Sleep Medicine 111 (2023) 86–93

Contents lists available at ScienceDirect

Sleep Medicine

journal homepage: www.elsevier.com/locate/sleep

Determinants of apnea-hypopnea index variability during home sleep testing

Jean-Benoît Martinot [a,b,*,1], Nhat-Nam Le-Dong [c,1], Renaud Tamisier [d,e], Sébastien Bailly [d,e], Jean-Louis Pépin [d,e]

[a] Sleep Laboratory, CHU Université Catholique de Louvain (UCL), Namur Site Sainte-Elisabeth, Namur, Belgium
[b] Institute of Experimental and Clinical Research, UCL Bruxelles Woluwe, Brussels, Belgium
[c] Sunrise, Namur, Belgium
[d] University Grenoble Alpes, HP2 Laboratory, Inserm U1300, Grenoble, France
[e] EFRC Laboratory, Grenoble Alpes University Hospital, Grenoble, France

ARTICLE INFO

Keywords:
Apnea-hypopnea index
Obstructive sleep apnea
Home sleep apnea testing
Mandibular jaw movements

ABSTRACT

Background: A single-night attended in-laboratory polysomnography or home sleep testing are common approaches for obstructive sleep apnea (OSA) diagnosis. However, internight variability in apnea-hypopnea index value is common, and may result in misclassification of OSA severity and inapropriate treatment decisions.
Objective: To investigate factors determining short-term apnea-hypopnea index variability using multi-night automated home sleep testing, and to determine how this variability impacts clinical decisions.
Patients/methods: Adults with suspected OSA who successfully performed three home sleep tests using measurements of mandibular jaw movements (Sunrise, Namur, Belgium) combined with automated machine learning analysis were enrolled. Data analysis included principal component analysis, generalized estimating equation regression and qualitative agreement analysis.
Results: 160 individuals who performed three sleep tests over a mean of 8.78 ± 8.48 days were included. The apnea-hypopnea index varied by −0.88 events/h (5th-95th percentile range: −14.33 to 9.72 events/h). Based on a single-night recording, rates of overtreatment and undertreatment would have been of 13.5% and 6.0%, respectively. Regression analysis adjusted for age, sex, body mass index, total sleep time, and time between home sleep tests showed that time spent in deep non-rapid eye movement sleep and with head in supine position were independent significant predictors of the apnea-hypopnea index variability.
Conclusions: At the individual level, short-term internight variability in the apnea-hypopnea index was significantly associated with time spent in deep non-rapid eye movement sleep and head in supine position. Clinical decisions based on a single-night testing may lead to errors in OSA severity classification and incorrect therapeutic decisions.

1. Background

Obstructive sleep apnea (OSA) is a major chronic health problem that has multi-organ consequences and is associated with a high economic and social burden. OSA affects nearly a billion people worldwide [1], often goes undiagnosed, and can have serious health consequences over time if left untreated.

The recommended procedure for establishing a diagnosis of OSA requires a single-night in-laboratory polysomnography (PSG) or home sleep apnea test (HSAT). The latter is suitable in individuals with high pre-test probability for OSA who do not have co-morbidities that could influence diagnosis performance [2]. A comprehensive evaluation with in-laboratory PSG is recommended for these individuals if the initial HSAT is negative [2].

It has been reported that there are no statistically or clinically significant differences in the apnea-hypopnea index (AHI) recorded on two consecutive PSG recordings (mean difference 0.14 events/h, 95% confidence interval −1.86 to 2.15 events/h) [2]. However, although

* Corresponding author. Centre du Sommeil et de la Vigilance, CHU Université catholique de Louvain, Namur Site Sainte-Elisabeth, 15, Place Louise Godin, 5000, Namur, Belgium.
E-mail address: martinot.j@respisom.be (J.-B. Martinot).
[1] Co-first authors.

https://doi.org/10.1016/j.sleep.2023.09.002
Received 7 June 2023; Received in revised form 14 August 2023; Accepted 3 September 2023
Available online 12 September 2023
1389-9457/© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

internight variability in AHI may be clinically negligible at the population level, recent publications have highlighted considerable short- or long-term night-to-night within-subject variability in AHI, especially (but not only) with HSAT [3–6]. This night-to-night variability in AHI can lead to a misclassification of OSA severity and either a delayed treatment or, conversely, overtreatment [5,7–9].

