

THE UNIVERSITY OF CALGARY

A Decision Rule for Diagnostic Testing in Obstructive Sleep

Apnea

by

Willis Harvey Tsai

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES  
IN PARTIAL FUFILLMENT OF THE REQUIREMENTS FOR THE  
DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF COMMUNITY HEALTH SCIENCES

CALGARY, ALBERTA

AUGUST, 1999

© Willis Harvey Tsai 1999



National Library  
of Canada

Bibliothèque nationale  
du Canada

Acquisitions and  
Bibliographic Services

Acquisitions et  
services bibliographiques

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

*Your file* *Votre référence*

*Our file* *Notre référence*

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-48052-6

**Canada**

## ABSTRACT

A consecutive series of 75 patients referred to a tertiary sleep centre underwent prospective evaluation with the Upper Airway Physical Exam Protocol. Predictors of obstructive sleep apnea included (OSA): age, snoring history, witnessed apneas, and hypertension, body mass index, neck circumference, mandibular protrusion, thyro-rami distance, sterno-mental distance, sterno-mental displacement, thyro-mental displacement, cricomenal space, pharyngeal grade, Sampsoon-Young classification, and overbite. A decision rule was developed: **cricomenal space  $\leq$  1.5 cm, pharyngeal grade  $>2$ , and the presence of overbite.** Patients with all 3 predictors had a positive predictive value of 95% (CI<sub>95%</sub>: 75-100%), negative predictive value: 49% (CI<sub>95%</sub>: 35-63%). A cricomenal space  $>$  1.5 cm excluded obstructive sleep apnea (negative predictive value: 100% (CI<sub>95%</sub>: 75-100%)). Comparable performance was obtained in an unfiltered validation sample of 50 patients referred for diagnostic testing. The inter-rater reliability was high. This decision rule provides a simple, reliable, and accurate method of identifying patients with and without OSA.

## **Acknowledgements**

I would like to acknowledge Colin MacArthur, my thesis supervisor. His insight and thoroughness have proven invaluable in training the critical thought processes so necessary to epidemiology.

To John Remmers, I owe many thanks. His guidance was appreciated on many a project. His enthusiasm stimulated a lasting interest in research. Not in the very least, special mention must be made with respect to introducing me to the joys of telemark skiing.

To Jan Davies, I thank for our many discussions concerning the development of this thesis; and to Rollin Brant for making those complex regression models seem a little less daunting.

I would also like to thank Ward Flemons, who provided valuable comments on this thesis; and was always helpful on other projects, as well as all aspects of clinical medicine.

Finally, I would like to thank my friends and family, who's presence made the "global" thesis experience so enjoyable.

# TABLE OF CONTENTS

Approval Page	ii
Abstract	iii
Acknowledgements	iv
Table of Contents	v
List of Tables	ix
List of Figures	x
<b>1. INTRODUCTION</b>	<b>1</b>
1.1 Health care burden associated with obstructive sleep apnea	1
1.2 Diagnosis of obstructive sleep apnea	4
1.2.1 Overnight polysomnography	4
1.2.2 Portable monitors	5
1.3 Clinical prediction instruments	7
1.3.1 Clinical prediction	7
1.3.2 Decision rules in OSA	8
1.3.3 Radiologic predictors of OSA	11
1.3.4 Physical examination based predictors of OSA	12
1.3.5 Upper airway scoring systems in anaesthesia	13
1.4 Rationale and study objectives	14
1.4.1 Rationale	14
1.4.2 Objective	15

<b>2. METHODS</b>	<b>16</b>
2.1 Sampling frame	16
2.2 The Upper Airway Physical Examination Protocol (UAPP)	17
2.3 Feasibility phase	25
2.4 Index development	26
2.4.1 Item selection	26
2.4.2 Index reliability	26
2.4.3 Measures of validity	28
2.4.4 Responsiveness	29
2.5 Development of the Decision Rule	29
2.5.1 Logistic regression	30
2.5.2 Assessment of criterion validity	31
2.5.3 Recursive partitioning for Decision Rule development	32
2.6 Reliability assessment	35
2.7 Decision Rule validation	35
2.8 Sample size determination	36

<b>3. RESULTS</b>	<b>37</b>
3.1 Feasibility assessment and item reduction	37
3.2 Sampling frame	38
3.3 Logistic regression: univariate predictors of OSA	44
3.4 Model building using logistic regression	46
3.4.1 "Significant p" approach	46
3.4.2 Biologically plausible approach	46
3.4.3 The parsimonious model	47
3.5 Model building using recursive partitioning	48
3.5.1 Classification and Regression Trees (CART)	48
3.5.2 Estimating optimal tree size	50
3.5.3 Pruning the tree	50
3.6 Developing the Decision Rule	52
3.6.1 Optimal cut-point selection by visual inspection	52
3.6.2 Optimal cut-point selection by CART	52
3.7 Diagnostic performance of the Decision Rule	55
3.8 Decision Rule reliability	58
3.9 Validation of the Decision Rule	59

<b>4. DISCUSSION</b>	<b>61</b>
4.1 Summary of results	61
4.2 Clinical predictors of obstructive sleep apnea	62
4.3 Statistical considerations during Decision Rule development	66
4.3.1 Logistic Regression	66
4.3.2 CART versus Regression: Moving forward or a step back	67
4.4 Decision Rule interpretation	69
4.5 Study critique	71
<b>5. CONCLUSION</b>	<b>74</b>
References	75
Appendix A: Upper Airway Physical Examination Protocol (UAPP)	80
Appendix B: The Epworth Sleepiness Score	84
Appendix C: Definitions	85
Appendix E: Consent form	86



## **List of Tables**

Table 1: A summary of decision rules for OSA	10
Table 2: A comparison of recursive partitioning with logistic regression	34
Table 3: Summary of patient characteristics	39
Table 4: Summary of clinical characteristics	40
Table 5: Summary of physical examination characteristics	42
Table 6: Univariate logistic regression analysis of predictive variables for OSA	45
Table 7: Cross-tabulation of predictive variables against a diagnosis of OSA	53
Table 8: Diagnostic performance of the predictive variables	56
Table 9: Diagnostic performance of the decision rule at a variety of RDI diagnostic criterion values	57
Table 10: Inter-rater agreement on the predictive variables	58
Table 11: Diagnostic performance of the decision rule at a variety of RDI criterion values (in the all patient referral sample)	60

## **List of Figures**

Figure 1: Assessment of facial profile	19
Figure 2: Assessment of the cricomental space	20
Figure 3: Palatopharyngeal anatomy	22
Figure 4: Sampsoon-Young grading system for tongue size	23
Figure 5: Pharyngeal grading system	24
Figure 6: Distribution of the respiratory disturbance index (RDI) in the model development population	41
Figure 7: Full decision tree for the diagnosis of obstructive sleep apnea	49
Figure 8: Deviance plot to estimate optimal tree size	51
Figure 9: A decision rule for diagnostic testing in obstructive sleep apnea	54

## CHAPTER 1: INTRODUCTION

### **1.1 Health care burden associated with obstructive sleep apnea**

Obstructive sleep apnea (OSA) is characterized by episodic upper airway obstruction during sleep. It is defined by the apnea hypopnea index (AHI), which is the number of apneas (complete cessation of airflow) and hypopneas (partial reduction in airflow) per hour of sleep. The obstructive sleep apnea syndrome (OSAS) consists of a physiologic event, namely OSA, plus OSA-related symptoms, of which, daytime hypersomnolence is the most frequently encountered presentation.

The obstructive sleep apnea syndrome is common in North America. In a random, community-based, adult sample of 602 government employees, the Wisconsin Sleep Cohort Study reported an OSAS prevalence of 2% and 4% in middle-aged women and men, respectively<sup>1</sup>. OSAS was diagnosed using polysomnography, with an AHI  $\geq 5$  hr<sup>-1</sup> plus associated daytime sleepiness, required to establish a diagnosis.

Using snoring as a proxy for OSA, several case-control and cross-sectional studies suggest a link between snoring and cardiovascular morbidity<sup>2-5</sup>. A limited number of retrospective case-control studies suggest that untreated OSA increases cardiovascular morbidity and all-cause mortality<sup>6-8</sup>. In a cross-sectional study of blood pressure measurements during wakefulness and sleep in the Wisconsin Sleep Cohort, Hla et al found an association between hypertension and sleep apnea independent of obesity, age, and sex<sup>2</sup>. Subjects with an AHI  $\geq 5$  hr<sup>-1</sup> were more likely to have hypertension

compared with those with an AHI  $< 5 \text{ hr}^{-1}$ . The probability of hypertension increased in a dose dependent fashion with AHI<sup>2,16</sup>.

The association between OSA, all-cause mortality and coronary artery disease is more controversial. He et al employed a cross-sectional design in a highly selected series of 706 male patients, evaluated at a tertiary sleep centre. An increased mortality rate was observed in patients with severe OSA (apnea index  $> 20 \text{ hr}^{-1}$ )<sup>6</sup>. Similarly, Partinen et al observed an increased rate of cardiovascular mortality in a small (n=198) group of patients treated at the Stanford Sleep Disorders Clinic, a highly specialized university-based sleep centre<sup>7</sup>. Hung et al employed a case control design using 101 consecutive male patients, who had been admitted for acute myocardial infarction. An increased risk of myocardial infarction was seen in patients with an apnea index  $> 5.3 \text{ hr}^{-1}$ <sup>8</sup>.

Furthermore, an increased rate of automobile collisions is observed in at least a subset of patients with OSA<sup>9-11</sup>. In a cohort study involving 913 subjects forming part of the ongoing Wisconsin Sleep Cohort, Young et al identified an increased risk of automobile accidents (OR=4.2) amongst subjects with OSA (AHI  $\geq 5 \text{ hr}^{-1}$ ), as compared with normal controls (AHI  $< 5 \text{ hr}^{-1}$ )<sup>11</sup>. More recently, in a prospectively followed cohort of 120 patients, Barbe et al also found an increased risk of automobile accidents in patients with OSA<sup>57</sup>.

Although cardiovascular morbidity is an important health issue, patients seek treatment for OSA primarily because of daytime sleepiness. Indeed, the definition of OSAS is predicated on the presence of OSA-related symptoms such as daytime hypersomnolence. In this regard, several large, well designed randomized control trials

have shown that treatment of OSA with continuous positive airway pressure (CPAP) improves sleep architecture, reduces daytime sleepiness, and improves performance<sup>12-15</sup>.

In a randomized controlled trial of CPAP compared with conservative therapy, Redline et al evaluated 111 subjects<sup>14</sup>. Patients with mild obstructive sleep apnea (AHI=5-30 hr<sup>-1</sup>) were randomized to receive either CPAP or conservative therapy. The latter consisted of behavioral and lifestyle counseling, treatment of nasal congestion, and nasal dilators. Patients in the CPAP group reported a greater improvement in hypersomnolence, mood, feeling of well being, and functional status. Similar results were observed by Engleman et al in a randomized placebo controlled trial comparing CPAP therapy with oral placebo<sup>12</sup>. Although differing somewhat in the instruments chosen to assess health outcomes, both studies used validated questionnaires and neurocognitive testing. Sleepiness was both self-reported and objectively identified by a shortened sleep latency on the multiple sleep latency test (MSLT). The MSLT objectively measures sleep predisposition by determining the mean time to polysomnographically confirmed sleep onset in a series of daytime naps.

Given the prevalence of OSA, its associated morbidity, and the effectiveness of treatment, identification of patients with OSA is an important public health issue.

## **1.2 Diagnosis of obstructive sleep apnea**

### **1.2.1 Overnight polysomnography**

Traditionally, OSA is diagnosed using overnight polysomnography (PSG), through the determination of an apnea hypopnea index (AHI). The AHI is the number of apneas (complete cessation of airflow) and hypopneas (reduction in airflow) per hour of sleep. While an  $AHI \geq 5 \text{ hr}^{-1}$  is considered the upper limit of normal, there is considerable controversy as to what constitutes a clinically significant OSA diagnostic criterion value.

In the Wisconsin Sleep Cohort Study, 24% of middle-aged men and 9% of middle-aged women had an  $AHI \geq 5 \text{ hr}^{-1}$ , but only 4% and 2% of these subjects reported symptoms of daytime hypersomnolence, respectively<sup>1</sup>. While epidemiological data suggest that adverse health outcomes such as hypertension, sleepiness, and motor vehicle collisions occur in persons with an  $AHI \geq 5 \text{ hr}^{-1}$ , this is not necessarily a threshold effect. Indeed, for health outcomes such as hypertension, the risk profile appears to follow a dose response relationship<sup>2,16</sup>. Moreover, the threshold value at which risks become significant depends on the health outcome of interest. For example, severe complications, such as cardiovascular morbidity and death are probably only associated with severe OSA<sup>6-8</sup>.

Given these data, it is unlikely that any single AHI diagnostic criterion value will be used in clinical decision making. As stated previously, the clinically relevant end-point is the obstructive sleep apnea syndrome, which consists of a physiologic process (OSA, as defined by the AHI), and associated OSA-related symptoms, many of which are

subjective. Consequently, the decision to pursue therapy depends on the absolute AHI value, OSA-related symptoms, and the perceived risks of adverse health outcomes. Therefore, the evaluation of a new diagnostic instrument requires validation at a variety of AHI cut-off values. AHI diagnostic criterion values of  $\geq 5$ , 10, 15, and 20  $\text{hr}^{-1}$  are the most commonly employed in the research literature.

Polysomnographic determination of AHI requires overnight admission to a sleep laboratory, and involves fitting the patient with cumbersome monitoring equipment: electroencephalogram, electrooculogram, chin electromyogram (EMG), airflow monitoring, inductance plethysmography to assess respiratory effort via thoracoabdominal movement, electrocardiogram, oxygen saturation, and leg EMG. As such, PSG is costly in terms of personnel, time, and money. Recently, a number of instruments have been developed as alternatives to PSG. In general, they either diagnose OSA, or identify “high risk” patients (i.e., those who should go on to full polysomnography).

### **1.2.2 Portable monitors**

Portable monitors are particularly useful in this regard. Monitors range from simple oximeters to multiple channel devices that approach PSG in terms of data acquisition and complexity of use<sup>17</sup>. One of the best validated ambulatory monitors is the SNORESAT<sup>18</sup>. This instrument determines a respiratory disturbance index (RDI) via off-line analysis of digitally recorded nocturnal oxygen saturation. Because portable monitors do not score apnea and hypopnea by conventional polysomnographic criteria, the

monitor-derived RDI is a surrogate measure for AHI. The RDI is defined as the number of respiratory events per hour of sleep. As such, it is a more general measure than the AHI, and its value is dependent on how a respiratory event is defined. It follows that the AHI is a specific form of RDI, where a respiratory event is defined as either an apnea or hypopnea. Historically, OSA has been diagnosed by polysomnographically determined AHI diagnostic criterion values, however, with the advent of portable monitors, RDI measures are being increasingly used.

The SNORESAT-derived RDI has been validated against the gold standard PSG-AHI at a variety of AHI values. Over a range of RDI values from 7-20  $\text{hr}^{-1}$ , SNORESAT had a positive predictive value of between 95%-96%, and a negative predictive value of between 4% to 12% for diagnosing OSA<sup>12</sup>. The prevalence of OSA in these studies ranged from 54% ( $\text{AHI} \geq 7 \text{ hr}^{-1}$ ) to 31% ( $\text{AHI} \geq 20 \text{ hr}^{-1}$ ) of the total sleep centre referral population. In a more recent study, on a prospectively selected group of patients referred to a tertiary sleep centre, using an improved analysis algorithm for defining respiratory events, SNORESAT had a sensitivity and specificity of 98% and 88% respectively for diagnosing OSA ( $\text{AHI} \geq 15 \text{ hr}^{-1}$ )<sup>19</sup>. The mean difference between monitor-derived RDI and PSG-AHI was 2.18  $\text{hr}^{-1}$ . The limits of agreement (2 standard deviations (SD) of the mean of the differences) was 12.34  $\text{hr}^{-1}$ . For comparison, PSG inter-rater scoring variability is associated with a mean AHI difference of 1.80  $\text{hr}^{-1}$ <sup>20</sup>.

