreted as favorably influencing health. However, increased weight following AT in children who are overweight at the time of AT may increase obesity-related health morbidities.

The most obvious response variable—change in the AHI—was identified as a secondary, mediating variable. This approach reflects our hypothesis that changes in OSA severity will mediate changes in the clinically relevant and other physiological outcomes. Objective and subjective measurements of sleep are made using standardized polysomnography and questionnaires. Acquisition of standardized sleep data required each participating sleep laboratory to adopt a standardized recording montage and set of sensors. Sleep technicians from each site underwent uniform training and certification prior to data collection for CHAT and all studies were scored centrally by research polysomnologists.

Study Procedures

Prior to randomization, an extensive series of screening procedures occur to ascertain that all eligibility criteria are met. During the informed consent procedure, special care is taken

to ensure that the guardians are comfortable with a randomization result that could lead to either a surgical or nonsurgical intervention in the near term or deferral of surgical treatment for 6 months.

Prior to enrollment, all children undergo a standardized polysomnography exam. In the majority of cases, such as for children recruited from otolaryngology practices, this study is performed after the child is consented for the research study as part of the research protocol. Children referred directly to a sleep clinic/laboratory may have undergone polysomnography studies prior to study enrollment as part of routine clinical care. In those cases, studies are obtained using a standardized montage established for use by all clinical sites participating in this study. Once enrolled in the study, all polysomnograms are transmitted to a central reading center where they are scored. Children who do not meet sleep study eligibility criteria are managed according to routine clinical care. At 7-month follow-up, all children undergo a repeat polysomnography study as part of the research protocol.

After determining that the polysomnography eligibility criteria are met and that the otolaryngologist has equipoise with regard to proceeding with or delaying AT, the child is scheduled for a baseline assessment exam in a clinical research unit. At this 5- to 6-hour visit, standardized assessments examine anthropometric characteristics, neurocognitive and behavioral functions, general health and functional status, metabolic profile, and morning blood pressure. The child and guardian are then provided with general instructions on sleep hygiene and use of the nasal saline spray. After the baseline testing is completed, unless contraindications arise during testing (e.g., high depression indices or low cognitive function suggesting a severe handicap), the child is randomized to one of the two treatment arms. Randomization, stratified by site, age (5-7 or 8-10), race (African American or other), and weight ($\leq$ or $>$ 95th percentile of BMI), is performed using a web-based procedure maintained by the DCC. Clinic sites do not have access to the randomization schedule, so the standard of allocation concealment is met.[30] If randomized to eAT, arrangements are made for AT to occur within 4 weeks. Approximately every 2 months, research coordinators make telephone contact with enrolled families for safety and adverse event monitoring and to reinforce general study participation. A brief ($<$ 1 h) interim visit is conducted at month 3 for safety and adverse event monitoring and to measure blood pressure, weight, and height, and to reinforce general study participation. At 7 months, a repeat sleep study is conducted. On a subsequent visit within several days of this study, a repeat research exam, nearly identical to the initial baseline research exam, is conducted. Participants in the WWSC arm are referred for re-evaluation by the otolaryngologist after they complete the 7-month exam and may then undergo any clinically recommended treatment.

Blinding

A major challenge of the CHAT study relates to the use of a surgical intervention that prevents blinding of the child, parent, and certain key staff members. Great efforts, however, are made to protect unblinding of the principal investigators, psychometricians, and others who directly collect or evaluate data or can otherwise influence the course of the study. Thus, an unusual single-blind situation is created, whereby the subjects and parents are unblinded but the sleep physicians, some of whom are responsible for the overall conduct of pediatric sleep medicine at their site, are blinded. At each site, a research coordinator is identified who is unblinded, while other staff, such as those who perform neuropsychological testing, are blinded. A structured format for communicating issues of potential clinical significance between the unblinded research coordinators, unblinded investigators, and physicians (e.g., the otolaryngologists and a medical monitor) was established to minimize the impact of unblinding on study outcomes and study progress, while still ensuring appropriate monitoring and care of study participants. However, a clear potential for unblinding exists at many contact points, especially when parents or children discuss their treatments with study personnel despite having been instructed not to do so. All such episodes are documented for use when interpreting study findings.