The within-subject variability in the clinical status of sleep-disordered breathing (SDB) across multiple nights of testing is due to various overlapping factors that can differ from one night to the other (Fig. 1). These factors include wake/sleep habits, alcohol or drug consumption, social jetlag, rostral fluid shift, natural night-to-night changes in sleep architecture (i.e., respective proportions of time spent in non-rapid eye movement [NREM] and rapid eye movement [REM] sleep), and changes in body and head positions during sleep [10–14].

Because the in-laboratory PSG is performed in a setting that differs from the home sleep environment, it may also affect sleep quality. This well-known phenomenon, called the first night effect, is characterized by lower sleep efficiency, decreased REM sleep and total sleep time (TST), and increased sleep and REM latencies. All of these can lead to underestimation of OSA severity on the first night of testing. The severity of OSA may also contribute to within-subject variability in the AHI, with the most important variability observed in those with milder disease. Other elements impacting the AHI variability can be controlled in a sleep study and relate to environmental effects (in-laboratory PSG versus HSAT), the type of recording system (more or less obtrusive), and the scoring approach (manual or automated) [15,16].

Recently, a new HSAT composed of a single point of contact sensor attached to the individual's chin (Sunrise, Namur, Belgium) has been introduced and showed to be a reliable home-based alternative in the evaluation of OSA [17–19]. With this device, the AHI is automatically derived from a machine learning-based analysis applied to mandibular jaw movements bio-signals. Mandibular jaw movements have been extensively studied over the last decade and were recently shown to be a reliable surrogate of the esophageal pressure gold standard signal to measure respiratory effort during sleep [20]. In addition to standard indices such as the AHI or respiratory disturbance index (RDI) also including respiratory effort-related arousals (RERAs), a new metric, the respiratory effort burden, or the proportion of sleep time spent in increased respiratory effort (REMOV) compared to normal breathing, is also provided. This unique index was recently identified as a stronger predictor of hypertension than common polysomnography-derived metrics in OSA patients [21]. Additionally, this device informs about the position of the head, provides accurate TST estimation by differentiating between wake and sleep states, and identifies sleep stages (light NREM, deep NREM and REM) [22].

The primary objective of this study was to explore interrelationships between different factors responsible for AHI variability in individuals with suspected OSA who were evaluated on three different occasions with a noninvasive tool using mandibular jaw movement measurements coupled with an automated scoring algorithm. A secondary objective was to evaluate the impact of repeated sleep recordings [5,23,24] on OSA diagnosis and severity classification that could impact on therapeutic decision-making.