Because of increasing financial limitations, there is a growing tendency to diagnose OSA using validated portable monitors, which generate RDI values, rather than PSG-determined AHI. This trend is observed in both clinical and research settings<sup>21-23</sup>.



The standard practice at the Alberta Lung Association Sleep Centre is to use SNORESAT as the primary OSA diagnostic instrument, with PSG being reserved for patients with negative SNORESAT studies, or for patients with clinical evidence of a primary sleep disorder other than OSA. For the purpose of this study, SNORESAT determined RDI cut-off values were used as the gold standard for OSA diagnosis.

### **1.3 Clinical prediction instruments**

#### **1.3.1 Clinical prediction**

Decision rules are prospectively validated algorithms consisting of sets of conditions that predict a particular clinical outcome or appropriate course of action. They differ from clinical practice guidelines in that guidelines represent an evidence-based consensus to guide clinical decision making. A commonly used decision rule is the Ottawa Ankle Rules<sup>56</sup>. The developers of this rule identified 3 variables, which if absent, effectively rule out the possibility of an ankle fracture. Consequently, an ankle x-ray series is only necessary if there is pain near the malleoli and either: (1) an inability to bear weight both immediately and in the emergency department (four steps), or (2) bone tenderness at the posterior edge or tip of either malleolus. The model was developed on 1032 patients, and was validated on 453 subsequent patients.

Although cheap and relatively easy to administer, decision rules aimed at diagnosing OSA have been limited by insufficient sensitivity and specificity for use as diagnostic instruments<sup>24</sup>. In general, most OSA decision rules have specificities around

90%, but sensitivities considerably less than 80%<sup>24-27, 53-55</sup>. However, decision rules could still have significant clinical value, specifically, for identifying patients at risk for OSA, who would then go on to more definitive diagnostic testing (e.g., PSG or portable monitoring). More importantly, a decision rule with a high negative predictive value could identify patients who would not benefit from further, more costly, evaluation. To date, most decision rules have focused on maximizing both sensitivity and specificity. By contrast, none have been designed with the intent of identifying patients who do not need further diagnostic testing.

### **1.3.2 Decision rules in OSA**

One of the better-validated decision rules for OSA was developed by Flemons et al<sup>24</sup>. In a randomly selected series of 180 patients referred to a tertiary sleep centre, increased neck circumference, hypertension, habitual snoring, and reports of nocturnal gasping/ choking were identified as being predictive of OSA (PSG-AHI  $\geq 10$  hr<sup>-1</sup>) using logistic regression modeling. A clinical prediction rule was then developed. Individuals with the highest clinical score (i.e., all 4 characteristics) had a likelihood ratio and post-test probability of OSA (AHI  $\geq 10$  hr<sup>-1</sup>) of 5.17 and 81%, respectively. By contrast, patients with the lowest clinical score had a likelihood ratio of 0.25 and a post-test probability of 17%.

Similarly, Viner et al evaluated 410 patients referred for suspected sleep apnea. Body mass index (BMI), age, male sex, and snoring were predictive of OSA (PSG-AHI  $\geq$

10 hr<sup>-1</sup>). The derived prediction model had a sensitivity and specificity of 28% and 95%, respectively<sup>28</sup>.

For the purposes of diagnosing OSA, the diagnostic performance of most decision rules fall within a similar range<sup>24-27,53-55</sup>. Most decision rules for OSA have high specificities, with sensitivities considerably less than 80%. A summary of decision rules for OSA is presented in Table 1. Although potentially useful measurement instruments, particularly in the context of a directed clinical assessment, most published decision rules are inadequate as stand-alone diagnostic instruments.

Table 1: A summary of decision rules for OSA

AUTHOR (reference)	OSA diagnosis criterion	Sample size	PREDICTIVE VARIABLES	Sensitivity	Specificity
Flemons <sup>24</sup>	AHI $\geq 10$ hr <sup>-1</sup>	180	Neck circumference, hypertension, snoring, gasping/choking	81%*	17% <sup>†</sup>
Viner <sup>28</sup>	AHI $\geq 10$ hr <sup>-1</sup>	410	Body mass index, snoring, age, male sex	28%	95%
Davies <sup>29</sup>	RDI correlation	150	Neck circumference, body mass index		NS
Kushida <sup>36</sup>	AHI $\geq 5$ hr <sup>-1</sup>	300	Body mass index, neck circumference, intermolar distance	100%*	98% <sup>†</sup>
Crocker <sup>53</sup>	AHI $\geq 15$ hr <sup>-1</sup>	100	Age, witnessed apneas, obesity, hypertension	79%	50%
Olson <sup>54</sup>	RDI $\geq 15$ hr <sup>-1</sup>	441	Snoring, apneas, gasping noises, bed covers in disarray	16%	99%

\*Positive predictive value; <sup>†</sup>Negative predictive value; NS not significant

### **1.3.3 Radiologic predictors of OSA**

Considerable work has been done in assessing the role of upper airway imaging in the setting of OSA, specifically through the use of cephalometry, computed tomography (CT), and magnetic resonance imaging (MRI). These radiologic techniques have been useful in advancing the understanding of upper airway pathophysiology<sup>31</sup>. On average, when compared with normal controls, patients with OSA have a small posteriorly placed mandible, a narrow posterior airway space, enlarged tongue and soft palate, and an inferiorly placed hyoid bone<sup>32,34</sup>. However, although several cephalometric variables may be predictive of OSA based on univariate analysis, using multivariate modeling, Davies et al found that only neck size and retroglossal space were independent predictors of OSA<sup>33</sup>. Therefore, the inclusion of cephalometric variables into a prediction model may not necessarily improve diagnostic performance. Furthermore, given the cost and logistic difficulty of performing radiologic imaging, these tests have limited value outside the research setting.

A major criticism of most radiographic techniques is that they study awake and upright patients, whereas, OSA typically occurs while the patient is asleep in the supine position<sup>32</sup>. Radiological imaging is also limited by an inability to dynamically assess soft tissue structures. By contrast, videoendoscopy permits dynamic visualization of the upper airway. Endoscopically observed pharyngeal narrowing has been repeatedly reported in OSA, and the site of narrowing is also a good predictor of surgical treatment success<sup>35</sup>. However, because of its logistical complexity, videoendoscopy remains more a research tool than a practical diagnostic instrument.

#### **1.3.4 Physical examination based predictors of OSA**

Neck circumference and BMI are the only physical examination characteristics that are consistently predictive of OSA<sup>27,29,30</sup>. However, with the notable exception of a study by Kushida et al, most studies have evaluated only basic characteristics. Therefore, the development of more sophisticated physical examination-based measurements may improve the diagnostic performance of existing decision rules.

Kushida et al developed one of the only well characterized morphometric models for diagnosing OSA<sup>36</sup>. They evaluated a consecutive sample of 300 patients referred to the Stanford University Sleep Centre, a university-based tertiary care centre that receives referrals from across the United States. The data set was split into a model development group, and a validation set. Body mass index, neck circumference, and intermolar distance were identified as predictive variables. A prediction index was developed, which had a sensitivity and specificity of 98% and 100% respectively.

However, there are serious questions as to whether the model was tested in a representative sample of patients, given that the prevalence of OSA was 85%, which is considerably higher than the approximately 50% prevalence rate observed at most sleep centres. More significantly, BMI alone had a diagnostic sensitivity and specificity of 93% and 74% respectively. Virtually all published articles report a much lower predictive value for BMI, if indeed predictive at all<sup>24,29,55</sup>. Selection bias is an important concern. Furthermore, clinical application of their model requires an arithmetic combination of a number of predictive variables with a variety of coefficients; a process which is

cumbersome, and unlikely to be used in routine practice. Nevertheless, the study represents one of the few systematic evaluations of physical examination-based measurements in OSA.

### **1.3.5 Upper airway scoring systems in anaesthesia**

Anaesthetists have developed and validated a variety of simple physical examination-based scoring systems that predict intubation difficulty through the assessment of upper airway structures. Several of these diagnostic approaches may have applicability to the OSA setting.

A “crowded” oropharynx may predispose individuals to intubation difficulties. Mallampati et al describe a scale for estimating relative tongue size<sup>37</sup>. The Sampsoon-Young classification is a refinement of the original Mallampati scale, and based on retrospective evaluation of over 1300 patients, is predictive of intubation difficulty<sup>38</sup>. It biologically is plausible that the soft tissue overshadowing of the larynx, which leads to intubation difficulty, may also predispose patients to OSA.

Wilson et al examined a number of measurements and rating systems during the process of developing a prospectively validated prediction index for intubation difficulty<sup>39</sup>. Significant predictors included body weight, head and neck movement, jaw movement, receding mandible and buckteeth. Other researchers have found a number of other clinical predictors: sternomental distance, thyromental distance, interincisor gap, mandibulohyoid distance, and mandibular angle<sup>40,41</sup>. In general, measurement of these characteristics is associated with a high level of inter-rater reliability<sup>42</sup>.

Davies and Eagle introduced the MOUTHS assessment protocol as a means of standardizing the approach to physical examination of the upper airway<sup>43</sup>. This protocol incorporates many of the predictive variables described above, and integrates them into a streamlined approach to physical examination.

## **1.4 Rationale and study objectives**

### **1.4.1 Rationale**

The obstructive sleep apnea syndrome is a relatively common condition in North America. Excessive daytime sleepiness, its most common symptom, may result in decreased quality of life, impaired performance, and an increased risk of automobile collisions. Moreover, there is limited evidence linking OSA with cardiovascular morbidity. Treatment of OSA with CPAP improves sleep architecture, reduces daytime sleepiness, and improves daytime performance and quality of life. Given the prevalence of OSA, its associated morbidity, and the effectiveness of treatment, identification of patients with OSA is an important health issue.

OSA has been traditionally diagnosed using overnight PSG, which is costly in terms of personnel, time and money. A number of portable monitors have been developed as alternatives to PSG. Decision rules also have particular appeal as diagnostic instruments because of their low cost. Current decision rules employ historical features and basic anthropomorphic measurements. In general, specificities are relatively high, but sensitivities are considerably less than 80%, thus limiting their use as stand-alone diagnostic instruments. A recent morphometric model had an OSA diagnostic sensitivity



and specificity of 98% and 100%, respectively, however, selection bias was a serious concern. Nevertheless, the study illustrated that physical examination-based decision rules may have sufficiently high performance characteristics to influence clinical decision making.

However, research on the diagnostic performance of physical examination in OSA is limited. By contrast, anaesthetists have developed a number of validated upper airways scoring systems. In particular, the MOUTHS assessment protocol provides a streamlined approach to physical examination of the upper airway.

There is a need for a standardized approach to the diagnosis of OSA, based on physical examination measurements. Such an approach would be cost-effective in screening patients at risk for OSA. Furthermore, physical examination-based measures are less invasive than traditional approaches to diagnosis such as PSG.

#### **1.4.2 Objective**

To develop a physical examination-based decision rule that will accurately identify patients at risk for OSA, and as importantly, identify patients who do not require further diagnostic testing.

## CHAPTER 2: METHODS

### **2.1 Sampling frame**

***Recruitment source:*** Alberta Lung Association (ALA) Sleep Centre

***Inclusion criteria:*** All referrals

***Exclusion criteria:***

1. Refusal to undergo SNORESAT evaluation
2. Previous assessment for a primary sleep disorder, or a specific referral for a sleep disorder other than OSA
3. Insomnia (and no suspicion of an underlying sleep disorder)

The Alberta Lung Association Sleep Centre is the major sleep centre in Southern Alberta, and draws from a wide variety of referral sources: family doctors, internists, otolaryngologists, and anaesthetists. The accrual population consisted of all referrals to the study investigators. Referrals to the sleep centre are received by fax, and subsequently assigned to one of five physicians by the sleep centre coordinator. The two physicians participating in the study managed approximately 40% of all patients seen at the sleep centre during the study period. There was no obvious reason to suspect systematic bias in patient allocation to individual sleep physicians, however, the potential for referral bias exists.

All patients referred to the two participating investigators, who did not meet the exclusion criteria, underwent a directed clinical assessment followed by SNORESAT monitoring. Aside from the addition of the upper airway physical examination protocol (UAPP) during initial assessment, no deviation from the usual standard of care existed. The Conjoint Ethics Committee of the University of Calgary approved the protocol.

Exclusion criteria were set up to prospectively identify patients who would not undergo SNORESAT evaluation. Patients in whom a primary sleep disorder other than OSA was suspected were sent directly for full polysomnography to establish a diagnosis. Similarly, patients with insomnia, and where another sleep disorder was not suspected, did not undergo further diagnostic testing. The diagnostic criteria for insomnia and other primary sleep disorders are summarized in the International Classification of Sleep Disorders (ICSD)<sup>44</sup>.

## **2.2 The Upper Airway Physical Examination Protocol (UAPP)**

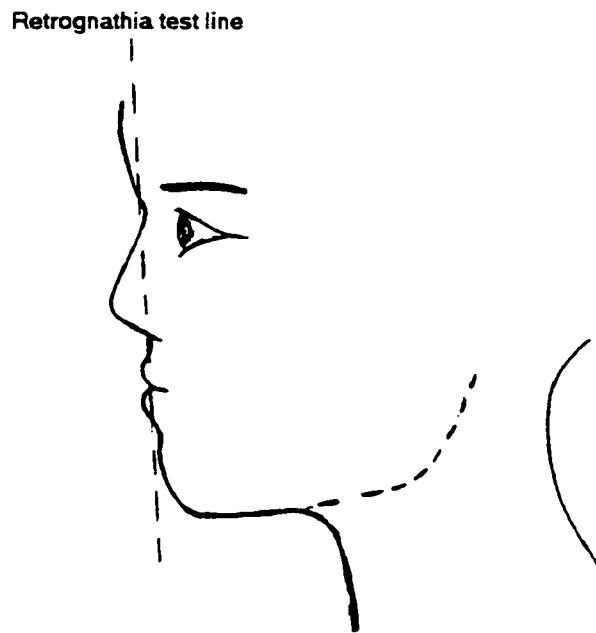
The UAPP is a structured physical exam protocol, modeled initially after MOUTHS<sup>43</sup>. The UAPP prototype was organized around the MOUTHS physical exam trait groupings: **Mandibular measurements, Opening (i.e., pharyngeal space), Uvula, Teeth, Head movement, and (body) Silhouette**. Three UAPP versions were used during the study. The UAPP-P (prototype version) was employed during the feasibility phase, UAPP-F (final version) for model development, and UAPP-SF (short form) for decision rule validation. Each UAPP version underwent progressive item reduction and thus the number of measurements and the groupings differed between versions. However, the individual measurement techniques remained the same. A description of all UAPP versions is provided in Appendix A.

"Mandibular" measurements consisted of: maximum mandibular advancement, mandibular length, thyro-mental, sterno-mental, temporal mandibular joint (TMJ)-ramus, ramus-ramus, thyro-rami, and mastoid-medial clavicle distance. Distances were

determined by using a measuring tape to take the linear distance between two bony points. *Thyro-* measurements were taken from the thyroid notch. *Mental-* measurements were taken from the posterior aspect of the inner mentum. *Sternal-* measurements were taken from the sternal notch. Mandibular length refers to the distance between the posterior ramus and the inner mentum.

