https://doi.org/10.5664/jcsm.9220

JCSM | Journal of Clinical Sleep Medicine

SCIENTIFIC INVESTIGATIONS

Interrater reliability between in-person and telemedicine evaluations in obstructive sleep apnea

Michael E. Yurcheshen, MD, FAASM[1]; Wilfred Pigeon, PhD[1]; Carolina Z. Marcus, MD[1]; Jonathan A. Marcus, MD[1]; Susan Messing, MS[1]; Kevin Nguyen, MD[2]; Jennifer Marsella, MD[1]

[1]*University of Rochester School of Medicine and Dentistry, Rochester, New York;* [2]*Saddleback Medical Center, Laguna Hills, California*

Study Objectives: We examined how telemedicine evaluation compares to face-to-face evaluation in identifying risk for sleep-disordered breathing.
Methods: This was a randomized interrater reliability study of 90 participants referred to a university sleep center. Participants were evaluated by a clinician investigator seeing the patient in-person, then randomized to a second clinician investigator who performed a patient evaluation online via audio-video conferencing. The primary comparator was pretest probability for obstructive sleep apnea,
Results: The primary outcome comparing pretest probability for obstructive sleep apnea showed a weighted kappa value of 0.414 (standard error 0.090, $P = .002$), suggesting moderate agreement between the 2 raters. Kappa values of our secondary outcomes varied widely, but the kappa values were lower for physical exam findings compared to historical elements.
Conclusions: Evaluation for pretest probability for obstructive sleep apnea via telemedicine has a moderate interrater correlation with in-person assessment. A low degree of interrater reliability for physical exam elements suggests telemedicine assessment for obstructive sleep apnea could be hampered by a suboptimal physical exam. Employing standardized scales for obstructive sleep apnea when performing telemedicine evaluations may help with risk-stratification and ultimately lead to more tailored clinical management.
Keywords: telemedicine, interrater reliability, obstructive sleep apnea, sleep-disordered breathing, clinical study, telehealth
Citation: Yurcheshen ME, Pigeon W, Marcus CZ, et al. Interrater reliability between in-person and telemedicine evaluations in obstructive sleep apnea. *J Clin Sleep Med*. 2021;17(7):1435–1440.

BRIEF SUMMARY
Current Knowledge/Study Rationale: Telemedicine is a promising technology that is now widely used in the practice of sleep medicine. The accuracy of telemedicine compared to in-person assessment in evaluating patients with obstructive sleep apnea is still unknown. The current study was an interrater reliability study comparing a telemedicine to an in-person evaluator in assessing pretest probability for obstructive sleep apnea in a community population.
Study Impact: The telemedicine and in-person investigators had moderate agreement in evaluating pretest probability for mild, moderate, or severe obstructive sleep apnea. The current study underscores the need to consider standardized processes that optimize telemedicine and support an online clinician's ability to accurately assess for sleep-disordered breathing.

INTRODUCTION

Telemedicine is changing health care delivery for all of clinical medicine. There is increasing demand for sleep medicine specialists, and online consultants have the potential to deliver care despite distance, transportation concerns, or pandemic conditions. Patients, sleep medicine providers, insurance carriers, and industry all have a stake in the development of telemedicine services.

Obstructive sleep apnea (OSA) is a common and expensive medical condition, estimated to afflict 2% to 20% of the United States population at an estimated annual cost of nearly 150 billion dollars to the US economy.[1–3] Suspicion for OSA is based largely on clinical history, with some additional detail gained from physical examination. Survey tools can also play a role in assessing for OSA.[4–7] Because the condition is common, and because it can be evaluated by history and examination suited for remote video-audio communication, OSA is a good candidate for telemedicine evaluation. Based on this promise, industry groups are looking to optimize this technology.[8]

