A prediction rule to screen patients with moderate-to-severe obstructive sleep apnea

Emma Grigor, MSc

A thesis submitted to the Faculty of Graduate and Postdoctoral Studies in partial fulfillment of the requirements for the M.Sc. degree in Epidemiology and Community Medicine

School of Epidemiology and Public Health

Faculty of Medicine

University of Ottawa

Abstract

Introduction: Obstructive sleep apnea (OSA) is a common breathing disorder with numerous health consequences, including greater risk of complications perioperatively. Undiagnosed OSA is known to place surgical patients at a higher risk of serious adverse events, including stroke and death. Polysomnography (PSG) assessment is the current gold standard test for diagnosing OSA. However, due to the significant time commitment and cost associated with PSG, a substantial number of OSA patients go undiagnosed before the perioperative period. Although the STOP-Bang questionnaire screening tool is currently used to help detect OSA patients, the low specificity to screen people without the disease is considered a major limitation. There is a clear need to develop a quick and effective prediction rule with higher overall accuracy to help streamline OSA diagnosis. Tracheal breathing sound analysis in awake patients at the bedside has shown potential to screen OSA patients with higher specificity compared to the STOP-Bang questionnaire. To date, no screening tools exist to detect OSA patients that combine the results of breathing sound analysis and STOP-Bang.

Objectives: The present study aimed to develop a prediction rule, using both breathing sound analysis and variables in the STOP-Bang questionnaire, to better streamline the diagnosis of OSA.

Methods: This prospective cohort study recruited patients referred for PSG at the Ottawa Hospital Sleep Centre from November 2016 to May 2017. The study conduct was approved by the Ottawa Health Science Network Research Ethics Board (#20160494-01H). After obtaining informed consent, anthropomorphic, breathing sound recordings, and STOP-Bang questionnaire data was collected from over 400 consenting patients. All patients that met the eligibility criteria were included. The breathing sound analysis and STOP-Bang results were utilized to design a

prediction rule using logistic regression. Sensitivity, specificity, and likelihood ratio were used to compare the diagnostic performance of the final model.

Results: Of the 439 consenting study participants, 280 study participants data were eligible for inclusion in the logistic regression analysis. Physician sleep specialists diagnosed 114 participants (41%) with moderate-to-severe OSA and 166 participants (59%) with normal-to-mild OSA. At a predicted probability of moderate-to-severe OSA greater than or equal to 0.5, breathing sound analysis had a similar sensitivity of 75.9 (95%CI; 65.4, 82.0) and higher specificity of 74.5% (95%CI; 68.5, 82.0) when compared to STOP-Bang with a sensitivity and specificity of 68.4% (95%CI; 58.9, 76.6) and 63.2% (95%CI: 55.0, 70.1), respectively. The sensitivity and specificity for the Safe-OSA rule, obtained by combining breathing sound analysis and STOP-Bang variables, were determined to be 75.4% (95%CI; 65.4, 82.0) and 74.5% (95%CI; 68.5, 82.0), respectively. A sensitivity analysis using a likelihood ratio test showed that breathing sound analysis contributed significantly to the performance of the Safe-OSA rule. The Safe-OSA rule was determined to be reasonably discriminative and well calibrated. The five-fold cross-validation showed similar results for the final model in the derivation and testing subsamples, which provides support for the internal validity of the Safe-OSA rule in our study population.

Conclusion: The present study lends further support for the future testing of tracheal breathing sound analysis as a potential method to screen for moderate-to-severe OSA to help streamline patient care in the perioperative setting.

Trial registration: ClinicalTrials.gov identifier NCT02987283.

Acknowledgments

I wholeheartedly thank my co-supervisors, Drs. Tim Ramsay and Sylvain Boet, for their supervision, support, and guidance throughout the completion of this thesis project.

I wholeheartedly thank Dr. Kaitlyn Duncan for her key role in the conduct of this research study as a principal investigator. Dr. Duncan provided invaluable support, continuous feedback, and contributions that assisted in the completion of this project.

I wholeheartedly thank all members of the research team that contributed to the administration, organization, and data collection for the project. Importantly, I would like to thank Sandy Lam for her help with coordination of various aspects of the project including administration, Mohammed Ibrahim for his help training research assistants and help coordinating data collection schedule, and the research volunteers who contributed in a major way to assisting in the data collection for this research project. Many thanks to Geoff Clarke who helped with data management and troubleshooting throughout the course of the project. Furthermore, a special thank you to the Ottawa Sleep Centre (The Ottawa Hospital Civic campus) for their support of the conduct of this research project throughout data collection. Thank you Sarah Schlievert and Jodi Peters for their administration help throughout the project.

I dedicate this thesis to my loving grandfather John Grigor and grandmother Angela Grigor for inspiring me to pursue my dreams and continue to reach higher success in my studies. I also dedicate this thesis to my beloved sister Kate Grigor for her unconditional love and support

throughout my graduate studies; my loving Grandmother Helen Merritt who helped me recognize the value of mindfulness in order to remain focused on achieving my goals; my Aunt and Godmother, Moneca Yardley, who has long been a source motivation to me to always work hard to achieve my goals as well as a key supporter throughout my University studies; and most importantly my loving parents Mary Ellen and Neil Grigor for their unconditional support throughout my academic endeavors and research work.

TABLE OF CONTENTS

CHAPTER 1: Introduction to thesis topic

1.1 Background

Obstructive sleep apnea (OSA) is a sleep-related breathing disorder. People with OSA experience *'apnea'* or periods of time where breathing stops during sleep. Specifically, this breathing disorder is characterized by the collapse of the upper airway during sleep that leads to periods of decreased blood oxygen levels (1). With pauses in breathing patients experience hypoxia, which means decreased blood oxygen levels. Hypoxia leads to other symptoms such as daytime fatigue, memory loss and attenuated cognitive function and concentration (2–4). In addition to these short-term side effects of sleep apnea, there are a number of serious health conditions associated with OSA such as type 2 diabetes and stroke (5,6).

Apnea-hypopnea index (AHI) is a scoring system that measures the number of apneas and hypopneas patients experience and is used to classify OSA severity (7). An AHI score of ≥ 5 indicates the presence of OSA. In general, the severity of OSA is considered to worsen as the AHI score increases. Previous studies have shown that severe OSA (defined as AHI ≥ 30) is statistically significantly associated with being overweight and obese (8–10). A recent re-examination of the Wisconsin sleep cohort suggested possible increases in the relative prevalence of moderate-to-severe OSA, ranging from 14 to 55% (in men and women), when taking into consideration the rising prevalence of people that are overweight or obese in the United States (11–14).

Based on population-level data, the prevalence of OSA accompanied by daytime fatigue is estimated to be between 3 and 7% for adult men and 2 and 5% for adult women (15). In Canada alone, the 2009 Canadian Community Health Survey estimated that 858,900 (3%) of Canadian adults (18 years and older) were told by a health professional that they have sleep apnea; an estimated 1 in 4 adults (26%) reported having other risk factors known to be associated with high risk of OSA (16). These other risk factors suggestive of OSA include loud snoring, choking or gasping during sleep, witnessed breathing pauses by a bed-partner, insomnia, and daytime sleepiness. It is estimated that approximately 90% of OSA cases remain undiagnosed in Canada (17).

On a population level, there are costs to the health care system that may be associated with undiagnosed OSA. One study estimated that there is currently about $150 billion cost burden associated with adults having undiagnosed OSA in the United States (18). Factors contributing to this burden from OSA among undiagnosed adults were a loss of productivity and absenteeism, costs due to increased risk of co-morbidities (*e.g.* heart disease, depression), motor vehicle accidents, and workplace accidents (18). The estimated cost burden associated with diagnosing and treating every adult living with undiagnosed OSA in the United States was $50 billion; however, this would also result in savings of approximately $100 billion (18). Thus, one could speculate that better screening people with OSA in the general population could help decrease burden to the health care system by providing patients with streamlined diagnosis and treatment. While no study on the economic burden of undiagnosed OSA exists for Canada, the American Association of Sleep Medicine reports that there are a number of downstream effects to leaving OSA undiagnosed (18).

1.2 Gold standard for diagnosing OSA

Polysomnography (PSG) is a full night study that requires collection and assessment of a number of measures (*e.g.* electroencephalogram, oxygen saturation, snoring sound, *etc*.) to confirm or rule out OSA diagnosis. One measure that is used to assist in obtaining an OSA diagnosis is the apnea-hypopnea index (AHI), which provides a score that increases with OSA severity. The apnea-hypopnea index (AHI) is the current framework for assessing the presence and severity of OSA (7). The AHI is computed as a single score based on the average number of apneic and hypopneic events. Obstructive apneic events are those that last longer than 10 seconds during sleep and hypopneic events those that result in a > 30% decrease in airflow from baseline lasting at least 10 seconds with 4% oxygen desaturation or a > 50% decrease in airflow from baseline (1,5,19). There are a number of AHI thresholds defined by the American Academy of Sleep Medicine (AASM, version 2.3, 2016) wherein a patient with a higher AHI score has a more severe form of the disease (7). These thresholds serve as a useful general guideline to help allocate patients into different OSA severity categories. The American Academy of Sleep Medicine (AASM) categorizes OSA severity into four categories based on the AHI score obtained from PSG: An AHI score of less than 5 indicates no OSA (normal breathing during sleep), 5 to less than 15 a mild form of OSA, 15 to less than 30 a more moderate form of OSA, and 30 or greater a severe form of OSA (7). The final diagnosis may also depend on other sleep study measures such as oxygen desaturation levels (19).

1.3 Diagnosis of OSA during perioperative period

There are a number of perioperative risks associated with OSA, especially moderate-to-severe forms of the disease, that when left undiagnosed can lead to a range of potentially serious adverse side effects in patients. A risk management plan developed by anesthesiologists in liaison with the surgical team during the preoperative period is used to help plan for these kinds of risks associated with undiagnosed OSA. Thus, the lack of preoperative diagnosis can pose serious perioperative risks (20). The presence of OSA alone has been shown to increase the risk of hypercapnia, cardiac arrhythmias, delirium, prolonged length of stay, and cardiac arrest perioperatively (21). The decreased blood oxygen levels that company OSA have been shown to increase strain on the heart (22). Furthermore, the consequences of allowing patients to forego OSA diagnosis and undergo surgery are numerous. Surgery and anesthetic have been shown to alter sleep patterns and increase the risk of hypoxic events in OSA patients (23). Opioids like morphine administered to patients with OSA have been shown to cause respiratory depression during sleep that can lead to obstructive apneas and worsened oxygen desaturation (21,23,24). Furthermore, a recent systematic review indicated that moderate-to-severe OSA may be at an even higher risk of complications like myocardial infarction and congestive cardiac failure postoperatively (24). This is particularly concerning given that 80 to 82% of men and 90 to 93% of women with moderate-to-severe OSA remain undiagnosed according to recent literature (21,25). The Society of Anesthesia and Sleep Medicine Guidelines (2016) recommend identifying patients at a higher risk for OSA preoperatively and states that the provision of interventions and targeted perioperative precautions could help minimize perioperative patient complications (26).

1.4 Diagnosis of OSA in the preoperative context

When a patient is suspected of OSA preoperatively, anesthesiologists will ideally have sufficient time to access resources and refer their patient for a polysomnography (PSG), the gold standard diagnostic test. However, due to the resource-intensive nature of PSG, it is not part of standard anesthesiology management guidelines. Rather, preoperative anesthesiology practice guidelines recommend making screening for OSA a standard component of preanesthetic evaluation during patient's perioperative risk management plan development (26). The diagnosis of OSA, especially moderate-to-severe OSA, has a significant influence on the anesthesiologist's perioperative management and has been recognized in the literature as a cut-off indicating a need for OSA treatment (27). The patient may be recommended to use positive airway pressure (PAP), which remains the most effective treatment of moderate-to-severe OSA (9,27).

