Margarida Pimenta Valério, Samuel Pereira, Joaquim Moita, Fátima Teixeira, Conceição Travassos, Ana Sofia Coutinho, Denny Marques Rodrigues

Pulmonology Department, Centro Hospitalar e Universitário de Coimbra, Coimbra, Portugal

Is the Nox-T3 device scoring algorithm accurate enough for the diagnosis of obstructive sleep apnea?

Abstract

Introduction: Obstructive sleep apnea (OSA) is highly prevalent. Home sleep apnea testing (HSAT) for OSA is rapidly expanding because of its cost effectiveness in the diagnosis of OSA. Type 3 portable monitors are used for this purpose. In most cases, these devices contain an algorithm for automatic scoring of events. We propose to study the accuracy of the automatic scoring algorithm in our population in order to compare it with the manually edited scoring of Nox-T3®.

Material and methods: For five months, a prospective study was performed. Patients were randomly distributed to the available HSAT devices. We collected the data of patients who performed HSAT with Nox-T3®. We used normality plots, the Spearman correlation, the Wilcoxon signed-rank test, and Bland–Altman plots.

Results: The sample consisted of 283 participants. The average manual apnea and hypopnea index (AHI) was 23.7 ± 22.1 events/h. All manual scores (AHI, apnea index, hypopnea index, and oxygen desaturation index) had strong correlations with their respective automated scores. When AHI > 15 and AHI > 30 the difference between the values of this index (automatic and manual) was not statistically significant. Also, for AHI values > 15 the mean difference between the two scoring methods was 0.17 events/h. For AHI values > 30, this difference was — 1.23 events/h.

Conclusions: When AHI is < 15, there may be a need for confirmation of automatic scores, especially in symptomatic patients with a high pretest probability of OSA. But, for patients with AHI > 15, automatic scores obtained from this device seem accurate enough to diagnose OSA in the correct clinical setting.

Key words: obstructive sleep apnea, home sleep apnea testing, automatic scoring algorithm

Adv Respir Med. 2021; 89: 262–267

Introduction

Obstructive sleep apnea (OSA) is prevalent and widespread as a result of the obesity epidemic and ageing populations. A recent study estimates that globally, nearly 1 billion adults between 30–69 years of age could have obstructive sleep apnea. Additionally, the number of people with moderate to severe obstructive sleep apnea, for which treatment is generally recommended, is estimated to be almost 425 million [1]. According to this study, the global burden of OSA is three to nine times higher than that presented in the FIRS report [2].

Home sleep apnea testing (HSAT) for OSA is rapidly expanding because of its sensitivity, specificity, and cost-effectiveness for OSA diagnosis, if associated with the patient's medical history and symptoms [3–5]. Type 3 portable monitors, as classified by the American Academy of Sleep Medicine (AASM), gather data through four to seven channels including airflow, respiratory effort, heart rate, and oxygen saturation. These devices also incorporate, in most cases, an algorithm for automatic scoring of events. There have been some studies to evaluate the accuracy of the automatically scored events in Type 3 monitors [6, 7]. A pilot study in the United States by Cairns *et al.* compared a simultaneous in-laboratory Nox-T3 (Nox Medical®, Inc., Reykjavik, Iceland) recording with polysomnography (PSG) in 32 adults. In this study, the portable

monitor recording had 100% sensitivity, 70% specificity, a positive predictive value of 88%, and a negative predictive value of 100% [8]. There are also some small validation studies for this device that include results comparing the automatic algorithm to polysomnography (PSG) [9, 10].

We propose to study the accuracy of the automatic scoring algorithm in our population as compared with the manually edited scoring of Nox-T3®.

Material and methods

Participants and procedure

For five months, from October 2019 to February 2020, we performed a prospective study in our Medicine Sleep Center. The patients were randomly distributed to the available HSAT devices (according to the usual procedure already in place). We collected the data of patients who performed HSAT with the Nox-T3® device and recorded the automatic generated report and the manually corrected report in a specific folder. We chose to use only this device to reduce bias related to possible differences in accuracy between different devices. The following signals were recorded: nasal pressure, rib cage and abdominal movement by inductance plethysmography, snoring, body position, activity, heart rate, and oxygen saturation by pulse oximetry. The participants received instructions on how to perform the recording. The test was performed at home over the course of one night and the equipment returned to the sleep center the following day.