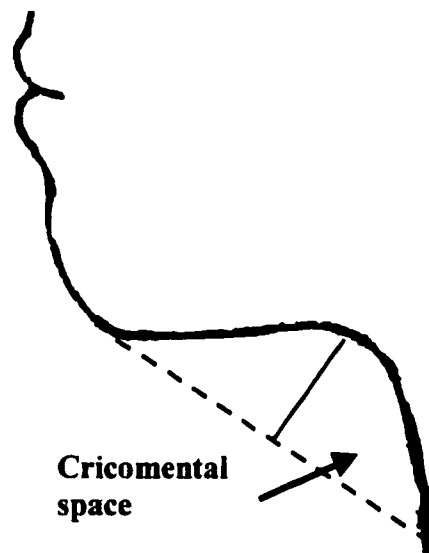
The facial profile was categorized as retrognathic, neutral, or prognathic (Figure 1). To classify a profile, an imaginary line was created, joining the brow and maxilla. If the anterior chin was behind the line, retrognathia was said to exist. If the chin lay in front of the line, prognathia was present.

The cricomental space was determined by using a thin ruler to connect the cricoid cartilage to the inner mentum. The cricomental line was bisected, and the perpendicular distance to the skin of the neck was measured (Figure 2). The use of a thin ruler ( $\leq 1\text{mm}$ ) was considered essential, because thicker straight edges (e.g., tongue depressors) may influence the measurement.



*Figure 1: Assessment of facial profile*

An imaginary line is created, joining the brow and maxilla. If the anterior chin is behind the line, *retrognathia* is said to exist. If the chin lies in front of the line, *prognathia* is present.



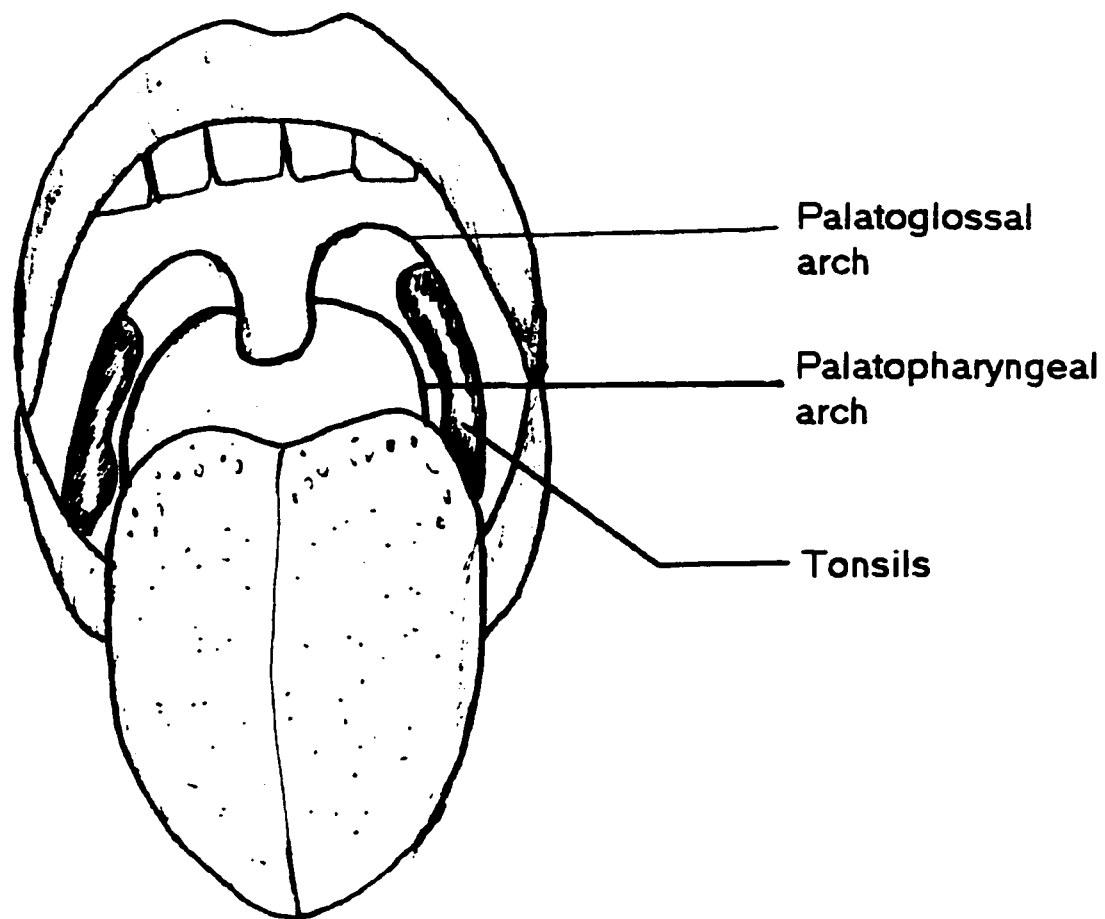
*Figure 2: Assessment of the cricomenal space*

Use a thin ruler to connect the cricoid cartilage to the inner mentum. The cricomenal line is bisected, and the perpendicular distance to the skin of the neck is measured

The extent of tonsillar enlargement (tonsillar grade) was assessed using a 4 point ordinal scale: Class I: tonsils absent, Class II: tonsils do not extend beyond the palatopharyngeal arch, Class III: tonsils are at the palatopharyngeal arch, Class IV: tonsils extend beyond the palatopharyngeal arch. Palatopharyngeal anatomy is illustrated in Figure 3.

Tongue size was assessed using the Sampsoon-Young classification system<sup>38</sup>. This scoring system depends on the relative position of the uvula and soft palate to the base of the tongue. Grading is summarized in Figure 4.

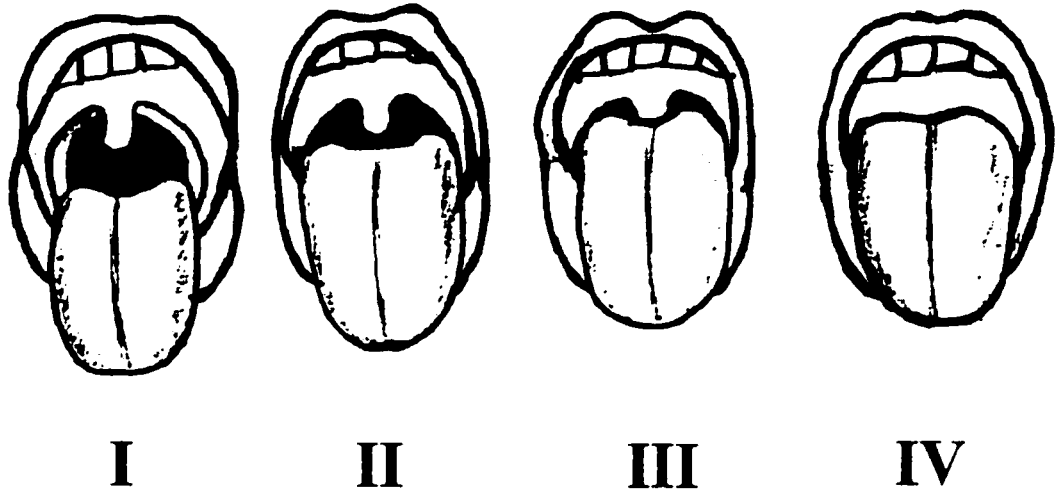
The pharyngeal space (pharyngeal grade) was also assessed using a 4 point ordinal scale: Class I: palatopharyngeal arch intersects at the edge of the tongue, Class II: palatopharyngeal arch intersects at  $\geq 25\%$  of the tongue diameter, Class III: palatopharyngeal arch intersects at  $\geq 50\%$  of the tongue diameter, Class IV: palatopharyngeal arch intersects at  $\geq 75\%$  of the tongue diameter (Figure 5).



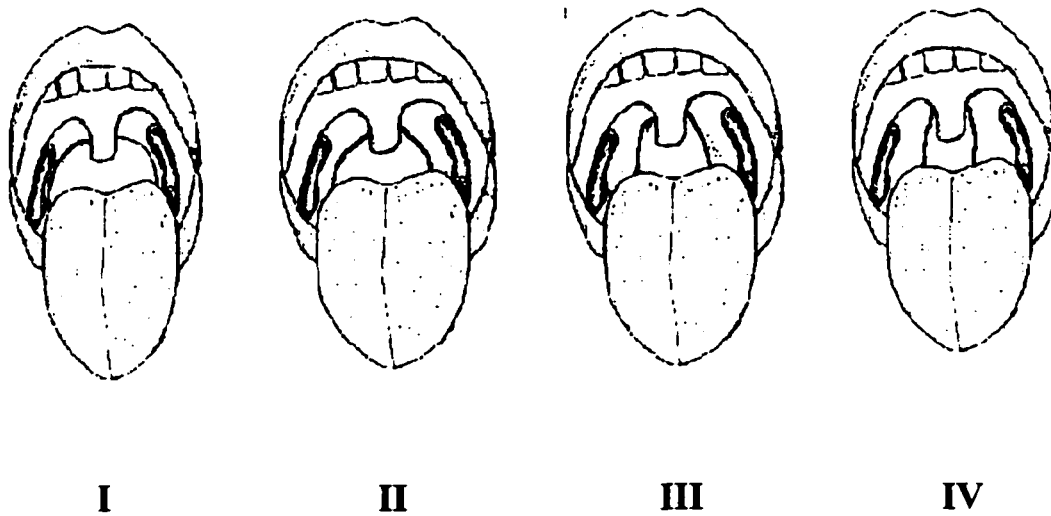
*Figure 3: Palatopharyngeal anatomy*

The extent of tonsillar enlargement (tonsillar grade) was assessed using a 4 point ordinal scale: Class I: tonsils absent, Class II: tonsils do not extend beyond the palatopharyngeal arch, Class III: tonsils are at the palatopharyngeal arch, Class IV: tonsils extend beyond the palatopharyngeal arch.





*Figure 4: Sampson-Young grading system for tongue size*



*Figure 5: Pharyngeal grading system*

Class I: palatopharyngeal arch intersects at the edge of the tongue

Class II: palatopharyngeal arch intersects at  $\geq 25\%$  of the tongue diameter

Class III: palatopharyngeal arch intersects at  $\geq 50\%$  of the tongue diameter

Class IV: palatopharyngeal arch intersects at  $\geq 75\%$  of the tongue diameter

## **2.3 Feasibility phase**

### **Goals:**

#### ***To determine:***

- (1) Whether clinicians were capable of performing the upper airways physical examination protocol (UAPP).
- (2) Whether the results obtained by the examiners were comparable
- (3) UAPP completion time

The emphasis was on developing an instrument acceptable to "real world" clinicians. Measurements that were considered cumbersome or excessively time consuming were eliminated, given that clinicians would be unlikely to use them in every day practice. The predictive value of each measurement was not assessed during this phase of decision rule development.

**Methods:** Twenty patients underwent routine clinical assessment, plus the upper airway physical examination protocol (UAPP-P, Appendix A), performed by one of two investigators. Both investigators independently assessed a randomly selected (n=15) subgroup of patients. Unreliable or time-consuming measurements were eliminated from the UAPP-P based on a consensus view.

## **2.4 Index development**

The development of any measurement instrument involves the following steps: item selection, item reduction, and determination of reliability, validity, and responsiveness<sup>45</sup>.

### **2.4.1 Item selection**

At the time of UAPP development, the morphometric measurements of Kushida et al were unavailable<sup>36</sup>. Aside from neck circumference and body mass index, a review of the OSA literature revealed little information on the predictive value of physical examination measurements for diagnosing OSA. Consequently, the selection of measurement variables was based on expert opinion and published upper airway physical exam scoring systems. For the purposes of model development, we also included clinical (historical) predictors of OSA: hypertension, habitual snoring, nocturnal choking/gasping, witnessed apneas, age, alcohol use, and smoking history. Measurement of these variables was based on self-report or via a history obtained from the subject's "bed partner".

### **2.4.2 Index reliability**

Reliability (or precision) is the degree of stability exhibited when a measurement is repeated under identical conditions<sup>59</sup>. Inter-rater reliability for categorical outcomes is often assessed using the kappa statistic<sup>60</sup>. This statistic takes into account agreement occurring by chance. Test-retest reliability refers to the reproducibility of a measure when

repeated on the same subject at two different points in time. This type of reliability may be influenced by changes in the subject's condition or external environment. In this study, where fixed anatomic measurements were evaluated, it was considered unlikely that anatomic characteristics would change either with treatment, disease status, or time.

Inter-rater reliability is usually assessed during the feasibility phase and also following development of the final model. The former facilitates item reduction, since unreliable measurement variables can be eliminated at the outset. The latter assessment ensures that the predictive outcome of the model is consistent across observers and time. We did not formally test reliability during the feasibility phase. The measure of agreement (kappa statistic) is dependent on both the number of categories being tested, and the prevalence of disease. Given the large number of variables tested during the feasibility phase, formal statistical testing of reliability would have been of questionable significance.

Consequently, only the final decision rule was subjected to formal statistical testing of reliability. Although this approach may have resulted in incomplete item reduction, it does not detract from the usefulness of the final model. From a clinical perspective, only the performance of the final decision rule is relevant.

### **2.4.3 Measures of validity**

A valid scale is one that measures what it says it is measuring<sup>45</sup>. Classically, validity has been assessed using the trinitarian standard of: (i) content validity (ii) criterion validity and (iii) construct validity.

(i) *Content validity* is a subjective judgment that the instrument appears to be measuring desired qualities. Content validity is usually determined by a panel of experts who are well versed in the subject literature.

(ii) *Criterion validity* measures the extent to which a scale produces similar results when compared with the current gold standard. In this study, OSA is defined using Snoresat-determined RDI as the gold standard. This will be discussed in more detail in the statistics section (see section 2.5).

(iii) *Construct validity* describes the extent to which a particular measure relates to other measures, under the assumption that the measures are associated according to theoretically derived hypotheses or concepts (constructs). Construct validity is employed in settings where a gold standard does not exist, e.g. quality of life. For example, we would expect the quality of life in patients with asthma to improve as objective measures of pulmonary function also improve, hence these objective measures form constructs against which quality of life is assessed. In this study, a gold standard instrument for diagnosing OSA already existed (i.e. SNORESAT-derived RDI), therefore the assessment of construct validity was considered unnecessary.

#### **2.4.4 Responsiveness**

Responsiveness is the ability of an instrument to change concurrently with changes in patient status<sup>61</sup>. The decision rule developed in this study was based predominantly on fixed anatomic measurements, so there was little expectation of change with treatment or time. Moreover, since the purpose of the decision rule was to diagnose OSA at a fixed point in time, temporal changes were less relevant. Consequently, an assessment of index responsiveness was not done.

### **2.5 Development of the Decision Rule**

Following the feasibility phase, all subsequent patients underwent assessment with the UAPP-F (final version), followed by home monitoring with SNORESAT for determination of their RDI. Data were collected prospectively, and a decision rule was developed using two techniques:

1. Multiple logistic regression
2. Recursive partitioning (Classification and Regression Trees, CART)

The initial *decision rule* was developed using multiple logistic regression, which was then compared to the *decision tree* generated by recursive partitioning.

### **2.5.1 Logistic Regression**

Predictors of OSA were identified by simple logistic regression, using a diagnosis of OSA ( $RDI \geq 10 \text{ hr}^{-1}$ ) as the dependent variable. The predictive model was then developed using two approaches:

- (1) "significant p" approach: Automated stepwise reduction on a full model consisting of variables identified as predictive by simple logistic regression ( $p < 0.10$ ).
- (2) "biologically plausible" approach: Investigator-driven construction of a model using all known and suspected predictors of OSA.

A parsimonious model is one with the fewest number of independent variables that does not differ significantly in its predictive ability from models with more variables. Differences in predictive ability between models are assessed using the likelihood ratio test, with p values  $> 0.05$  considered to be non-significant.

During manual reduction, the parsimonious model was obtained by investigator-driven reduction of variables. Items were eliminated based on an expectation of low predictive value, either from a statistical (i.e. low odds ratio) or biologically plausible standpoint. This approach to model building was then compared with the automated stepwise reduction procedure. The manual approach to model building was conducted prior to the automated stepwise reduction procedure.