1.5 Statement of the current problem in the perioperative context

There are a number of perioperative risks associated with OSA, especially moderate-to-severe forms of the disease, that when left undiagnosed can lead to a range of potentially serious adverse side effects in patients. A risk management plan developed by anesthesiologists in liaison with the surgical team during the preoperative period is used to help plan for these kinds of risks associated with undiagnosed OSA. The current problem in the perioperative context is that the gold standard PSG used to diagnose OSA is highly resource intensive and many patients referred for testing are forced to wait from anywhere between several weeks to months to be tested. Due to this limitation, PSG is not a routine test in the preoperative period. Although a

number of screening tools currently exist, they are not without limitations. One of the most commonly used screening tools for OSA is the STOP-Bang questionnaire. Although STOP-Bang is quick and easy to use with a high sensitivity to rule out moderate and severe OSA (93 and 100%), it has a low specificity (37 and 47%) (24). In light of this limitation, there is a need to develop a screening tool with high accuracy to better screen for undiagnosed moderate-to-severe OSA patients during the preoperative period. Although other similarly feasibly to use screening tools have been developed to rule out OSA, including the Berlin Questionnaire and the Epworth Sleepiness Scale, previous studies suggest that STOP-Bang has the higher accuracy (28,29). A prediction rule with higher specificity and similar sensitivity to STOP-Bang could decrease the number of unnecessary referrals for PSG and allow higher risk patients quicker access to much needed therapeutic interventions (*e.g.* CPAP). Tracheal breathing sound analysis (BSA) in awake patients at the bedside has shown promising results to predict the presence of OSA in previous studies (38,39).

A quick and easy to administer screening tool that could minimize referrals for the gold standard PSG has significant potential and could help improve OSA diagnosis guidelines. The early identification of moderate-to-severe OSA has important implications for surgical patients. Early detection allows anesthesiologists to develop an early risk management to help minimize perioperative adverse effects. Thus, there is a need to design a less resource intensive screening tool to help better streamline diagnosis of moderate-to-severe OSA.

1.6 Overview of thesis structure

The first chapter of the present thesis gives an overview of the epidemiological background and impact of OSA at a population level, the need for an effective screening tool to identify patients at the highest risk for moderate-to-severe OSA, and the importance of early identification of patients with moderate-to-severe OSA in the perioperative period. The second chapter presents an article for our prospective cohort study aimed at developing a prediction rule centered at The Ottawa Hospital. Our study is the first to attempt to develop a prediction rule using both BSA and STOP-Bang. In chapter three, the potential implications of the prediction rule in clinical practice are discussed.

CHAPTER 2: Article component

2.1 Preface

Aim of the article

The objective of the article is to give an overview of the findings and implications of a new prediction rule designed to improve screening and management of OSA during the perioperative period.

Author contributions

Emma Grigor, MSc

Emma Grigor was a co-investigator on the study and led all aspects of prospective data collection and data management. Emma led the creation of the prediction rule and drafted the article presented in chapter two of this thesis in partial fulfillment of the requirements for the M.Sc. degree in Epidemiology and Community Medicine (1st author).

Kaitlin Duncan, MD

Dr. Kaitlin Duncan was a co-principal investigator and was responsible for obtaining funding for the study as well as obtaining ethics approval. Dr. Duncan also provided feedback on the research

design, helped to train research assistants for data collection and assisted in providing content expertise and feedback during the writing of the article presented in chapter 2.

Tim Ramsay, PhD

Dr. Tim Ramsay collaborated as a co-investigator on the study and co-supervised the completion of this thesis project. Dr. Ramsay contributed content expertise in the design of the decision-making tool using logistic regression. Dr. Ramsay provided guidance on the statistical analyses and statistical approach to assist in developing the decision-making tool.

Zahra Moussavi, PhD

Dr. Zahra Moussavi was a co-investigator and provided engineering expertise to support the analysis of tracheal breathing sound recordings. Dr. Moussavi has a patent for the Awake-OSA breath sound recording technology.

Ahmed Khalil, BSc

Ahmed Khalil is a graduate student working with Dr. Zahra Moussavi who developed an Awake-OSA algorithm to produce the breathing sound analysis results used in the present study.

Frances Chung, MD PhD

Dr. Frances Chung was a co-investigator on the study and contributed expertise in the area of sleep medicine and OSA diagnosis.

George Chandy, MD

Dr. George Chandy was a co-investigator and contributed expertise in the area of sleep medicine and OSA diagnosis.

Sylvain Boet, MD PhD

Dr. Sylvain Boet was a co-principal investigator and co-supervised the completion of this thesis project. Dr. Boet supervised all aspects of the study and thesis project. Dr. Boet provided clinical expertise and guidance in the area of sleep apnea research, data collection, and design and conduct of the research study.

Other contributions

Sandy Lam: Assistance with data collection; and assistance interviewing and training research assistants.

Mohammed Ibrahim: Assistance with data collection; assistance training of the research assistants; and assistance coordinating data collection schedules.

Katie Chechalk, Jordan Taylor, Emilie Chan, Manvinder Kaur, Chau Huynh and Emil Prikryl: Research assistant volunteers that helped with data collection.

Sarah Schlievert and Jodi Peters: Administrative assistance.

Trial registration

This prospective study was registered prior to the start of participant recruitment at ClinicalTrials.gov (identifier NCT02987283).

Ethics approvals

The protocol was approved (protocol #20160494-01H) by the Ottawa Health Science Network Research Ethics Board. Participants signed an informed consent document that was approved by the Ottawa Health Science Network Research Ethics Board prior to participation in the study (**Appendix 1**).

Appendices

The thesis report provides supplemental information in the following appendices:

Appendix 1. Ethics approval

Appendix 2. Informed consent

Appendix 3. Participant questionnaire

Appendix 4. STOP-Bang questionnaire and scoring criteria

Appendix 5. Standardized tracheal breathing sound script for research assistants

Appendix 6. Standards for Reporting of Diagnostic Accuracy Studies (STARD) research checklist

Appendix 7. Classification tables for BSA, STOP-Bang, and the Safe-OSA rule (N=280)

2.2 Background and rationale

Obstructive sleep apnea (OSA) is a sleep disorder characterized by the collapse of the upper airway during sleep that leads to periods of decreased blood oxygen levels and increased strain on the heart (22). The consequences of allowing patients to forego OSA diagnosis and undergo surgery are numerous. Surgery and anesthesia have been shown to alter sleep patterns and increase the risk of hypoxic events in OSA patients (23). Furthermore, opioids like morphine administered to patients with OSA have been shown to cause respiratory depression during sleep that can lead to obstructive apneas and worsened oxygen desaturation (21,23,24).

The presence of OSA alone has been shown to increase patient risk for adverse side effects in the perioperative period. This includes an increased risk of hypercapnia, cardiac arrhythmias, delirium, prolonged length of stay, and cardiac arrest perioperatively (21). Furthermore, a recent systematic

review indicated that moderate-to-severe OSA may be at an even higher risk of complications like myocardial infarction and congestive cardiac failure postoperatively (24). This is particularly concerning given that 80 to 82% of men and 90 to 93% of women with moderate-to-severe OSA remain undiagnosed according to recent literature (21,25). The SASM Guideline (2016) recommends identifying patients at a higher risk for OSA preoperatively and states that the provision of interventions and targeted perioperative precautions could help minimize perioperative patient complications (26). When a patient is suspected of OSA preoperatively, anesthesiologists will ideally have sufficient time to access resources and refer their patient for a polysomnography (PSG), the gold standard test. The diagnosis of OSA, especially moderate-to-severe OSA, has a significant influence on the anesthesiologist's perioperative management and has been recognized in the literature as a cut-off indicating a need for OSA treatment (30).

In order to prevent postoperative complications associated with surgery and anesthesia, preoperative diagnosis is essential. If a patient has been diagnosed with moderate-to-severe OSA in clinical practice, often they will already be using a recommended therapy, such as CPAP. Patients are known to have moderate-to-severe OSA normally continue to use such therapies throughout the perioperative period. For example, during postoperative recovery in the post-anesthesia care unit moderate-to-severe OSA patients are often placed on CPAP post-extubation. A systematic review of suggested that OSA patients treated with CPAP perioperatively had a lower level of some postoperative complications, including admission to the intensive care unit and pneumonia (31). Thus, the lack of an OSA diagnosis can place patients at a higher risk for certain complications during the perioperative period (20).

PSG is a full night study that requires collection and assessment of a number of measures (*e.g.* electroencephalogram, oxygen saturation, snoring sound, etc.) to obtain an apnea-hypopnea index (AHI) score related to OSA severity. The American Academy of Sleep Medicine (AASM) categorizes OSA severity into four categories based on the AHI score obtained from PSG. An AHI score of less than 5 indicates no OSA, 5 to less than 15 mild OSA, 15 to less than 30 moderate OSA, and 30 or greater severe OSA (7). The final diagnosis may depend on other sleep study measures such as oxygen desaturation levels (19). However, given the resource-intensive nature of PSG, it can often take weeks to receive the sleep study results. Thus, in spite of the perioperative risks associated with undiagnosed OSA, PSG is not part of the anesthesiology practice guideline recommendations.

Current preoperative anesthesiology practice guidelines recommend making screening for OSA a standard component of preanesthetic evaluation as well as identification of patients with higher OSA risk (*e.g.* moderate-to-severe) before surgery (1,26). Although a number of screening tools currently exist, they are not without limitations. One of the most well-known and widely used screening tools is STOP-Bang. STOP-Bang is a questionnaire that is quick and easy to implement. It contains both subjective and objective questions and measurements, resulting in an eight-point score that indicates the likelihood of OSA (24). Although STOP-Bang has demonstrated a high sensitivity to detect moderate and severe OSA (93 and 100%), its low specificity to detect moderate and severe OSA (37 and 47%) is a major limitation. Due to the low specificity, a number of patients that actually have a low risk for OSA are unnecessarily treated with added precaution during the perioperative period, which poses an increased strain on the health care system (24). Previous studies have investigated a new potential screening tool technology for OSA based on breathing

sound analysis performed in awake patients at the bedside (32,33). Awake-OSA is a new technology that incorporates BSA and other anthropomorphic variables. Recent pilot studies including a limited number of subjects have demonstrated that Awake-OSA has a sensitivity (83.8%) similar to STOP-Bang, but with a higher specificity (83.9%) for detecting individuals without OSA compared to STOP-Bang (34).

If the present study demonstrates that breathing sound analysis is able to screen moderate-to-severe OSA with higher accuracy compared to STOP-Bang, it could help support future use of screening tools that incorporate breath sound analysis, such as Awake-OSA. A higher accuracy tool would help to decrease the number of patients requiring a referral for PSG. Furthermore, it could be very useful to rule out undiagnosed moderate-to-severe OSA patients during the perioperative period. Surgical patients suspected of moderate-to-severe OSA are given added precaution in the operating room (e.g. different anesthesia administered) and perioperative therapeutic precautions (e.g. CPAP pre- and postoperatively), which can increase health care costs during the perioperative period and burden (19,30). A higher accuracy tool for ruling out moderate-to-severe OSA would help to decrease the burden to health care system resources. Thus, designing a new prediction rule to rule out patients with moderate-to-severe obstructive sleep apnea based on BSA and STOP-Bang with reasonable accuracy has the potential to provide earlier diagnosis and improve management of OSA. This could be particularly useful implications for patients and health care providers during the perioperative period.

The objective of the present study is to develop a prediction rule that incorporates BSA and STOP-Bang.

2.3 Research methods

2.3.1 Study design and setting

A single center prospective observational cohort study was performed to collect participant baseline, tracheal breathing sounds, STOP-Bang tool questionnaire, and polysomnography data. The study adheres to the Standards for Reporting of Diagnostic Accuracy Studies (STARD) research checklist to ensure transparency and ensure reporting in the present study (**Appendix 5**).

2.3.2 Patient recruitment and screening process

The protocol was approved (protocol #20160494-01H) by the Ottawa Health Science Network Research Ethics Board. Participants scheduled for a PSG at the Ottawa sleep center were eligible to participate in the study. They were admitted by a sleep technician for their sleep study according to the existing outpatient protocol (**Table 1**). Eligible patients were recruited to participate consecutively upon arrival. After a research assistant obtained informed consent, they administered the STOP-Bang questionnaire and performed tracheal breathing sound recording prior to the start of the participant's sleep study. BMI, previous medical history, comorbidities, and smoking history were collected at the study baseline prior to PSG.

