Inter-Observer Reliability of Candidate Predictive Morphometric Measurements for Women with Suspected Obstructive Sleep Apnea

John A. Gjevre, M.D., M.Sc.¹; Regina M. Taylor-Gjevre, M.D., M.Sc.²; John K. Reid, M.D., F.A.A.S.M.¹; Robert Skomro, M.D., F.A.A.S.M.¹; David Cotton, M.D.¹

¹Division of Respiratory, Critical Care and Sleep Medicine, and

²Department of Medicine, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

Objective: Obstructive sleep apnea (OSA) is increasingly recognized as a public health concern. Definitive diagnosis is by overnight polysomnographic (PSG) examination. Identification of clinical predictors would be beneficial in helping prioritize high-risk patients for assessment. Practical application of morphometric predictive variables would require a high level of reproducibility in a clinical setting. In this study, our objective was to evaluate reliability between observers in measurements of candidate morphometric parameters in women.

Design and Methods: This was a prospective study of 71 women who had been referred for PSG with suspected OSA. Selected morphometric parameters were measured independently in the sleep laboratory by two trained sleep physicians.

Results: Neck circumference and truncal measurements for lower costal, midabdominal, and hip circumferences had higher reliability coefficients (intraclass correlation coefficients [ICC] of 0.78, 0.95, 0.95, and 0.81) than the smaller dimension

measurements, including cricomenal distance or retrognathia (ICC of 0.04 and 0.17). Of the women participating in this study, 50 of 71 had apnea-hypopnea indexes (AHI) ≥ 5 . Body mass index (BMI), neck circumference, lower costal girth, midabdominal girth, and hip girth were all significantly higher ($p < 0.001$ -0.004) in women with AHI ≥ 5 .

Conclusions: There was wide variation in inter-observer reliability for different physical dimensions. We propose that any clinical morphologic measurement employed in predictive modeling should be reliably reproducible in clinical setting conditions. Our findings support the use of several truncal measures, BMI, and neck circumference as predictive measures in women undergoing evaluation for OSA.

Keywords: Women, sleep, body measures

Citation: Gjevre JA; Taylor-Gjevre RM; Reid JK; Skomro R; Cotton D. Inter-observer reliability of candidate predictive morphometric measurements for women with suspected obstructive sleep apnea. *J Clin Sleep Med* 2013;9(7):695-699.

There is growing recognition of obstructive sleep apnea (OSA) as a public health concern, with up to 20% of adults estimated to have at least mild OSA.¹ Untreated, OSA may pose substantial risks, contributing to development of systemic hypertension and mortality from cardio/cerebrovascular disease.² If associated with hypersomnolence, OSA may contribute to an increase in motor vehicle or work-related accidents.^{3,4}

The definitive diagnostic tool for OSA is the overnight polysomnogram (PSG). Increasing awareness of this disorder among both physicians and the general public is leading to greater numbers of referrals for sleep physician assessment and for PSG. Timely access to PSG varies by geographic region. Because of PSG access issues, alternative home-based sleep diagnostic testing has been validated for use.⁵ Clinical predictors for significant OSA are useful tools enabling the physician to prioritize referrals and diagnostic test scheduling by degree of urgency. In men, increased BMI (body mass index) and neck circumference are strong predictors of OSA.⁶ However, gender differences in body fat distribution and OSA associated symptoms have been reported, and it is unclear that recognized predictors of OSA in men would also carry the same significance in women.⁷⁻¹⁰

BRIEF SUMMARY

Current Knowledge/Study Rationale: Physical parameters may serve as clinical predictors of obstructive sleep apnea, assisting physicians in the diagnostic process and in prioritization of patient referrals for polysomnography. A high level of reproducibility of such physical measurements would be required for practical application; this study evaluates level of agreement between physicians in such measurements.