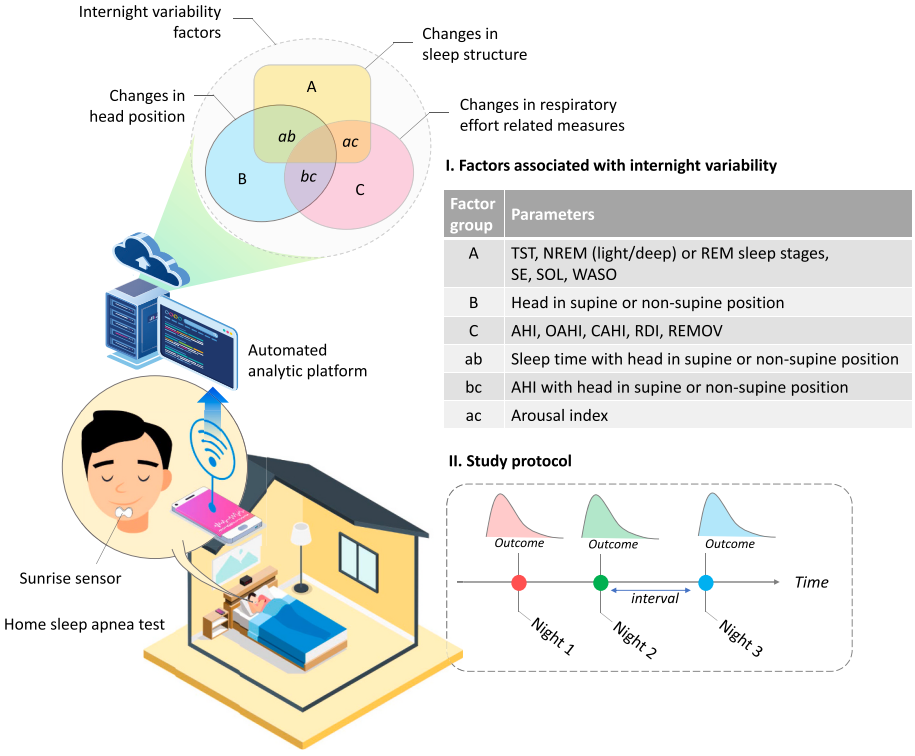
2. Methods

2.1. Study design

Data was acquired from 160 adult participants from a database of people who registered to use the home sleep apnea test based on automated mandibular jaw movement analysis (Sunrise, Namur, Belgium) between January 1, 2020, and March 30, 2022. This selected sample was composed of participants from European countries. All participants provided informed consent for their de-identified data to be used for research purposes. This study was approved by the Ethics Committee of the Faculty Hospital of Liège University, Belgium (ref 2023/96).

2.2. Data acquisition

Mandibular jaw movement bio-signals were captured using the Sunrise device (Sunrise, Namur, Belgium) [17–22]. This device comprises a single point of contact sensor weighing 3 g that is attached to the



Factor group	Parameters
A	TST, NREM (light/deep) or REM sleep stages, SE, SOL, WASO
B	Head in supine or non-supine position
C	AHI, OAHI, CAHI, RDI, REMOV
ab	Sleep time with head in supine or non-supine position
bc	AHI with head in supine or non-supine position
ac	Arousal index

Fig. 1. Factors associated with AHI variability and study protocol. A diagram summarizing the factors potentially associated with apnea-hypopnea index variability that were evaluated in the current analysis and details of the study protocol in which three sleep tests were performed at home using the Sunrise device. AHI, apnea-hypopnea index; CAHI, central apnea-hypopnea index; NREM, non-rapid eye movement; OAHI, obstructive apnea-hypopnea index; RDI, respiratory disturbance index; REM, rapid eye movement; REMOV, sleep time spent in increased respiratory effort; SE, sleep efficiency; SOL, sleep onset latency; WASO, wake time after sleep onset.

mentolabial sulcus of the individual. It includes an embedded inertial measurement unit that records the rotational movement of the mandible using a gyroscope and the position of the mandible using an accelerometer. The sensor communicates with a smartphone application for external control and enables mandibular jaw movement data transfer overnight via Bluetooth. Recorded raw data is then transferred to a cloud-based infrastructure at the end of the night using a secured data protocol, where it is pre-processed and automatically analyzed by a machine learning-based algorithm. Detailed information about the validity of automated mandibular jaw movement analysis is provided in the supplementary material. The position is labelled as supine when the head rests against the occipital area and non-supine when outside the occiput (Fig. S1).

2.3. Data analysis

Data processing and analysis was conducted using statistical computing packages in Python and R programming languages numpy [26], scipy [27] and FactoMineR [28].