Table 1. Participant eligibility criteria

Criteria	Details
Inclusion	• Referral to the Ottawa Sleep Centre for diagnostic obstructive sleep apnea polysomnography or sleep study • Able to provide verbal and written informed consent
Exclusion	• Less than 18 years old • Pregnant • Previous obstructive sleep apnea diagnosis • Patients referred for therapeutic obstructive sleep apnea polysomnography, such as titration of Continuous positive airway pressure (CPAP)

2.3.3 Informed consent and patient confidentiality

Participants signed an informed consent document that was approved by the Ottawa Health Science Network Research Ethics Board prior to participation in the study (**Appendix 1**). Each participant had sufficient opportunity to discuss the study and consider the information in the consent process prior to and during the study. Voluntary and ongoing consent was monitored and ensured for the duration of the study.

2.3.4 Data collection methods, recruitment, and participant timeline

Prior to initiation of data collection, research assistants were trained and familiarized with the method of administration of informed consent, the STOP-Bang questionnaire, and tracheal breathing sound recordings.

Research assistants described the study to patients that were referred to the sleep center for a diagnostic PSG. Only patients that met the full eligibility criteria were invited to participate in the study. Research assistants provided participants a clear description of the breathing sound data collection when first obtaining informed consent and prior to performing breathing sound recordings to ensure ongoing consent. Prior to tracheal breathing sound data collection, research assistants collected data for the STOP-Bang questionnaire tool from consenting participants (**Appendix 2**).

Research assistants used a standardized procedure to collect participant tracheal breathing sound data (**Appendix 3**) (33). This involved placement of a microphone (ECM77B Sony, Toronto, Canada) over the skin above the suprasternal notch of the trachea on participants as shown in figure 1. Four BSA recordings in the supine position with and without pillow breathing through the nose and mouth separately were taken for each study participant. Participants were guided by a research assistant using non-verbal hand cues through a total of five normal and five deeper breaths. A PreSonus AudioBox iOne (Los Angeles, United States) was used for sound acquisition. The Audacity program (Audacity 2.1.3, open source) captured and saved the breath sound files in WAV format. We used audio recording equipment that was capable of signal resolution and sampling rates of 24 bits per second and 11025 Hz, respectively. WAV files were

sent to Manitoba for breathing sound analysis to determine predicted participant OSA outcomes (normal-to-mild versus moderate-to-severe OSA). Audio WAV files (containing breath sound recordings) were uploaded to a secure server hosted at the University of Manitoba, for pre-processing by their laboratory. If audio recordings were found to be suboptimal for analysis during pre-processing (*e.g.* sound artifacts), the corresponding participant data was excluded from future analyses at that time by listwise deletion.

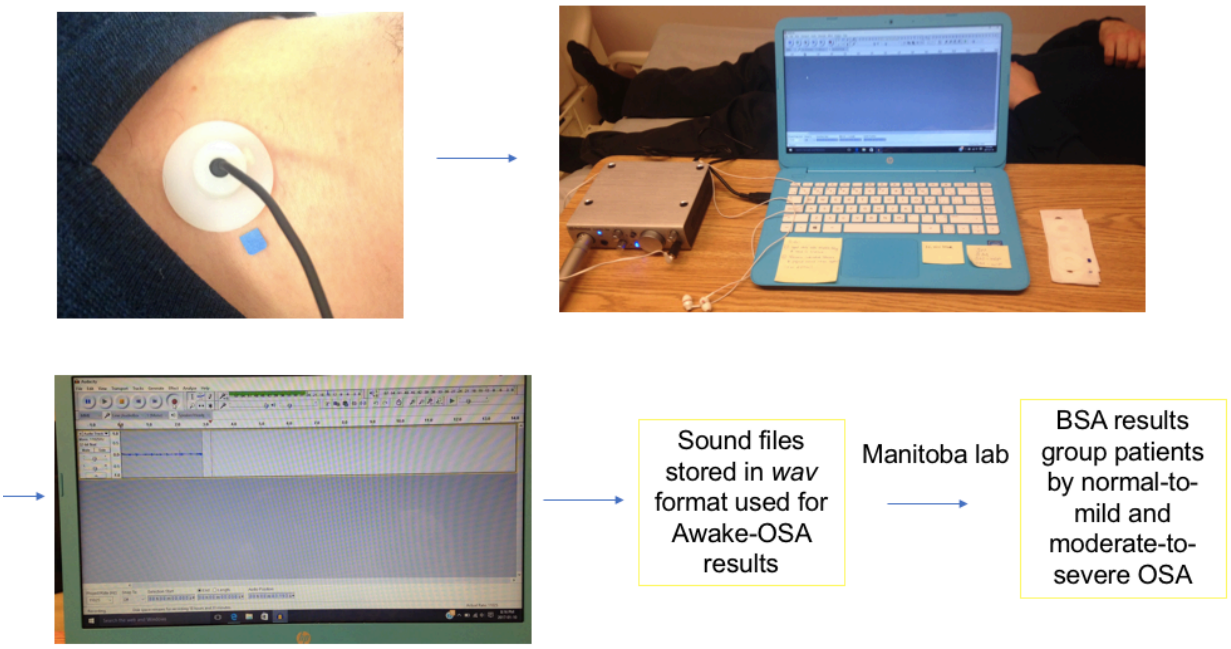

Figure 1. Sound acquisition for breathing sound analysis.

The STOP-Bang questionnaires and tracheal breathing sound data were obtained the same evening and immediately prior to the participants scheduled PSG. The approximate duration to obtain informed consent and STOP-Bang questionnaire data was ten minutes while the tracheal breathing sound data collection was fifteen minutes. The STOP-Bang questionnaires and tracheal breathing sound data were obtained the same evening and immediately prior to the participants

scheduled PSG. The STOP-Bang questionnaire results and tracheal breathing sound recordings were obtained for all patients immediately after informed consent was obtained (within one-hour maximum).

Certified sleep technicians collected PSG data (Sandman Elite PSG Software, Pleasanton, United States) over the course of an 8-hour period during the evenings at the Ottawa Hospital Sleep Centre from November 2017 to May 2018. The PSG results were scored by Sleep Medicine Specialists using the 2016 American Academy of Sleep Medicine criteria (7). Polysomnograms were scored using an apnea-hypopnea index (AHI) score, which was based on the number of apneic and hypopneic events that occur during sleep (19). Apneic events are those that last longer than 10 seconds during sleep and hypopneic events those that result in a $\geq$ 30% decrease in airflow from baseline lasting at least 10 seconds with 4% oxygen desaturation or a $\geq$ 50% decrease in airflow from baseline (35). The final physician diagnosis (normal, mild, moderate, severe OSA) was collected in addition to the total AHI score from the electronic Clinical Information System (OACIS, The Ottawa Hospital). The AHI scores were then categorized into normal-to-mild OSA and moderate-to-severe OSA physician diagnosis groups. Sleep Medicine Specialists were blinded to the results of breathing sound analysis to avoid bias in the final physician diagnosis results.

Our team in Manitoba performed an analysis on the breathing sound recordings using the WAV files to predict patients physician diagnosis of OSA. After the visual inspection of the signals in time and frequency domains, features were extracted from the four recordings. The features went through a feature reduction process to keep only the most robust, reliable, discriminative and

non-correlated features. The training (59% of the dataset) and testing (41% of dataset) data were selected randomly from our data with constraints (required that the datasets were balanced by gender, age, and AHI prior to starting the analysis). Using the training data which involved a cross-validation process, the best feature combinations giving the highest overall accuracies were selected for each anthropometric information subset of data. These combinations were used to generate a multi-classifier process, each sub-classifier contributed with a classification decision per subject. These decisions went through a voting algorithm to result in the final decision. The training data were used to train the classifiers for classifying the testing data. PSG results were provided to our team in Manitoba only during the training phase, but not during the testing phase to ensure proper blinding when validating BSA algorithm. The final BSA algorithm classified our study participants into categories based on a clinically relevant AHI cut-off of less than 15 (normal-to-mild OSA) and greater than or equal to 15 (moderate-to-severe OSA) that indicates a need for intervention (*e.g.* PAP) and need for further testing by PSG. All data were saved in a secure password protected database (24,25). Only the AHI classifications (AHI < 15 or AHI $\geq$ 15) determined by the BSA algorithm obtained from our participant tracheal breathing sound results were used in the present study.

2.3.5 Stop-Bang questionnaire

The STOP-Bang questionnaire is a previously validated tool that is used to screen patients for OSA. STOP-Bang answers eight yes or no questions based on patients self-reported symptoms and physical characteristics. An additional point is counted when the patient responds ‘yes’ to a question. The overall sum is then used to classify patients by their risk for OSA (36,37). If

patients respond yes to two or fewer questions they are considered a 'low risk' for OSA, yes to 3 to 4 questions an 'intermediate risk', and five to eight questions a 'high risk' (**Appendix 4**). In the present study, variables from the STOP-Bang questionnaire were considered for inclusion in the prediction rule.

2.3.6 Breathing sound analysis

There has been emerging research relying on tracheal breathing sound analysis (BSA) in awake patients at the bedside to predict the presence of OSA (38,39). A diagnosis of moderate-to-severe OSA, consistent with an apnea-hypopnea index (AHI) scores greater than 15, has been demonstrated to be clinically relevant for further therapeutic intervention (19). In the present study, BSA was used to attempt to categorize patients by normal-to-mild and moderate-to-severe OSA.

2.3.7 Data analysis

2.3.7i Descriptive statistics

The outcome of interest was OSA physician diagnosis and baseline characteristics were collected for study participant's, which included gender, age, BMI, hypertension, loud snoring during sleep, observed apneas during sleep, smoking status, daytime fatigue, and neck circumference. Univariate analyses were performed to provide an overview of the distribution of variables of interest for included participant data. A χ^2 test of independence was utilized for categorical

variables and an independent *t*-test for continuous variables to explore their association with OSA. Study participants with tracheal breathing sound data of poor quality or incomplete survey data were excluded from all analyses by listwise deletion. Definitions for variable measurement, coding, and reference groups are provided in **Table 2**. All univariate analyses were considered statistically significant for *p*-values less than 0.05.

Table 2. Definitions of the study variables included in logistic regression

Variable	Measurement definition	Coding	Reference group
OSA Physician Diagnosis	Physician (sleep specialist) diagnosis based on results of polysomnography test (sleep study)	OSA physician diagnosis classified as normal-to-mild OSA, an AHI < 15, (0) and moderate-to-severe OSA, an AHI > 15, (1)	Not applicable[a]
Breathing sound analysis	Predicted diagnosis of OSA based on results of breathing sound analysis	Predicted diagnosis of OSA classified as normal-to-mild OSA, an AHI < 15, (0) and moderate-to-severe OSA, an AHI > 15, (1)	Normal-to-mild OSA (AHI < 15)
Gender	Participants reported gender at time of self-report	Gender status classified as female (0) or male (1)	Females

Variable	Measurement definition	Coding	Reference group
Age	Respondents reported age (years) at time of self-report	Age dichotomized to less than or equal to 50 years old (0) and greater than 50 years old (1)	Less than or equal to 50 years old
Neck circumference	Respondents had neck circumference (cm) measured at time of self-report	Neck circumference dichotomized to less than or equal to 42 cm (0) and greater than 42 cm (1)	Less than or equal to 42 cm
Daytime fatigue	Respondents reported presence of daytime fatigue at the time of self-report	Daytime fatigue classified as absent (0) or present (1)	Absent (0)
Loud snoring	Respondents reported presence of loud snoring during sleep at the time of self-report	Loud snoring classified as absent (0) or present (1)	Absent (0)
BMI	Respondents reported BMI (kg/m^2) at time of self-report	BMI dichotomized to less than or equal to 35 kg/m^2 (0) and greater than 35 kg/m^2 (1)	Less than or equal to 35 kg/m^2

Variable	Measurement definition	Coding	Reference group
Observed apneas	Respondents reported presence of observed apneas (*e.g.* by spouse) during sleep at the time of self-report	Observed apneas classified as absent (0) or present (1)	Absent (0)
Hypertension	Respondents reported presence of hypertension of self-report	Hypertension classified as absent (0) or present (1)	Absent (0)

Abbreviations: BP, blood pressure; BMI, body mass index; OSA, obstructive sleep apnea

[a]OSA physician diagnosis was the outcome of interest (independent variable)

2.3.7ii Logistic regression analysis

A prediction model for the dichotomous outcome of interest physician diagnosis of OSA by polysomnography (normal-to-mild OSA versus moderate-to-severe OSA) was developed using multiple logistic regression [SAS 9.4, Cary NC USA].