Study Impact: The results from this study demonstrate wide variation in inter-observer reliability for different physical measurements. Our findings support the utilization of body mass index, neck circumference and several truncal measures in predictive modeling for women undergoing evaluation for obstructive sleep apnea.

There have been efforts made to identify specific morphometric measurements, which could be employed as clinical predictors for OSA, either alone or in conjunction with other parameters.^{11,12}

In order for such dimensions to be employed as practical screening tools, it is crucial to understand the reliability of measurements in a clinical environment. A physical dimension measurement that in a particular range is predictive for OSA may prove to be misleading should the measurement itself be subject to substantial variation between physicians. In

this study, we address this concern by examination of inter-observer agreement in candidate predictive morphometric measures in women who are undergoing polysomnography for suspected OSA.

METHODS

Consecutive women scheduled for routine PSG testing for evaluation of clinically suspected OSA were invited to participate in this study. Informed consent was obtained. This study was approved by the institutional research ethics board and is in accordance with the Helsinki Declaration.

Inclusion criteria included age ≥ 21 years and ability to provide informed consent. Exclusion criteria were: referring sleep physician's strong suspicion of another primary sleep disorder (primary insomnia, narcolepsy, restless legs syndrome, a parasomnia, or nocturnal seizures) as indicated on the patient's referral form.

Morphometric variables were assessed by standardized, focused upper airway examination and general examination. For each participant, this was performed by 2 independent sleep physician observers to assess inter-observer reliability. Measurements included: neck circumference, lower costal or chest circumference (lower margin of the antero-lateral ribs, while standing at functional residual capacity [FRC]), umbilical abdominal (mid-abdominal) circumference (peri-umbilical circumference of the abdomen with abdominal muscles relaxed at FRC while standing), hip circumference (widest circumference of the buttocks while standing), lateral pharyngeal space narrowing (grading by Tsai et al.¹¹), vertical pharyngeal space narrowing (modified Mallampati score),¹² cricomental space (in mm), maxillary over jet (in mm), and retrognathia (in mm). The presence or absence of tongue ridging, macroglossia, and tonsillar enlargement were also recorded.^{11,13}

The pharyngeal grading system described by Tsai included a 4-class categorization, with class I designated when the palatopharyngeal arch intersects at the edge of the tongue, class II when the palatopharyngeal arch intersects at $\geq 25\%$ of the tongue diameter, class III when the palatopharyngeal arch intersects at $\geq 50\%$ of the tongue diameter, and class IV when the palatopharyngeal arch intersects at $\geq 75\%$ of the tongue diameter.¹¹

All 5 sleep physicians assessing patients for this study participated in a standardized training session on morphometric measurement techniques prior to study initiation.

Patients were studied overnight in the sleep lab using the standard 15-channel PSG (Sandman Elite version 8.0 sleep diagnostic software, Ottawa, Canada). Established protocols were used for all PSG studies.¹⁴ This included electroencephalogram (EEG, 3-channel), electrooculogram (2-channel), electromyogram (chin and leg), electrocardiogram, heart rate, snoring, thermistor airflow, nasal pressure airflow, oxygen saturation, chest wall motion, and abdominal motion.

Statistical Analysis

SPSS v.17.0 was employed for data entry and analysis. Means and standard deviations were calculated for continuous data. Proportions were calculated for categorical data. Between group comparisons of continuous data were evaluated with independent 2-tailed *t*-tests. Between group comparisons of cat-

egorical data were evaluated with χ^2 testing and Fisher exact test when the cell size was < 5 . Measures of agreement between observers were calculated for morphometric measurements, κ coefficients for categorical data, and intraclass correlation coefficients for continuous data.^{15,16}

For this reliability study using intraclass correlations, we used conventional values for α of 0.05 and β of 0.20. As each patient underwent separate measurements by 2 different observers, using a minimum acceptable level of agreement of $\rho_0 = 0.4$, and an expected level of agreement of $\rho_1 = 0.9$, then approximately 7 subjects would be the estimated sample required per reliability assessment.¹⁷

For 2 group comparisons based on an apnea hypopnea index (AHI) cutoff of 5, mean morphometric values derived from the 2 observers were utilized. Receiver operator characteristic (ROC) curves were plotted for the predictive relationships between abnormal range AHI (≥ 5) and morphometric parameters.