A two-step analysis plan was undertaken. First, a principal component analysis (PCA) [29] was performed to explore the structural correlation between fifteen parameters that could contribute to the variability in AHI, including sleep quality and architecture (TST, sleep onset latency [SOL], wake time after sleep onset [WASO], sleep efficiency [SE], and proportion of TST spent in REM, light and deep NREM sleep), proportion of TST with head in supine and non-supine positions, and indices of OSA severity including respiratory effort related measures (AHI, obstructive AHI [OAHI], central AHI [CAHI], RDI, sleep time spent in increased respiratory effort [REMOV], arousal index [ArI]). Assuming that a decision to treat OSA is made when AHI is >15 events/h regardless of symptoms, data structure was divided into four distinct subgroups based on the change in AHI from one night to another (increase or decrease, with or without crossing an AHI cut-off of 15 events/h), and whether or not the change in AHI would result in a change in clinical decision-making [30]. Second, the effect of 9 sleep parameters (TST, SOL, WASO, SE, proportion of REM, light and deep NREM sleep, and proportion of sleep time with head in supine/non-supine positions), three anthropometric parameters (body mass index [BMI], age, neck circumference) and time interval between sleep tests on the night-to-night AHI variability was estimated by generalized estimating equation (GEE) regression analysis [31].

Then, the impact of internight AHI variability on the diagnosis and severity classification of OSA was evaluated. Four OSA categories were defined as follows: non-OSA (AHI <5 events/h); mild OSA (AHI 5 to <15 events/h); moderate OSA (AHI 15 to <30 events/h); and severe OSA (AHI ≥30 events/h) [32]. The probability of transition from one clinical category to another based on recordings from one night to another was determined and visualized in a network. The impact of a two-night sleep testing compared to a single-night testing on the accuracy of OSA diagnosis was also determined. The average AHI obtained over the three recordings was considered the reference clinical OSA status. A permutation test was then used to compare diagnostic performance metrics (sensitivity, specificity, F1 score, negative/positive predictive values and likelihood ratios) between a single-night and a two-night sleep testing for the detection of OSA at AHI cut-off values of ≥5, ≥15 and ≥ 30 events/h. The permutation test consisted of building a theoretical distribution of the mean between two groups from all possible permutations, then of verifying how frequent each difference was based on random grouping (permutation) and whether this was large or more extreme than the factual difference.

Statistical inferences were based on null-hypothesis testing (Friedman andConover post-hoc tests, Wald test and permutation tests), at a significance level of $p = 0.005$.

3. Results

3.1. Study population

A total of 160 participants (53 women, 107 men) were included. The mean ± standard deviation age of the participants was 47.9 ± 12.6 years, mean BMI was 28.1 ± 6.1 kg/m^2 and mean neck circumference was 39.7 ± 5.5 cm. The gender-specific distribution of sleep parameter variations across the three nights is outlined in the supplementary material, Tables S1 and S2.

All participants successfully completed three sequential sleep tests at home, over an interval of 8.78 ± 8.48 days. The mean interval between recording night 1 and night 2 was 3.90 ± 5.18 days, and between night 2 and night 3 was 4.50 ± 5.38 days. Of all parameters recorded, the only one that showed a statistically significant change from night 1 to night 2 and from night 1 to night 3 was SOL ($p < 0.005$) (Table 1).

3.2. AHI variability

Variability in AHI was common at the individual level, with only 5.6%–11.3% of observations having a stable AHI (defined as a relative change of <5% compared with the preceding night, or night 1). The AHI increased across the three nights of assessment in 33.1–47.5% of observations and decreased in 46.9–57.5% of observations. The within-subject variation was randomly distributed and there was no clinical pattern of increasing or decreasing AHI over the 3 nights.

Pairwise analysis showed that night-to-night AHI variations were randomly distributed within a 95% percentile range of −13.43 to +7.96

Table 1
Variations in key home sleep apnea testing parameters across different nights.