Safety Monitoring

Safety concerns require special consideration in this study because: (1) subjects of the study are children and thus are vulnerable; (2) one intervention exposes the child to general anesthesia and surgery with known, albeit small, perioperative morbidity and mortality; and (3) one intervention can be perceived as withholding a clinically accepted standard therapy. In addition to the usual monitoring by an external DSMB and local institutional review boards, the study established a central independent MM charged with the responsibility for "real time" review of all serious and unexpected adverse events. A class of outcomes was established, denoted as "Treatment Failures (TFs)," defined as changes in clinical status interpreted by the MM as requiring a change in the assigned therapy (e.g., move to AT for children assigned to WWSC; referral for possible additional treatments such as positive airway pressure for children having already undergone AT). Operationally, *potential* TFs are identified by local research coordinators during interim adverse event monitoring. The research coordinators prepare reports containing relevant information and present these to the MM who makes a final adjudication of the status of a TF. To help standardize these judgments and to assist with quality assurance, a sample set of scenarios of potential TFs was generated and discussed by the study staff and MM, with consensus recommendations made for each case. For example, if the research coordinator identified a situation where a mother of a child assigned to WWSC had initiated efforts to have her child receive early AT, the research coordinator would collect information on the clinical situation surrounding this and provide this to the MM. If the parent had initiated efforts to have her child receive treatment other than that assigned by the study due to concerns about changes in health insurance, the MM would classify the event as a "crossover" to the alternative treatment, but not as a TF. In contrast, if early surgery was sought due to worsening sleepiness and snoring while on WWSC, the MM would likely classify the event as a TF if supporting data were consistent with a change in clinical status. Along with adverse event review, the DSMB periodically reviews TFs as an additional safety parameter. Serious adverse events and TFs are transmitted to the DSMB as they occur, and are presented in summary at DSMB meetings. It should be noted that since the primary

analysis will be based on "intention to treat," that every effort is made to continue the child in the study, regardless of treatment crossover or TF assignment.

Quality Control

Quality Control in this study involves multiple levels of training, monitoring, and feedback activities, including: central training of site coordinators, polysomnologists, data managers, psychometricians, and investigators; certification of research personnel for all specialized testing procedures and for data entry; documentation of all procedures in a written and web-accessible manual of procedures; ongoing central monitoring of study quality (e.g., including ongoing reporting of data quality, centralized re-scoring of selective tests); monthly reports on rates of overdue and late visits, outstanding queries and missing values on submitted forms; and site visits to each clinical site. All procedures require certification of staff prior to their interaction with study participants. Requirements differ per procedure, but generally include documentation of successful performance during central training and observation, completion of a written exam, and submission of successfully completed studies during pilot studies (meeting standards for quality and completeness when evaluated by the relevant Quality Control group). After initial certification, each site's performance is monitored on an ongoing basis. Quality control is ultimately the responsibility of the Steering Committee, but several subcommittees, including the Surgical Quality, Neuropsychological Quality, and Polysomnography Subcommittees, are charged with developing and monitoring specific procedures relevant to their area. The Recruitment and Operations Subcommittee, largely composed of the research coordinators from each site, also conducts regular telephone conference calls where quality issues are discussed and exercises are conducted to enhance consistency in performance—e.g., review of case scenarios of children with potential adverse events and development of consensus on how to classify specific scenarios.

Quality assurance for surgical procedures performed as part of routine care is particularly challenging. In CHAT, participating otolaryngologists view a mandatory training video summarizing the surgical protocol and review the CHAT study manual of procedures. Adenotonsillectomies are performed by, or under the direct supervision of board-certified otolaryngologists. To ensure surgical uniformity across participating sites, intraoperative photographs are obtained on a representative sample of subjects (every tenth consecutive patient at each site) and are reviewed for adequacy of lymphoid tissue removal by the surgical core director. A designated lead otolaryngologist from each clinical site participates in a monthly telephone conference wherein surgical related adverse events, recruiting, equipoise, and accuracy of data reporting are discussed.