Telemedicine consultation has been studied in various disease processes, including Parkinson disease, emergency ocular disorders, and acute pediatric illnesses.[9–13] This type of assessment has led to substantial satisfaction with medical care, as well as reduced travel time and distance.[9] Regarding telemedicine management of OSA, the literature lacks large, community-based outcome studies, but there does seem to be promise in the technology based on a number of studies. In 1 study, over 60% of OSA patients felt comfortable engaging in virtual consultations.[14] Likewise, a community population showed equal satisfaction with in-person and telemedicine evaluations for their sleep condition.[15] One study showed a small decrease in continuous positive airway pressure compliance among participants managed by teleconsultation.[16] In contrast, continuous positive airway pressure participants

recruited from a veteran population had similar functional outcomes with telemedicine compared to in-person management.[17]

These studies are encouraging but have limitations. Some lacked a control group, and others had limited numbers of participants. Even the most compelling of these studies has limited generalizability as it was in a veteran population only.[17] Although the sleep field is rapidly moving toward adoption of telemedicine, there are still gaps in our understanding of this care model. For instance, there are no community studies examining the accuracy of remote assessments compared to the gold standard of in-person evaluation.[18] Many treatment decisions, including the decision to order at-home or in-lab sleep testing, are based on a patient's pretest probability for sleep apnea.[19] Telemedicine will be most useful if patient assessments are sufficiently accurate to drive this type of clinical decision-making.

The present randomized clinical trial aimed to compare telemedicine to in-person assessment for sleep-disordered breathing. We hypothesized that there would be high interrater reliability between telemedicine and in-person assessors in determining pretest probability for OSA. Secondary aims included interrater reliability for historical and physical exam elements suggestive of this condition, as well as a comparison between raters in interpreting home sleep apnea testing.

METHODS

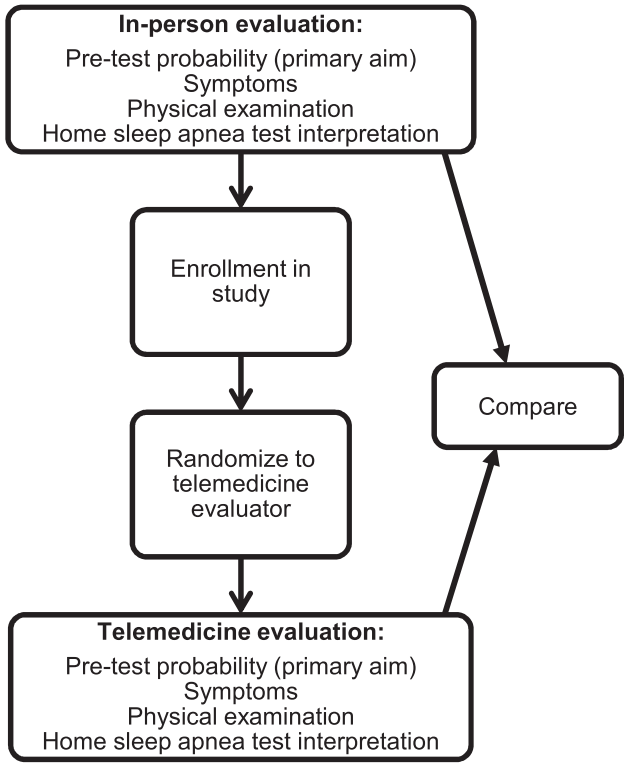
Study design and randomization

We conducted a randomized, blinded, interrater reliability study comparing the impressions of a clinician seeing a patient in-person to a clinician seeing the same patient via telemedicine. Subjects were recruited between March 2017 and January 2019. The study design is outlined in **Figure 1**. The study design is based on comparing assessments between in-person and telemedicine physicians, rather than between a single telemedicine rater at different time points, or between 2 different telemedicine raters. This design was chosen as a compromise, given the rapid pace of clinical assessment/testing and the impracticality of purposely delaying a repeat assessment that would minimize evaluator memory bias.

Three American Board of Medical Specialties sleep-board eligible/certified clinicians (M.E.Y., C.Z.M., J.M.) underwent group training to familiarize themselves with the study protocol and outcome assessments. During this training, the raters reviewed 20 theoretical case histories to reach consensus about pretest probability for OSA. Training included a description of the physical exam but did not include simulated video footage of the oral cavity. We elected to use 3 different raters in our study to expedite recruitment (each rater was responsible for recruiting roughly one-third of the cohort) and to minimize bias that may have stemmed from using only 2 raters.