The multiple logistic regression model was generated using a manual backward selection procedure to attempt to create a prediction rule for OSA physician diagnosis (normal-to-mild versus moderate-to-severe OSA). The BSA and STOP-Bang questionnaire variables were explored as potential independent predictors of interest to be included in the model building. The STOP-Bang variables included loud snoring during sleep, daytime fatigue, observed apnea during sleep, hypertension, gender, body mass index (BMI $\leq$ or $>$ 35 kg/m^2), age ($\leq$ or $>$ 50 years old), and neck circumference size ($\leq$ or $>$ 42 cm). A χ^2 test of independence was used to assess possible associations between predictor variables with the outcome of interest as well as to determine any potential collinearity between predictor variables prior to model building. Variables entered the model at the preliminary model if they were statistically significantly associated with the outcome of interest at a cut-off of $p < 0.1$. During backward selection, variables with the highest p-values were removed in iterative steps until all included predictor variables were statistically significant with $p < 0.05$. Finally, all removed predictor variables were individually entered back into the model determine if they became significant (defined as $p < 0.05$). Multiple logistic regression odds ratios were determined to be statistically significant for 95% confidence intervals that did not cross the null hypothesis value.

Cross-validation is an alternative validation method to external validation when there is a lack of an independent population. Cross-validation allows for evaluation of the internal validity for a given prediction model (*e.g.* whether the results can be generalized to the population used to fit the model) (40). Cross-validation is a process of testing a prediction model in an independent subsample of data that was excluded from the original population of interest during model derivation or building. In k-fold cross-validation, the process of testing (or internally validating) the prediction model in this independent subsample (made up of data randomly removed from the original population prior to model building) is repeated *k* number of times. If the model shows similar overall accuracy (sensitivity and specificity) in the derivation and testing datasets, this will provide support for the internal validity of the final model in our population of interest.

The final model performance was assessed by determining both discrimination (*C* statistic, 95% CIs) and calibration (Hosmer and Lemeshow goodness-of-fit test) (41–43). The *C* statistic is equivalent to the area under the receiver operating characteristic (ROC) curve. An area under the curve of 0.5 indicates that the model performs similarly to predicting the outcome due to chance alone. Model predictive performance is usually considered reasonable where the *C* statistic is greater than 0.70 and strong when greater than 0.80 (44).

The final model accuracy was assessed through analysis of sensitivity and specificity (at a range of different test thresholds). Statistical significance was determined for *p*-values less than 0.05 (two-tailed).

2.3.8 Sample size calculation

A prevalence of 40% moderate-to-severe OSA participants is predicted among the study population being recruited at the Ottawa Sleep Centre based on previous literature (25). It was determined that at least 250 participants needed to be recruited in order to determine sensitivity and specificity within a 5% confidence interval, based on the predicted 40% prevalence of OSA. In order to account for potential loss of breathing sound data (*e.g.* poor sound quality) a sample size of at least 40% greater than the calculated 250 was sought (*i.e.* 400 or more patients).

2.4 Results

Out of 439 consenting participants, 280 eligible study participants were included in the BSA and prediction rule analysis. All participants had been referred for polysomnography at the Ottawa Sleep Centre between November 2016 and May 2017 (**Figure 2**). The data for 159 participants was excluded by listwise deletion where 136 had poor sound data, 18 were missing survey data, and 5 were missing sound data. Of the 280 participants, 114 study participants (41%) were diagnosed with moderate-to-severe OSA and 166 participants (59%) with normal-to-mild OSA based on the polysomnography results evaluated by a physician sleep specialist.

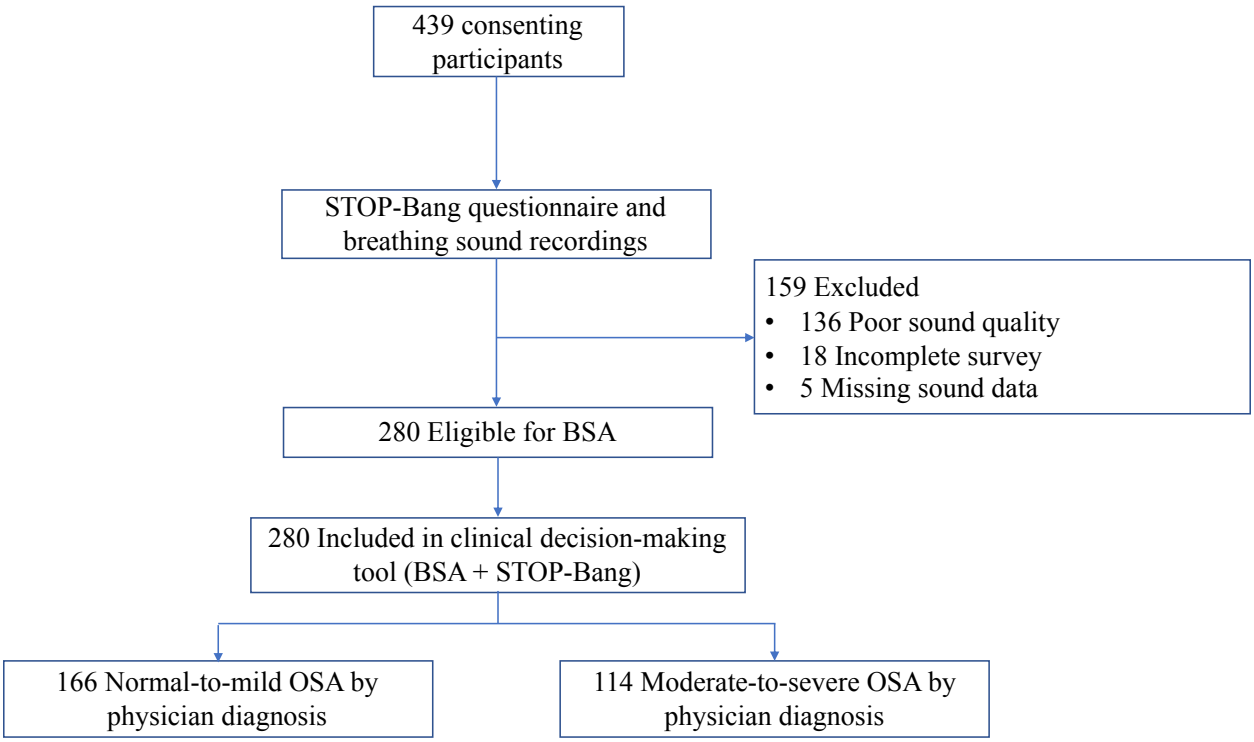

Figure 2. Flow diagram of study participant recruitment.

2.4.1 Descriptive statistics

Univariate analysis performed on patient anthropomorphic variables of interest were grouped according to the outcome of interest, physician diagnosis for OSA (grouped according to normal-to-mild or moderate-to-severe OSA) (**Table 3**). Continuous variables age and BMI and the outcome were described as a mean (SD) and analyzed using an independent *t*-test. Categorical variables, sex, race, smoking status, and co-morbidities were described as the number of patients and percentages and were analyzed using a χ^2 test of independence. Of the 280 patients with good quality sound data, there was a statistically significantly higher age, BMI and percentage of participants who self-reported being male, having a history of a stroke or mini-stroke, and having

a history of asthma in the group diagnosed with moderate-to-severe OSA compared to the group with normal-to-mild OSA ($p < 0.05$) (**Table 3**).

Table 3. Baseline characteristics of study participants (N = 280) grouped according to the presence of normal-to-mild and moderate-to-severe OSA by physician diagnosis

Variable[a]	**Normal-to-mild OSA (total 166)**	**Moderate-to-severe OSA (total 114)**	***p*-value**[*]
Age, mean years (SD)	46.9 (13.9)	52.2 (14.1)	**0.002**[*]
BMI, mean kg/m^2 (SD)	29.7 (6.1)	33.2 (8.2)	**<0.001**[*]
Sex, male no. %	73 (44)	78 (68)	**<0.001**[*]
Ethnicity, no. %			
Caucasian	139 (84)	95 (83)	
Asian	11 (7)	6 (5)	
African American	10 (6)	4 (4)	0.599
Arabian	5 (3)	6 (5)	
Other	4 (2)	1 (1)	
Smoking status, no. %			
Current	18 (11)	7 (6)	
Former	43 (26)	32 (28)	0.596
Never	106 (64)	71 (62)	
Unknown	2 (1)	2 (2)	

Variable[a]	Normal-to-mild OSA (total 166)	Moderate-to-severe OSA (total 114)	*p*-value[*]
Co-morbidities, no. %			
Hypertension	37 (22)	37 (32)	**0.052**[*]
Diabetes	10 (6)	11 (10)	0.251
MI	7 (4)	5 (4)	1.00
CAD	6 (4)	4 (2)	1.00
Kidney disease	2 (1)	2 (2)	0.652
COPD	1 (1)	2 (2)	0.448
Stroke or mini-stroke	1 (1)	5 (4)	**0.039**[*]
Asthma	34 (20)	11 (10)	**0.030**[*]
GERD	54 (32)	31 (27)	0.508

Abbreviations: BMI, Body Mass Index; CAD, coronary artery disease; COPD, Chronic obstructive pulmonary disease; GERD, Gastroesophageal reflux disease; MI, myocardial infarction; no., number; SD, standard deviation

[a]Independent *t*-tests and chi-square tests used for continuous and categorical variables, respectively

[*]Considered statistically significant for *p*-values less than 0.05

2.4.2 Comparing the baseline characteristics for included and excluded patient data

The distribution of important patient anthropomorphic variables and the outcome physician diagnosis of OSA were similar for the data in the group of excluded versus included participant data. Age alone showed a statistically significant difference ($p = 0.025$) across included and excluded patient groups (**Table 4**). However, this statistically significant difference in the mean

age comparing included and excluded patient groups, between 48.9 and 52.4 years old, was not deemed a clinically relevant difference.

Table 4. Baseline characteristics of study participants (N = 439) grouped according to inclusion or exclusion due to poor sound quality

Variable[a]	**Excluded, poor sound quality (total 159)**	**Included (total 280)**	***p*-value**[*]
Age, mean years (SD)	52.4 (15.6)	48.9 (14)	**0.025***
BMI, mean kg/m^2 (SD)	30.8 (7.3)	31.1 (7.2)	0.709
Sex, male no. %	80 (57)	151 (54)	0.437
Co-morbidities, no. %			
Hypertension	45 (32)	74 (26)	0.338
COPD	5 (3.6)	3 (1.1)	0.198
Physician diagnosis, no. %			
Normal-to-mild OSA	73 (52)	166 (59)	0.210
Moderate-to-severe OSA	68 (48)	114 (41)	

Abbreviations: BMI, Body Mass Index; COPD, Chronic obstructive pulmonary disease; no., number; OSA, obstructive sleep apnea; SD, standard deviation

[a]Independent *t*-tests and chi-square tests used for continuous and categorical variables, respectively

[*]Considered statistically significant for *p*-values less than 0.05

2.4.3 Accuracy results for BSA and STOP-Bang

At a threshold probability of greater than or equal to 0.5, BSA had a higher specificity of 74.5% (95%CI; 68.5, 82.0) and comparable sensitivity of 75.9% (95%CI; 65.4, 82.0) when compared to the STOP-Bang, which had a specificity of sensitivity and specificity of 68.4% (95%CI; 58.9, 76.6) and 63.2% (95%CI: 55.0, 70.1) as shown in **Table 5**. The classification tables for BSA alone and STOP-Bang are available in **Appendix 7**.