RESULTS

Of 95 consecutive female patients who attended the sleep lab during the study period, 71 consented to participate. The means and standard deviations of morphometric measurements by 2 physicians and the measures of agreement are detailed in **Table 1**. A greater degree of inter-observer agreement, as represented by the intra-class correlation coefficients and κ coefficients,^{15,16} was observed for some measurements compared to others. The greatest agreement was observed for the truncal measures of lower costal girth, midabdominal girth, and hip circumferences. The lowest degree of agreement was evident for the smaller dimension measurements for cricomental distance and retrognathia. Subjective dichotomous observations for the presence or absence of tongue enlargement, tongue ridging, or tonsillar enlargement also had lower measures of agreement between observers.

Of the 71 participants, 50 had AHI ≥ 5 . Comparisons of morphometric continuous measurements between the 2 groups are detailed in **Table 2**. Mean measurement scores derived from the 2 observers were employed for this 2-group comparison. There were no significant differences in proportions of participants designated to have tongue ridging, macroglossia, or tonsillar abnormalities between those with elevated AHI and those with normal AHI values. Predictive relationships for abnormal AHI (≥ 5) with morphometric measures are described in **Figures 1 and 2** and **Table 3**.

DISCUSSION

Although obesity has been linked to OSA and there are comparable frequencies of obesity between genders, the prevalence of OSA in women has been lower than in men.¹ In men, increased BMI and increased neck circumference are predictive of OSA.⁶ It is less clear that these variables are predictive in women with OSA.^{8-10,18} Whittle et al. have demonstrated in magnetic resonance imaging (MRI) studies of men and women that there are differences in neck fat deposition distribution between the genders and greater overall soft tissue volume around the airway in men. They speculate that these or other anatomic factors may contribute to the

Table 1—Comparison of morphometric measurements between observers

Measure	Observer 1, mean (SD)	Observer 2, mean (SD)	Measure of Agreement
Neck circumference (cm)	38.33 (4.76)	38.16 (4.49)	0.780
Lower costal girth (cm)	100.91 (14.53)	100.25 (15.77)	0.952
Midabdominal girth (cm)	104.85 (18.98)	105.23 (19.84)	0.949
Hip girth (cm)	121.47 (19.22)	120.89 (18.61)	0.810
Retrognathia (mm)	5.72 (4.56)	5.00 (5.83)	0.174
Cricomental distance (mm)	5.45 (5.84)	5.06 (7.05)	0.044
Maxillary overjet (mm)	2.50 (2.01)	2.38 (1.78)	0.579
Lateral pharynx (mm)	2.54 (0.92)	2.61 (0.80)	0.327
Vertical pharynx (mm)	2.32 (0.96)	2.35 (0.95)	0.246
% Tongue ridging	63.4%	76.1%	0.329*
% Tongue enlarged	40.8%	42.3%	0.488*
% Tonsils abnormal	9.9%	9.9%	0.363*

SD, standard deviation; cm, centimeter; mm, millimeter. *k coefficient for categorical variables, intraclass correlation coefficient for continuous variables.