	Home sleep apnea testing		
	Night 1	Night 2	Night 3
TST, min	399.67 [281.78–507.70]	406.00 [252.85–519.55]	408.75 [267.93–503.55]
SOL, min	10.75 [2.50–45.72] *	8.75 [0.16–24.53]	8.92 [2.48–23.92]
WASO, min	47.00 [10.00–135.70]	38.00 [9.32–116.68]	41.25 [10.95–110.15]
SE, %	86.97 [69.08–95.89]	88.82 [71.74–96.00]	88.66 [72.62–96.68]
Deep NREM, % TST	6.99 [0.42–15.83]	6.60 [0.16–15.90]	6.80 [0.11–18.55]
Light NREM, % TST	81.54 [67.48–93.64]	81.04 [66.40–95.53]	80.12 [66.01–94.15]
REM, % TST	11.02 [3.58–22.13]	11.28 [2.53–22.82]	10.71 [3.52–21.96]
NREM, % TST	88.98 [77.87–96.42]	88.72 [77.18–97.47]	89.29 [78.03–96.47]
Head non-SUP, % TST	86.75 [45.34–100.00]	88.13 [55.67–100.00]	86.82 [47.86–100.00]
Head SUP, % TST	13.25 [0.00–54.66]	11.87 [0.00–44.33]	13.18 [0.00–52.14]
AHI, events/h	12.03 [4.94–32.88]	11.86 [3.68–35.82]	10.98 [4.38–30.82]
ArI, events/h	25.10 [13.67–46.32]	25.50 [12.68–48.59]	25.10 [12.49–41.92]
OAHI, events/h	7.17 [1.60–29.92]	7.25 [1.35–31.24]	6.24 [1.02–24.03]
CAHI, events/h	3.41 [0.00–13.63]	3.10 [0.17–14.13]	3.16 [0.26–12.51]
RDI, events/h	15.40 [5.83–39.34]	15.61 [4.41–43.92]	14.25 [5.23–39.04]
REMOV, % TST	43.50 [12.95–85.20]	42.00 [12.00–83.00]	41.00 [10.90–84.00]

Data were described as median [5th - 95th centiles]; *$p < 0.005$ vs. other nights. AHI, apnea-hypopnea index; ArI, arousal index; CAHI, central apnea-hypopnea index; Head non-SUP; head in non-supine position; Head SUP, head in supine position; NREM, non-rapid eye movement; OAHI, obstructive apnea-hypopnea index; RDI, respiratory disturbance index; REM, rapid eye movement; REMOV, sleep time spent in increased respiratory effort; SE, sleep efficiency; SOL, sleep onset latency; TST, total sleep time; WASO, wake time after sleep onset.

events/h (night 1 to night 2), −15.40 to +7.25 events/h (night 2 to night 3) and −13.66 to +8.59 events/h (night 1 to night 3), with very low median values (+0.05, −0.63 and −1.48 events/h, respectively). Except for two individuals with extreme variations (−33.42 and +27.52 events/h), the distribution of internight AHI variability was similar across the three nights and within each OSA severity subgroup. Internight AHI variability at the individual level is visualized in Fig. S2.

3.3. Correlations between variations in sleep indices across recordings

A PCA showed the inter-relationship among the internight variation of sleep parameters derived from measurement of mandibular jaw movements, overall and in the four clinical subgroups (Fig. 2). Along with OAHI and RDI, variations in time spent in light NREM sleep, REMOV, ArI and time with head in supine position contributed to the definition of the two extreme clinical subgroups (A [overtreated patients, n = 65; 14%] and D [undertreated patients, n = 29; 6%]). In contrast, variations in WASO and SOL did not correlate with AHI variability. Specific values for variation in all these indices are presented in Table S1.