We excluded patients with recording failures in any of the channels essential for the recognition of respiratory events during sleep - oximetry, airflow, and respiratory effort (rip cage and abdominal movement). We also excluded those with recordings with a duration of less than 4 hours. All other patients were included.

The scoring was completed according to the AASM manual for the scoring of sleep and associated events. The manual scoring was performed by a trained sleep technician.

Data analysis

We analyzed the data obtained using the Statistical Package of the Social Sciences (SPSS®) version 23 from IBM®. Age, sex, and comorbidities were analyzed using descriptive statistical analysis. Quantitative variables were reported as mean ± SD. The primary outcome variables were the apnea-hypopnea index (AHI), the oxygen--desaturation index (ODI), and the apnea and hypopnea indexes (AI and HI) derived from automatic and manual scoring. Normality of quantitative variables was assessed by using skewness and kurtosis ($p < 0.05$). Correlations between automatic and manual scores of these variables (AHI, ODI, AI, and HI) were assessed using Spearman's coefficient. Measured automatic and manual AHI values were compared using Wilcoxon-related samples signed-rank test. The levels of agreement between automatic AHI and manual AHI values were assessed by Bland-Altman plots with agreement lines defined by the formula: mean difference ± 1,96 × standard deviation (SD) of the measured differences. We used 15 and 30 as cutoff values for AHI taking into account OSA categories (moderate and severe).

Results

The final sample consisted of 283 participants (27 were excluded). As presented in Table 1, 60.1% were male, and the mean age was 57.1 ± 14.3 years. Most patients (92.9%) had an intermediate or high risk of OSA (according to STOP-BANG score: ≥ 3).

Mean values of AHI, AI, HI, and ODI, as well as AHI category in both scoring systems are represented in Table 2. Average manual AHI was 23.7 ± 22.1, whereas automatically scored AHI was 24.6 ± 20.7.

All represented manual scores (AHI, AI, HI, and ODI) had strong correlations with their respective automated scores ($\rho = 0.97$, $\rho = 0.89$, $\rho = 0.92$, $\rho = 0.99$, respectively; $p < 0.001$), as represented in Figure 1.

Table 1. Demographic characteristics

|	
Age (mean, years)	57.1 ± 14.3
Min	22
Max	88
Gender	n (%)
Female	113 (39.9)
Male	170 (60.1)
STOP-BANG	n (%)
≥ 3	263 (92.9)
Comorbidities	n (%)
Obesity	168 (59.3)
Hypertension	167 (59.0)
Cardiovascular disease	77 (27.2)
Respiratory disease	25 (8.8)

Table 2. Sleep parameters for patients who underwent HSAT

|	Manual HSAT (mean, events/h)	Autoscored HSAT (mean, events/h)
AHI	23,5 ± 25.6	24.3 ± 23.7
AI	11.6 ± 18.5	10.8 ± 15.1
HI	11.9 ± 11.6	13.4 ± 12.2
ODI	23.9 ± 23.9	24.2 ± 23.9
AHI category	n (%)	n (%)
Normal (<5)	52 (18,4)	41 (14,5)
Mild (5.0-14.9)	81 (28,6)	80 (28,3)
Moderate (15.0-29.9)	67 (23,7)	72 (25,4)
Severe (≥30)	83 (29,3)	90 (31,8)

HSAT — home sleep apnea testing; AHI — apnea and hypopnea index; AI — apnea index; HI — hypopnea index; ODI — oxygen-desaturation index

In all patients, when compared to the automatic scores previously mentioned, the respective manually corrected values (Wilcoxon signed-rank test) were significantly different ($p < 0.001$). They were also different if manual AHI ≤ 15 ($p < 0.001$). However, when considering only the cases in which the manual AHI > 15, the difference between the values of this index was not statistically significant ($p = 0.098$). This was also true for AHI values > 30 ($p = 0.454$).