Table 5. Sensitivity and specificity of BSA and STOP-Bang alone

	Sensitivity	**95% CI**	**Specificity**	**95% CI**
BSA alone	74.5	(65.4, 82.0)	75.9	(65.4, 82.0)
STOP-Bang alone	68.4	(58.9, 76.6)	63.2	(55.0, 70.1)

Abbreviations: BSA, breathing sound analysis; CI, confidence interval

*Predicted probability threshold of 0.5

2.4.4 Safe-OSA rule model derivation

It was determined that the BSA variable and a number of STOP-Bang variables were statistically significantly associated with the presence of moderate-to-severe OSA by physician diagnosis (**Table 6**). STOP-Bang questionnaire variables that demonstrated an association with the physician diagnosis of OSA at a significance level of $p < 0.1$ were entered into the model. It was demonstrated using a χ^2 test of independence that loud snoring during sleep, observed apnea during sleep, hypertension, gender, body mass index (BMI $\leq$ or > 35 kg/m), age ($\leq$ or > 50 years old), and neck circumference size ($\leq$ or > 42 cm) were statistically associated with OSA at the cut-off of $p < 0.1$ (**Table 6**).

Table 6. Distribution of variables of interest and OSA status by physician diagnosis

Variable	Normal-to-mild OSA, % (total 166)	Moderate-to-severe OSA, % (total 114)	*p*-value*
BSA, moderate-to-severe OSA (AHI ≥ 15)	40 (24)	84 (74)	**<0.0001***
Age, > 50 years old	73 (44)	48 (55)	**0.0689***
BMI, > 35 kg/m^2	22 (19)	34 (30)	**0.0319***
Loud snoring during sleep, present	120 (72)	98 (86)	**0.0081***
Daytime fatigue, yes	125 (75)	82 (72)	0.5801
Observed apneas during sleep, present	46 (28)	42 (37)	**<0.0001***
Hypertension, yes	37 (22)	38 (33)	**0.0383***
Neck circumference, > 42 cm	37 (22)	46 (40)	**0.0014***
Gender, male	95 (57)	36 (32)	**<0.0001***

Abbreviations: BSA, breathing sound analysis; BMI, body mass index; no., number; OSA, obstructive sleep apnea

[a]Chi-square test used to determine significance of potential statistical associations

*Entered into the model for *p*-values less than 0.1

After entering variables that were statistically significantly associated with the outcome of interest into the model, backward stepwise regression was used to obtain a final model with the highest predictive value (**Table 7**). The final parameters in the prediction rule, called the Safe-OSA rule , included BSA, BMI (≤ and > 35 kg/m^2), and observed apneas during sleep. The odds ratio statistics for the Safe-OSA rule are provided in **Table 7**. None of the other variables listed in **Table 6** met the criteria for inclusion in the final model.

Table 7. Safe-OSA rule odds ratio statistics

Variables	Unadjusted OR Statistics	95% CI	Adjusted OR Statistics	95% CI
BSA (AHI < or ≥ 15)	9.233	(5.319, 16.027)	8.067	(4.601, 14.142)
Observed apneas (yes or no)	3.002	(1.819, 4.954)	2.057	(1.164, 3.634)
BMI (≤ or > 35 kg/m^2)	1.851	(1.057, 3.239)	1.761	(0.904, 3.428)

Abbreviations: BSA, breathing sound analysis; BMI, body mass index; CI, confidence interval, OR, odds ratio

2.4.5 Safe-OSA rule model performance

The final logistic model, the Safe-OSA rule, which contained the BSA (AHI < or ≥ 15), BMI (≤ or > 35 kg/m^2), and observed apnea (yes or no) variables, was determined to be reasonably discriminative (*C* statistic 0.794, 95% CI 0.782 to 0.814) and well calibrated (Hosmer and Lemeshow goodness of fit statistic 3.54, 2 degrees of freedom, $p = 0.9125$). Sensitivity and specificity were determined over a range of thresholds (**Table 8**). At a predicted probability of moderate-to-severe OSA greater than or equal to 0.5 (approximately the same as the prevalence of moderate-to-severe OSA in our sample population at 41% and maximizes the Youden's index), the sensitivity and specificity for the final prediction rule were determined to be 74.5 (95%CI; 65.4, 82.0) and 75.9% (95%CI; 68.5, 82.0), respectively (45). A sensitivity analysis

using a likelihood ratio test showed that overall the Safe-OSA rule performed better compared to a model with BSA alone (Deviance statistic = 9.508; df = 2; $p = 0.002$) (**Table 8**).

Table 8. Sensitivity and specificity of the Safe-OSA rule

	Sensitivity	**95% CI**	**Specificity**	**95% CI**	***C* statistic**
Safe-OSA rule[a]	74.5	(65.4, 82.0)	75.9	(68.5, 82.0)	0.794

Abbreviations: BSA, breathing sound analysis; CI, confidence interval

[*]Predicted probability threshold of 0.5

[a]Safe-OSA rule included BSA (AHI < or ≥ 15), BMI (≤ or > 35 kg/m^2), and observed apneas during sleep (yes or no) predictive variables

2.4.6 Model cross-validation (internal validation)

The final model was cross-validated using a five-fold method. First, the data was pseudo-randomly divided into 5 different subgroups. Then, 80% (4 of the five-folds) of the data was used for derivation of the model and the remaining 20% (remaining 1 fold) used for testing the model performance. This model was refitted five times using the same process. The model sensitivity, specificity, and discrimination (*C* statistic) for the five-fold cross-validation were averaged in the derivation and testing datasets (**Table 9**).

Table 9. Performance results from five-fold cross-validation procedure*

Average	Sensitivity[a]	Specificity[a]	*C* statistic[b]
Derivation subsample	73.7	75.9	0.790
Testing subsample	71.0	78.5	0.810

*Five-fold cross-validation was used perform internal validation of final logistic regression model

[s]Sensitivity and specificity were compared to support the validity of our model performance

[b]*C* statistic was used to assess the discriminative ability of the model in the derivation (training) and testing subsamples

2.5 Discussion

2.5.1 Interpretation of the results

In the present study, we performed a prospective cohort study in order to develop a prediction rule from BSA and STOP-Bang. A polysomnography (PSG), a sleep study, is currently the gold standard to obtain a formal diagnosis for patients suspected of OSA. However, due to the resource-intensive nature of PSG it is not a recommended test to be performed routinely during the preoperative period (3-4). The results from our study support future research to further explore breathing sound analysis to rule out OSA as a potential screening tool. A validated screening tool incorporating BSA that has higher specificity compared to STOP-Bang, with reasonable sensitivity, could be especially useful for anesthesiologists in the perioperative setting to help evaluate the need for therapeutic intervention.

At a threshold probability of greater than or equal to 0.5, BSA alone had a higher specificity of 74.5% (95%CI; 65.4, 82.0) and comparable sensitivity of 75.9% (95%CI: 57.0, 69.2) when compared to the STOP-Bang alone, which had a sensitivity and specificity of 68.4% (95%CI; 58.9, 76.6) and 63.2% (95%CI: 55.0, 70.1) respectively. The Safe-OSA rule was developed from the BSA and STOP-Bang results using logistic regression methods. The final Safe-OSA rule model demonstrated reasonable discriminative performance (*C* statistic 0.794, 95% CI 0.782 to 0.814) and good accuracy to predict the presence of normal-to-mild or moderate-to-severe OSA in participants referred for PSG the Ottawa Sleep Centre using BSA and variables from the STOP-Bang questionnaire. A sensitivity analysis (comparing our final Safe-OSA rule model with and without BSA) indicated that our prediction rule, which had a reasonable discriminative ability, performed significantly better overall with BSA included in the model. However, at a probability threshold of 0.5, BSA had the same sensitivity and specificity as the Safe-OSA rule. This means that BSA alone has the potential to rule out OSA at a predicted probability of 0.5 with similar accuracy when compared to our more complex Safe-OSA rule.

External validation is considered to be a highly reliable method to test the validity and generalizability of a predictive model. However, when an independent population is not available for external validation, there are several internal validation methods that can be implemented as an alternative. Two common methods for internal validation of predictive models, including models developed using logistic regression, include bootstrap resampling and cross-validation. Bootstrap resampling involves fitting the logistic regression model to the bootstrap sample of n study participants taken from the original study sample. The averages for the measures of

performance are taken over a large number of repetitions of this approach. Using the performance of the bootstrap sample and the original sample you obtain the apparent performance and test performance, respectively. The difference between these performances, referred to as the estimate of the optimism in the apparent performance, can be determined. The estimated performance, which is used to help indicate the internal validity of the model, is determined by subtracting the optimism from the apparent performance of the bootstrap sample (46). Cross-validation is another technique to internally validate a predictive model that involves fitting the model to a derivation sample of n participants taken from the original sample. In K-fold cross-validation, the original sample is divided into multiple subsamples. The derivation sample is generated by excluding one of the subsamples. Then the model is derived on the derivation sample and tested on the excluded subsample. This same process is repeated for multiple procedures, excluding a different subsample each time. The performance for all derivation samples and testing samples is averaged and compared to assess the internal validity. In the present study, a five-fold cross-validation procedure was performed and revealed a similar average sensitivity, specificity, and discriminative performance in the training and testing subsamples. This strengthened the validity of the original model performance in the population studied. Both bootstrap resampling and cross-validation would have been an option for internal validation of the logistic regression prediction rule. Familiarity with the cross-validation technique was the main rationale for its use in this study. The five-fold cross-validation showed similar results for the final model in the derivation and testing subsamples, which provides support for the internal validity of the Safe-OSA rule in our study population.

2.5.2 Strengths and limitations

This was the first study to develop a prediction rule to incorporate the results of tracheal breathing sound analysis and variables from the STOP-Bang questionnaire. However, prior to beginning this study, our initial plan was to perform a validation study of the Awake-OSA technology, developed in Manitoba previously. Due to unforeseen differences in the amplifier used in the present study, a substantial amount of frequency signals were lost in the sound recordings obtained in our population. This meant that only about 64% of the data we collected had any analyzable signals. This forced us to change our original plan from a validation study to a developmental study whereby we essentially created a new screening tool based on usable breathing sound samples and variables collected from STOP-Bang. There is still value in the findings from the present study. In spite of the significant loss of data and poor quality of the usable data, the results from breathing sound analysis alone still performed better than the STOP-Bang screening tool with a higher specificity and similar sensitivity. This lends further support to the promise that the Awake-OSA tool, which incorporates a breathing sound analysis component, will perform well when tested on good quality data.

A previous systematic review of biases in diagnostic accuracy studies indicated that retrospective design and nonconsecutive recruitment of patients was associated with overestimated accuracy (47). Thus, the prospective nature of our study is a strength. We also aimed to decrease biases associated with BSA outcome ascertainment through blinding. Our team in Manitoba that obtained the BSA results were blinded to the PSG results during the testing phase to prevent detection bias when validating the BSA algorithm (normal-to-mild versus moderate-to-severe OSA). Our study also attempted to minimize systematic biases in terms of data collection. We

trained numerous research assistants to assist with collecting the breathing sound recordings (following a standard protocol) as well as had multiple physician sleep specialists evaluate polysomnography results. Furthermore, we compared levels of patient demographic variables between the group that was included for BSA and the group excluded due to missing data or poor sound quality to rule out possible selection bias due to sound quality. We demonstrated that there were no substantial differences and thus increased our certainty that selection bias was not an influence on participant inclusion in the development of the prediction rule (47).