Statistical Considerations

A sample size of 400 was calculated to provide 90% power to detect an effect size of 0.32 in the primary outcome of change in NEPSY attention/executive function between baseline and the month 7 evaluation. The target sample size was set at 460 to maintain high power given a small proportion of study dropouts and treatment crossovers.

The primary analysis comparing the change in the NEPSY Attention/Executive Functioning score between the eAT and WWSC groups will be performed using analysis of covariance (ANCOVA) adjusting for the stratification factors of age, race, weight status, and site. Change will be defined as the difference between the 7-month and baseline responses. A number of secondary analyses will be conducted, both to evaluate the secondary outcomes and to supplement the primary outcome comparison. The most important secondary outcomes include change in AHI, total score from Pediatric Sleep Questionnaire, child self-report of daytime sleepiness from Epworth Sleepiness Scale, percentage of total sleep time with $SpO_2 < 92\%$ (sleep domain), the General Conceptual Ability from the Differential Abilities Scale (DAS)-II (neurocognitive domain), Behavior Regulation from the BRIEF and the Attention Deficit Hyperactivity Disorder (ADHD) subscale from the Connors Rating Scale (behavioral domain), CRP and HOMA (metabolic domain), and the total scores from the PedsQL and OSAS-18 (health quality of life domain). Standard regression diagnostics will be used to assess model adequacy, and to examine potential outlying or influential data points. Given interest in subgroup differences in treatment responses, exploratory analyses also will be performed stratifying by overweight status at baseline and race, and fitting interaction terms as appropriate.

Primary and secondary analyses will follow the "intention-to-treat" principle: individuals will be analyzed according to their assigned treatment group, whether or not they remain on the assigned treatment. Every effort will be made to obtain follow-up data on all children randomized, whether or not they follow their assigned treatment.

Two formal interim analyses will be performed, after one-quarter and one-half of subjects have completed follow-up. These analyses will be assessed by the study DSMB, who will view data by uncoded treatment arms at their discretion. Early stopping is to be considered only on the basis of safety considerations.

DISCUSSION

Design of the CHAT study present several challenges, including: (1) evaluation of a "clinically accepted" treatment; (2) variations in equipoise among the clinical sites; (3) masking key personnel while ensuring safety and responsiveness to parent concerns; (4) standardizing approaches for adjudicating "treatment failures" as an endpoint; (5) utilizing and standardizing physiological data collected from a variety of settings and with various equipment; (6) providing appropriate levels of feedback of study test results to participating families and their physicians; (7) standardizing approaches and data collection related to surgical procedures performed in routine clinical settings; (8) and standardizing adverse event reporting for conditions such as respiratory illnesses that may be difficult to identify consistently through interim participant contact. Clinical trials involving children and evaluating surgical procedures also present special challenges. Children are considered a vulnerable population and decision making often is complex,[31] with two guardians potentially differing in their level of enthusiasm for enrolling their children in studies. Despite these challenges, close coordination and engagement of a multidisciplinary team in CHAT established procedures and systems for balancing many competing study and ethical needs.

During the past 40 years, numerous trials have evaluated clinically accepted or standard-of-care treatments, often iden-

tifying unanticipated benefits or harm. Three prominent examples are the Cardiac Arrhythmia Suppression Trial (CAST),[32] which showed that the repression of ventricular ectopic beats resulted in increased mortality; the Women's Health Initiative (WHI), which showed that hormone replacement therapy increased both cardiovascular and cancer mortality in women[33]; and several studies that established the safety of surgically conservative procedures for breast cancer despite initial standards that called for more radical procedures.[34] Clinical trials in children, however, have been fewer, in part due to the general reluctance of families to enroll their children in investigational research[31] and to challenges associated with the special concerns in studying vulnerable populations. Thus, launching studies that challenge standard practices are less familiar to the pediatric clinical and scientific communities.

