SLEEP, 2022, 1–2

https://doi.org/10.1093/sleep/zsac188
Advance Access Publication Date: 23 August 2022
Letter to the Editor

LETTER TO THE EDITOR

Near-boundary double-labeling-based classification: the new standard when evaluating performances of new sleep apnea diagnostic solutions against polysomnography?

Jean-Benoit Martinot[1,2,*], Jean-Louis Pépin[3], Atul Malhotra[4] and Nhat-Nam Le-Dong[5]

[1]Sleep Laboratory, CHU Université Catholique de Louvain (UCL) Namur Site Sainte-Elisabeth, Namur, Belgium; [2]Institute of Experimental and Clinical Research, UCL Bruxelles Woluwe, Brussels, Belgium; [3]HP2 Laboratory, Inserm U1300, University Grenoble Alpes, Grenoble, France; [4]Division of Pulmonary, Critical Care and Sleep Medicine, School of Medicine, University of California San Diego, La Jolla, CA 92121, USA; [5]Sunrise, Namur, Belgium

*Corresponding author. Jean-Benoit Martinot, Centre du Sommeil et de la Vigilance, CHU-UCL Namur Site Sainte-Elisabeth, 15 Place Louise Godin, 5000 Namur, Belgium. Email martinot.j@respisom.be.

Dear Editor:

Van Pee et al. [1] and the accompanying editorial [2] reported on a very promising methodological approach called near-boundary double-labeling (NBL) method classification for the evaluation of new sleep apnea diagnostic tools. Beyond the usual statistical methods (Bland-Altman, intraclass correlation coefficient, etc.), this complementary analysis led, at least in their peripheral arterial tonometry home sleep apnea test (HSAT) data set (derived from NightOwl technology), to a better agreement with polysomnography (PSG) scored by two independent investigators. This innovative approach allows to assign individuals with apnea hypopnea index (AHI) in predefined near-boundary zones to two different categories of AHI grades of severity. For example, one individual with an AHI close to the threshold of 15 events/hour was labeled with both mild and moderate grades of severity. This approach avoided reporting overly pessimistic results between two diagnostic methods for small AHI differences that are not clinically relevant. We have recently reported that the agreement between a mandibular movement (MM) HSAT and in-home PSG was similar to the agreement for the scoring of PSG between two expert centers (London and Grenoble) [3]. This finding demonstrated that unconscious biases associated with manual scoring are equivalent to some differences that can be found between simplified diagnostic techniques and ostensibly gold-standard PSG.

In this context, in a clinical study aiming to validate a machine learning–based algorithm for mandibular movement signals (Sunrise, Namur, Belgium), we applied the innovative NBL approach. We postulated that the risk of AHI-based severity misclassification due to inter-human PSG rating could be reduced

Table 1. Definition and proportions of seven categories obtained by NBL classification rule.

Class	Interval	*n*	%
Non-OSA	0–1.59	3	1.0
Non-OSA or mild	1.59–7.38	44	15.2
Mild	7.38–10.63	32	11.1
Mild or moderate	10.63–18.47	84	29.1
Moderate	18.47–25.96	55	19.0
Moderate or severe	25.96–36.10	36	12.5
Severe	≥ 36.10	35	12.1