Continuous variables that were identified as predictive in the parsimonious model were cross-tabulated against a diagnosis of OSA, and cut-points were visually selected.



All independent predictors were thus modeled as dichotomous variables. For comparison, CART was used to automatically determine cut-points (see section 2.5.3)

The data were also analyzed using an RDI diagnostic criterion value of  $\geq 15 \text{ hr}^{-1}$  to define OSA to determine if any new predictors or parsimonious models were identified.

Classification and Regression Tree (CART) software from S-Plus 4.0 (MathSoft, Cambridge, MA) was used for modeling the decision tree. All other statistical analyses were performed using Stata 5.0 (Stata Corporation, College Station, TX).

### **2.5.2 Assessment of criterion validity**

A decision rule was created using the binary predictors derived from the logistic regression model. Sensitivity, specificity, positive and negative predictive values were then determined. The diagnostic performance of the decision rule was compared with the diagnostic performance of the tree model.

Sensitivity is defined as the number of patients with a true positive test divided by the number of patients with the disease (ie. OSA). Specificity is defined as the number of patients with a true negative test divided by the number of patients without the disease. Positive predictive value is defined as the number of patients with a positive test and the disease divided by the number of positive tests. Negative predictive value is the number of patients with a negative test and no disease divided by the number of negative tests.

### **2.5.3 Recursive Partitioning for Decision Rule development**

Recursive partitioning is a type of regression analysis used to develop binary prediction trees. Tree-based modeling is useful for developing prediction rules, identifying screening variables, determining the adequacy of linear models, and summarizing large multivariate data sets<sup>48</sup>.

The strategy involves splitting complex data sets into progressively smaller subgroups, while a computationally intensive algorithm produces a sequence of increasingly “pure” binary splits, termed “nodes”. The data are recursively split until either each node is homogenous or contains too few observations (usually  $\leq 5$ ). A pruning algorithm then cuts off branches of the tree that impair overall accuracy. In other words, partitions resulting in the greatest reduction in deviance (i.e. the ability of a model to predict an outcome when compared to a model with perfect prediction) are removed. It follows that increasing the size of a tree will increase the accuracy because of increased degrees of freedom, but in so doing, the data becomes increasingly imprecise because of increased variance. Therefore, optimal deviance occurs when the tree size is such that an increase in accuracy is offset by decreased precision.

“Terminal nodes” form the base of the tree and contain the two possible predicted outcomes of interest (e.g. OSA/ no OSA). The predictive ability of the decision tree is determined by algebraically combining the “terminal nodes” and reducing them to a “simple” Boolean expression using a computationally intensive process<sup>46</sup>.

Classification and Regression Tree (CART) software by Breiman et al is revolutionary, in that it automates the laborious process of recursive partitioning<sup>47</sup>. Moreover, it introduces a penalty function for tree complexity to offset the increasing purity of subgroups as the sample size decreases.

Recursive partitioning has a number of theoretical advantages over other multivariate analytic approaches such as logistic regression. These are summarized in Table 2. From a practical standpoint, the use of boolean, rather than arithmetic expressions, mimics the clinical decision-making process employed by physicians; thus tree models are often easy to understand. Clinicians look for the presence or absence of a variety of conditions or states, integrate these findings, and then decide on an appropriate course of action. By contrast, logistic regression models frequently require the cumbersome and computationally difficult process of combining weighted coefficients.

*Table 2: A comparison of recursive partitioning with logistic regression*

**RECURSIVE PARTITIONING AS AN ALTERNATIVE TO LOGISTIC MULTIVARIATE ANALYSIS**

**FAVOURABLE CHARACTERISTICS:**

1. The interpretation is intuitive, and mimics the clinical decision making process.
2. It can easily identify synergistic interactions.
3. May identify nonlinear relationships (i.e. non-additive behavior).
4. Provides a simple format for constructing "homogeneous" risk strata, or for the detailed matching of patients.
5. Provides more satisfactory treatment of missing values.
6. Allows easy interpretation when predictors are a mix of numeric variables and factors.

**UNFAVOURABLE CHARACTERISTICS:**

1. May miss additional predictive factors during the later stage of selection process.
2. May increase the problem of "multiple testing".
3. May increase the problem of "over-training".
4. May not account for the full predictive ability of a continuous factor.

- adapted from Cook and Goldman<sup>48</sup>

## **2.6 Reliability assessment**

The reliability of the decision rule was tested in twenty patients distinct from those used during other phases of the study. The sample consisted of patients referred to one of three sleep centre physicians. Each patient underwent two independent evaluations using the UAPP-SF. An unweighted kappa statistic was used to assess inter-observer agreement for each measurement.

## **2.7 Decision Rule validation**

The final predictive model was validated on an “all SNORESAT referral” population, distinct from the sample used for model development. This population consisted of a consecutive series of patients referred for SNORESAT diagnostic testing at the ALA sleep centre.

SNORESAT referrals are completely at the discretion of physicians with referral privileges (sleep physicians (n=5), otolaryngologists (n=3)). Although over 90% of the referrals are from sleep centre physicians, validation of the decision rule in the “all SNORESAT referral” sample alleviates some of the concerns with respect to patient allocation bias amongst study investigators.

Sensitivity, specificity, positive and negative predictive values were determined and compared with the values obtained from the model development sample.

## **2.8 Sample size determination**

Sample size calculations were based on a minimum event per variable (EPV). Using a simulation study of forward stepwise multiple linear regression, Freedman and Pee demonstrated a significant increase in Type I error when the EPV was less than 4<sup>49</sup>. More recently, Peduzzi and Feinstein performed a Monte Carlo simulation to determine the optimum EPV in multiple logistic regression<sup>50</sup>. For EPV  $\geq 10$ , no major problems occurred. Moreover, this appeared to be a “threshold” effect. In other words, increasing the EPV above 10 did not have a dramatic effect on the validity of the logistic regression results.

However, this is a conservative estimate. Simulation studies indicate that an EPV  $\geq 10$  represents an upper limit. For example, other authors have used an EPV  $\geq 4$ , and the Peduzzi and Feinstein simulation suggests that an EPV  $\geq 5$  is acceptable for some measures. Therefore, based on the use of 15 variables, and an EPV  $\geq 4$ , approximately 60 events were required (i.e. 60 patients diagnosed with OSA).

## CHAPTER 3: RESULTS

### **3.1 Feasibility assessment and item reduction**

Twenty consecutive patients were assessed using the upper airway physical examination protocol (UAPP-P, Appendix A). It was considered feasible to perform all measurements, but the complete protocol (UAPP-P) was cumbersome and time-consuming. Because the UAPP had to be acceptable to bedside clinicians, items were removed based on the subjective impression of unreliability or excessive complexity. Consensus agreement between clinicians was used to select items for removal. No formal statistical testing was performed during the feasibility assessment.

Head movement measurements were eliminated because of time constraints. Chin protrusion was eliminated because of difficulty in mastering the technique and the subjective impression of unreliability. Assessment of the cricomental space replaced chin protrusion. Pharyngeal space measurements were converted to a 4 point ordinal scale for greater ease of assessment. Aside from the thyro-mental and sterno-mental distances, “mandibular” measurements were assessed in the neutral head position after it became apparent that there was no difference between measurements with the head in the neutral position or with full neck extension.

The reduced UAPP-F (final version) was then used for decision rule development. Physical examination measurements included: mandibular length, thyro-rami distance, mastoid-medial clavicle distance, TMJ-rami distance, rami-rami distance, thyro-mental distance (neutral position and with neck extended), thyro-mental displacement, sternal-mental distance (neutral position and with neck extended), sternomental displacement,

inter-incisor distance, cricomental space, mandibular advancement, facial profile, pharyngeal class, Sampsoon-Young classification, presence of overbite or overjet.

### **3.2 Sampling frame**

A total of 99 patients were evaluated, with 75 patients eligible for study. Of the 24 excluded patients, 14 met the International Classification of Sleep Disorders (ICSD) criteria for insomnia and did not undergo diagnostic testing. Ten patients proceeded directly to polysomnography because they presented with symptoms suggestive of a primary sleep disorder other than OSA: restless leg syndrome/ periodic leg movement syndrome (n=6), idiopathic hypersomnolence (n=2), severe COPD (n=1), and narcolepsy (n=1). None of the excluded patients, who underwent PSG, had a diagnosis of OSA.

The patient population was predominantly middle-aged, male, and obese. Patient characteristics are summarized in Table 3. As expected, the patients reported excessive daytime sleepiness, as assessed by a mean Epworth Sleepiness Score (ESS) of 11.7. The ESS is a self-administered questionnaire that assesses the patients predisposition towards falling asleep. An ESS score of greater than 6 is considered abnormal<sup>61</sup>.

Clinical characteristics and physical examination findings are presented in Tables 4 and 5, respectively.



*Table 3: Summary of patient characteristics (n=75)*

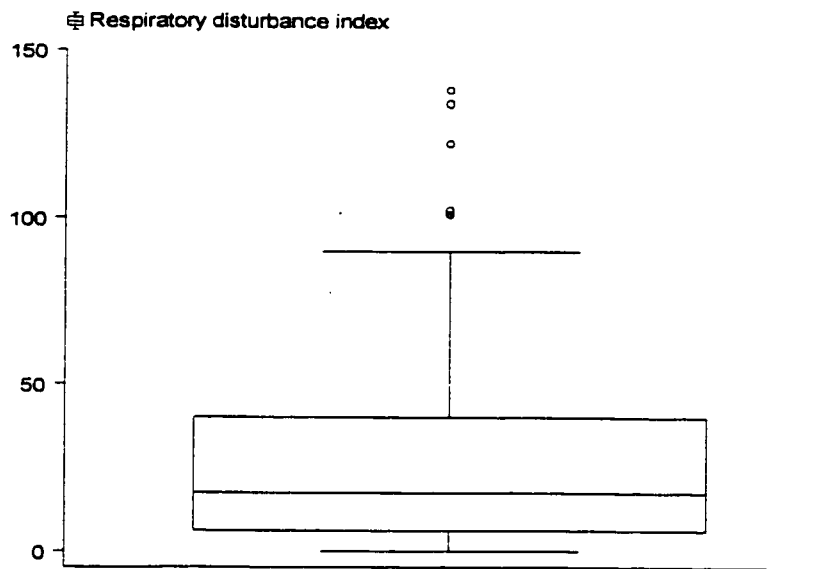
Patient Characteristic	Mean	Standard Error	Range
Age	47.5	1.33	26-74
Sex (M/F)			75% / 25%
Body mass index (BMI)	33.1	0.83	19-51
Neck circumference (cm)	42.1	0.56	30-58
Epworth Sleepiness Scale	11.7	0.71	0-22
Respiratory disturbance index (hr <sup>-1</sup> )	16.0	1.17	0-138

**Table 4: Summary of clinical characteristics (n=75)**

Characteristic	N (%)
Snoring (yes)	68 (91%)
Choking sensation (yes)	29 (39%)
Witnessed apneas (yes)	48 (64%)
Hypertension (yes)	14 (19%)
Alcohol use (yes)	19 (25%)
Smoker	
Yes	9 (12%)
Ex	17 (23%)
No	49 (65%)

The respiratory disturbance index was skewed in distribution, with a mean value of  $30 \text{ hr}^{-1}$  and a median value of  $17 \text{ hr}^{-1}$ . The RDI distribution for the group is graphically demonstrated in a box plot (Figure 6).

*Figure 6: Distribution of the respiratory disturbance index (RDI) in the model development population (n=75)*



*Table 5: Summary of physical examination characteristics (n=75)*

Characteristic	Mean	Standard Error (SE)	Range
Mandibular length (cm)	11.9	0.15	9-15
Thyro-rami distance (cm)	11.2	0.15	8-14
Mastoid-medial clavicle (cm)	18.4	0.20	14-22
TMJ-rami distance (cm)	5.47	0.13	3-9
Rami-rami distance (cm)	9.65	0.14	7.5-15
Thyro-mental distance (cm)			
Neutral position	6.18	0.14	4-12
Neck extended	7.60	0.20	4.5-7
Thyromental displacement (cm)	1.42	0.11	0-5
Sternal-mental distance (cm)			
Neutral position	12.5	0.25	8-17
Neck extended	16.8	0.29	11-23
Sternomental displacement (cm)	4.31	0.21	1-9
Inter-incisor distance (cm)	5.76	0.08	4-7.5
Cricomental space (cm)	0.56	0.10	0-4

**Table 5 continued: Summary of physical examination characteristics (n=75)**

Characteristic	N	%
<b>Mandibular advancement</b>		
0-5 cm	11	15
5.1-10 cm	14	19
10.1-15 cm	18	24
>15 cm	32	42
<b>Profile</b>		
Neutral	33	45
Retrognathia	27	36
Prognathia	14	19
<b>Pharyngeal class</b>		
Neutral		
Class I	17	23
Class II	17	23
Class III	14	18
Class IV	27	16
With phonation		
Class I	16	21
Class II	18	24
Class III	14	19
Class IV	27	36
<b>Sampsoon-Young (I / II / III /IV)</b>		
Neutral		
Class I	6	8
Class II	14	19
Class III	12	16
Class IV	43	57
With phonation		
Class I	28	37
Class II	10	13
Class III	27	37
Class IV	10	13
<b>Overbite (yes/no)</b>	39	52
<b>Overjet (yes/no)</b>	36	48

### **3.3 Logistic regression: univariate predictors of OSA**

In the model development cohort of 75 patients, the prevalence of OSA was 81%, 63%, 57%, or 44%, depending on whether an RDI diagnostic criterion value of greater than 5, 10, 15, or 20 hr<sup>-1</sup> was employed.

Simple logistic regression was performed using clinical and physical examination features as independent variables. The presence of OSA (yes/no) was the dependent variable. OSA was defined by an RDI  $\geq 10$  hr<sup>-1</sup>. A variable was considered predictive if the p-value was <0.10.

The following clinical features were identified as predictive of OSA: age, snoring history, witnessed apneas, and hypertension. The physical examination measurements predictive of OSA were: body mass index, neck circumference, mandibular length, thyro-ramus distance, thyro-mental displacement, sterno-mental displacement, cricomental space, pharyngeal grade, Sampsoon-Young class, and overbite. These univariate results are summarized in Table 6, with predictive variables highlighted in bold. Odds ratios and 95% confidence intervals are also displayed.

The data were also analyzed using an RDI diagnostic criterion value of greater than 15 hr<sup>-1</sup> to define OSA. No new predictive variables were identified.