While quantitative agreement analysis at the population level did not show any significant change in AHI distribution across the three nights

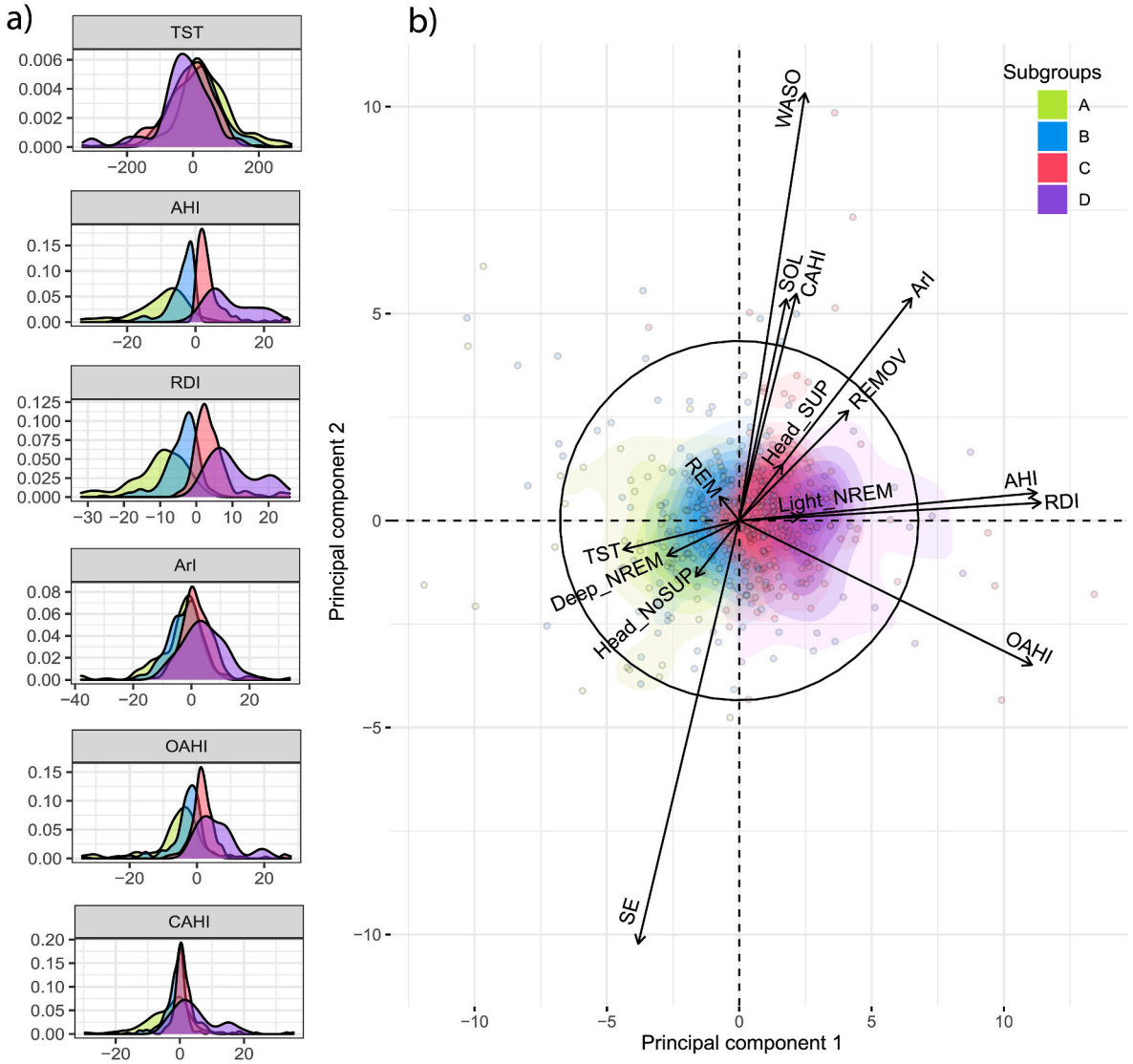


Fig. 2. Distribution and correlation structure of the variations in mandibular jaw movement-derived indices. Density plots representing the distribution of night-to-night variation in mandibular jaw movement-derived indices and total sleep time (**a**); principal component analysis graphs consisting of a bi-dimensional density and scatter plot layer, representing the joint distribution of two first principal components (1 and 2), and coordinates of the 15 input variables as vectors (**b**). In part (**b**), each dimension forms an apparent scale which is proportional to the standardized values of input variables. Thus, a positive coordinate value indicates an increase while a negative value indicates a decrease. Each point represents an observation unit (variation from a pair of two sequential nights, n = 480). Each vector represents a variable, and its length and orientation with respect to a principal component axis indicate how much the variable contributes to that component; angles between the vectors and their direction allow evaluation of correlations (small angles indicate a strong positive correlation; opposite angles indicate a negative correlation). In both (**a**) and (**b**), data distribution is stratified into four subgroups based on the AHI tendency to change from a night to another and whether or not this change resulted in a difference in OSA therapeutic decision (at an apnea-hypopnea index [AHI] cut-off of 15 events/h). Group A (n = 65): lower AHI from one night to the next with a change in therapeutic decision; Group B (n = 207): lower AHI from one night to the next without any change in therapeutic decision; Group C (n = 179): higher AHI from one night to the next without any change in therapeutic decision; Group D (n = 29): higher AHI from one night to the next with a change in therapeutic decision. AHI, apnea-hypopnea index; ArI, arousal index; CAHI, central apnea-hypopnea index; Head_NoSUP; head in non-supine position; Head_SUP, head in supine position; NREM, non-rapid eye movement; OAHI, obstructive apnea-hypopnea index; RDI, respiratory disturbance index; REM, rapid eye movement; REMOV, sleep time spent in increased respiratory effort; SE, sleep efficiency; SOL, sleep onset latency; TST, total sleep time; WASO, wake time after sleep onset.

of assessment, a clinical-based stratification allowed the identification of two subgroups (A and D) that showed uncertainty in OSA severity classification or a false negative diagnosis and/or in appropriate treatment. The time interval between two consecutive nights of assessment was not significantly associated with internight variation in any parameter.