Consenting participants were asked to complete a demographic profile and an Epworth Sleepiness Scale during their in-person clinical encounter. The in-person investigator reviewed the patient's electronic medical record and conducted a history and physical examination. The investigator completed a clinical impression battery that included responses relating to the primary and secondary endpoints. These participants, upon consent, were randomized to 1 of 2 other raters by utilizing a randomized block design of size 4, and an online clinical encounter was conducted through an audiovisual conferencing application (Zoom; San Jose, CA). The telemedicine encounter was scheduled within 5 business days. This tele-evaluation included record review, history, and a brief, noninvasive examination of the oral cavity using the patient's web-enabled camera and an incandescent light source. The telemedicine investigator was blinded to the in-person assessment, and blinding was ensured by a third-party audit of a sample of 5 study participants. The participant completed a repeat Epworth Sleepiness Scale around the time of the telemedicine encounter and completed a survey questionnaire regarding their experience with telemedicine.

Figure 1—Study design.



For participants who completed home sleep apnea testing (HSAT) a type 3 home sleep testing device [Braebon Medibyte (Kanata, ON, Canada) or Respironics Alice NightOne (Murrysville, PA)] was used. The devices recorded the following signals: snoring, body position, heart rate, oxygen saturation, airflow (nasal pressure), and thoracoabdominal movement. The studies were scored by an experienced New York State licensed sleep technologist who was unaware of participant randomization. Respiratory scoring was performed in accordance with published criteria.[20] The study was interpreted by both the in-person and telemedicine clinicians independently, and an impression of mild, moderate, or severe sleep apnea (or uninterpretable study) was recorded.

The protocol for this study was approved by the University of Rochester Institutional Review Board and complied with the

ethical principles of the Declaration of Helsinki and were in accordance with the International Conference on Harmonization Good Clinical Practice guideline. All participants provided written informed consent before study enrollment.

Participants

Eligible participants were recruited from general referrals to the University of Rochester Sleep Center. Men and women 30–70 years old referred for any reason were asked to participate. Potential participants needed sufficient computer literacy, access to high-speed internet (minimum 384 kbps), a computing device with appropriate video camera (minimum 640 × 480 resolution at 30 frames/s), and microphone. Patients with dementia, severe psychiatric or developmental illness, complete hearing or visual loss, or not fluent in English were excluded. Participants conducted the telemedicine evaluation in their domicile or other private area. Participants were offered a $25 gift card if they completed both phases of the study. All participants were also provided an incandescent pen light, by which the evaluator examined the oral cavity. This incandescent light source helped to minimize "white out" contrast issues in the oral cavity seen when brighter LED lighting is used.

Measures

The study's primary outcome was to assess the pretest probability for sleep apnea (high, moderate, or low). Secondary outcome measures included interrater reliability for both historical and physical exam findings, including snoring volume (none, mild, moderate, severe, not sure), witnessed apneas (yes, no, not sure), degree of daytime sleepiness (none, mild, moderate, severe), modified Mallampati class (1–4) of airway crowding, overjet distance of maxillary-mandibular relationship (< 0 mm, 0–3 mm, > 3 mm), tonsil size (0–4+), and severity of sleep apnea based on HSAT (inconclusive, mild, moderate, severe).