Although AT is considered a standard intervention for pediatric OSA, no randomized controlled trial previously has evaluated the efficacy of this treatment. Changes in the clinical spectrum of patients referred for surgery (i.e., an increasing proportion of overweight children) further necessitate a careful examination of possible heterogeneity in treatment response according to the patients' risk factors and comorbidities. Similar to the Netherlands study of AT for treatment of recurrent infection,[22] it is possible that generally held assumptions about the effectiveness of AT for treatment of sleep disordered breathing are not supported by rigorous evidence. However, when designing CHAT, we recognized that despite community-wide equipoise regarding the role of given treatments for pediatric AT, "uncertainty" regarding the clinical value of given treatments may not be endorsed by individual practitioners nor by patients/guardians whose personal experiences and strong preferences for conservative or more aggressive treatments may influence decision making. Thus, on an ongoing basis, the CHAT investigator team spends considerable effort at assessing issues influencing equipoise at each clinical site.

We established study eligibility criteria identified by our multi-disciplinary team to describe a sample of children in whom clinical equipoise exists. For example, we exclude children with severe OSA or evidence of medical morbidity that is serious and potentially exacerbated by untreated OSA (see Table 1). A recent review of ethical questions in OSA clinical trials concluded that in the absence of any proven, serious morbidity from a medical condition, randomized clinical trials that do not exclude patients with that morbidity are still permissible and often warranted.[35] Overall, in the sample of children without serious comorbidity or hypoxemia, the CHAT study has been judged ethical and approved by all relevant IRBs.

Although there are large numbers of children who undergo AT at each of the participating clinical sites, our initial projections were that only 50% of these children would meet the study eligibility criteria (PSG findings of mild to moderate OSA and absence of severe comorbidities), and of these potentially eligible children, only 40% of their guardians would be comfortable enrolling in a study with equal likelihood of assignment to a surgical or a conservative intervention. Like other clinical trials, the smaller number of children who meet study eligibility criteria (established to ensure safety and internal validity) and who are agreeable to participate in a clinical trial contrasts with the larger number of patients undergoing AT or diagnosed with OSA, which may limit the generalizability of study findings while also reducing the efficiency of meeting study recruitment goals. Thus, we identified a need to screen approximately 3000 children referred for AT or OSA evaluation to meet our targeted recruitment goal of 460 children, requiring active recruitment by a minimum of 4 clinical sites

Anecdotally, we find that attitudes about OSA treatment vary widely among families. Many parents feel strongly that their child needs a timely surgical intervention due to loud snoring or due to concerns about behavior or academic performance. Others have concerns about the convenience of scheduling surgery within the context of this trial or have concerns over insurance coverage. An important issue identified early in the study is the need to resolve potential differences in interests in study participation by each guardian before randomization, to reduce the likelihood that a guardian will withdraw the child if the randomization assignment is not what was desired.

Although we anticipated a great deal of between-family variability in equipoise, we were surprised to find that even within a consortium of experienced academic pediatric sleep centers, there is substantial variation among sites with respect to clinical approaches for using PSG and AT in children referred for evaluation of suspected OSA. The CHAT clinical sites also vary in the proportion of children who are recruited from otolaryngology, sleep, and primary care settings. This variability further underscores the need to generate evidence relevant to clinical practice.

The challenge of maintaining objectivity in study assessments across the duration of a 4- to 5-year trial in which interventions cannot be easily blinded is a special concern. The CHAT study solution is to segregate, as much as feasible, roles for various investigators and research personnel and to create clear algorithms for dealing with communications. To minimize unblinding of Principal Investigators, other medical safety monitoring need to be established, including identification of local physicians and a central MM to assist with patient-related safety issues. However, this process also needs to accommodate the key regulatory responsibilities of the Principal Investigators.