There were some potential limitations that warrant consideration. One major limitation was that a significant amount of participant data had to be excluded due to the poor quality of the tracheal breathing sounds recorded. We attempted to minimize any errors in the recording of tracheal breathing sounds by standardizing the training of research assistants for data collection. However, 159 of 439 participants (36%) were still excluded after recruitment due to poor sound recording quality to determine their BSA score. A number of these exclusions were due to artifacts in the environment outside the participant's room during recording external environmental noise (*e.g.* noises outside the participant room captured in the recording) or rubbing of physical artifacts against the microphone during recording (*e.g.* participant chest hair or collar of their shirt rubbing against microphone).

Another limitation is that our population included in the study was not limited to surgical patients, but rather any patients eligible that were referred for PSG. Thus, the results may not be generalizable to surgical populations, which is one of the main populations of interest in the present study. Therefore, external validation of this model in independent surgical patient

populations is imperative in order to improve certainty around our prediction rule generalizability to populations suspected of OSA during the perioperative period.

Some possible biases associated with our study may have included selection bias and small sample bias. Selection biases might have resulted from the single-center design of our study where participants were recruited from The Ottawa Hospital Sleep Centre. Multicenter trial designs have higher heterogeneity within the study sample and thus increase generalizability (or external validity) of results to the population of interest at large (48). Repeating the study at other sleep centers would help to assess the external validity of our model. Also, our tool was developed based on a study sample of 280 eligible participants and may be subject to small sample bias (49,50).

2.6 Clinical significance

Polysomnography (or sleep study) is the gold standard test to diagnose OSA. However, this test is very resource-intensive and current guidelines recommend screening tools to streamline the diagnosis of OSA in the perioperative period. Our study developed a new prediction rule called the Safe-OSA rule to improve the detection of patients with moderate-to-severe OSA by combining the results of BSA and the STOP-Bang questionnaire. This new prediction rule was demonstrated to have reasonable discriminative performance and a fairly high accuracy.

Currently, this prediction rule is not ready for implementation into clinical practice or anesthesia practice guidelines (*e.g.* ASA guidelines). The tool must be validated in other external

populations to ensure validity and generalizability. Furthermore, the feasibility of using BSA must be improved to make this tool easier to administer. In the present study, we had to rely on performing a 15-minute recording to obtain four different samples of breathing sound data per study participant in a quiet environment. Afterwards, we had to send this data for further processing and analysis by our team in Manitoba. Although the tool is not currently usable in clinical practice, the ultimate aim of future research would be to develop a higher accuracy screening tool to better streamline the early identification and management of patients with OSA in the perioperative period (51).

2.7 List of abbreviations

AE; adverse event; AHI, apnea-hypopnea index; BMI, body mass index; CPAP, continuous positive airway pressure; OSA, obstructive sleep apnea; PSG, polysomnography; REB, research ethics board; ROC, receiver operator curve; STARD, standards for reporting diagnostic accuracy studies; TOHAMO, The Ottawa Hospital Medical Organization.

2.8 Funding and declaration of interests

TOHAMO funded this clinical trial. The funding sponsor was not involved in the study and collection, analysis, or interpretation of data and writing the manuscript.

The TOHAMO grant provided the MSc stipend funding for Emma Grigor. Emma Grigor received a Canada Graduate Scholarship for her Master's Program, which was funded by the Canadian Institutes of Health Research.

Dr. Sylvain Boet was supported by The Ottawa Hospital Anesthesia Alternate Fund Association.

The 3rd author, ZM, holds two patents on "Acoustic System and Methodology for identifying the risk of obstructive sleep apnea during wakefulness", US 0140142452 A1, and US 20130253357 A1; both are licensed to BomiMed Ltd.

2.9 References

1. Kapur VK, Auckley DH, Chowdhuri S, Kuhlmann DC, Mehra R. Clinical Practice Guideline for Diagnostic Testing for Adult Obstructive Sleep Apnea : An American Academy of Sleep Medicine Clinical Practice Guideline. 2017;13(3):479–504.
2. Yarlagadda A, Kaushik S, Clayton AH. Obstructive Sleep Apnea , Hypoxia , and Metabolic Syndrome in Psychiatric and Nonpsychiatric Settings. 2008;
3. Gottlieb JD, Schwartz AR, Marshall J, Ouyang P, Kern L, Shetty V, et al. Hypoxia , Not the Frequency of Sleep Apnea , Induces Acute Hemodynamic Stress in Patients With Chronic Heart Failure. JAC [Internet]. 2009;54(18):1706–12. Available from: http://dx.doi.org/10.1016/j.jacc.2009.08.016
4. Xie J, Kuniyoshi FHS, Covassin N, Singh P, Gami AS, Wang S, et al. Nocturnal Hypoxemia Due to Obstructive Sleep Apnea Is. :1–9.

5. Tasali E, Mokhlesi B, Van Cauter E. Obstructive sleep apnea and type 2 diabetes: Interacting epidemics. Chest [Internet]. 2008;133(2):496–506. Available from: http://dx.doi.org/10.1378/chest.07-0828

6. Yaggi, H Klar and Concato, John and Kernan, Walter N and Lichtman, Judith H and Brass, Lawrence M and Mohsenin V. Obstructive sleep apnea as a risk factor for stroke and death. N Engl J Med. 2005;353(19):2034–41.

7. Berry RB, Brooks R, Gamaldo CE, Harding SM, Lloyd RM, Marcus CL, et al. The AASM Manual for the Scoring of Sleep and Associated Events. 2015;1–7.

8. Pinto JA, Ribeiro DK, Da Silva Cavallini AF, Duarte C, Freitas GS. Comorbidities associated with obstructive sleep apnea: A retrospective study. Int Arch Otorhinolaryngol. 2016;20(2):145–50.

9. Kaw R, Gay PC. Perioperative Management of Obstructive Sleep Apnea. 2015;251–6.

10. Williams JS, Tacon A, Carter R. Losing sleep over weight: obesity and OSA secondary to obesity represent major health hazards. Untreated OSA is associated with cardiovascular disease, hypertension, mood/memory disorders, and motor vehicle accidents. RT Decis Makers Respir Care. 2008;21(5):20–3.

11. Peppard PE, Young T, Barnet JH, Palta M, Hagen EW, Hla KM. Original Contribution Increased Prevalence of Sleep-Disordered Breathing in Adults. 2013;177(9):1006–14.

12. Newman AB, Foster G, Givelber R, Nieto FJ, Redline S, Young T. Progression and Regression of Sleep-Disordered Breathing With Changes in Weight. 2005;165.

13. Vgointzas A, Tan T, Bixler E, Martin L, Shubert D, Kales A. Sleep Sleep. Arch intern Med. 1994;154:1705–11.

14. Lopez P, Stefan B, Schulman C, Byers P. Prevalence of Sleep Apnea in Morbidly Obese

Patients Who Presented ... Am Surg. 2008;74(9):834–8.

15. Punjabi NM. The Epidemiology of Adult Obstructive Sleep Apnea. 2008;5:136–43.

16. Public Health Agency of Canada. What is the Impact of Sleep Apnea on Canadians? 2009;900:1–3.

17. Singh M, Liao P, Kobah S, Wijeysundera DN, Shapiro C, Chung F. Proportion of surgical patients with undiagnosed obstructive sleep apnoea. 2012;1–8.

18. Watson NF. Health Care Savings: The Economic Value of Diagnostic and Therapeutic Care for Obstructive Sleep Apnea. J Clin Sleep Med [Internet]. 2016;12(8):1075–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27448424%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4957182

19. Adesanya AO, Lee W, Greilich NB, Joshi GP. Perioperative management of obstructive sleep apnea. Chest [Internet]. 2010;138(6):1489–98. Available from: http://dx.doi.org/10.1378/chest.10-1108

20. Liao P, Yegneswaran ÆB, Vairavanathan MÆS, Mbbs Æ, Zilberman P, Chung ÆF. Postoperative complications in patients with obstructive sleep apnea : a retrospective matched cohort study ´ ratoires chez des patients souffrant Complications postope ´ e obstructive du sommeil : une e ´ tude de cohorte d ' apne ´ e re ´ trospective app. 2009;819–28.

21. Osa A, Strategic PR, Team W. The ASPAN Obstructive Sleep Apnea in the Adult Patient Evidence-Based Practice Recommendation. J PeriAnesthesia Nurs [Internet]. 2012;27(5):309–15. Available from: http://dx.doi.org/10.1016/j.jopan.2012.07.003

22. Warner DS, Warner MA, Management PA. VI REVIEW ARTICLES Obstructive Sleep

Apnea of Obese Adults. 2009;(4):908–21.

23. Ramachandran SK, Kheterpal S, Consens F, Shanks A, Doherty TM, Morris M, et al. Derivation and Validation of a Simple Perioperative Sleep Apnea Prediction Score. 2010;110(4):1007–15.

24. Chung SA. A Systemic Review of Obstructive Sleep Apnea and Its Implications for Anesthesiologists. 2008;107(5):1543–63.

25. Chung F, Subramanyam R, Liao P, Sasaki E, Shapiro C, Sun Y. High STOP-Bang score indicates a high probability of obstructive sleep apnoea. 2012;108(March):768–75.

26. Chung F, Memtsoudis SG, Ramachandran SK, Nagappa M, Opperer M, Cozowicz C, et al. Society of Anesthesia and Sleep Medicine Guidelines on Preoperative Screening and Assessment of Adult Patients with Obstructive Sleep Apnea. Anesth Analg. 2016;123(2):452–73.

27. Wickwire EM, Lettieri CJ, Cairns AA. Adherence in Adults A Common-Sense Approach. Chest [Internet]. 2013;144(2):680–93. Available from: http://dx.doi.org/10.1378/chest.12-2681

28. Amra B, Javani M, Soltaninejad F, Penzel T, Fietze I, Schoebel C, et al. Comparison of Berlin Questionnaire , STOP-Bang , and Epworth Sleepiness Scale for Diagnosing Obstructive Sleep Apnea in Persian Patients. 2018;1–5.

29. Suliman LA, Shalabi NM, Elmorsy AS, Moawed MK. Value of STOP-BANG and Berlin questionnaires in the diagnosis and severity prediction of obstructive sleep apnea hypopnea syndrome. 2016;367–71.

30. Kushida CA, Morgenthaler TI, Littner MR, Alessi CA, Bailey D, Coleman Jr. J, et al. Practice parameters for the treatment of snoring and Obstructive Sleep Apnea with oral

appliances: an update for 2005. Sleep [Internet]. 2006;29(2):240–3. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16494092

31. Chung F, Nagappa M, Singh M, Mokhlesi B. CPAP in the Perioperative Setting. Chest [Internet]. 2016;149(2):586–97. Available from: http://dx.doi.org/10.1378/chest.15-1777

32. Moussavi Z, Karimi D. Acoustic system and methodology for identifying the risk of obstructive sleep apnea during wakefulness. [Internet]. US20140142452 A1. Available from: http://www.google.com/patents/US20140142452

33. Montazeri A, Moussavi Z. System and methods of acoustical screening for obstructive sleep apnea during wakefulness. [Internet]. WO2011154791 A3. Available from: http://incenter.medical.philips.com/doclib/enc/fetch/2000/4504/577242/577256/588821/5050628/5313460/7063104/11_MM_543_Subramanyam.pdf%3Fnodeid%3D7063595%26vernum%3D-2

34. Montazeri A, Giannouli E, Moussavi Z. Assessment of obstructive sleep apnea and its severity during wakefulness. Ann Biomed Eng. 2012;40(4):916–24.

35. Lee W, Nagubadi S, Kryger MH, Mokhlesi B. Epidemiology of obstructive sleep apnea: A population-based perspective. Expert Rev Respir Med. 2008;2(3):349–64.

36. Chung F, Yegneswaran B, Liao P, Chung SA, Vairavanathan S, Islam S, et al. STOP Questionnaire. 2017;(5):812–21.

37. Chung F, Yegneswaran B, Liao P, Chung SA, Vairavanathan S, Islam S, et al. STOP Questionnaire. 2008;(5):812–21.

38. Soltanzadeh R, Member IS, Moussavi Z, Member IS. Breathing Sounds Spectral and Higher Order Statistics Changes from Wakefulness to Sleep in apneic and non-apneic People. 2014;4228–31.