Statistical methods

Sample size

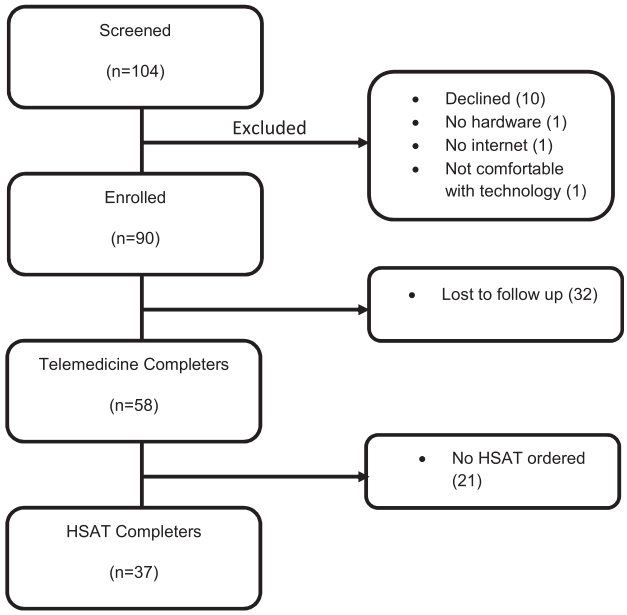
A sample size of 80 participants was selected based on 80% power at a significance level of .05 to detect a kappa of 0.60 in a test of kappa > 0.4, assuming an overall clinic population risk profile of 30% low, 30% moderate, and 40% high pretest probability for OSA. We recruited 90 participants for the study, based on an assumed 10% dropout.

Statistical analysis

Kappa and weighted kappa were evaluated for in-person vs telemedicine evaluations. Weighted kappa utilized the Cicchetti-Allison weighting scheme. For this study, there were no missing data from our raters.

When score values are relatively equally exercised along the ordinal scale, the percent agreement and kappa demonstrate concordance; when some of the categories are very sparsely populated, kappa values can drop precipitously. To evaluate the amount of agreement, we followed the Landis and Koch suggestions, where: < 0 is poor agreement, 0.0–0.20 is slight agreement, 0.21–0.40 is fair agreement, 0.41–0.60 is moderate agreement, 0.61– 0.80 is substantial agreement, and 0.81–1.00 is almost perfect agreement.[21]

Figure 2—CONSORT (Consolidated Standards of Reporting Trials) diagram. HSAT = home sleep apnea testing.



RESULTS

Ninety (90) participants enrolled in the study, and 58 participants completed the entire protocol. The CONSORT (Consolidated Standards of Reporting Trials) diagram in **Figure 2** depicts participant progress throughout the study. Based on the post-hoc audit of the electronic health record, none of the telemedicine investigators accessed the evaluation of the primary investigator during the embargoed timeframe between the 2 evaluations. One participant discovered the HSAT results and shared them with the telemedicine investigator. Since the telemedicine investigator did not access the in-person investigator's assessments, we decided to include this participant in the analysis. The only reason for noncompletion was loss-to-follow-up with the participant not showing up for the telemedicine appointment (32 participants). There were 37 participants who underwent HSAT evaluation. No evaluation was stopped for technical concerns.

Participant characteristics are included in **Table 1**. The 58 completers had a mean age of 49.9 years. Of enrolled participants 11% were African-American, 4% were Hispanic, and 1% were Asian American. Of completing participants, 7% were African-American, 4% were Hispanic, and 0% were Asian American. The baseline Epworth Sleepiness Scale score was 9.4.

Of those who completed both appointments, 13 participants completed their telemedicine appointment later than the 5-business-day target window (ranging from 6–24 days). Any

delay in the telemedicine appointment was due to patient factors/preferences and rescheduling.

Statistics about patient and evaluator satisfaction with the process, timing, and duration of the telemedicine appointment are in preparation for another manuscript and are not reported here.

Interrater reliability

The results for our primary and secondary outcomes are included in **Table 2**, including percentage agreement, kappa, and weighted kappa where appropriate.

Based on a sample size of all protocol completers, the quadratic weighted kappa value was 0.414 [standard error (SE) 0.090, $P = .002$] in determining pretest probability of OSA based on clinical evaluation. This value is consistent with moderate agreement between evaluators. In a post-hoc analysis, when the sample categories were compressed to high (which includes high- and moderate-risk categories) and low pretest probability, the linear weighted kappa value was calculated at 0.28 (SE 0.17, 95% confidence interval 0–0.62). In this case, the value suggests fair agreement between evaluators.

Of our secondary clinical endpoints, the historical element of witnessed apneas had the highest weighted kappa (0.702, SE 0.079, $P < .0001$), and assessment of the physical exam finding of overjet had the lowest weighted kappa but did not reach statistical significance (–0.044, SE 0.014, $P = .496$).