For AHI values > 15, automatic AHI values (mean of 38.7 events/h) and manual AHI values (mean of 38.6 events/h) did not differ significantly, with a mean difference of 0.17 events/h (95% CI: −0,9 to 1,2 events/h).
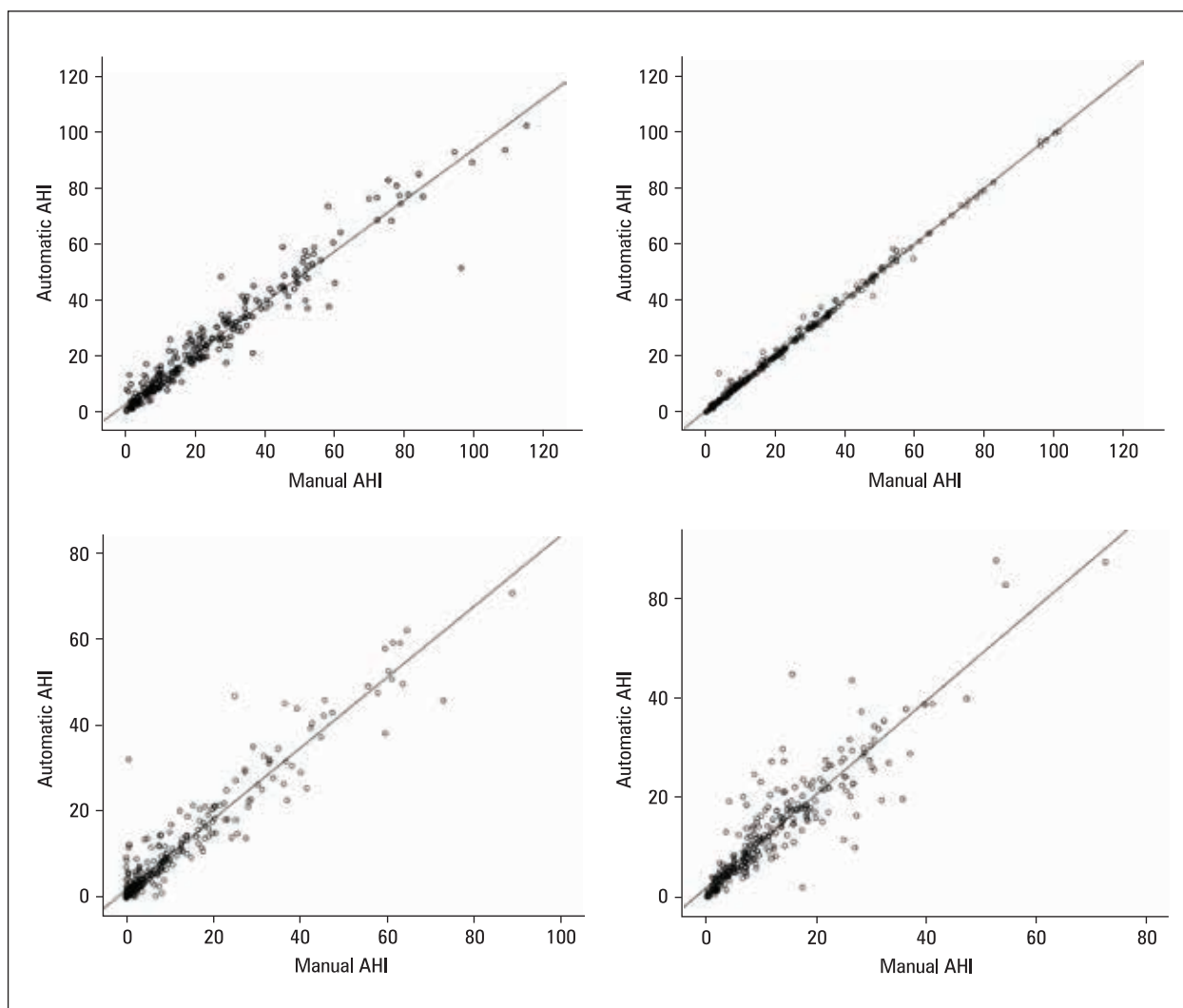
Figure 1. Scatter plot. Correlation between the automatic and manual scores (AHI, AI, HI, and ODI) using Spearman's coefficient. For the group of patients with respiratory