the AHI value, the really important question from a clinical viewpoint would be this: In what percentage of cases could this variability lead physicians to change their therapeutic decision (e.g., the indication of CPAP or alternative treatments)? To make this calculation, the authors would need access to individual patients' clinical data. We do not know how many of the studied patients were treated with CPAP (or alternatives), but even without this information, the authors could make an estimate based on the indications for treatment specified in the different international guidelines (5).

On the other hand, the authors could offer additional clinically relevant information by assessing the minimum number of consecutive sleep studies needed to minimize their variability and thus reduce costs. The authors observe that the more sleep tests that are performed, the more the AHI stabilizes, according to an analysis of the different areas under the receiver operating curve (ROC) curves at different moments in time. Thus, it can be seen that, although the variability was very high in the first days, it was much lower after 1 week and even more so after 2 weeks, although the results obtained on Days 7 and 14 were very similar. The authors could calculate the relevant differences between the various areas under the ROC curves from clinical and statistical viewpoints. This information would be of enormous clinical relevance, as it would indicate the minimum number of days required for sleep tests to obtain an optimal balance between the least variability that would allow a minimum number of clinically acceptable misclassifications (and, above all, a minimum number of relevant therapeutic changes) and lower costs, less time to make an accurate diagnosis, and fewer resources.

Therefore, we believe that it would be very interesting and enriching for the study, and for clinicians who care for patients with OSA, if the authors could contribute these suggested new analyses to their already excellent study. ■

Author disclosures are available with the text of this letter at www.atsjournals.org.

Miguel Angel Martinez-Garcia, M.D.*
Hospital Universitario y Politécnico La Fe
Valencia, Spain

and

Instituto de Salud Carlos III
Madrid, Spain

Grace Oscullo, M.D.
Jose Daniel Gómez-Olivas, M.D.
Hospital Universitario y Politécnico La Fe
Valencia, Spain

ORCID ID: 0000-0002-7321-1891 (M.A.M.-G.).

*Corresponding author (e-mail: mianmartinezgarcia@gmail.com).

References

1. Lechat B, Naik G, Reynolds A, Aishah A, Scott H, Loffler KA, *et al.* Multinight prevalence, variability, and diagnostic misclassification of obstructive sleep apnea. *Am J Respir Crit Care Med* 2022;205:563–569.
2. Anitua E, Duran-Cantolla J, Almeida GZ, Alkhraisat MH. Predicting the night-to-night variability in the severity of obstructive sleep apnea: the case of the standard error of measurement. *Sleep Sci* 2019;12: 72–78.
3. Roeder M, Bradicich M, Schwarz EI, Thiel S, Gaisl T, Held U, *et al.* Night-to-night variability of respiratory events in obstructive sleep apnoea: a systematic review and meta-analysis. *Thorax* 2020;75: 1095–1102.
4. Leppänen T, Myllymaa S, Kulkas A, Töyräs J. Beyond the apnea-hypopnea index: alternative diagnostic parameters and machine learning solutions for estimation of sleep apnea severity. *Sleep* 2021;44: zsab134.
5. Epstein LJ, Kristo D, Strollo PJ Jr, Friedman N, Malhotra A, Patil SP, *et al.*; Adult Obstructive Sleep Apnea Task Force of the American Academy of Sleep Medicine. Clinical guideline for the evaluation, management and long-term care of obstructive sleep apnea in adults. *J Clin Sleep Med* 2009;5:263–276.

How Many Nights Are Really Needed to Diagnose Obstructive Sleep Apnea?

To the Editor:

We read with great interest the report by Lechat and colleagues (1) on characterizing the prevalence, variability, and diagnostic misclassification of obstructive sleep apnea (OSA) using multinight testing. The authors are to be commended on leveraging observations from the largest community-based sample with home recordings to address an issue of immense clinical relevance. The amassed data are impressive given the number of people included and the volume of nocturnal data used to describe the variability and misclassification of OSA. The authors were indeed crafty in using crowdsourced data from scalable technology and have thus paved the way for future studies that can leverage the ongoing explosive growth in sensors. Without doubt, the report by Lechat and colleagues (1) adds to the accepted notion that one night of monitoring, which is common in clinical decision making, is insufficient to case identify and classify OSA severity. Because the data on OSA diagnosis were derived at home, the issues of variability and misclassification, a phenomenon that is well known with in-lab studies, has been further addressed in the home setting (2).

Despite the many valuable insights, however, their report also raises several issues. First, the terminology used to describe the prevalence, variability, and misclassification uses "OSA" without further qualification. In their methods, the authors state OSA was defined as an apnea–hypopnea index (AHI) of ⩾15 events/h. However, the qualifier, " … at least moderate severity … ", does not consistently permeate the report, particularly with regard to the global estimate of OSA prevalence. It is important to recognize that the estimate of 22.6% is for moderate to severe OSA and not just OSA. This is not a trivial issue, because the prevalence of OSA of any severity will be much higher than 22.6%. In fact, analyses presented in Figure 2 show that data on prevalent mild, moderate, and severe

Supported by NIH grant HL146709.