Table 6: Univariate logistic regression analysis of predictive variables for OSA

Variable	Odds ratio	P-value	Confidence Interval <sub>95%</sub>
<b>Age (years)</b>	<b>1.10</b>	<b>0.001</b>	<b>[1.03, 1.16]</b>
Epworth Sleepiness Scale	1.03	0.558	[0.93, 1.13]
<b>Snoring history</b>	<b>12.5</b>	<b>0.023</b>	<b>[1.42, 110.6]</b>
Choking episodes	2.02	0.169	[0.74, 5.49]
<b>Witnessed apneas</b>	<b>3.37</b>	<b>0.016</b>	<b>[1.25, 9.06]</b>
<b>Hypertension</b>	<b>10.3</b>	<b>0.029</b>	<b>[1.27, 83.9]</b>
Alcohol use	1.20	0.658	[0.53, 2.74]
Smoker	1.28	0.482	[0.64, 2.56]
<b>Body mass index (kg/m<sup>2</sup>)</b>	<b>1.13</b>	<b>0.009</b>	<b>[1.03, 1.24]</b>
<b>Neck circumference (cm)</b>	<b>1.36</b>	<b>0.000</b>	<b>[1.15, 1.61]</b>
Mandibular advancement (cm)	0.69	0.107	[0.43, 1.08]
<b>Mandibular length (cm)</b>	<b>1.83</b>	<b>0.005</b>	<b>[1.20, 2.79]</b>
<b>Thyro-rami distance (cm)</b>	<b>1.59</b>	<b>0.020</b>	<b>[1.07, 2.35]</b>
Mastoid-medial clavicle (cm)	1.25	0.129	[0.94, 1.65]
TMJ-ramus distance (cm)	1.39	0.164	[0.88, 2.19]
Ramus-ramus distance (cm)	0.97	0.891	[0.67, 1.42]
Thyro-mental (neutral, cm)	1.23	0.359	[0.79, 1.90]
<b>Thyromental displacement (cm)</b>	<b>0.59</b>	<b>0.059</b>	<b>[0.35, 1.02]</b>
Sterno-mental (neutral, cm)	0.86	0.180	[0.68, 1.07]
<b>Sternomental displacement (cm)</b>	<b>0.75</b>	<b>0.041</b>	<b>[0.57, 0.99]</b>
Retrognathia	0.89	0.706	[0.48, 1.65]
<b>Cricomental space (cm)</b>	<b>0.15</b>	<b>0.000</b>	<b>[0.06, 0.38]</b>
Tonsillar grade (I-IV)	0.85	0.415	[0.57, 1.26]
<b>Pharyngeal grade (I-IV)</b>	<b>1.52</b>	<b>0.046</b>	<b>[1.01, 2.30]</b>
<b>Sampsoon-Young class (I-IV)</b>	<b>1.77</b>	<b>0.018</b>	<b>[1.10, 2.86]</b>
Palatal elevation	1.41	0.303	[0.73, 2.71]
Inter-incisor distance (cm)	0.86	0.673	[0.44, 1.71]
<b>Overbite</b>	<b>2.19</b>	<b>0.044</b>	<b>[1.02, 4.70]</b>

### **3.4 Model building using logistic regression**

#### **3.4.1 "Significant p" approach**

A "significant p" model was constructed using variables identified as potentially predictive by simple logistic regression. The dependent variable was OSA (yes/no), as defined by an RDI  $\geq 10$  hr<sup>-1</sup>. Independent variables consisted of both clinical and physical exam characteristics: age, snoring history, witnessed apneas, hypertension, body mass index, neck circumference, thyro-ramus distance, sternomental displacement, cricomental space, pharyngeal grade, Sampsoon-Young class, and overbite. The full model was progressively reduced using automated stepwise reduction. A significance level of  $p=0.1$  was selected for item elimination. Using this approach, the model was reduced to 3 predictive variables: cricomental space, pharyngeal grade, and overbite. The likelihood ratio (LR) test compared the reduced model to the full "significant p" model. The parsimonious model was not significantly different from the full model ( $p=0.14$ , LR test). Identical results were obtained when an RDI  $\geq 15$  hr<sup>-1</sup> was used to define OSA.

#### **3.4.2 Biologically plausible approach**

A model was also constructed using known and suspected predictors of OSA: snoring, choking episodes, witnessed apneas, hypertension, neck circumference, body mass index, thyromental displacement, sternomental displacement, retrognathia, cricomental space, tonsillar grade, pharyngeal grade, overbite, Sampsoon-Young class, and palatal elevation (change in Sampsoon-Young class on phonation). Investigator-driven elimination was used to reduce the model. Items were removed based on an



expected minimal contribution to the overall model, either from a statistical (i.e. low odds ratio) or biologically plausible standpoint. Once again, the parsimonious model incorporated the same 3 variables, namely, cricomental space, pharyngeal grade, and the presence of overbite. The parsimonious model was also not significantly different from the full "biologically plausible" model ( $p=0.13$ , likelihood ratio test).

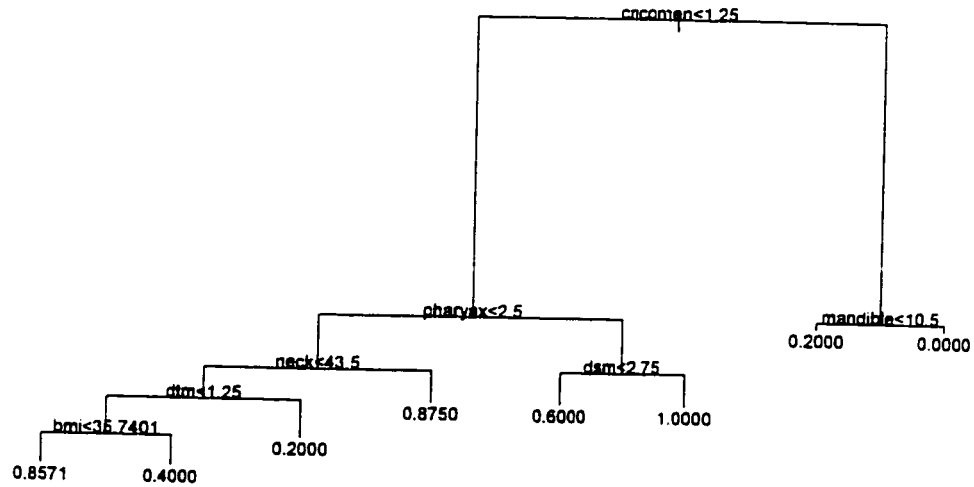
### **3.4.3 The parsimonious model**

In summary, the investigator-driven, biologically plausible approach and the automated "significant p" approach produced identical results. Regardless of the modeling approach, the final parsimonious models consisted of the same 3 predictive variables, namely cricomental space, pharyngeal grade, and the presence of overbite.

### **3.5 Model building using recursive partitioning**

#### **3.5.1 Classification and Regression Trees (CART)**

A recursive tree model was developed using the same variables employed to develop the "significant p" model, namely age, snoring history, witnessed apneas, hypertension, body mass index, neck circumference, mandibular length, thyro-ramus distance, thyromental displacement, sternomental displacement, cricomental space, pharyngeal grade, Sampsoon-Young class, and the presence of overbite. The CART algorithm was set using the following parameters: minimum number of observations before split: 5, minimum node size: 10, minimum deviance: 0.010, pruning method: deviance. The CART algorithm identified the following predictive variables: cricomental space, mandibular length, pharyngeal grade, sternomental displacement, thyromental displacement, and body mass index (Figure 7).



*Figure 7: Full decision tree for the diagnosis of obstructive sleep apnea*

Each node contains binary predictors of OSA. Branches to the left of a node means that the node determined condition (predictor) exists, while branches to the right of a node indicate the absence of the condition. Nodes at the base of the tree form terminal nodes. The probability of OSA given a series of conditions is determined by combining a series of contiguous nodes to the terminal node. For example, if a patient has a cricomenal space  $> 1.25$  cm and a mandibular length  $> 10.5$  cm, there is 0% chance of having OSA.

### **3.5.2 Estimating optimal tree size**

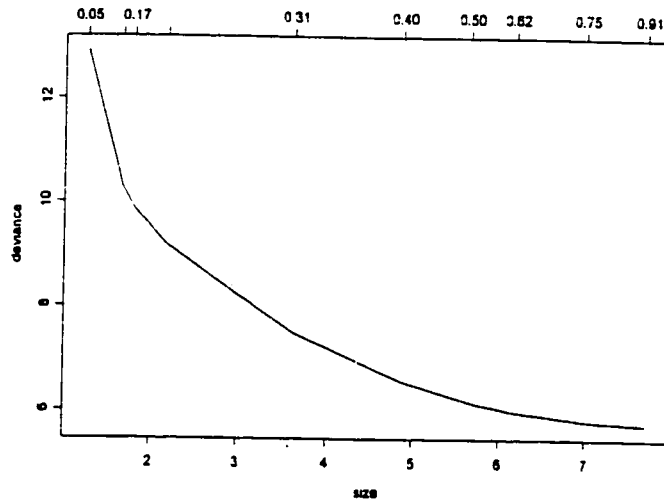
The optimal tree size is frequently a compromise between clinical acceptance and precision of the estimate. Increasing the tree size increases potential accuracy, but may lead to over-fitting, thus limiting generalizability to only the data set from which the model was derived. Moreover, the increased complexity will limit its acceptability to clinicians. The full tree illustrated in Figure 7 was deemed too large for use in clinical practice. A deviance plot was generated to estimate optimal tree size (Figure 8). Deviance (inaccuracy) decreases with increasing tree size, thus optimal tree size is also a compromise between deviance and tree complexity. The slope of the deviance plot changes drastically for tree sizes  $<2$  or  $>4$ . Therefore, a tree size of between 2 and 4 terminal nodes was thought to be ideal.

### **3.5.3 Pruning the Tree**

Using a tree size of 4, cricomental space, pharyngeal grade, neck circumference, and thyromental displacement were predictive of OSA. A tree size of 3 yielded only two predictive variables: cricomental space and pharyngeal grade.

The reduced tree model shared two predictive variables with the regression model, namely cricomental space and pharyngeal grade. The regression and recursive partitioning models differed in their use of overbite (regression model only) versus thyromental displacement and neck circumference (tree model only). However, in the pruned tree model, none of the terminal nodes contained a positive predictive value approaching 100%, nor did the tree identify a cricomental space cut-point that excluded

the possibility of OSA. Consequently, the logistic regression-derived model was used for derivation of the final decision rule.



*Figure 8: Deviance plot to estimate optimal tree size*

The curve has the greatest change in slope between tree sizes of 2 and 4. Therefore, the optimal tree size is assumed to be between 2-4 terminal nodes.

## **3.6 Developing the Decision Rule**

### **3.6.1 Optimal cut-point selection by visual inspection**

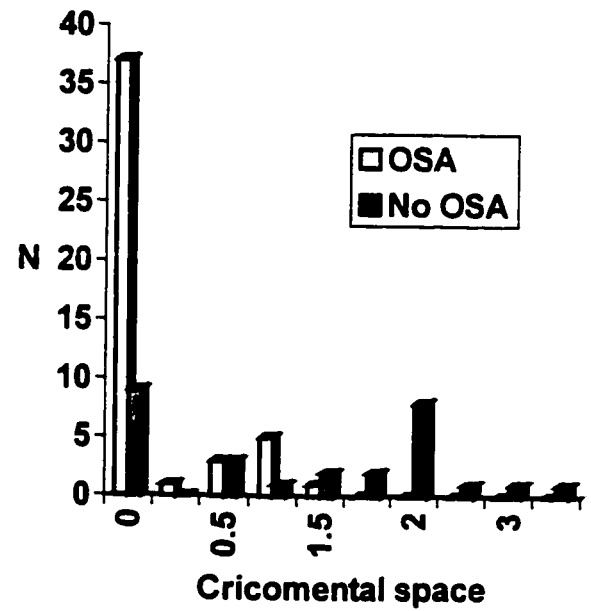
Cricomental space and pharyngeal grade were measured as continuous variables. These variables were cross-tabulated against a diagnosis of OSA, and optimal cut-points were visually selected (see Table 7). A cricomental space  $> 1.5$  cm and a pharyngeal grade  $> II$  were chosen as optimal cut-points. Consequently, the decision rule was based on three binary variables: **a cricomental space  $\leq 1.5$  cm, a pharyngeal grade  $> II$ , and the presence of overbite**. Table 7 reveals that a cricomental space  $> 1.5$  cm effectively excludes the possibility of OSA.

### **3.6.2 Optimal cut-point selection by CART**

CART automates the process of optimal cut-point determination for continuous variables. We used CART to select cut-points for predictive variables identified in the parsimonious logistic model (cricomental space, pharyngeal grade, and the presence of overbite). CART selected cut-points were similar to those prospectively selected by visual inspection of cross-tabulations. The CART approach determined optimal cut-points of 1.25 cm for cricomental space and 2.5 for pharyngeal grade. A pharyngeal grade 2.5 is equivalent to a grade of 2, given that the grading system is ordinal. Similarly, a cricomental space of  $>1.25$  cm and  $>1.50$  cm are identical, since no intervening values were measured.

*Table 7: Cross-tabulation of predictive variables against a diagnosis of OSA*

Cricomental Space (mm)	OSA +	OSA -
0	37	9
0.25	1	0
0.50	3	3
1.0	5	1
1.50	1	2
1.75	0	2
2.0	0	8
2.5	0	1
3.0	0	1
4.0	0	1



Pharyngeal Grade	OSA +	OSA -
I	8	9
II	9	8
III	10	4
IV	20	7

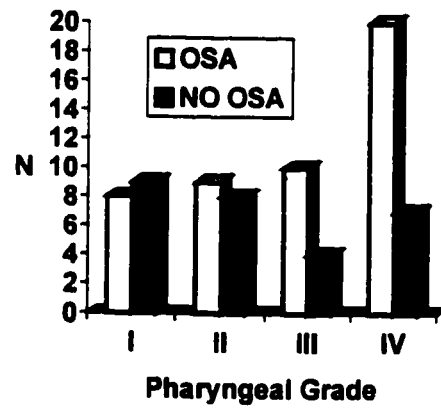
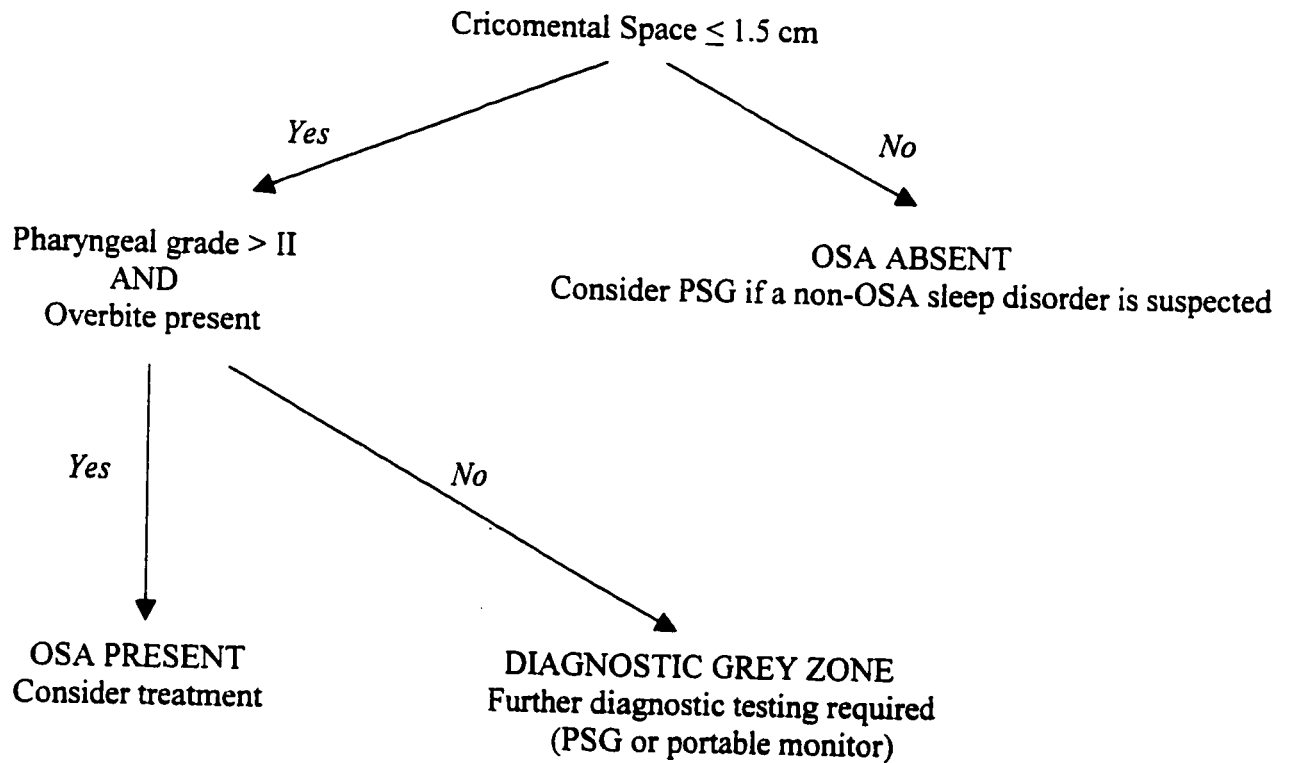


Figure 9: A decision rule for diagnostic testing in obstructive sleep apnea





### **3.7 Diagnostic performance of the Decision Rule**

The diagnostic performance of various combinations of clinical predictors is summarized in Table 8. Depending on the combination of variables, sensitivities ranged from 40-100%, with specificities between 46-96%. No single combination of variables simultaneously provided near perfect sensitivity and specificity. However, the presence of all three predictors: a narrowed cricomenal space, a high pharyngeal grade, and overbite is highly predictive of OSA (positive predictive value 95% (CI<sub>95%</sub>: 75-100%) at an RDI cut-off value of 10 hr<sup>-1</sup>). However, the negative predictive value is relatively low 49% (CI<sub>95%</sub>: 35-63). By contrast, a cricomenal space of > 1.5 cm effectively excludes the possibility of OSA (negative predictive value of 100% (CI<sub>95%</sub>: 75-100%) at an RDI diagnostic cut-off value of 10 hr<sup>-1</sup>), but is not very specific (specificity: 46% (CI<sub>95%</sub>: 26-66%)).