Table 2—Comparison of morphometric measures between groups based on AHI

Measure	AHI ≥ 5 , mean (SD)	AHI < 5 , mean (SD)	Significance	95% CI
Body mass index (kg/m ²)	39.37 (10.50)	30.33 (5.14)	< 0.001	4.230, 13.845
Neck circumference (cm)	39.39 (4.55)	35.54 (2.24)	< 0.001	1.766, 5.935
Lower costal girth (cm)	104.65 (15.18)	90.89 (8.96)	< 0.001	6.664, 20.850
Midabdominal girth (cm)	110.58 (18.67)	91.86 (13.14)	< 0.001	9.772, 27.672
Hip girth (cm)	125.07 (18.06)	111.93 (14.38)	0.004	4.277, 21.996
Retrognathia (mm)	4.95 (3.61)	6.33 (4.80)	0.186	-3.453, 0.684
Cricomental distance (mm)	4.22 (4.25)	7.73 (4.80)	0.003	-5.799, -1.213
Maxillary overjet (mm)	2.30 (1.73)	2.88 (1.56)	0.200	-1.479, 0.315
Lateral pharynx (mm)	2.59 (0.76)	2.57 (0.56)	0.934	-0.354, 0.384
Vertical pharynx (mm)	2.41 (0.77)	2.14 (0.71)	0.175	-0.122, 0.656

SD, standard deviation; cm, centimeter; mm, millimeter; kg, kilogram; m, meter; CI, confidence interval.

Figure 1—ROC curve for prediction of AHI ≥ 5 from truncal measures **Figure 2**—ROC curve for prediction of AHI ≥ 5 from upper airway/mandibular measures

Table 3—Area under the curve (AUC) for prediction of abnormal AHI

Measure	AUC	Significance	95% CI
Body mass index	0.785	< 0.001	0.679, 0.890
Neck circumference	0.785	< 0.001	0.679, 0.891
Lower costal girth	0.782	< 0.001	0.672, 0.893
Midabdominal girth	0.778	< 0.001	0.668, 0.888
Hip girth	0.715	< 0.004	0.586, 0.844
Retrognathia	0.410	0.247	0.263, 0.557
Cricomental distance	0.313	0.016	0.176, 0.449
Maxillary overjet	0.396	0.183	0.255, 0.537
Lateral pharynx	0.500	1.000	0.356, 0.644
Vertical pharynx	0.603	0.187	0.456, 0.750

CI, confidence interval.

gender disparity in OSA prevalence.⁷ Identification of clinical predictors for OSA in women would help prioritize PSG evaluation. Morphometric measures have been examined for identification of candidate variables to aid in the screening process. In this group of female patients who had been referred for polysomnography we found significantly different mean values for a number of measures (BMI, neck circumference, lower costal girth, midabdominal girth, hip girth, and the cricomental distance) between groups based on AHI category. However, of these potentially predictive measurements, the cricomental distance has a quite low measure of inter-observer agreement.

As illustrated by **Figure 1** and 2 ROC curves, there is substantial overlap in the predictive relationship between truncal measures for an abnormal AHI, whereas the upper airway/mandibular measures have a lesser area under the curve and are observed to be clustered about the reference line, which implies lack of predictive contribution. The truncal parameters each provide a statistically significant predictive measure, with BMI and neck circumferences having the highest area under the curve. It is possible that greater predictive capacity may be achieved by an additive combination of physical measures, or alternatively by ratio (through adjustment for height as an example). However, identification of such predictive models was outside the scope of this study.

The extent of inter-observer reproducibility would be expected to influence utilization of any predictive morphometric parameter. In this study evaluating measure of agreement between observers in a variety of morphometric assessments, we observe the greatest degree of agreement for truncal dimensions and neck circumference, and the lowest agreement for smaller upper airway/mandibular dimensions, such as cricomental distance. We propose that any clinical morphologic measurement employed in a predictive capacity would need to be one reliably reproduced in clinical settings in order to be of practical value. Our findings support the use of truncal measures, BMI, and neck circumference as predictive measures in women undergoing evaluation for OSA.