Regarding home sleep testing, based on a sample size of 37 participants (all HSAT completers with dual impressions), we calculated a weighted kappa value of 0.899 (SE 0.056, $P < .0001$) in determining the severity of OSA based on home sleep testing. This value is consistent with almost perfect agreement between evaluators.

Table 1—Participant characteristics.

	All (n = 90)	Completers (n = 58, 65%)
Age (mean ± SD), y	48.84 ± 8.04	49.93 ± 8.10
Female, %	53.3	55.2
Race/ethnicity, %		
Asian	1.1	0.0
Black/African American	11.1	6.9
Hispanic	4.4	3.4
White/Caucasian	83.3	89.7
Income, %		
< $50,000	20.0	19.0
$50,000–100,000	33.3	37.9
> $100,000	36.7	34.5
Preferred not to answer	10.0	8.6
Average business days to enrollment	n/a	4.98 ± 2.98
Average days to enrollment	n/a	6.48 ± 4.68

n/a = not available, SD = standard deviation.

DISCUSSION

Although telemedicine is a promising tool for the delivery of sleep care, its accuracy compared to in-person evaluation has been uncertain. This present study is the first to evaluate the accuracy of telemedicine in determining pretest probability for obstructive sleep apnea in a community population.

Our results show moderate agreement between an in-person and a telemedicine evaluator in determining pretest probability for obstructive sleep apnea. A much higher level of agreement was noted for our secondary endpoint of witnessed apneas but was low for all of the elements of physical examination. The agreement between raters in ultimately determining the degree of sleep apnea when looking at HSAT results was almost perfect, according to published criteria.[21]

Tele-sleep-medicine is becoming increasingly visible, and adoption is happening quickly in the middle of a coronavirus pandemic. In a mid-decade review of patient attitudes toward sleep telemedicine, 63% of respondents surveyed stated they would be comfortable or willing to try telemedicine visits for their sleep appointments.[14] When considering the shortage of sleep medicine providers, and the time, expense, and safety of traveling to and conducting in-person appointments, telemedicine evaluation is helping to improve access. It is important that the sleep field continue to optimize both the technology and the clinical standards for this tool.

Table 2—Interrater reliability of in-person vs telemedicine physician evaluations for obstructive sleep apnea.

Evaluation	% Agreement	Kappa (SE)	Weighted Kappa (SE)	*P* Value
Pretest probability of OSA (low, moderate, high)	53.44	0.286 (.100)	0.414 (.090)	.002
Level of daytime sleepiness	44.83	0.225 (.092)	0.391 (.083)	.004
Snoring at maximum level	63.16	0.354 (.107)	0.353 (.109)	.0002
Apneas witnessed by third party	77.43	0.628 (.086)	0.702 (.079)	< .0001
Apneas witnessed by third party (Y/N)	91.30	0.825 (.084)	—	< .0001
Modified Mallampati score	48.28	0.179 (.093)	0.200 (.091)	.067
Tonsils (Y/N)	64.29	–0.134 (.105)	—	.326
Overjet	82.46	–0.069 (.022)	–0.044 (.014)	.496
After reviewing home sleep apnea test (mild, moderate, severe)	91.89	0.872 (.070)	0.899 (.056)	< .0001

OSA = obstructive sleep apnea, SE = standard error, Y/N, yes/no.

Pretest probability for sleep apnea was selected as the primary endpoint for this study because it is a major determinant in developing evaluation and management plans.[19] Uncomplicated patients with moderate or high pretest probability for significant OSA may be well suited for home testing, whereas patients with low pretest probability may not be referred for testing at all. Still other patients may have a significant pretest probability for mild sleep apnea, and in-lab testing may be more appropriate. As such, pretest probability drives clinical decision-making.