disease, the manual AHI as well as the manual ODI had a strong correlation with their automatic values ($\rho = 0.94$, $\rho = 0.96$, respectively; $p < 0.001$)
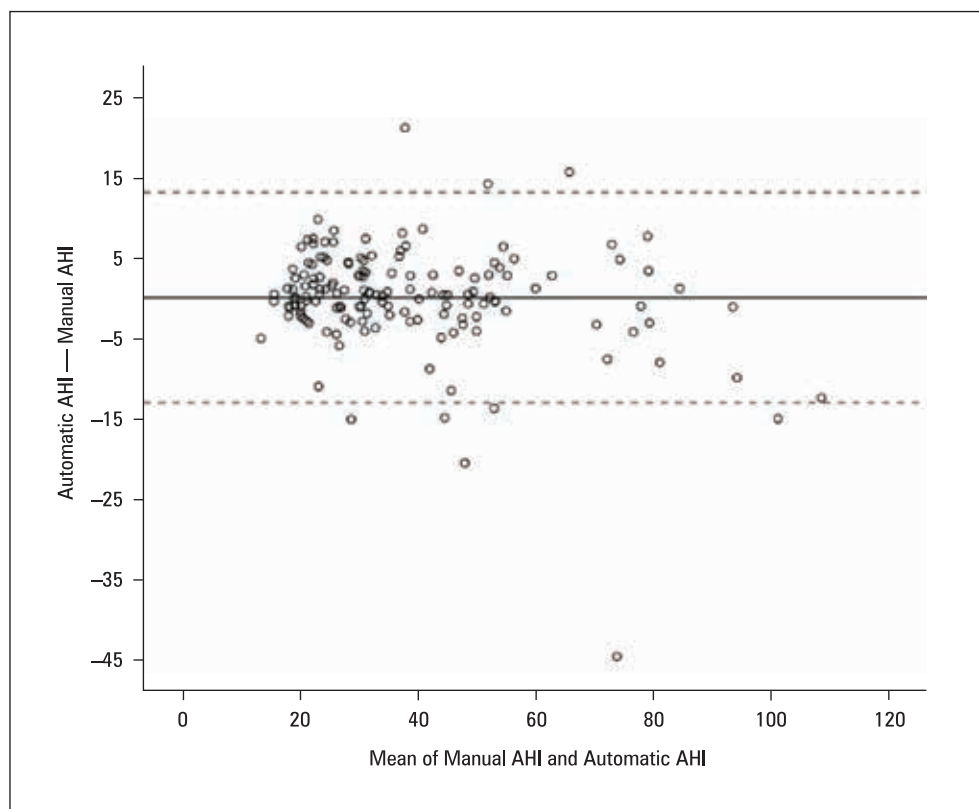
Figure 2. Bland-Altman plot, representing the agreement between manually scored AHI and automatically scored AHI for manual AHI > 15 events/h. Solid line shows mean; dotted lines denote upper and lower limits of agreement (± 1.96 SD)