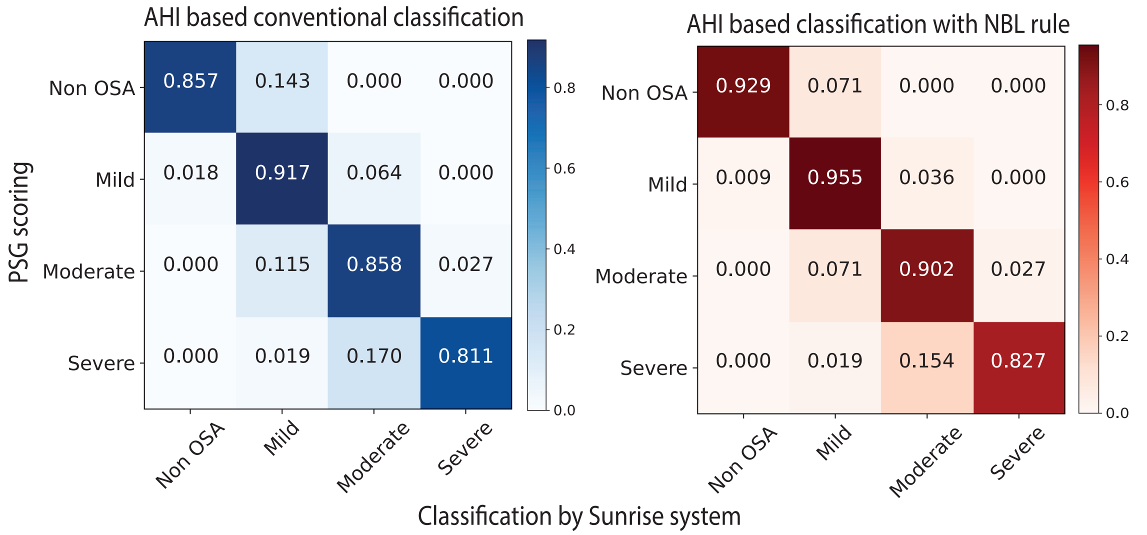

Figure 1. Distribution of PSG-AHI scores within four conventional severity levels and NBL-based categories.

when considering borderline zones around the traditional fixed AHI thresholds. A sample including 289 consecutive participants presenting with obstructive sleep apnea (OSA) suspicion was evaluated with an in-laboratory PSG coupled with simultaneous MM recordings using Sunrise device. The PSG data were then manually scored by two experienced and blinded investigators. The double-scored AHI data sets were used to establish the intervals for NBL method classification rule, by using the same procedure described in the original Van Pee article [1].

The collected MM data were automatically analyzed by a machine learning algorithm developed by Sunrise, which has the capability to classify sleep/wake states, obstructive and central respiratory events, thus allowing for estimation of AHI values [4, 5].

Based on the conventional rules for severity grading, the participants could be categorized into non-OSA (n = 14; 4.8%), mild (n = 109; 37.7%), moderate (n = 113; 39.1%), and severe OSA (n = 53; 18.4%). Corresponding proportions of the seven categories in the NBL classification are presented in Table 1.

As shown in Figure 1, applying the NBL method resulted in a significant improvement of agreement between the PSG readings and Sunrise automatic scoring. The four-way Cohen's Kappa coefficient was improved from 0.80 to 0.86. The NBL rule had also a positive impact on the accuracy in OSA severity grading. The F1 score, indicating the harmonic mean of precision and sensitivity was improved from 0.93 to 0.96 for detecting moderate OSA, and from 0.87 to 0.88 for severe OSA. As Sunrise algorithmic scoring represents a constant approximation of the true clinical status without any human factor-induced variability, we can assume that the uncertainty in PSG scoring has been improved when assessing performances by the NBL method. Therefore, the application of the NBL approach could help to make better diagnostic and therapeutic decisions in patients who present PSG results within near-boundary intervals. We also suggest that applying NBL procedure in clinical research would allow for a better standardization, transparency, and reproducibility in PSG data analysis.

Disclosure Statement

J.B.M. reports being a scientific advisor to Sunrise and being an investigator in pharmacy trials for Jazz Pharmaceuticals, Theranexus, and Desitin. J.L.P. reports being a scientific advisor to Sunrise; receiving grants and/or personal fees from ResMed, Philips, Fisher & Paykel, Sefam, AstraZeneca, AGIR à dom, Elevie, VitalAire, Boehringer Ingelheim, Jazz Pharmaceuticals, Night Balance and Itamar Medical; and receiving research support for clinical studies from Mutualia and Air Liquide Foundation. A.M. reports income related to medical education from Sunrise, Livanova, Equillium, Jazz Pharmaceuticals, and Corvus; ResMed provided a philanthropic donation to UCSD. N.N.L.D. is an employee of Sunrise. Non-financial disclosure: none.

References

1. Van Pee B, *et al*. A multicentric validation study of a novel home sleep apnea test based on peripheral arterial tonometry. *Sleep* 2022;**45**(5). doi:10.1093/sleep/zsac028
2. Ioachimescu OC. On PAT, patterns and paths. *Sleep* 2022;**45**(5). doi:10.1093/sleep/zsac057
3. Kelly JL, *et al*. Diagnosis of sleep apnoea using a mandibular monitor and machine learning analysis: one-night agreement compared to in-home polysomnography. *Front Neurosci*. 2022;**33**:3.
4. Pépin JL, *et al*. Assessment of mandibular movement monitoring with machine learning analysis for the diagnosis of obstructive sleep apnea. *JAMA Network Open* 2020;**3**(1):e1919657–e1919657.
5. Le-Dong NN, *et al*. Machine learning–based sleep staging in patients with sleep apnea using a single mandibular movement signal. *Am J Respir Crit Care Med*. 2021;**204**(10):1227–1231.