39. Schwab R, Pasirstein M, Pierson R, Mackley A, Hachadoorian R, Arens R, et al. Identification of upper airway anatomic risk factors for obstructive sleep apnea with volumetric magnetic resonance imaging. Am J Respir Crit Care Med. 2003;168(5):522.

40. Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. 2006;8:1–8.

41. Gonen, Mithat. Single Continuous Predictor. In: Analyzing receiver operating characteristic curves with SAS. Cary, NC SAS Inst Inc. 2008;32(Mar):15–36.

42. Hosmer DW, Hosmer T, Cessie SLE, Lemeshow S. A Comparison of Goodness-of-fit Tests for the Logistic Regression Model. 1997;16(May 1995):965–80.

43. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Obuchowski N, Pencina MJ, et al. Assessing the performance of prediction models: a framework for some traditional and novel measures. 2013;21(1):128–38.

44. Hosmer DW, Lemeshow S. Assessing the Fit of the Model. In: Applied Logistic Regression. John Wiley and Sons, Inc. 2000. 143-202 p.

45. Habibzadeh F, Habibzadeh P, Yadollahie M. Lessons in biostatistics Responsible writing in science continuous results. 2016;26(3):297–307.

46. Steyerberg EW, Harrell FE, Borsboom GJJM, Eijkemans MJCR, Vergouwe Y, Habbema JDF. Internal validation of predictive models : Efficiency of some procedures for logistic regression analysis. 2001;54:774–81.

47. Whiting P, Rutjes A, Reitsma J, Flas A, Bossuyt P, Jos Keijnen. Academia and Clinic Sources of Variation and Bias in Studies of Diagnostic Accuracy. 2004;189–203.

48. Chung KC, Song JW, Grewal R, Haase SC, Hammert W, Hausman M, et al. A Guide on Organizing a Multicenter Clinical Trial: the WRIST study group. 2011;126(2):515–23.

49. Nemes S, Jonasson JM, Genell A, Steineck G. Bias in odds ratios by logistic regression modelling and sample size. 2009;5:1–5.

50. Sterne JAC, Gavaghan D, Egger M. Publication and related bias in meta-analysis : Power of statistical tests and prevalence in the literature. 2000;53:1119–29.

51. Articles S. Practice Guidelines for the Perioperative Management of Patients with Obstructive Sleep Apnea. 2014;(2).

Chapter 3: Discussion

Previous literature suggested that Awake-OSA, a new technology that relies on breathing sound analysis, was a promising potential screening tool for OSA as an alternative to the currently used STOP-Bang questionnaire (1,2). One of the limitations associated with STOP-Bang was that it has a low specificity (3,4). The main problem with this low specificity is that a number of people screened as having a more severe form of OSA are actually at a low risk. Specifically, this means that a portion of the population, who have a low risk for OSA, are unnecessarily sent for PSG in the general population to rule out OSA or unnecessarily treated with added precautions in the perioperative context as a result of the low specificity with using the STOP-Bang screening tool. Thus, undiagnosed OSA may be resulting in unnecessary added costs to the health care system. Early trial findings that investigated Awake-OSA as a possible screening tool demonstrated a higher specificity to rule out OSA, which is promising (1,2). Our study was the largest to test breathing sound analysis to data and the first to combine the results from breathing sound analysis and STOP-Bang into a single prediction rule. The high accuracy from the BSA results demonstrated in the present study, as compared to STOP-Bang in our study population, lends further support to future testing of Awake-OSA.

Although the STOP-Bang tool is relatively easy to use, it is not always possible to collect all of the required measurements at the bedside (*e.g.* measuring neck circumference). Breathing sound analysis (BSA) shows promise in its ability to accurately rule out patients with OSA. However, although breathing sounds were collected in awake patients, the analysis of these recordings required a complex and long process at another lab. This significant limitation means the

feasibility of BSA for use in future research must be seriously considered when preparing to conduct another study. Ideally, a future study will be able to collaborate with both computer and mechanical engineers. A computer engineer could test the Awake-OSA on a subset of the data early on in the study to ensure there are no issues with the quality of the sound recordings. Furthermore, one of the other limitations of this technology is that it requires no residual background noise. This will pose difficulty for future studies collecting data in a busy clinic setting. Thus, perhaps working with an engineer to enhance the algorithm such that residual noise does not significantly impact the results could be investigated. Furthermore, the cost associated with this added time and resources needed to actually collect breathing sounds could be studied to further evaluate the cost and resources associated with BSA technology in comparison to STOP-Bang.

Ultimately, the aim of our research is to lead to the implementation of a feasible and accurate prediction rule that health care providers (*e.g.* anesthesiologists) would be able to in a clinical setting at the bedside. The implementation of a screening tool that is feasible and easy to use could help to improve health outcomes of Canadian's living with OSA across different contexts, particularly in rural areas where access to health resources like polysomnography is limited. A screening tool that is easy to use at the bedside could also mean that health care clinics, including those in remote areas, could better screen people for OSA and minimize unnecessary PSG referrals.

The prevalence of undiagnosed OSA in the perioperative context remains unknown (5). There may be additional hidden costs associated with undiagnosed OSA in surgical patient populations

could be explored. Although STOP-Bang is the most widely used screening tool for OSA perioperatively, a number of people without high risk are treated as high risk during perioperative period, which is likely adding health care costs. Thus, another possible study could investigate whether the use of a tool that incorporates BSA (using STOP-Bang as a control) saves any costs to the health care system through an economic analysis. Furthermore, at the same time, investigators could also evaluate whether a tool that incorporates BSA is effective at improving relevant health outcomes (*e.g.* wait times for PSG after referral) to allow for a cost-benefit analysis. Perhaps a prospective quasi-randomized clinical trial that compares two sleep study sites with a similar prevalence of OSA could assess possible differences in hospital costs and relevant health outcomes for sites using a tool that incorporates BSA (test site) compared to usual care using STOP-Bang (control site). The results from this kind of study would help inform future practice guideline implementation as it could show a potential cost benefit to the Canadian health care system by using a tool that incorporates BSA.

Future research should also investigate the potential of BSA technology to assist in screening OSA in non-surgical populations. Of note, Canada has been shown to have an increased prevalence of OSA in recent years that has been associated with rising obesity rates (6,7). Therefore, the need for a tool that is accurately able to rule out moderate-to-severe OSA has significant implications to help streamline referrals for PSG among people living with undiagnosed OSA.

References

1. Montazeri A, Giannouli E, Moussavi Z. Assessment of obstructive sleep apnea and its severity during wakefulness. Annals of Biomedical Engineering. 2012;40(4):916–24. MacGregor CA,
2. Moussavi Z. A novel expert classifier approach to pre-screening obstructive sleep apnea during wakefulness. Conf Proc IEEE Eng Med Biol Soc. IEEE; 2014;2014:4236–9.
3. Chung F, Subnyam R, Liao P, Sasaki E, Shapiro C, Sun Y. High STOP-Bang score indicates a high probability of obstructive sleep apnoea. *Br J Anaesth*. 2012; 108: 5. 768–75.
4. Chung SA. A Systemic Review of Obstructive Sleep Apnea and Its Implications for Anesthesiologists. 2008;107(5):1543–63
5. Singh M, Liao P, Kobah S, Wijeysundera DN, Shapiro C, Chung F. Proportion of surgical patients with undiagnosed obstructive sleep apnoea. 2012;1–8.
6. Lopez P, Stefan B, Schulman C, Byers P. Prevalence of Sleep Apnea in Morbidly Obese Patients Who Presented ... Am Surg. 2008;74(9):834–8.
7. Warner DS, Warner MA, Management PA. VI REVIEW ARTICLES Obstructive Sleep Apnea of Obese Adults. 2009;(4):908–21.

Appendices

Appendix 1. Ethics approval

Ottawa Hospital Research Institute
Institut de recherche de l'Hôpital d'Ottawa

uOttawa

UNIVERSITY OF OTTAWA HEART INSTITUTE
INSTITUT DE CARDIOLOGIE DE L'UNIVERSITÉ D'OTTAWA

Ottawa Health Science Network Research Ethics Board/ Conseil d'éthique de la recherche du Réseau de science de la santé d'Ottawa

Civic Box 411 725 Parkdale Avenue, Ottawa, Ontario K1Y 4E9 613-798-5555 ext. 14902 Fax : 613-761-4311
http://www.ohri.ca/ohsn-reb

September 23, 2016

Re: Protocol #	**20160494-01H**	**Safer Sleep: A Diagnostic Accuracy Study Evaluating Breath Sound Recordings and STOP-Bang In Screening For Obstructive Sleep Apnea While Awake.**
Protocol approval valid until -		**November 23, 2016**

I am pleased to inform you that this protocol underwent delegated review by the Ottawa Health Science Network Research Ethics Board (OHSN-REB) and is approved for two months to begin recruitment of English speaking participants. No changes, amendments or addenda may be made to the protocol or the consent form without the OHSN-REB's review and approval.

Approval includes the following:
- Protocol, version dated 22-08-2016
- English Participant Informed Consent Form, version dated 22-08-2016
- English STOP-Bang Questionnaire, uploaded February 19, 2016
- Case Report Form, version 3.0 (no previously approved versions 1.0 and 2.0 on file)

We acknowledge the Patient Data Key, version 1, dated 2016-05-13.

Upon receipt of the French Consent Form and questionnaire, ethical approval will be extended to September 22, 2017 and the recruitment of French speaking participants may begin.

When submitting the French documentation to the OHSN-REB, please be sure to indicate who did the translation by providing evidence in the form of a certificate of translation or email. Refer to SOP '#701Addendum' for the translation options when submitting your French documents to the REB for approval. The SOP can be found in the Translation section of the OHSN-REB website (www.ohri.ca/OHSN-REB). The OHSN-REB does offer a fee for service translator, Mr. Eric Lépine (elepine@ohri.ca).

The Ottawa Health Science Network Research Ethics Board (OHSN-REB) was created by the merger of both the Ottawa Hospital Research Ethics Board (OHREB) and the Human Research Ethics Board (HREB) for meetings held at the University of Ottawa Heart Institute.

OHSN-REB complies with the membership requirements and operates in compliance with the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans; the International Conference on Harmonization - Good Clinical Practice: Consolidated Guideline and the provisions of the Personal Health Information Protection Act 2004.

Appendix 2. Participant Informed Consent Form

Title of Study: Safer Sleep: A Diagnostic Accuracy Study Evaluating Breath Sound Recordings and STOP-Bang in Screening for Obstructive Sleep Apnea while Awake

Local Site Principal Investigator (PI): Dr. Sylvain Boet

Funding Agency: The Ottawa Hospital Academic Medical Organization (TOHAMO), 2015-2016 Innovation Fund

Involvement in this study is voluntary. Please read this Consent Form before you decide if you would like to participate. Ask the study team as many questions as you like.

Why am I being given this form?

You are being asked to take part in this study because you have an appointment at The Ottawa Hospital Sleep Centre. This study will help create a new test for obstructive sleep apnea (OSA).

Why is this study being done?

This study will compare a tool that records breath sounds ("Awake-OSA") to a survey ("STOP-Bang"). We want to know if this tool can detect OSA. OSA affects nearly 25% of people living in North America. As obesity rates increase, this number will grow over time. OSA increases risk of stroke, cardiac events and death as well as diabetes and hypertension.

The best way to detect OSA is with a sleep study. However, access to sleep labs is limited. This causes longer wait times and limited access. Surveys such as STOP-Bang can also help detect OSA, but these surveys are not perfect.

We want to use a new tool to help detect OSA. Awake-OSA is a simple, rapid and non-invasive test. Awake-OSA records breath sounds while a patient is awake. The sounds are analyzed, and changes in these sounds may indicate OSA.

For this study, we will compare breathing sound analysis to the STOP-Bang survey. We will look at each tool's ability to detect OSA. Our goal is to build a decision-making tool that could be used in different settings, such as in the hospital and in family doctors' offices.

About 400 people from The Ottawa Hospital, Civic Campus Sleep Centre will be enrolled in the study.