The mean differences between the AHI values measured also did not differ significantly from zero in AHI values > 30. The observed automatically scored measures (mean of 50.4 events/h) were underestimated compared to the manual measures (mean of 51.6 events/h) by an average of 1.23 events/h (95% CI: −2.9 to 0.4 events/h).

Figure 2 and 3 illustrate the measurement agreement between the two scoring systems using Bland–Altman plots and show that the majority of data fell within the lines of agreement.

Discussion

Considering the growing incidence of OSA, it is essential to use diagnostic methods that are accurate and cost-effective. With this study, we tried to understand whether the values that resulted from the automatic scoring software of the Nox-T3 portable device were reproducible. The aim was to reduce the time and resources needed to achieve a diagnosis in the future.

There is a strong and statistically significant correlation in all variables from the two scoring methods even though there was a statistical differ-

ence between paired values of all scores. However, for scores of AHI > 15 and > 30 events/h, there is no statistically significant difference. Also, the mean difference between the two scores, in both cases (> 15 and > 30), is low (< 1.5 events/h). Therefore, when AHI is < 15 events/h, there may be a need for confirmation of automatic scores by manual editing or PSG, especially in symptomatic patients with a high pretest probability of OSA. However, for patients with AHI > 15 events/h, automatic scores obtained from this device seem accurate enough to diagnose OSA in the correct clinical setting.

These results are similar to the ones of previous studies. As in Cachada *et al.* [11] and Kristiansen *et al.* [12], there is a strong correlation between manual and automatic AHI. Also, in Kristiansen *et al.* [12], the largest overestimation and underestimation from automatic scoring, in terms of the number of misclassified recordings, were found for patients with an AHI < 20 events/h. The present study confirms that the difference between the paired values of the two scoring methods is statistically significant when AHI ≤ 15 events/h. Finally, as in Xu *et al.* [9], the mean difference between automatically *vs* manually
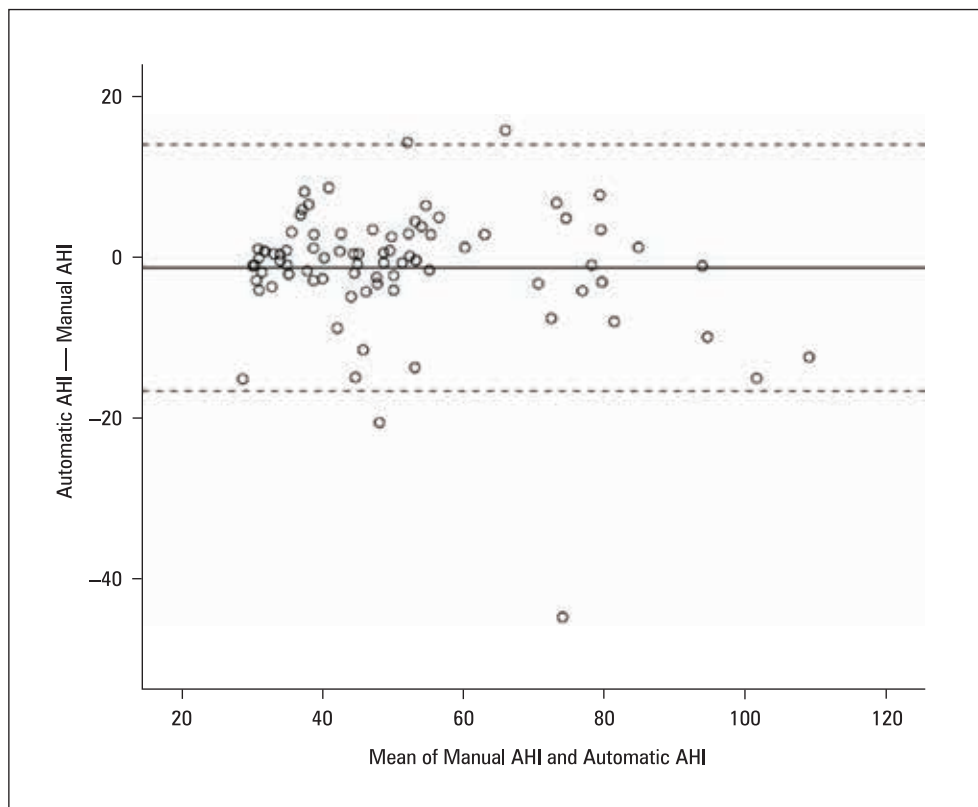
Figure 3. Bland-Altman plot, representing the agreement between manually scored AHI and automatically scored AHI for manual AHI > 30 events/h. Solid line shows mean; dotted lines denote upper and lower limits of agreement (± 1.96 SD)

edited scoring was less than 2 events/h (in our study for values of AHI > 15).