Originally Published in Press as DOI: 10.1164/rccm.202112-2837LE on April 27, 2022

OSA from the contactless sensor are available. Given that the previous validation study (3) and the supplementary data comparing the contactless sensor and polysomnography showed no difference in AHI between the two tests, reporting the prevalence of different OSA categories using AHI thresholds would be of value.

Second, the authors have opted to use the mean AHI of all available nights to calculate the reference AHI against which the reliability of a subset of the nights is compared. It could be easily argued that the median AHI may be a better estimator of central tendency than the mean AHI, particularly if a person has extreme AHI values that may result from factors such as being in the supine position only or consumption of alcohol on any particular night. A possible alternative to the median could be the mode of the AHI distribution from each person. Although we are not proponents, an argument could also be made that the "diagnosis of OSA" should be based on the highest AHI value. Did the authors examine whether the prevalence and misclassification of OSA would be different if the median or mode were used for the reference AHI instead of the mean?

Third, the data on operating characteristics of multinight testing suggest that the increase in positive predictive and the drop in negative predictive values when comparing 7 with 14 nights is relatively small. Thus, what is the minimum number of nights of monitoring necessary to reliably estimate AHI in clinical practice within a ±5% margin of error? Having such information would help change the paradigm of clinical testing in which 1 night is always used despite the capability for multinight testing. It is time that multinight testing became mainstream practice, because the body of empirical evidence on AHI variability is unquestionable (4). One night is just not enough! ■

Author disclosures are available with the text of this letter at www.atsjournals.org.

Alexandre Abreu, M.D.
Naresh M. Punjabi, M.D., Ph.D.*
University of Miami Miller School of Medicine
Miami, Florida

*Corresponding author (e-mail: npunjabi@miami.edu).

References

1. Lechat B, Naik G, Reynolds A, Aishah A, Scott H, Loffler KA, *et al.* Multinight prevalence, variability, and diagnostic misclassification of obstructive sleep apnea. *Am J Respir Crit Care Med* 2022;205: 563–569.
2. Punjabi NM, Patil S, Crainiceanu C, Aurora RN. Variability and misclassification of sleep apnea severity based on multi-night testing. *Chest* 2020;158:365–373.
3. Edouard P, Campo D, Bartet P, Yang RY, Bruyneel M, Roisman G, *et al.* Validation of the Withings Sleep Analyzer, an under-the-mattress device for the detection of moderate-severe sleep apnea syndrome. *J Clin Sleep Med* 2021;17:1217–1227.
4. Punjabi NM, Aurora RN, Patil SP. Home sleep testing for obstructive sleep apnea: one night is enough! *Chest* 2013;143:291–294.

Reply to Martinez-Garcia *et al*. and to Abreu and Punjabi

From the Authors:

We thank Dr. Martinez-Garcia and colleagues, as well as Drs. Abreu and Punjabi, for their positive comments and thoughtful insights on our recent research on multinight prevalence and the potential impact of night-to-night variability in obstructive sleep apnea (OSA) severity on misdiagnosis (1). Some of the clinically relevant discussion points raised have been eloquently outlined in the accompanying editorial by Dr. Simonds (2). To provide further insights on this important topic as outlined in the letters to the editor, additional commentary and key analyses are provided below.

Use of Apnea–Hypopnea Index of 15 or More Events per Hour

Clinical guidelines indicate that an apnea–hypopnea index (AHI) of ⩾15 events/h even in the absence of symptoms is sufficient for the diagnosis of OSA and initiation of therapy (3). Community-based cohort studies also indicate that an AHI of ⩾ 15 events/h is associated with adverse cardiometabolic outcomes (4). Thus, the focus of our analyses was primarily on a cutoff of ⩾ 15 events/h. However, prior OSA prevalence estimates have used different AHI thresholds, including as low as 5 events/h (5). To allow for comparison with these findings, OSA prevalence estimates per country based on an AHI of ⩾ 5 events/h from our data are provided in Table 1. The estimated global prevalence of ∼55% based on this definition is higher than the estimated ∼37% in the study by Benjafield and colleagues (5). This may be, at least in part, owing to selection bias of the current consumer sample. Interestingly, however, our estimates appear more consistent across countries, which may be an advantage of the standardized, objective, and long-term data collection approach.

Mean versus Median AHI

We elected to use mean AHI as the reference value rather than median values, which could potentially yield different results. However, this was not the case, with comparable misclassification rates when mean versus median values were used. For example, 21% of OSA diagnoses were estimated to be false negatives on a single-night study based on mean AHI as the reference versus 18.4% (SD = 0.15) for median AHI. Similarly, receiver operating curves, F1-scores, and detection-error curves remained comparable when mean or median was used as the reference AHI. Ultimately, the optimal multinight OSA severity metrics will need to be determined empirically on the basis of predictive performance in relation to health outcomes and/or treatment response.

Supported by a National Health and Medical Research Council of Australia Leadership Fellowship (1196261) (D.J.E.).

Originally Published in Press as DOI: 10.1164/rccm.202202-0228LE on April 27, 2022