Formulation of pretest probability is driven by both (1) history and (2) physical exam. Analysis of our primary aim suggests reasonable but imperfect interrater reliability in deciding pretest probability for OSA, on par with interrater reliability for some other medical conditions, but less than others.[22–23] The reliability of our primary aim stands in contrast to substantial agreement in one of our historical elements (witnessed apneas), and the poor or unclear agreement in the physical exam findings (modified Mallampati class, overjet distance, tonsil size). These results suggest that uncertainty introduced by the physical exam may have tempered the clinical picture generated by history.

In our study, once participants had home sleep testing, there was excellent agreement in determining the severity of sleep apnea based on HSAT. This result supports a similar finding in a veteran population and suggests that a telemedicine provider is unlikely to miss sleep apnea on HSAT once appropriate patients are identified.[17]

The results suggest challenges and a way forward in developing evaluation/management plans via telemedicine. To minimize uncertainty introduced by the telemedicine assessment, a standardized, protocol-driven approach with predictive survey tools (ie, STOP-Bang, Berlin Questionnaire) could help stratify patients by risk.[4–7] Although these tools were not employed in this study, they have value in predicting sleep apnea, can be performed remotely, and might increase the accuracy of telemedicine evaluations. More recently, structured interviews to assess for a wide variety of sleep disorders have been developed.[24–25] These too could play a role in telemedicine assessment, although they are more time intensive than the aforementioned, sleep-apnea focused questionnaires.

Also, consideration should be given to how to optimize the physical examination portion of the remote assessment. There is renewed interest in the role of the physical exam in sleep medicine and how best to unify descriptions of the airway.[26] For telemedicine, there are additional considerations that might include better lighting, higher resolution cameras, or the use of in-person patient presenters. New technologies will certainly play a role in improving the remote physical exam.[27] Further, telemedicine evaluation opens unconventional avenues for a clinician, not least of which include the "physical examination" of the patient's sleep environment. An evaluator may be able to get additional information via telemedicine that he or she could not directly appreciate in the office (ie, an easy chair in the bedroom, continuous positive airway pressure at bedside, etc). This capability was not directly studied in this trial.

There were several limitations to our study. The most significant of these was the absence of an in-person–to–in-person, or telemedicine-to-telemedicine comparison for interrater reliability. The published literature includes some limited examples of in-person interrater reliability for sleep disorders. One study, using a structured interview template, showed a kappa coefficient of 0.73 for obstructive sleep apnea, substantially better than our weighted kappa of 0.414.[24] Another study, using a different structured interview template, assessed kappa as a secondary outcome measure, and had a result closer to ours (kappa for obstructive sleep apnea = 0.38).[25] Still, these studies employed designs different from the present study, so a direct comparison is imperfect.

Likewise, an interrater reliability study could have been designed with delayed repeated encounters using a single rater, or timely assessments between 2 raters, with both assessments employing telemedicine.[28] Given practical concerns of scheduling a patient for 2 different assessments before sleep testing, a study design that employed 2 different raters across 2 different settings (in-person vs telemedicine) was employed. This type of mixed methodology has been used by other telemedicine interrater reliability studies.[22–23,29–30] We imagine the introduction of an extra variable (2 different raters in 2 different settings) negatively impacted our kappa coefficients, and we might expect that our values would have been even stronger if both raters had utilized the telemedicine platform. Other limitations include individual evaluator practices. Temporal dispersion between the 2 evaluations was also a concern; we tried to minimize this by encouraging a 5-business-day timeframe, We also had a sample size smaller than anticipated due to a relatively high number of patients who were lost to follow-up. The "no shows" for the second visit were higher than expected, despite confirmation of adequate technology, multiple attempts to schedule the teleassessment by phone and email, and a small financial incentive. Perhaps a post-pandemic/telemedicine-familiar mindset, or a higher financial incentive, could have aided retention. We do not feel that this no-show rate is necessarily reflective of missed telemedicine appointments in clinical practice. In addition, although our raters received standardized case training for the protocol, this training did not include video review of the airway. This shortcoming may have impacted our raters' impressions. Due to study design limitations, the in-person evaluation always predated the telemedicine evaluation, potentially introducing patient bias into the telemedicine assessment. Also, the number of participants in this study was too small to determine with confidence if 2 of the raters were more closely aligned than the other 2 pairings. Last, the population recruited for this study was mostly White and of higher socioeconomic status, although the impact of this demographic is uncertain.