Because of the high specificity of the 3 variable model and the high sensitivity of the cricomenal space measurement, these two conditions formed the basis of the decision rule (see Figure 9). The decision rule will be discussed in more detail later, however, it can be summarized as follows:

- (1) A cricomenal space  $\leq$  1.5 cm, a pharyngeal grade > II, and the presence of overbite is highly suggestive of OSA.
- (2) A cricomenal space of > 1.5 cm effectively rules out the possibility of OSA.

As summarized in Table 9, the diagnostic performance of the decision rule remained stable across a range of RDI diagnostic criterion values.

*Table 8: Diagnostic performance of the predictive variables*

Variable combination	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
Cricomental & pharyngeal narrowing, & overbite	40 (27-56)	96 (82-100)	95 (75-100)	49 (35-63)
Cricomental & pharyngeal narrowing	64 (49-77)	82 (63-94)	86 (70-95)	58 (41-73)
Cricomental narrowing & overbite	60 (44-74)	75 (55-89)	80 (63-92)	52 (36-69)
Pharyngeal narrowing & overbite	40 (26-56)	93 (76-99)	90 (70-99)	48 (34-62)
Cricomental narrowing	100 (92-100)	46 (28-66)	76 (63-86)	100 (75-100)

PPV: positive predictive value

NPV: negative predictive value

Cricomental narrowing is defined by a cricomental space  $\leq 1.5$  cm

Pharyngeal narrowing is defined by a pharyngeal grade  $> II$

Table 9: Diagnostic performance of the decision rule at a variety of RDI diagnostic criterion values

OSA diagnostic criterion value	Cricomental occlusion			
	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
RDI $\geq$ 5 hr <sup>-1</sup>	90 (80-96)	50 (23-77)	89 (78-95)	54 (25-80)
RDI $\geq$ 10 hr <sup>-1</sup>	100 (92-100)	46 (26-66)	76 (63-86)	100 (75-100)
RDI $\geq$ 15 hr <sup>-1</sup>	100 (92-100)	41 (24-59)	69 (56-80)	100 (75-100)
RDI $\geq$ 20 hr <sup>-1</sup>	100 (89-100)	31 (18-47)	53 (40-60)	100 (75-100)

OSA diagnostic criterion value	Three variable model			
	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
RDI $\geq$ 5 hr <sup>-1</sup>	33 (21-46)	100 (77-100)	100 (83-100)	25 (15-39)
RDI $\geq$ 10 hr <sup>-1</sup>	40 (27-56)	96 (82-100)	95 (75-100)	49 (35-63)
RDI $\geq$ 15 hr <sup>-1</sup>	37 (23-53)	88 (71-96)	80 (56-94)	51 (37-65)
RDI $\geq$ 20 hr <sup>-1</sup>	42 (26-61)	86 (71-95)	80 (56-94)	65 (51-78)

**PPV: positive predictive value**

**NPV: negative predictive value**

### **3.8 Decision Rule reliability**

Twenty patients underwent two independent assessments using the UAPP-SF. Agreement between measurement variables identified as predictive by either the decision tree (CART) or decision rule (logistic regression model) was assessed. These variables were cricomental space > 1.5 cm, presence of overbite, presence of retrognathia, tonsil enlargement, pharyngeal narrowing (pharyngeal grade > II), and thyromental displacement. As shown in Table 10, agreement was high for all variables (kappa coefficient range: 0.58-1.00) other than retrognathia (kappa=0.22). Inter-observer agreement for decision rule measurement variables, namely cricomental space, overbite, and pharyngeal narrowing was high.

**Table 10: Inter-rater agreement on the predictive variables**

Measurement variable	Kappa
Cricomental space > 1.5 cm	1.0
Overbite present	0.61
Retrognathia	0.22
Tonsil enlargement	0.73
Pharyngeal narrowing (pharyngeal grade > II)	0.78
Thyromental displacement < 1.25 cm	0.58

### **3.9 Validation of the Decision Rule**

The diagnostic performance of the decision rule was validated in the “all SNORESAT referral” sample. Fifty consecutive patients, referred to the ALA Sleep Centre for ambulatory monitoring, were assessed using the UAPP-SF prior to diagnostic testing. Decision rule performance is summarized in Table 11. The presence of all three predictors: a narrowed cricomenal space, a high pharyngeal grade, and overbite was highly predictive of OSA. The diagnostic performance was similar to that observed in the model development cohort (positive predictive value 100% (CI<sub>95%</sub>: 63-100%) and specificity of 100% (CI<sub>95%</sub>: 84-100%) at an RDI cut-off value of 10 hr<sup>-1</sup>). Similarly, a cricomenal space of > 1.5 cm effectively eliminated the possibility of obstructive sleep apnea (sensitivity: 100% (CI<sub>95%</sub>: 88-100%), negative predictive value: 100% (CI<sub>95%</sub>: 63-100%)). In summary, the diagnostic performance of the decision rule was virtually identical in both the model development sample and the validation sample.

**Table 11: Diagnostic performance of the decision rule at a variety of RDI criterion values (in the all patient referral sample)**

OSA diagnostic criterion value	Three variable model			
	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
RDI $\geq$ 5 hr <sup>-1</sup>	22 (10-38)	100 (75-100)	100 (63-100)	31 (18-47)
RDI $\geq$ 10 hr <sup>-1</sup>	28 (13-47)	100 (84-100)	100 (63-100)	50 (34-66)
RDI $\geq$ 15 hr <sup>-1</sup>	28 (12-49)	96 (80-100)	88 (47-100)	46 (30-61)
RDI $\geq$ 20 hr <sup>-1</sup>	32 (13-57)	94 (79-99)	75 (35-100)	69 (53-83)

OSA diagnostic criterion value	Cricomental occlusion			
	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
RDI $\geq$ 5 hr <sup>-1</sup>	97 (86-100)	54 (25-81)	86 (71-95)	88 (47-100)
RDI $\geq$ 10 hr <sup>-1</sup>	100 (88-100)	38 (18-62)	69 (53-82)	100 (63-100)
RDI $\geq$ 15 hr <sup>-1</sup>	100 (86-100)	32 (15-53)	60 (43-74)	100 (63-100)
RDI $\geq$ 20 hr <sup>-1</sup>	100 (83-100)	26 (12-45)	45 (30-61)	100 (63-100)

**PPV: positive predictive value**

**NPV: negative predictive value**

## CHAPTER 4: DISCUSSION

### 4.1 Summary of results

In a consecutive series of 75 patients referred to a tertiary sleep centre, a number of predictors of obstructive sleep apnea were identified. Clinical predictors included age, snoring history, witnessed apneas, and hypertension. Physical examination-based predictors included: body mass index, neck circumference, mandibular protrusion, thyro-rami distance, sterno-mental distance, sterno-mental displacement, thyro-mental displacement, cricomental space, pharyngeal grade, Sampsoon-Young classification, and overbite. A decision rule was subsequently developed: **cricomental space  $\leq$  1.5 cm, pharyngeal grade  $>2$ , and the presence of overbite**. In patients with all 3 predictors, the decision rule had the following performance characteristics: positive predictive value: 95% (CI<sub>95%</sub>: 75-100%), negative predictive value: 49% (CI<sub>95%</sub>: 35-63%), sensitivity: 40% (CI<sub>95%</sub>: 27-56%), specificity: 96% (CI<sub>95%</sub>: 82-100%). A **cricomental space  $>$  1.5 cm** excluded the possibility of OSA (negative predictive value: 100% (CI<sub>95%</sub>: 75-100%)). Comparable performance was obtained in an unfiltered validation sample of 50 patients referred for diagnostic testing. The inter-rater reliability of decision rule measurement variables was high. This decision rule provides a simple, reliable, and accurate method of identifying patients with, and perhaps more importantly, patients without OSA.

## **4.2 Clinical predictors of obstructive sleep apnea**

Although cheap and relatively easy to administer, decision rules aimed at diagnosing OSA have been limited by insufficient sensitivity and specificity for use as diagnostic instruments<sup>24</sup>. In general, most OSA decision rules have specificities around 90%, but sensitivities considerably less than 80% (see Table 1)<sup>24-27, 53-55</sup>.

Neck circumference and BMI are the only physical examination characteristics that are consistently predictive of OSA<sup>27,29,30</sup>. In this study, both body mass index and neck circumference, were confirmed as being predictive of OSA. However, BMI may be susceptible to selection bias, given that its relative contribution to predictive models is study site dependent<sup>24,28,29,36,55</sup>. In this sample, although predictive of OSA, BMI was not of sufficient predictive value to be included in the final decision rule.

A major limitation to physical examination-based prediction is that, with the notable exception of a study by Kushida et al, previous studies only evaluated basic morphometric characteristics. Kushida et al evaluated a consecutive sample of 300 patients referred to the Stanford University Sleep Centre, a university-based tertiary care centre that receives referrals from across the United States<sup>36</sup>. Body mass index, neck circumference, and intermolar distance were identified as predictive variables. A prediction index was developed, which had a sensitivity and specificity of 98% and 100% respectively. Unfortunately, there are serious questions as to whether the model was tested on a representative sample of patients. At their centre, the prevalence of OSA was 85%, which is considerably higher than the approximately 50% prevalence rate observed at most sleep centres. More significantly, BMI had a diagnostic sensitivity and specificity



of 93% and 74% respectively. In effect, in their sample of patients, the use of BMI alone exceeded the diagnostic performance of all previous decision rules, and some portable monitors. Virtually all published articles report a much lower predictive value for BMI, if indeed predictive at all<sup>24, 29, 55</sup>. Selection bias is clearly a concern.

Furthermore, there are serious considerations concerning ease of use. A major limitation to the adoption of decision rules into routine clinical practice has been time constraints. In the primary care setting, there is little time to perform complex measurements or calculations requiring the arithmetic combination of predictive variables with a variety of coefficients. To achieve widespread acceptability, a clinical decision rule must be easy to interpret, and executable without extraneous equipment. This decision rule makes use of only 3 clinical predictors (cricomenal space, pharyngeal grade, overbite), all of which can be assessed with no more than a ruler.

In contrast to physical examination-based predictors, clinical predictors are remarkably consistent. In a randomly selected series of 180 patients referred to a tertiary sleep centre, Flemons et al reported increased neck circumference, hypertension, habitual snoring, and nocturnal gasping/ choking as being predictive of OSA (PSG-AHI  $\geq 10$  hr<sup>-1</sup>) using logistic regression modeling. Similarly, Viner et al evaluated 410 patients referred for suspected sleep apnea. Body mass index (BMI), age, male sex, and snoring were predictive of OSA (PSG-AHI  $\geq 10$  hr<sup>-1</sup>). The derived prediction model had a sensitivity and specificity of 28% and 95%, respectively<sup>23</sup>. This study confirmed the results of previous investigators. Snoring history, witnessed apneas, and hypertension have been repeatedly identified as being predictive of OSA. Although not observed in this study,

choking/ gasping episodes have also been reported to have predictive value by some investigators<sup>24,54</sup>.

Several physical examination features that have been presumed predictive of OSA, but never subjected to formal evaluation were assessed. For example, clinicians have long suspected that pharyngeal narrowing, a low-lying palate, and overbite are associated with OSA. The predictive value of pharyngeal grade, Sampsoon-Young class, and overbite supports these impressions. Moreover, these measurements have a high level of inter-rater reliability. Interestingly, despite the commonly held belief, retrognathia, tonsil size, and extent of palatal elevation (change in Sampsoon-Young classification with phonation) were not predictive of OSA. Similarly, none of the mandibular measurements were of predictive value. Other measurements such as retrognathia could not be reliably determined between investigators.

While the identification of predictors of OSA allows for a more directed patient assessment, it does not necessarily influence clinical decision-making. To be useful, a predictive variable must not only be associated with the outcome of interest, but also be of high predictive value. Many of the clinical measurements correlated with each other. As such, they were not independent predictors; and many were eliminated from the final decision rule (Figure 9). For example, although BMI was a significant predictor of OSA ( $p < 0.01$ ), it correlated with both cricomental space and OSA, and as such, did not contribute to the final decision rule.

Although both clinical and physical examination-based predictors were incorporated into the initial regression model, only physical examination-based predictors

formed the final decision rule. This suggests that for those patients referred to a tertiary sleep centre, the inclusion of clinical features adds minimal predictive value for diagnosing OSA beyond that of physical examination alone.

This result may appear surprising given that the clinical predictors had high odds ratios on univariate regression (Table 6), yet were not included in the parsimonious model. However, odds ratios derived from univariate regression are of little interpretive value when there are multiple predictors. In other words, with univariate analysis, the odds ratio is unadjusted for other predictive variables. In contrast, multiple logistic regression assesses the independent contribution of each predictive variable following adjustment for the other variables. In essence, multiple logistic regression adjusts for mathematical confounding, thus providing an “honest” assessment of the independent predictive value of each predictive variable. Similarly, an examination of the confidence intervals around each univariate odds ratio reveals a very wide range amongst clinical predictors (Table 6). For example, hypertension has an OR of 10.3, but a confidence interval of between 1.27 to 83.9. Therefore, the true OR may in fact be relatively close to 1. Finally, when independent variables are continuous, an odds ratio generated using logistic regression is not as easily interpreted as with dichotomous independent variables, since the odds ratio will change by varying the unit of measurement.

An ideal predictive test combines several independently predictive variables into a single measurement. The cricomental space may be one such predictor. It is a multi-dimensional measurement that incorporates diverse characteristics such as neck circumference, BMI, hyoid bone position, neck posture, mandibular positioning, and

possibly pharyngeal length. The integration of several independent predictive variables into a single measurement is the likely explanation for its high sensitivity.

### **4.3 Statistical considerations during Decision Rule development**

Validation involves the assessment of stochastic (statistical) validity, internal validity and external validity. Stochastic validity refers to whether the appropriate statistical tests were used, and if pertinent assumptions were taken into consideration. Internal and external validity assess the extent of study bias.

This study was considered statistically robust. The decision rule was developed using two different modeling approaches, both of which provided the same results. Also, for comparison, a CART decision tree was created, with the final results being similar to those generated by logistic regression.

#### **4.3.1 Logistic regression**

Although considered a valid statistical technique, model building using logistic regression has sometimes been likened to an art as well as a science. Depending on the approach used to build a model, different results may be obtained from the same data set. Differences in modeling technique probably explain some of the variation amongst the decision rules for OSA described in the literature.

Consequently, two different, and independent, approaches to logistic modeling were employed for decision rule development: the automated "significant p" approach using a stepwise reduction procedure, and an investigator-driven, biologically plausible

approach. Regardless of the approach used to develop the model, the final parsimonious models included the same 3 variables: cricomental space, pharyngeal grade, and presence of overbite. Identical results were also obtained when OSA was defined by an RDI  $\geq 15$  hr<sup>-1</sup>. The convergence of results argues for the robustness of the final decision rule.