A number of primary, secondary, and mediation analyses are specified. The data for these analyses are derived from multiple sources, including the operative record, the polysomnogram, parent/teacher ratings of child behavior, sleep, and quality of life, and the results of neuropsychological testing. Data collected from clinical encounters, such as surgical procedures, are standardized by use of common data collection instruments (an intraoperative surgical form), consensus to utilize certain procedures (such as excluding subcapsular tonsillectomies), and central review of a sample of intraoperative photographs. In contrast, given the known inter-laboratory variability of PSG data, all PSG data are prospectively standardized using the same or comparable sensors, PSG technicians undergo common core training and certification, and all PSGs are scored at a central Reading Center. Finally, given the high level of interest in the neuropsychological outcomes, these data are all prospectively collected by trained research neuropsychologists in a research setting following a standardized protocol with central over-reading of a sample of records. To better standardize potentially subjective safety assessments, such as certain Adverse

Event adjudication and Treatment Failures, a set of teaching case scenarios was developed and are used to train the MM, investigators, and research coordinators.

The issue of what "control" condition was most appropriate for this study was a topic of great debate during design of the study. Whereas a conservative supportive control arm was considered a reasonable and ethical alternative to AT, concerns were raised that such a control would be perceived by guardians as insufficiently "medicalized," given the unblinded design. Thus, children in both groups were provided nasal saline spray to use nightly. Although emerging data suggest a role for nasal and systemic anti-inflammatory medications for treatment of pediatric OSA,[36] insufficient data were available at initiation of CHAT to justify a comparative effectiveness study.

In summary, as a randomized, controlled study of a commonly accepted pediatric surgical intervention, the CHAT study wrestled with numerous challenges. Given the recognized need for evidenced-based data on treatment effects, the study framework should provide a useful template for other pediatric controlled studies.

ACKNOWLEDGMENTS

Supported by grants: HL083075, HL083129,UL1-RR-024134, UL1 RR024989. For the **Ch**ildhood **A**denotonsillectomy **T**rial (CHAT): Boston Children's Hospital, Harvard University, Boston MA (Eliot Katz, MD; Janice Ware, PhD; Dwight Jones, MD); Brigham and Women's Hospital and Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA (Susan Redline, MD MPH; Susan Surovec); Cardinal Glenn Children's Hospital, St. Louis University, St. Louis MO (Ron Mitchell, MD; Shalini Paruthi, MD; Karen Snyder, MS); Children's Hospital of Philadelphia,(Carole Marcus, MB-BCh; Nina H. Thomas, PhD; Lisa Elden, MD); Cincinnati Children's Medical Center, University of Cincinnati, Cincinnati, OH (Raouf Amin, MD; Dean Beebe, PhD; Paul Willging, MD); Montefiore Children's Hospital, Albert Einstein College of Medicine, Yeshiva University, Bronx, NY (Raanan Arens, MD; Hiren Muzumdar, MD; Shelby Harris, PsyD, CBSM); Rainbow Babies and Children's Hospital, Case Western Reserve University School of Medicine, Cleveland, OH (Carol Rosen, MD; H. Gerry Taylor, PhD; Robert Sprecher, MD); University of Kentucky, Louisville, Kentucky (David Gozal, MD); University of Michigan, Ann Arbor, MI (Ronald Chervin, MD; Susan Garetz, MD, Bruno Giordani, PhD); University of Pennsylvania, Philadelphia, PA (Susan Ellenberg PhD; Renee´ H. Moore, PhD; Kim Lacy, RN, BSN).

DISCLOSURE STATEMENT

This was not an industry supported study. Dr. Ellenberg collaborated with Pfizer and Abbott Laboratories. She has also consulted for Wyeth (now Pfizer) Millennium, Novavax, Inc., Bristol Myers, Squibb, and Lux Biosciences, Inc. She received honoraria from BMS and Shane Pharmaceuticals. Dr. Marcus has received grant support from Elan Corporation, Eli Lilly, Philips Respironics, and Wyeth unrelated to this study. Dr. Chervin has served on the advisory board for Pavad Medical and as a consultant for Arena Pharmaceuticals. He is named as an inventor of intellectual property related to diagnosis and treatment of sleep disorders and owned by the University of Michigan. Dr. Redline has received research support from Dymedix Inc., use of equipment from Philips Respironics, and is the incumbent of an endowed professorship donated to Harvard Medical School by Dr. Peter Farrell, the founder and Board Chairman of ResMed Inc. Dr. Rosen has received grant support from the NIH and the ASMF. Dr. Amin has received research funding from Proctor & Gamble. Dr. Gozal has received a research grant from the NIH and serves on the National Speaker Bureau for Merck Co. Dr. Moore has received grant support from Ethicon Endo-Surgery, Inc. The other authors have indicated no financial conflicts of interest.