The present study demonstrated a promising signal in determining the accuracy of telemedicine encounters for OSA. Ultimately, outcome and cost-analysis studies are needed to determine the utility of this promising technology.

ABBREVIATIONS

HSAT, home sleep apnea test
OSA, obstructive sleep apnea
SE, standard error

REFERENCES

1. Young T, Palta M, Dempsey J, Skatrud J, Weber S, Badr S. The occurrence of sleep-disordered breathing among middle-aged adults. *N Engl J Med*. 1993; 328(17):1230–1235.
2. Frost & Sullivan. *Hidden Health Crisis Costing America Billions. Underdiagnosing and Undertreating Obstructive Sleep Apnea Draining Healthcare System*. Darien, IL: American Academy of Sleep Medicine, 2016. https://aasm.org/advocacy/initiatives/economic-impact-obstructive-sleep-apnea/. Accessed April 5, 2021.
3. Peppard PE, Young T, Barnet JH, et al.. Increased prevalence of sleep-disordered breathing in adults. *Am J Epidemiol*. 2013;177(9):1006–1014.
4. Vana KD, Silva GE, Goldberg R. Predictive abilities of the STOP-Bang and Epworth Sleepiness Scale in identifying sleep clinic patients at high risk for obstructive sleep apnea. *Res Nurs Health*. 2013;36(1):84–94.
5. Netzer NC, Stoohs RA, Netzer CM, Clark K, Strohl KP. Using the Berlin Questionnaire to identify patients at risk for the sleep apnea syndrome. *Ann Intern Med*. 1999;131(7):485–491.
6. Fenton ME, Heathcote K, Bryce R, et al. The utility of the elbow sign in the diagnosis of OSA. *Chest*. 2014;145(3):518–524.
7. Prasad KT, Sehgal IS, Agarwal R, Nath Aggarwal A, Behera D, Dhooria S. Assessing the likelihood of obstructive sleep apnea: a comparison of nine screening questionnaires. *Sleep Breath*. 2017;21(4):909–917.
8. Singh J, Badr MS, Diebert W, et al. American Academy of Sleep Medicine (AASM) position paper for the use of telemedicine for the diagnosis and treatment of sleep disorders. *J Clin Sleep Med*. 2015;11(10):1187–1198.
9. Bowman RJ, Kennedy C, Kirwan JF, Sze P, Murdoch IE. Reliability of telemedicine for diagnosing and managing eye problems in accident and emergency departments. *Eye (Lond)*. 2003;17(6):743–746.
10. Siew L, Hsiao A, McCarthy P, Agarwal A, Lee E, Chen L. Reliability of telemedicine in the assessment of seriously ill children. *Pediatrics*. 2016;137(3):e20150712.
11. Dorsey ER, Venkataraman V, Grana MJ, et al. Randomized controlled clinical trial of "virtual house calls" for Parkinson disease. *JAMA Neurol*. 2013;70(5):565–570.
12. Venkataraman V, Donohue SJ, Biglan KM, Wicks P, Dorsey ER. Virtual visits for Parkinson disease: a case series. *Neurol Clin Pract*. 2014;4(2):146–152.
13. Schneider R, Dorsey ER, Biglan K. Telemedicine care for nursing home residents with Parkinsonism. *J Am Geriatr Soc*. 2016;64(1):218–220.
14. Kelly JM, Schwamm LH, Bianchi MT. Sleep telemedicine: a survey study of patient preferences. *ISRN Neurol*. 2012;2012:135329.
15. Parikh R, Touvelle MN, Wang H, Zallek SN. Sleep telemedicine: patient satisfaction and treatment adherence. *Telemed J E Health*. 2011;17(8):609–614.
16. Coma-Del-Corral MJ, Alonso-Alvarez ML, Allende M, et al. Reliability of telemedicine in the diagnosis and treatment of sleep apnea syndrome. *Telemed J E Health* 2013;19(1):7-12.
17. Fields BG, Behari PP, McCloskey S, et al. Remote ambulatory management of veterans with obstructive sleep apnea. *Sleep*. 2016;39(3):501–509.
18. Zia S, Fields BG. Sleep telemedicine: an emerging field's latest frontier. *Chest*. 2016;149(6):1556–1565.
19. Rosen IM, Kirsch DB, Chervin RD, et al; American Academy of Sleep Medicine Board of Directors. Clinical use of a home sleep apnea test: an American Academy of Sleep Medicine position statement. *J Clin Sleep Med*. 2017;13(10):1205–1207.
20. Berry RB, Quan SF, Abreu AR, et al; for the American Academy of Sleep Medicine. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*. Version 2.6. Darien, IL: American Academy of Sleep Medicine; 2020.
21. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–174.
22. Awadallah M, Janssen F, Körber B, Breuer L, Scibor M, Handschu R. Telemedicine in general neurology: interrater reliability of clinical neurological examination via audio-visual telemedicine. *Eur Neurol*. 2018;80: 289–294.
23. Handschu R, Littmann R, Reulbach U, et al. Telemedicine in emergency evaluation of acute stroke: interrater agreement in remote video examination with a novel multimedia system. *Stroke*. 2003;34(12):2842–2846.
24. Taylor DJ, Wilkerson AK, Pruiksma KE, et al; STRONG STAR Consortium. Reliability of the structured clinical interview for DSM-5 sleep disorders module. *J Clin Sleep Med*. 2018;14(3):459–464.
25. Merikangas KR, Zhang J, Emsellem H, et al. The structured diagnostic interview for sleep patterns and disorders: rationale and initial evaluation. *Sleep Med*. 2014; 15(5):530–535.
26. Yu JL, Rosen I. Utility of the modified Mallampati grade and Friedman tongue position in the assessment of obstructive sleep apnea. *J Clin Sleep Med*. 2020; 16(2):303–308.
27. Teslong. NTE390W/430W Wi-Fi Digital Otoscope for iPhone/iPad/Android. https://www.teslong.com/Ear-Otoscope/WiFi-Ear-Otoscope. Accessed April 5, 2021.
28. Russell TG, Martin-Khan M, Khan A, Wade V. Method-comparison studies in telehealth: study design and analysis considerations. *J Telemed Telecare*. 2017; 23(9):797–802.
29. Cabana F, Boissy P, Tousignant M, Moffet H, Corriveau H, Dumais R. Interrater agreement between telerehabilitation and face-to-face clinical outcome measurements for total knee arthroplasty. *Telemed J E Health*. 2010;16(3): 293–298.
30. Basaran A, Ozlu O, Das K. Telemedicine in burn patients: reliability and patient preference. *Burns*. 2020.