#### **4.3.2 CART versus Regression: Moving forward or a step back**

The logistic regression-derived decision rule was compared to a decision tree developed using recursive partitioning. Both the tree model and the logistic regression model incorporated cricomental space and pharyngeal grade as part of the final decision rule, but differed on their inclusion of overbite (logistic regression model only) and the inclusion of neck circumference and thyromental displacement (reduced tree model only). Although similar, the final decision rule was derived using the logistic regression model because it contained predictive values (“terminal nodes”) for OSA approaching 100%, while the tree model did not.

The inconsistencies between CART and logistic modeling may be related to differences in how each statistical approach deals with data. CART takes a large data set and automatically partitions it into a series of increasingly homogenous nodes. However, recursive partitioning may miss predictive factors during the late stages of the selection process. By contrast, with directed modeling in logistic regression, clinically important variables can be identified at the outset, and weighted accordingly.

CART has a provision for assigning weights or penalties, usually for tree complexity, thus offsetting the increasing purity of subgroups as sample size decreases.

Weightings can also be arbitrarily assigned to specific variables or goals. Currently, this is a laborious process, and CART does not offer any clear advantage over the logistic regression approach.

Theoretically, CART has a major practical advantage over logistic regression, namely, the lack of need to arithmetically combine weighted coefficients. Most physicians want a simple decision rule: “given a set of conditions, what is the probability that my patient has the condition of interest?”. However, by assigning cut-points to continuous variables identified as predictive by logistic regression, a binary decision rule can also be created from a regression model.

CART may be of value in determining optimal binary cut-points. When the decision rule variables (cricomental space, pharyngeal grade, and overbite) were used as dependent variables, CART identified similar cut-points to those chosen by visual inspection of cross-tabulations. In small data sets, CART does not have any obvious advantage over simple visual inspection; however, automated cut-point selection has greater value when dealing with larger data sets.

In summary, for the purposes of identifying parsimonious models, CART does not offer any clear advantage over conventional logistic modeling. Like all automated selection procedures, CART offers a high degree of inter-rater reproducibility, but does not permit directed modeling. Nevertheless, CART may be helpful in choosing optimal cut-points during decision rule development.

#### **4.4 Decision Rule interpretation**

Previously reported decision rules for OSA attempted to simultaneously maximize sensitivity and specificity. Because of inter-patient variability in clinical features, it is unlikely that near perfect sensitivity and specificity can be achieved. In general, decision rules using clinical criteria have relatively high specificities, but sensitivities below 80%. Consequently, most investigators have abandoned clinical prediction in OSA as interesting in theory, but of limited practical value. Indeed, in this study, the use of a variety of bivariate models (see Table 8) yielded sensitivities and specificities similar to those in the existing literature. When analyzing the model development cohort using a combination of snoring, choking episodes, neck circumference and hypertension, the results were similar to the existing literature: specificity: 96%, sensitivity: 13%. A combination of snoring, witnessed apneas, hypertension, and body mass index yielded a specificity of 96%, but a sensitivity of 27%.

Rather than focusing on a universally perfect diagnostic instrument, this decision rule applies to a subset of patients. The Ottawa Ankle Rule is a classic example of this approach. Although the specificity of the ankle rule is only 50%, it has a sensitivity of 100%. Therefore, not all patients meet the decision rule criteria, but in those who do, the need for an ankle radiograph can be eliminated. It is estimated that the rule has reduced the need for ankle radiography by 30%<sup>56</sup>.

Similarly, it was not possible to identify a combination of variables that had a high sensitivity and specificity. However, the use of a 3 variable model to *rule in* a

diagnosis of OSA (PPV=95%), and the use of cricomental space  $> 1.5$  cm to *rule out* OSA (NPV=100%) holds considerable promise.

Patients with a cricomental space  $>1.5$  cm are unlikely to have OSA, therefore, depending on the initial reason for referral, such patients either require no further testing, or else direct referral for polysomnography if a non-respiratory sleep disorder is suspected. By contrast, patients with cricomental space  $\leq 1.5$  cm, pharyngeal grade  $>2$ , and the presence of overbite are likely to have OSA, and can proceed directly to a trial of therapy. Remaining patients fall into a diagnostic gray zone, and require further investigation.

Patients who fit the decision rule criteria, namely, a cricomental space  $>1.5$  cm, or those meeting all criteria of the 3 variable model, accounted for 17% and 27% of the study population, respectively. Clearly, most (~60%) patients fell into a diagnostic grey zone. However, because of the high cost of diagnostic testing in OSA, if even a subset of patients can be appropriately triaged, important economic gains may be realized.



## **4.5 Study critique**

Rarely is a diagnostic instrument cheap, easy to use, and highly accurate, albeit in a subset of patients. The diagnostic performance of this decision rule may degrade as more patients are evaluated, particularly in populations where the prevalence of OSA is low.

Criticism could be raised with respect to the selection of subjects, since the study population consisted only of patients referred to the study investigators. Although no systematic triaging of referrals existed, recruitment may have occurred from a non-representative sample of the clinic population. However, the study prevalence of OSA (63% at an RDI  $\geq 10$  hr<sup>-1</sup>) was similar to that reported in a previous study that had recruited patients from the entire clinic population at the same institution<sup>19</sup>. Moreover, the performance of the decision rule was virtually identical to that observed in the “all SNOREAT referral” sample, which consisted of all patients referred for SNORESAT diagnostic testing.

Furthermore, a large proportion of patients (n=24) were excluded from the study because of insomnia or suspected non-respiratory sleep disorders. The exclusion rules, however, are well established and operationalised in the International Classification of Sleep Disorders (ICSD).

A more problematic issue relates to referral centre bias. This issue has already been illustrated in the study by Kushida et al. The Alberta Lung Association Sleep Centre in Calgary is also a tertiary, university-based referral centre, however, it is the only major sleep centre in Southern Alberta; a geographic region encompassing Calgary and a

number of smaller municipalities. Patients seen at the sleep centre range from highly specialized cases to uncomplicated snorers. As a result, the case mix is often similar to that seen at non-academic sleep centres.

Nevertheless, the decision rule developed in this research requires prospective evaluation in different settings, most specifically, at the primary care level. This decision rule is most likely to have the biggest clinical impact in settings where other sleep disorders are not under consideration (i.e. no further testing is necessary once OSA is excluded). With growing public awareness of OSA, an increasing number of patients seek medical attention because of witnessed apneic episodes. As observed in the Wisconsin Sleep Cohort, a large proportion of the population will have at least some apneic episodes, but these are of questionable clinical significance. A simple decision rule for excluding these patients would obviate the need for further, expensive, diagnostic testing.

Similarly, otolaryngologists perform uvulopalatopharyngoplasty (UPPP) for treatment of uncomplicated snoring. However, it is difficult to determine if such a patient has OSA, for which UPPP is not optimal treatment. For this reason, the current practice is to refer patients to sleep centres for the purpose of ruling out OSA. A simple decision rule for excluding OSA would reduce the need for diagnostic testing. Additionally, as more is learned about OSA-related cardiovascular complications, there is a growing number of cardiac patients referred specifically for the purposes of ruling out OSA.

It is likely that the diagnostic performance of the decision rule will change when deployed in non-sleep centre settings. The decision rule was validated in a setting where

the OSA prevalence was relatively high. As such, the positive predictive value would be expected to fall in lower prevalence populations, specifically in the primary care setting. However, in contrast, the negative predictive value (or ability to exclude OSA) should increase. As outlined previously, most practical applications for the decision rule would be in ruling out, rather than ruling in, OSA.

The choice of gold standard may also pose a threat to internal validity. The use of SNORESAT as the gold standard diagnostic instrument may be challenged, given that some investigators feel that only full overnight polysomnography is acceptable for OSA diagnosis. Although SNORESAT is now the standard OSA diagnostic instrument at the ALA Sleep Centre, acceptance of portable monitors is far from universal. However, SNORESAT has been well characterized and there is very close correlation and agreement between the SNORESAT-derived RDI and PSG-derived AHI<sup>19</sup>. The use of SNORESAT was thus considered an acceptable option. Moreover, with the decline in financial incentives to perform PSG, particularly with the introduction of Health Maintenance Organizations in the United States, there is a growing, evidence-based move away from PSG for the routine diagnosis of OSA<sup>52</sup>.

Finally, from a practical standpoint, the cost effectiveness of OSA-specific decision rules has never been adequately assessed. In theory, the ability to identify a subset of patients with OSA using the decision rule should allow clinicians to refer patients directly for therapy. However, the value of this approach may be limited by the willingness of third party payers to fund OSA treatment in the absence of “objective” baseline testing. Similarly, the ability to exclude patients with OSA based on the decision

rule may also have limited value in many settings. Any cost savings associated with the decision rule is based on two premises: (1) portable monitors are used for routine OSA diagnosis; (2) a non-respiratory sleep disorder is not suspected. Patients frequently present to the sleep centre with excessive daytime sleepiness, and even if OSA is excluded, full polysomnography must still need to be performed to rule out other primary sleep disorders. However, in sleep centres that make extensive use of portable monitors, the ability to identify patients without OSA would allow direct referral for polysomnography; thus bypassing portable monitoring, which has not been well validated for diagnosing non-respiratory sleep disorders.

Ultimately, the true value of a decision rule lies in its ability to have an impact on clinical practice. The effects of a decision rule on referral patterns for diagnostic testing, and subsequent economic impact, remains to be assessed.

### ***CONCLUSION***

In a subset of patients, this decision rule provides a simple, reliable, and highly accurate method of identifying patients with and without OSA. Its validity in the primary care setting, and its effect on changing clinical practice remains to be determined.

## REFERENCES

1. Young T, Palta M, Leder R, Dempsey J, Skatrud J, et al. The occurrence of sleep disordered breathing among middle-aged adults. *NEJM* 1993; 328:1230-5.
2. Hla KM, Young TB, Bidwell T, Palta M, et al. Sleep apnea and hypertension. *Ann Intern Med* 1994; 120:382-388.
3. Koskenvuo M, Kaprio J, Telakivi T, Partinen M, et al. Snoring as a risk factor for ischemic heart disease and stroke in men. *BMJ* 1987; 294:16-19.
4. Koskenvuo M, Kaprio J, Partinen M, et al. Snoring as a risk factor for angina pectoris. *Lancet* 1985; 1: 893-96.
5. Fletcher EC, Debehnke RD, Lavoie S, Gorin AB. Undiagnosed sleep apnea in patients with essential hypertension. *Ann Intern Med* 1985; 103:190-5.
6. He J, Kryger MH, Zorick FJ, Conway W, Roth T. Mortality and apnea index in obstructive sleep apnea. *Chest* 1988; 94(1): 9-14.
7. Partinen M, Jamieson A, Guilleminault C. Long-term outcome for obstructive sleep apnea syndrome patients: mortality. *Chest* 1988; 94: 1200-4.
8. Hung J, Whiteford EG, Parsons RW, Hillman D. Association of sleep apnea with myocardial infarction in men. *Lancet* 1990; 336:261-4.
9. George CFP, Nickerson P, Hanly P, Millar T, Kryger M. Sleep apnea patients have more automobile accidents. *Lancet* 1987; 1:477.
10. Findley LJ, Unverzagt ME, Suratt PM. Automobile accidents involving patients with obstructive sleep apnea. *Am Rev Respir Dis* 1988; 138: 337-40.
11. Young T, Blustein J, Finn L, Palta M. Sleep disordered breathing and motor vehicle accidents in a population based sample of employed adults. *Sleep* 1997; 20: 608-13.
12. Engleman HM, Martin SE, Deary IJ, Douglas NJ. Effect of continuous positive airway pressure treatment on daytime function in sleep apnea/hypopnea syndrome. *Lancet* 1994; 343: 572-5.
13. Kribbs NBA, Pack AI, Kline LR, Getsy JE, Schuett JS, Henry JN, Maislin G, Dinges DF. Effects of one night without nasal CPAP treatment on sleep and sleepiness in patients with obstructive sleep apnea. *Am Rev Respir Dis* 1993; 14: 1162-68.

14. Redline S, Adams N, Strauss ME, Roebuck T, Winters M, Rosenberg C. Improvement of mild sleep-disordered breathing with CPAP compared with conservative therapy. *Am J Respir Crit Care Med* 1998; 157: 858-65.
15. George CF, Boudreau AC, Smiley A. Effects of nasal CPAP on simulated driving performance in patients with obstructive sleep apnea. *Thorax* 1997; 52(7): 648-53.
16. Young T, Finn L, Hla KM, Morgan B, Palta M. Snoring as part of a dose-response relationship between sleep-disordered breathing and blood pressure. *Sleep* 1996; 19: 5202-5
17. Standards of Practice Committee. ASDA Standards of Practice: Portable recording in the assessment of obstructive sleep apnea. *Sleep* 1994; 17(4): 378-92
18. Issa FG, Morrison DL, HaJduk E, Remmers JE, et al. Digital monitoring of sleep disordered breathing using snoring sound and arterial oxygen saturation. *Am Rev Respir Dis* 1993; 148: 1023-29.
19. Vazquez JC, Tsai WH, Flemons WW, Masuda A, Brant R, Hadjuk E, Whitelaw WA, Remmers JE. Automated analysis of digital oximetry in the diagnosis of obstructive sleep apnea. *Thorax* (submitted)
20. Tsai WH, Flemons WW, Whitelaw WA, Remmers JE. The effects of arousal and desaturation based scoring criteria on the apnea hypopnea index and the prevalence of obstructive sleep apnea. *Am J Respir Crit Care Med* 1999; 159(1): 43-48.
21. Ancoli-Israel S, Kripke DF, Mason W, Messin S. Comparisons of home sleep recordings and polysomnograms in older adults with sleep disorders. *Sleep* 1981; 4: 283-91.
22. White DP, Gribb TJ, Wall JM, Westbrook PR. Assessment of accuracy and analysis time of a novel device to monitor sleep and breathing in the home. *Sleep* 1995; 18(2): 115-26.
23. Redline S, Tosteson T, Boucher MA, Millman RP. Measurement of sleep-disordered breathing disturbances in epidemiologic studies. *Chest* 1991; 100: 1281-86.
24. Flemmons WW, Whitelaw WA, Brant R, Remmers JE. Likelihood ratios for a sleep apnea clinical prediction rule. *Am J Respir Crit Care Med* 1994; 150:1279-85.
25. Gyulay S, Olson LG, Hensley MJ, King MT, Allen KM, Saunders NA. A comparison of clinical assessment and home oximetry in the diagnosis of obstructive sleep apnea. *Am Rev Respir Dis* 1993; 147: 50-53.

26. Williams AJ, Yu G, Santiago S, Stein M. Screening and sleep apnea using pulse oximetry and a clinical score. *Chest* 1991; 100: 631-35.
27. Kump K, Whalen C, Tishler P, Redline S, et al. Assessment of the validity and utility of a sleep-symptom questionnaire. *Am J Respir Crit Care Med* 1994; 150: 735-41.
28. Viner S, Szalai JP, Hoffstein V. Are history and physical examination good screening test for sleep apnea. *Ann Int Med* 1991; 115: 356-9.
29. Davies RJO, Ali HK, Stradling JR. Neck circumference and other clinical features in the diagnosis of the obstructive sleep apnea syndrome. *Thorax* 1992; 47:101-5.
30. Katz I, Stradling J, Slutsky AS, Zamel N, Hoffstein V. Do patients with obstructive sleep apnea have thick necks? *Am Rev Respir Dis* 1990; 141:1228-31.
31. Shepard JW, Gefer WB, Guilleminault C, White DP, et al. Evaluation of the upper airway in patients with obstructive sleep apnea. *Sleep* 1991; 14(4): 361-71.
32. Fleetham JA, Upper airway imaging in relation to obstructive sleep apnea. *Clinics in Chest Medicine* 1992; 13(3): 399-416.
33. Davies RJO, Stradling JR. The relationship between neck circumference, radiologic pharyngeal anatomy, and obstructive sleep apnea syndrome. *Eur Respir J* 1990; 3: 509-14.
34. Pepin JL, Levy P, Veale D, Ferretti G. Evaluation of the upper airway in sleep apnea syndrome. *Sleep* 1992; 15: S50-5.
35. Launois SH, Feroah TR, Campbell WN, Issa FG, et al. Site of pharyngeal narrowing predicts outcome of surgery for obstructive sleep apnea. *Am Rev Respir Dis* 1993; 147: 71-94.
36. Kushida CA, Efron B, Guilleminault C. A predictive morphometric model for the obstructive sleep apnea syndrome. *Ann Intern Med* 1997; 127: 581-7.
37. Mallampati SR, Gatt SP, Gugino L, Liu PL, et al. A clinical sign to predict difficult tracheal intubation: a prospective study. *Can Anaesth Soc J* 1985; 32(4): 429-34.
38. Samssoon GLT, Young JRB. Difficult tracheal intubation: a retrospective study. *Anaesthesia* 1987; 42: 487-90.
39. Wilson ME, Spiegelhalter D, Robertson JA, Lesser P. Predicting difficult intubation. *Br J Anaesth* 1988; 61:211-6.