Even though the results of this study, in combination with the previous studies mentioned, show us how well automatic scoring works, it must be taken into consideration that to achieve a high quality of automatic analysis, it is important to have a good quality of the recording. Therefore, the patients should receive very clear instructions on how to perform the recording. As well, the signal quality of the oximetry, airflow, and respiratory effort channels should always be checked.

The main limitation of this study is that the manually corrected scores obtained were not confirmed through PSG.

Conclusions

Even though automatic scoring algorithms may not be accurate enough to replace manual reviewing in all situations, it may be a useful time saving tool. With these results, we hope to expedite the diagnosis of OSA in specific clinical settings which will allow to reduce the significant number of patients with undiagnosed OSA.

Conflict of interest:

None declared.

References:

1. Benjafield A, Ayas N, Eastwood P, et al. Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. The Lancet Respiratory Medicine. 2019; 7(8): 687–698, doi: 10.1016/s2213-2600(19)30198-5.
2. Forum of International Respiratory Societies. The Global Impact of Respiratory Disease – Second Edition. European Respiratory Society, Sheffield 2017: 26–27.
3. Flemons W, Littner M, Rowley J, et al. Home diagnosis of sleep apnea: a systematic review of the literature. Chest. 2003; 124(4): 1543–1579, doi: 10.1378/chest.124.4.1543.
4. 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, doi: 10.5664/jcsm.6774, indexed in Pubmed: 28942762.
5. Kapur VK, Auckley DH, Chowdhuri S, et al. Clinical practice guideline for diagnostic testing for adult obstructive sleep apnea: an American Academy of Sleep Medicine clinical practice guideline. J Clin Sleep Med. 2017; 13(3): 479–504, doi: 10.5664/jcsm.6506, indexed in Pubmed: 28162150.
6. Aurora RN, Swartz R, Punjabi NM. Misclassification of OSA severity with automated scoring of home sleep recordings. Chest. 2015; 147(3): 719–727, doi: 10.1378/chest.14-0929, indexed in Pubmed: 25411804.
7. Zhao YY, Weng J, Mobley DR, et al. Effect of manual editing of total recording time: implications for home sleep apnea testing. J Clin Sleep Med. 2017; 13(1): 121–126, doi: 10.5664/jcsm.6404, indexed in Pubmed: 27707441.

8. Cairns A, Wickwire E, Schaefer E, et al. A pilot validation study for the NOX T3(TM) portable monitor for the detection of OSA. Sleep Breath. 2014; 18(3): 609–614, doi: 10.1007/s11325-013-0924-2, indexed in Pubmed: 24442914.

9. Xu L, Han F, Keenan BT, et al. Validation of the Nox-T3 portable monitor for diagnosis of obstructive sleep apnea in chinese adults. J Clin Sleep Med. 2017; 13(5): 675–683, doi: 10.5664/jcsm.6582, indexed in Pubmed: 28356181.

10. Chang Y, Xu L, Han F, et al. Validation of the Nox-T3 portable monitor for diagnosis of obstructive sleep apnea in patients with chronic obstructive pulmonary disease. J Clin Sleep Med. 2019; 15(4): 587–596, doi: 10.5664/jcsm.7720, indexed in Pubmed: 30952218.

11. Cachada N, Thomas M, Wharton S. Comparison of manual and automatic scoring of limited channel sleep studies: nocturnal software correlates well with manual scoring in severe OSA. European Respiratory Journal . 2017; 50: PA2301, doi: 10.1183/1393003.congress-2017.pa2301.

12. Kristiansen S, Traaen GM, Øverland B, et al. Comparing manual and automatic scoring of sleep monitoring data from portable polygraphy. J Sleep Res. 2021; 30(2): e13036, doi: 10.1111/jsr.13036, indexed in Pubmed: 32430962.