SUBMISSION & CORRESPONDENCE INFORMATION

Submitted for publication November 4, 2020
Submitted in final revised form February 19, 2021
Accepted for publication February 19, 2021
Address correspondence to: Michael E. Yurcheshen, MD, FAASM, UR Sleep Center, 2337 South Clinton Avenue, Rochester, NY 14610; Email: Michael_Yurcheshen@urmc.rochester.edu

DISCLOSURE STATEMENT

All authors have seen and approved the manuscript. Work for this study was performed at the UR Sleep Center of the University of Rochester in Rochester, NY. The study was funded by a grant from the American Academy of Sleep Medicine Foundation (AASM Foundation grant #163-FP-17). Drs. Yurcheshen, C. Marcus, J. Marcus, and Messing received financial support from this grant. Dr. Yurcheshen has served as a clinical trials consultant for Jazz Pharmaceuticals and Harmony Biosciences; none of these consulting activities involve the subject matter for this present study. Dr. Pigeon has been a subinvestigator on observational trials funded by Pfizer, Inc., and by Abbvie, Inc., that are unrelated to this manuscript. He is an employee of the US Department of Veterans Affairs (VA); the views or opinions expressed herein do not necessarily represent those of the VA or the US government. Drs. C. Marcus, J. Marcus, Marsella, Messing, and Nguyen report no conflicts of interest.