REFERENCES

1. Redline S, Tishler PV, Hans MG, Tosteson TD, Strohl KP, Spry K. Racial differences in sleep-disordered breathing in African-Americans and Caucasians. Am J Respir Crit Care Med 1997;155:186-92.
2. Rosen CL, Larkin EK, Kirchner HL, et al. Prevalence and risk factors for sleep-disordered breathing in 8- to 11-year-old children: association with race and prematurity. J Pediatr 2003;142:383-9.
3. Cullen KA, Hall MJ, Golosinskiy A. Ambulatory Surgery in the United States, 2006.In: Statistics National Center for Health Statistics. Hyattsville, MD; 2009.
4. Croft CB, Brockbank MJ, Wright A, Swanston AR. Obstructive sleep apnea in children undergoing routine tonsillectomy and adenoidectomy. Clin Otolaryngol 1990;15:307-14.
5. Ali NJ, Pitson D, Stradling JR. Sleep disordered breathing:effects of adenotonsillectomy on behaviour and psychological functioning. Eur J Pediatr 1996;155:56-62.
6. Suen JS, Arnold JE, Brooks LJ. Adenotonsillectomy for treatment of obstructive sleep apnea in children. Arch Otolaryngol Head Neck Surg 1995;121:525-30.
7. Mitchell RB, Kelly J. Adenotonsillectomy for obstructive sleep apnea in obese children. Otolaryngol Head Neck Surg 2004;131:104-8.
8. Apostolidou MT, Alexopoulos EI, Chaidas K, et al. Obesity and persisting sleep apnea after adenotonsillectomy in Greek children. Chest 2008;134:1149-55.
9. Mitchell RB, Kelly J. Outcome of adenotonsillectomy for obstructive sleep apnea in obese and normal-weight children. Otolaryngol Head Neck Surg 2007;137:43-8.
10. Tauman R, Gulliver TE, Krishna J, et al. Persistence of obstructive sleep apnea syndrome in children after adenotonsillectomy. J Pediatr 2006;149:803-8.
11. Rudnick EF, Walsh JS, Hampton MC, Mitchell RB. Prevalence and ethnicity of sleep-disordered breathing and obesity in children. Otolaryngol Head Neck Surg 2007;137:878-82.
12. Conlon BJ, Donnelly MJ, McShane DP. Improvements in health and behaviour following childhood tonsillectomy: a parental perspective at 1 year. Int J Pediatr Otorhinolaryngol 1997;41:155-61.
13. Wolfensberger M, Haury JA, Linder T. Parent satisfaction 1 year after adenotonsillectomy of their children. Int J Pediatr Otorhinolaryngol 2000;56:199-205.
14. Stradling JR, Thomas G, Warley ARH, Williams P, Freeland A. Effect of adenotonsillectomy on nocturnal hypoxaemia, sleep disturbance, and symptoms in snoring children. Lancet 1990;335:249-53.
15. Friedman BC, Hendeles-Amitai A, Kozminsky E, et al. Adenotonsillectomy improves neurocognitive function in children with obstructive sleep apnea syndrome. Sleep 2003;26:999-1005.
16. Montgomery-Downs HE, Crabtree VM, Gozal D. Cognition, sleep and respiration in at-risk children treated for obstructive sleep apnoea. Eur Respir J 2005;25:336-42.
17. Gozal D. **S**leep-disordered breathing and school performance in children. Pediatrics 1998;102:616-20.
18. Chervin RD, Ruzicka DL, Giordani BJ, et al. Sleep-disordered breathing, behavior, and cognition in children before and after adenotonsillectomy. Pediatrics 2006;117:e769-78.
19. Dillon JE, Blunden S, Ruzicka DL, et al. DSM-IV diagnoses and obstructive sleep apnea in children before and 1 year after adenotonsillectomy. J Am Acad Child Adolesc Psychiatry 2007;46:1425-36.