The independent effects of variations in nine sleep, three anthropometric parameters and interval between nights on AHI variability are shown in Fig. 3. Time spent in deep NREM sleep and with head in supine position were independent significant predictors of AHI variation. After adjustment for TST and time interval between nights, regression models showed that there was an internight change in AHI of 1.0 event/h for each 10% increase in sleep time with head in supine position, and an internight decrease in AHI of 2.9 events/h for each 10% increase in the proportion of time spent in deep NREM sleep ($p < 0.001$). In addition, a decrease in TST had a slight but significant effect on internight AHI variability.

3.4. Internight AHI variability and clinical decision-making

As shown in Fig. 4, night-to-night variation in AHI may lead to uncertainty in the clinical classification of OSA severity, although the pattern of transitions in clinical severity grading was consistent across the three nights.

Individuals with mild OSA showed the best within-subject reproducibility in clinical OSA severity classification (Fig. 4). In this population, AHI remained within the original range in most cases (from 75.3% between nights 1 and 3 to 83.2% between nights 2 and 3). However, 40.0–70.0% of participants with a normal AHI (<5 events/h) on the first night transitioned to having an AHI of greater OSA severity on the second or third night, indicating that the first sleep test provided a false negative diagnosis. A transition between OSA severity categories was most common from non-OSA to mild OSA on the first night of assessment. Of those having mild OSA initially, 9.0–13.5% and 1.2–2.3% of individuals had an AHI indicating moderate and severe OSA at the next measurement, respectively. A decline in OSA severity classification between two sleep tests occurred in all OSA severity subgroups (i.e., from severe to moderate, moderate to mild, or mild to non-OSA) but was most common between two adjacent categories of OSA severity. Skipping a category occurred infrequently, and no cases with transition from non-OSA to severe OSA were observed.

3.5. OSA severity classification based on a single-night versus two-night testing

For this analysis, the AHI average value based on 3 nights of assessment was considered to indicate the true clinical condition, and appropriate treatment was prescribed if an AHI score greater than 15 events/h was measured. Clinical decision-making based on a single-night AHI score (from night 1, 2 or 3) was different to the decision based on the AHI score obtained by averaging the score from the three nights in 19.6% of cases (94/480 observations). Moreover, an over-treatment due to an overestimated AHI on a single-night test would have been prescribed in 13.5% of cases (65/480 observations) and 6.0% (29/480 observations) would not have received treatment due to underestimation of AHI during a single-night test.