REFERENCES

1. Young T, Palta M, Dempsey J, Skatrud J, Weber S, Badr S. The occurrence of sleep-disordered breathing among middle-aged adults. *N Engl J Med* 1993;32:1230-5.
2. Peker Y, Carlson J, Hedner J. Increased incidence of coronary artery disease in sleep apnoea: a long-term follow-up. *Eur Respir J* 2006;28:596-602.
3. Ulfberg J, Carter N, Edling C. Sleep-disordered breathing and occupational accidents. *Scand J Work Environ Health* 2000;26:237-42.
4. Mulgrew AT, Nasvadi G, Butt A, et al. Risk and severity of motor vehicle crashes in patients with obstructive sleep apnoea/hypopnoea. *Thorax* 2008;63:536-41.
5. Gjevre JA, Taylor-Gjevre RM, Skomro R, Reid J, Fenton M, Cotton D. Comparison of polysomnographic and home-based Embletta assessments of obstructive sleep apnea in Saskatchewan women. *Can Respir J* 2011;18:271-4.
6. Hoffstein V, Szalai JP. Predictive value of clinical features in diagnosing obstructive sleep apnea. *Sleep* 1993;16:118-22.
7. Whittle AT, Marshall I, Mortimore IL, Wraith PK, Sellar RJ, Douglas NJ. Neck soft tissue and fat distribution: comparison between normal men and women by magnetic resonance imaging. *Thorax* 1999;54:323-8.
8. Wahner-Roedler DL, Olson EJ, Narayanan S, et al. Gender-specific differences in a patient population with obstructive sleep apnea-hypopnea syndrome. *Genet Med* 2007;4:329-38.
9. Valipour A, Lothaller H, Rauscher H, Zwick H, Burghuber OC, Lavie P. Gender-related differences in symptoms of patients with suspected breathing disorders in sleep: a clinical population study using the sleep disorders questionnaire. *Sleep* 2007;30:312-9.
10. Millman RP, Carlisle CC, McGarvey ST, Eveloff SE, Levinson PD. Body fat distribution and sleep apnea severity in women. *Chest* 1995;107:362-6.
11. Tsai WH, Remmers JE, Brant R, Flemons WW, Davies J, Macarthur C. A decision rule for diagnostic testing in obstructive sleep apnea. *Am J Respir Crit Care Med* 2003;167:1427-32.
12. Samsom GLT, Young JRB. Difficult tracheal intubation: a retrospective study. *Anaesthesia* 1987;42:487-90.
13. Schellenberg JB, Maislin G, Schwab RJ. Physical findings and the risk for obstructive sleep apnea—The importance of oropharyngeal structures. *Am J Respir Crit Care Med* 2000;162:740-8.
14. Iber C, Ancoli-Israel S, Chesson A, Quan S. *The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications*. Westchester, IL: American Academy of Sleep Medicine; 2007.
15. Müller R, Büttner P. A critical discussion of intraclass correlation coefficients. *Stat Med* 1994;13:2465-76.
16. Rigby AS. Statistical methods in epidemiology. V. Towards an understanding of the kappa coefficient. *Disabil Rehabil* 2000;22:339-44.
17. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. *Stat Med* 1998;17:101-10.
18. Dancsey DR, Hanly PJ, Soong C, Lee B, Shepard J Jr., Hoffstein V. Gender differences in sleep apnea. The role of neck circumference. *Chest* 2003;123:1544-50.

ACKNOWLEDGMENTS

The authors thank Dr. Brian McNab for his contribution to this study and also the Saskatchewan Health Research Foundation for their support of this work. This study was supported by a grant from the Saskatchewan Health Research Foundation.

SUBMISSION & CORRESPONDENCE INFORMATION

Submitted for publication November, 2012

Submitted in final revised form December, 2012

Accepted for publication January, 2013

Address correspondence to: John A. Gjevre M.D., M.Sc., FRCP(C), Division of Respiratory, Critical Care and Sleep Medicine, Department of Medicine, University of Saskatchewan, Saskatoon, SK, Canada, S7N 0W8; Tel: (306) 966-8299; Fax: (306) 966-8694; E-mail: john.gjevre@usask.ca

DISCLOSURE STATEMENT

This was not an industry supported study. The authors have indicated no financial conflicts of interest.