20. Giordani B, Hodges EK, Guire KE, et al. Neuropsychological and behavioral functioning in children with and without obstructive sleep apnea referred for tonsillectomy. J Int Neuropsychol Soc 2008;14:571-81.
21. Van Den Akker EH, Hoes AW, Burton MJ, Schilder AG. Large international differences in (adeno)tonsillectomy rates. Clin Otolaryngol Allied Sci 2004;29:161-4.
22. van Staaji BK, van den Akker EH, Rovers MM, Hordijk GJ, Hoes AW, Schilder AG. Effectiveness of adenotonsillectomy in children with mild symptoms of throat infections or adenotonsillar hypertrophy: open, randomised controlled trial. Clin Otolaryngol 2005;30:60-3.
23. Buskens E, van Staaij B, van den Akker J, Hoes AW, Schilder AG. Adenotonsillectomy or watchful waiting in patients with mild to moderate symptoms of throat infections or adenotonsillar hypertrophy: a randomized comparison of costs and effects. Arch Otolaryngol Head Neck Surg 2007;133:1083-8.
24. Korkman M, Kirk U, Kemp S. Developmental Neuropsychological Assessment (NEPSY): Harcourt Assessment, Inc.; 1998.
25. Archbold KH, Giordani B, Ruzicka DL, Chervin RD. Cognitive executive dysfunction in children with mild sleep-disordered breathing. Biol Res Nurs 2004;5:168-76.
26. Beebe DW, Gozal D. Obstructive sleep apnea and the prefrontal cortex: towards a comprehensive model linking nocturnal upper airway obstruction to daytime cognitive and behavioral deficits. J Sleep Res 2002;11:1-16.
27. Beebe DW, Groesz L, Wells C, Nichols A, McGee K. The neuropsychological effects of obstructive sleep apnea: a meta-analysis of norm-referenced and case-controlled data. Sleep 2003;26:298-307.
28. Gottlieb DJ, Chase C, Vezina RM, et al. Sleep-disordered breathing symptoms are associated with poorer cognitive function in 5-year-old children. J Pediatr 2004;145:458-64.
29. Larkin EK, Rosen CL, Kirchner HL, et al. Variation of C-reactive protein levels in adolescents: association with sleep-disordered breathing and sleep duration. Circulation 2005;111:1978-84.
30. Schulz KF, Grimes DA. Allocation concealment in randomised trials: defending against deciphering. Lancet 2002;359:614-8.
31. Shaddy RE, Denne SC. Clinical report--guidelines for the ethical conduct of studies to evaluate drugs in pediatric populations. Pediatrics 2010;125:850-60.
32. Pratt CM, Moye LA. The Cardiac Arrhythmia Suppression Trial: background, interim results and implications. Am J Cardiol 1990;65:20B-9B.
33. Rossouw JE, Anderson GL, Prentice RL, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. JAMA 2002;288:321-33.
34. Fisher B, Anderson S, Bryant J, et al. Twenty-year follow-up of a randomized trial comparing total mastectomy, lumpectomy, and lumpectomy plus irradiation for the treatment of invasive breast cancer. N Engl J Med 2002;347:1233-41.
35. Brown DL, Anderson CS, Chervin RD, et al. Ethical issues in the conduct of clinical trials in obstructive sleep apnea. J Clin Sleep Med 2011;15;7:103-8.
36. Kheirandish-Gozal L, Serpero LD, Dayyat E, et al. Corticosteroids suppress in vitro tonsillar proliferation in children with obstructive sleep apnoea. Eur Respir J 2009;33:1077-84.