40. Chou HC, Wu TL. Mandibulohyoid distance in difficult laryngoscopy. *Br. J Anaesth* 1993; 71: 335-9.
41. Savva D. Prediction of difficult tracheal intubation. *Br J Anaesth* 1994; 73: 149-53.
42. Karkouti K, Rose DK, Ferris LE, Lee H, et al. Inter-observer reliability of ten tests used for predicting difficult tracheal intubation. *Can J Anaesth* 1996; 43(6): 554-9.
43. Davies JM, Eagle CJ. M.O.U.T.H.S. *Can J Anaesth* 1991; 38(5): 687-8.
44. Diagnostic Classification Steering Committee of the American Sleep Disorders Association. *The International Classification of Sleep Disorders- Revised: Diagnostic and Coding Manual*. Davies Printing Company. 1997.
45. Kirshner B, Guyatt G. A methodological framework for assessing health indices. *J Chron Dis* 1985; 38(1): 27-36.
46. Marshall RJ. Partitioning methods for classification and decision making in medicine. *Statistics in Medicine* 1986; 5: 517-26.
47. Breiman L, Friedman JH, Olshen RA, Stone CJ. 1984. *Classification and regression trees*. Woodsworth and Brooks/Cole, Monterey, CA.
48. Cook EF, Goldman L. Empiric comparison of multivariate analytic techniques: advantages and disadvantages of recursive partitioning analysis. *J Chron Dis* 1984; 37: 721-31.
49. Freedman LS, Pee D. Return to a note on screening regression equations. *Am Statistician* 1989; 43: 279-82.
50. Peduzzi P, Concato J, Kemper E, Feinstein AR, et al. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996; 49(12): 1373-1379.
51. Moser NJ, Phillips BA, Berry DTR, Harbinson L. 1994. What is a hypopnea, anyway? *Chest* 105: 426-28.
52. *The Diagnosis of Sleep Apnea, 1998 APSS (Associated Professional Sleep Societies) Conference*, New Orleans, LA, USA.
53. Crocker BD, Olson LG, Saunders NA, Hensley MJ, McKeon JL, Allen KM, Gyulay SG. 1990. *Am Rev Respir Dis* 142: 14-18.



54. Olson LG, King MT, Hensley MJ, Saunders NA. A community study of snoring and sleep-disordered breathing: Symptoms. 1995. *Am J Respir Crit Care Med* 152: 707-10.
55. Kump K, Whalen C, Tishler PV, Browner I, Ferrette V, Strohl KP, Rosenberg C, Redline S. Assessment of the validity and utility of a sleep-symptom questionnaire. *Am J Respir Crit Care Med* 1994; 150: 735-41.
56. Stiell IG, Greenberg GH, McKnight RD, Nair RC, McDowell, Reardon M, Stewart JP, Maloney J. Decision rules for the use of radiography in acute ankle injuries: refinement and prospective validation. 1993. *JAMA* 269: 1127-32.
57. Barbe F, Pericas J, Munoz A, Findley L, Anto J, Agusti AGN. Automobile accidents in patients with sleep apnea syndrome: an epidemiological and mecahnistic study. *Am J Respir Crit Care Med* 1998; 158: 18-22.
58. Jenkinson C, Stradling J, Petersen S. How should we evaluate health status? A comparison of three methods in patients presenting with obstructive sleep apnea. *Quality of Life Research* 1998; 7: 95-100.
59. Last, JM. *A dictionary of epidemiology*. 3<sup>rd</sup> ed. Toronto: Oxford University Press, 1995: 145.
60. Streiner DL, Norman GR. *Health measurement scales: a practical guide to their development and use*. 2<sup>nd</sup> ed. Oxford: Oxford University Press. 1995: 116-118.
61. Johns MW. Daytime sleepiness, snoring, and obstructive sleep apnea: The Epworth sleepiness scale. *Chest* 1993; 103: 30-36.

**APPENDIX A: UPPER AIRWAY PHYSICAL EXAM PROTOCOL (UAPP)**  
**- UAPP-F (FINAL VERSION) -**

Name: \_\_\_\_\_ ID#: \_\_\_\_\_

Date of assessment: \_\_\_\_\_ Phone number: \_\_\_\_\_

**Patient Characteristics**

Sex: Male Female Age: \_\_\_\_\_ ESS: \_\_\_\_\_

Symptoms: snorer nocturnal choking witnessed apneas hypertension

**Drugs and medications:**

Alcohol: NO YES: \_\_\_\_\_ Smoker: NO EX CURRENT

Medications:

**Past Medical History:**

**Polysomnography:** [ ] PSG [ ] Snoresat

Date: \_\_\_\_\_ total sleep time: \_\_\_\_\_ min

RDI: \_\_\_\_\_ supine: \_\_\_\_\_ lateral: \_\_\_\_\_

Snoring index: \_\_\_\_\_ supine: \_\_\_\_\_ lateral: \_\_\_\_\_

Mean oxygen saturation: \_\_\_\_\_ %

Time spent (%): <90%: \_\_\_\_\_ <85%: \_\_\_\_\_ <80%: \_\_\_\_\_ <70%: \_\_\_\_\_

**RATER:** [ ] WHT [ ] JER [ ] other: \_\_\_\_\_

Anthromorphics

Height: \_\_\_\_\_ cm Weight: \_\_\_\_\_ kg Neck circumference: \_\_\_\_\_ cm

<b><u>Mandible:</u></b>	neutral	Full extension		neutral	Full extension
max mandibular advancement (mm)			Mandibular length (cm)		
thryo-rami (cm)			Mastoid-medial clavicle (cm)		
TMJ-ramus (cm)			ramus-ramus (cm)		
thryo-mental (cm) (natural position)			Sterno-mental (cm) (natural position)		
thryo-mental (cm) (neutral position)			Sterno-mental (cm) (neutral position)		

Profile:  Retrognathia  neutral  prognathiaCricomental space:  occluded distance @ mid-point: \_\_\_\_\_ cmChin:  Dimpled  widened**Opening:** neutral with phonation

Tonsils (Class I-IV) I II III IV

Pharyngeal space: I II III IV I II III IV

inter-incisor gap: \_\_\_\_\_ cm

**Uvula:** neutral phonation

Samsoon-Young Class: I II III IV I II III IV

**Teeth:**

Overbite: \_\_\_\_\_ mm

Overjet: \_\_\_\_\_ mm

**- UAPP-P (PROTOTYPE VERSIONS) -**

**RATER:**                       WHT       JER       other: \_\_\_\_\_

<u><b>Mandible:</b></u>	neutral	full extension		neutral	full extension
max mandibular advancement (mm)			mandibular length (mm)		
thryo-mental (mm)			Sterno-mental (cm)		
TMJ-ramus (mm)			ramus-ramus (mm)		
thryo-rami (mm)			Mastoid-medial clavicle (cm)		

Profile:  Retrognathia     neutral     prognathia    Chin protrusion: \_\_\_\_\_ mm

**Opening:**                      *upright*                      *supine*

Tonsils (Class I-IV)              I II III IV                      I II III IV

Pharyngeal space:              \_\_\_\_\_ mm                      \_\_\_\_\_ mm

inter-incisor gap:              \_\_\_\_\_ mm

**Uvula:**                      *neutral*                      *phonation*                      *best view*

Samsoon-Young Class:              I II III IV                      I II III IV                      I II III IV

**Teeth:** \_\_\_\_\_

Overbite:              \_\_\_\_\_ mm                      Overjet: \_\_\_\_\_ mm

<u><b>Head Movement</b></u>	<i>Right</i>	<i>Left</i>
Neck lateral flexion (cm)		
Flexion (cm)		
Extension (cm)		

**Nose:**                       deviated septum       nasal polyps       nasal stuffiness

- UAPP-SF (SHORT-FORM) -

NAME: \_\_\_\_\_

Rater: [ ] WHT [ ] JER [ ] WWF

Height: \_\_\_\_\_ cm Weight: \_\_\_\_\_ kg Neck circumference: \_\_\_\_\_ cm

Thyromental displacement: \_\_\_\_\_ cm

Cricomental space: [ ] occluded distance at midpoint: \_\_\_\_\_ cm

Profile: [ ] retrognathic [ ] neutral [ ] prognathic

Tonsils: [ ] enlarged [ ] normal or small

Overbite: [ ] present [ ] absent

Pharynx:

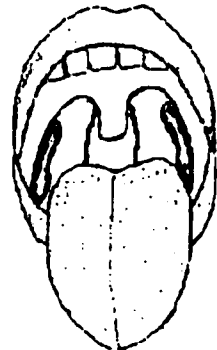
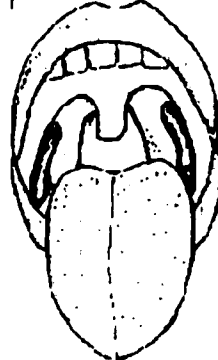
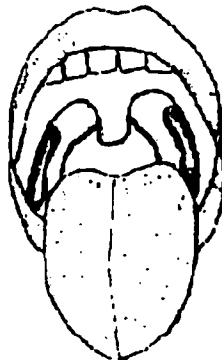
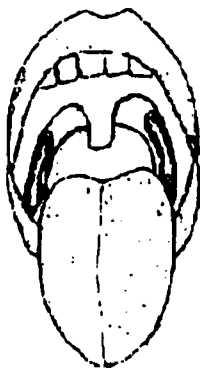
tongue diameter

100%

75%

50%

25%



CIRCLE ONE

I

II

III

IV

## APPENDIX B: The Epworth Sleepiness Score

YOUR NAME: \_\_\_\_\_

How likely are you to doze off or fall asleep in the following situation, in contrast to feeling just tired. This refers to your usual way of life in recent times. Even if you have not done some of these things recently, try to work out how they would have affected you.

	CHANCE OF DOZING OFF			
	<i>NEVER</i>	<i>SLIGHT</i>	<i>MODERATE</i>	<i>HIGH</i>
Sitting and reading	0	1	2	3
Watching TV	0	1	2	3
Sitting inactive in a public space (eg. Theatre or a meeting)	0	1	2	3
As a passenger in a car for an hour without a break	0	1	2	3
Lying down to rest in the afternoon when circumstances permit	0	1	2	3
Sitting and talking to someone	0	1	2	3
Sitting quietly after a lunch without alcohol	0	1	2	3
In a car, stopped for a few minutes in traffic	0	1	2	3

## APPENDIX C: Definitions

**Apnea:** absence of airflow for more than 10 seconds

central apnea- respiratory effort is absent

obstructive- cessation of airflow with persistent continued respiratory effort (as evidenced by thoracoabdominal movement)

**Apnea hypopnea index (AHI):** # of apneas & hypopneas/ hour of sleep

**Body mass index:** weight (kg)/ height (m)<sup>2</sup>

**Hypopnea:** a >10 second reduction in airflow, usually assessed by a reduction in respiratory effort +/- an associated oxygen desaturation or EEG-based arousal. There is significant controversy over the hypopnea definition<sup>20,51</sup>.

**Respiratory disturbance index (RDI):** a surrogate measure of the AHI. The definition of RDI differs according to the monitor used to derive it. However, it should have close correlation and agreement with the AHI. When scored off a polysomnogram, the RDI is considered synonymous with the AHI.

**Obstructive sleep apnea (OSA):** episodic interruption in airflow during sleep. The diagnosis is established by an AHI exceeding a threshold value. Considerable controversy exists over which AHI criterion value is diagnostic of OSA, if indeed one is even appropriate.

**Sleep apnea hypopnea syndrome (SAHS):** the presence of obstructive sleep apnea and OSA related symptoms.



## APPENDIX E: CONSENT FORM FOR UAPP

**Research Project:** UAPP Project  
**Investigators:** Drs. J.E. Remmers, W. Tsai, C. McArthur, J. Davies, R. Brant  
**Funding Agency:** Alberta Heritage Foundation for Medical Research

This consent form, a copy of which has been given to you, is only part of the process of informed consent. It should give you the basic idea of what the research is about and what your participation will involve. If you would like more detail about something mentioned here, or information not included here, you should feel free to ask. Please take the time to read this carefully and to understand any accompanying information.

### **What is obstructive sleep apnea (OSA)?**

Obstructive sleep apnea is a condition where people episodically stop breathing during their sleep (when people stop breathing for a long time, we call this an **apnea**). The brain is pretty smart, and realizes that you've stopped breathing, and eventually wakes you up. People are frequently unaware that this is happening, however, the repetitive arousals from sleep result in daytime sleepiness. In more severe cases, OSA is associated with increased rate of car accidents, high blood pressure, heart disease, and strokes. It may occur in up to 4% of people, however, most of these people are unaware that they have it. Your physician suspects that you may have OSA, and that is why you were referred to the sleep centre.

### **So what's involved in this study?**

Traditionally, OSA is diagnosed by an overnight sleep study at a sleep centre. This is expensive in terms of time and money. We are looking at ways of determining if we can predict whether people have OSA based on physical appearance. We will be doing simple measurements of your mouth and jaw area. This takes no more than 5 minutes, and it is often part of the standard physical exam for sleep apnea. You will then undergo a home sleep study, which is part of the standard test for people suspected of having OSA. Aside from taking a few simple head and neck measurements, your evaluation will be no different than that of any other patient referred to the sleep centre.



The information collected will be added to your patient file. Some of the data may be reported in a scientific report of research results, but your name will not be connected with this. In the unlikely event that you suffer injury as a result of participating in this research, no compensation or treatment (beyond that routinely provided by Alberta Health) will be provided to you by the funding agencies, investigators, the University of Calgary, or the Calgary Regional Health Authority. You still have all your legal rights. Nothing said here about treatment or compensation in any way alters your right to pursue legal recourse to recover damages.

Your signature on this form indicates that you have understood to your satisfaction the information regarding participation in the research project and agree to participate as a subject. In no way does this waive your legal rights nor release the investigators, sponsors, or involved institutions from their legal and professional responsibilities. You are free to withdraw from the study at any time without jeopardising your health care. Your continued participation should be informed as your initial consent, so you should feel free to ask for clarification or new information throughout your participation.

If you have further questions concerning matter related to this research, please contact: Dr. Willis Tsai (670-2540)

If you have any questions concerning your rights as a possible participant in this research, please contact the Office of Medical Bioethics, Faculty of Medicine, The University of Calgary, at 220-7990. Please sign below.

\_\_\_\_\_  
Participant's Signature

\_\_\_\_\_  
Date

\_\_\_\_\_  
Investigator and/or Delegate's Signature

\_\_\_\_\_  
Date

\_\_\_\_\_  
Witness' Signature

\_\_\_\_\_  
Date

A copy of this consent form has been given to you to keep for your records and reference.