How is the study designed?

Our study will compare two tools- (1) breath sound recording and (2) a screening survey. We will compare the results of these tools to your scheduled sleep test results.

Breath sounds and de-identified health information will be analyzed by Dr. Moussavi in her biomedical engineering laboratory at the University of Manitoba. All other data will be analyzed here at The Ottawa Hospital.

What is expected of me?

If you consent to take part in this study, you will first be prepared for your scheduled sleep test by a technician from the Sleep Centre. The technician will apply the monitors for your sleep test as they normally would even if this study wasn't being done. Once you are fully prepared for your sleep test, the study research assistant will join you to begin the study.

You will first complete one STOP-Bang survey. STOP-Bang looks at factors that can indicate OSA. The questions ask about snoring, tiredness, and breathing patterns. There will also be a few other questions about your health. The survey will take about five minutes to complete. You may skip any questions that you do not wish to answer.

The research assistant will then record your breath sounds using a small microphone placed against your neck, just above your collarbone. You will be asked to take three to five large breaths while lying flat, first with a pillow under your head and then without. The research assistant will guide you through how to take these breaths. You are free to do this at your own pace and to stop at any time. Breath sound recording will take about ten minutes.

Once this step is complete, the study is over. You will continue on with your sleep test. You will not need to have any further involvement with the study team.

Will my samples or research data be used in future research?

Your results may be used in future studies to help diagnose sleep apnea. These results will all remain de-identified. The breath results and some de-identified health information will be stored on a secure server at the University of Manitoba for analysis by the biomedical engineering laboratory. All results will also be stored on a secure platform at The Ottawa Hospital.

How long will I be involved in the study?

The entire study will last about twenty four months. Your involvement in the study will only last for about a half hour before you have your scheduled sleep test. You will not need to visit the Sleep Centre again with regard to this study.

Your involvement in the study may be stopped if:

- Any time you wish to stop.
- The study team feels it is in your best interest.
- You have been unable to follow the study team's instructions.

What are the potential risks I may experience?

The study doctors know of no harm they may come to you by contributing to this study.

Questionnaires:

You might not like some of the questions you are asked. You do not have to answer any questions that make you uncomfortable.

Breath Sounds:

You might feel mildly short of breath during or after the breath sound recording. You may stop or take a break at any time.

Can I expect to benefit from participating in this research study?

You will not directly benefit from participating in this study. Your contribution may help the study doctors build tools that could improve early detection of OSA. This may help future patients.

Do I have to participate? What alternatives do I have? If I agree now, can I change my mind and withdraw later?

Your involvement in this study is voluntary. You may choose not to participate.

You will have your sleep test whether or not you partake in our study. Involvement in our study will not affect the outcome of your sleep test in any manner. Any treatment decisions based on your sleep test results will not be affected. Your sleep doctor is responsible for these decisions.

You may decide not to be in this study, or to be in the study now, and then change your mind later without affecting the medical care or other services to which you are entitled or are presently receiving at this institution.

You may withdraw from the study at any time without any impact on your current or future care at this institution.

If you withdraw your consent, the study team will no longer collect your personal health data for research purposes. If you decide to withdraw, you may also request that any data associated with you up to that point in the study be deleted.

Will I be paid for my participation or will there be any additional costs to me?

You will not be paid for contribution to this study. There will be no additional costs to you associated with the study.

How is my personal data being protected?

- All data collected during your involvement in this study will be identified with a unique study number, and will not contain data that identifies you, such as your name, address, etc.
- The link between your unique study number and your name and contact data will be stored securely and separately from your study records. This link will not leave this site.
- Any documents or recordings leaving The Ottawa Hospital will contain only your unique study number. This includes articles or presentations resulting from this study.

- Data that identifies you will be released only if it is required by law.
- For audit purposes only, your original study records may be reviewed under the supervision of Dr. Sylvain Boet's staff by:
    - The Ottawa Health Science Network Research Ethics Board (OHSN-REB),
    - The Ottawa Hospital Research Institute.
- Research records will be kept for 10 years. After this time they will be destroyed.

Do the study doctors have any conflicts of interest?

Dr. Moussavi has a patent for the breath sound recording technology. Dr. Zahra Moussavi is the Director of the Biomedical Engineering Program at the University of Manitoba. She is an expert on breath sound recording for sleep apnea. She is a partner in this research project.

Will I be informed about any new information that might affect my decision to continue participating?

You will be told in a timely fashion of any new findings during the study that could affect your willingness to continue in the study. You may be asked to sign a new consent form.

Who do I contact if I have any further questions?

If you have any questions about this study, please contact Dr. Sylvain Boet

The Ottawa Health Science Network Research Ethics Board (OHSN-REB) has reviewed the plans for this research study. The Board considers the ethical aspects of all research studies involving human participants at the Ottawa Hospital. If you have any questions about your rights as a study participant, you may contact the Chairperson at 613-798-5555, extension 16719.

Appendix 3. Participant questionnaire

Data collection questionnaire

INFORMED CONSENT/ELIGIBILITY CRITERIA

Informed Consent Signed: ☐ Yes ☐ No Date: _____Time: _____
Eligibility Criteria Determined? ☐ Yes ☐ No

DEMOGRAPHICS

Birth month and year (mm/yr): _____ / _____ Sex (M/F):_____
Age: _____
Race (Check all that applies):
☐ Caucasian ☐ Black or African American ☐ Asian ☐ Native Indian
☐ Arabian ☐ Other: _____

ANTHROPOMETRIC DATA

Weight (kg): _____ Height (cm): _____
BMI: _____
Mallampati class (1-4): _____
Neck circumference (cm): _____

QUESTIONNAIRE (Has the subject ever been told/experienced/treated for?)

Snoring (loud)	☐ Yes	☐ No
Daytime tiredness	☐ Yes	☐ No
Observed apneas	☐ Yes	☐ No
High blood pressure	☐ Yes	☐ No

GENERAL MEDICAL HISTORY

Pre-existing conditions/Diagnosis	Y/N
Diabetes	☐ Yes ☐ No
Coronary artery disease	☐ Yes ☐ No

History of heart attack	☐ Yes ☐ No
Stroke or Mini-Stroke	☐ Yes ☐ No
Kidney disease on dialysis	☐ Yes ☐ No
Hypertension	☐ Yes ☐ No
COPD	☐ Yes ☐ No
Asthma	☐ Yes ☐ No
GERD	☐ Yes ☐ No

TOBACCO

Has the Subject ever used cigarettes? ☐ Unknown ☐ Never ☐ Current ☐ Former
If current or former user, specify
Number of years: _____ Number of packs/day: _____

Appendix 4. STOP-Bang Questionnaire and Scoring Criteria

QUESTIONNAIRE

Snoring?

Do you Snore Loudly (loud enough to be heard through closed doors or your bed-partner elbows you for snoring at night)?

Tired?

Do you often feel Tired, Fatigued, or Sleepy during the daytime (such as falling asleep when driving)?

Observed?

Has anyone Observed you Stop Breathing or Choking/Gasping during your sleep?

Pressure?

Do you have, or are being treated for, High Blood Pressure?

Body Mass Index more than 35 kg/m2?

Age older than 50 year old?

Neck size large? (Measured around Adams apple)

For male, is your shirt collar 17 inches/43 cm or larger? For female, is your shirt collar 16 inches/41 cm or larger?

Gender = Male?

SCORING CRITERIA

Low risk of OSA: Yes to 0-2 questions

Moderate risk: Yes to 3-4 questions

High risk:

Yes to 5-8 questions

Yes to 2 or more of 4 STOP questions + male gender

Yes to 2 or more of 4 STOP questions + BMI > 35 kg/m^2 Yes to 2 or more of 4 STOP questions + neck circumference >40cm or 16 inches.

Appendix 5. Standardized tracheal breathing sound script for research assistants

Standardized Procedure for Tracheal Breathing Sound Analysis

1. Settings on Audacity:
    - First, put the frequency setting at >10,000 Hz
    - Second, put the recording setting to 'Line (Audacity)'

2. Place a very sensitive microphone on neck, below Adam's apple, using double sided tape:
    - Other considerations: must prevent objects rubbing against the wire. Wire must be stationary (use tape to secure the wire).
    - Listen to the sound quality before starting to ensure no crackling. Adjust wire position as needed.
    - Shut the door to participant hospital room to block any residual noise.

3. Inform the participant you are taking a total of 4 different recordings. Each recording will consist of 5 regular and 5 deep breaths. An Inhalation and exhalation is 1 count.

4. First recording is through the nose with pillow. Guide patient through the recording using the following procedure:
    - "Breath in....breath out" -> watch patient chest rise and fall.
    - Then hold up hand for patient to see and count on fingers 5 normal/ regular breaths.
    - "Deep breath in..... and exhale" -> motion increased depth of breathing using arm/hand up and down motions.
    - Then hold up hand for patient to see and count on fingers 5 deep breaths.

- On 5th deep breath, bring BOTH hands up in unison to show they must hold breath after inhalation.
- After 2-3 seconds say "Release" to signal exhalation to the patient.
- Stop the recording.
- Label the recording according to naming scheme.

5. Second recording is the exact same routine, but now the participant breathes through the mouth:
    - Ask participant to plug nose.
    - Repeat same procedure as in step 4.

6. Third and fourth recording without the pillow using the same steps from 4-5.

7. If you make a mistake at any point, or notice some sort of error (i.e. microphone not flat, the settings on audacity are incorrect) just repeat the recording.

8. Remove the microphone off CAREFULLY (hold the skin tense while peeling off double sided tape).

Appendix 6. Standards for Reporting of Diagnostic Accuracy Studies (STARD) research checklist

Section & Topic	No	Item	Page No
TITLE OR ABSTRACT			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	2
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	2-3
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	20-23
	4	Study objectives and hypotheses	24
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	24
Participants	**6**	Eligibility criteria	25

Section & Topic	No	Item	Page No
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	25
	8	Where and when potentially eligible participants were identified (setting, location and dates)	24-25
	9	Whether participants formed a consecutive, random or convenience series	24
Test methods	**10a**	Index test, in sufficient detail to allow replication	29
	10b	Reference standard, in sufficient detail to allow replication	29
	11	Rationale for choosing the reference standard (if alternatives exist)	-
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	29
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	29
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	29
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	29

Section & Topic	No	Item	Page No
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	34
	15	How indeterminate index test or reference standard results were handled	30
	16	How missing data on the index test and reference standard were handled	30
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	30
	18	Intended sample size and how it was determined	34-35
RESULTS			
Participants	**19**	Flow of participants, using a diagram	36
	20	Baseline demographic and clinical characteristics of participants	36-41
	21a	Distribution of severity of disease in those with the target condition	36-39
	21b	Distribution of alternative diagnoses in those without the target condition	-
	22	Time interval and any clinical interventions between index test and reference standard	-

Section & Topic	No	Item	Page No
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	43
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	45
	25	Any adverse events from performing the index test or the reference standard	-
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	47-49
	27	Implications for practice, including the intended use and clinical role of the index test	46, 50-52
OTHER INFORMATION			
	28	Registration number and name of registry	3
	29	Where the full study protocol can be accessed	Not available
	30	Sources of funding and other support; role of funders	52-53

Appendix 7. Classification tables for BSA, STOP-Bang, and the Safe-OSA rule (N=280)

		PSG diagnosis of moderate-to-severe OSA from PSG	
		Present	Absent
BSA predicted outcomes of moderate-to-severe OSA	Yes	85	40
	No	29	126

Abbreviations: OSA, obstructive sleep apnea; PSG, polysomnography.

		Diagnosis of moderate-to-severe OSA from PSG	
		Present	Absent
STOP-Bang score >3	Yes	78	62
	No	36	105

Abbreviations: OSA, obstructive sleep apnea; PSG, polysomnography.

		PSG diagnosis of moderate-to-severe OSA from PSG	
		Present	Absent
Safe-OSA rule predicted outcomes of moderate-to-severe OSA	Yes	85	40
	No	29	126

Abbreviations: OSA, obstructive sleep apnea; PSG, polysomnography.