For the three OSA categories at AHI cut-off values of 5, 15 or 30 events/h, the diagnostic performance was significantly improved when two nights of assessment were performed compared with a single-night recording (Table 2).

4. Discussion

This study analyzed the complex relationship between various contributing factors and determined short-term internight AHI variability using an automated device across three nights of recording for the first time. The availability of multidimensional data about the variations in different parameters derived from measurement of mandibular jaw movements allowed principal component analysis that was used to provide a comprehensive description of these variations and their interrelationship. The resulting loading plot also facilitated intuitive understanding of how variations in sleep parameters contribute to the variation of AHI.

It is known that being monitored can trigger a “first night effect”, but this study aimed to capture short-term variability without imposing a strict protocol on the patients. To ensure comfort and engagement, patients were allowed to choose the nights of recording. The study presented successive recordings instead of consecutive ones to capture natural inter-night variability while minimizing interference in behavioral and lifestyle factors such as nutrition, physical activity, alcohol, caffeine use, and tobacco intake. This approach focused on short-term variability rather than night-to-night variability.

The findings of this study indicate that internight AHI variability could have important consequences regarding clinical and treatment decision-making. Data analysis also indicates that sleeping with the

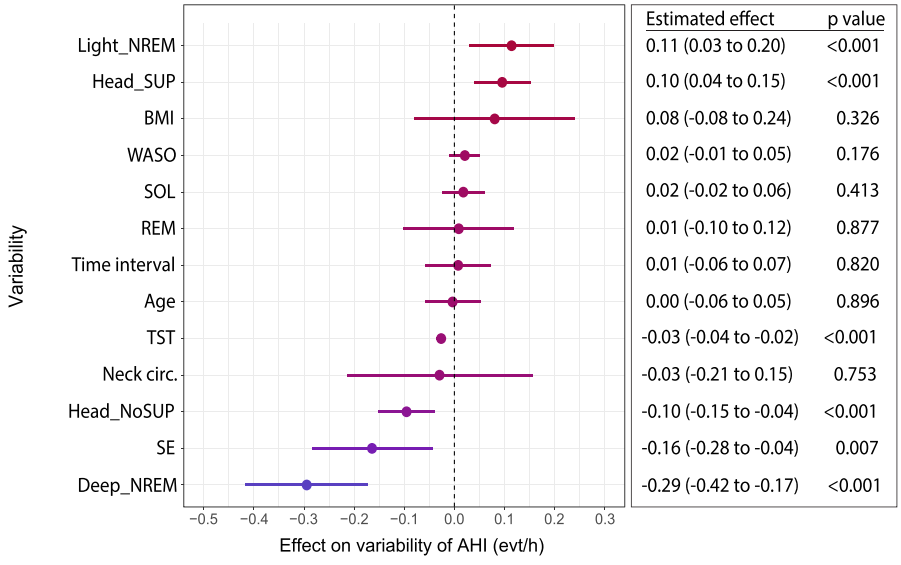


Fig. 3. Effect of variation of sleep and anthropometric parameters on the night-to-night AHI variability. Marginal effect of variation in 10 sleep and three anthropometric parameters on night-to-night variability in the apnea-hypopnea index (AHI) determined using general estimating equation (GEE) models. These models estimated AHI variability depending on the variability of a target parameter, after adjustment for the effect of variation in total sleep time (TST) and interval between testing nights. Dots represent the estimated value and horizontal lines represent the 95% confidence interval (absolute values are shown to the right of the graph). Statistical inferences were based on Wald test at significance threshold of $p = 0.005$. BMI, body mass index; evt, events; Head_NoSUP; head in non-supine position; Head_SUP, head in supine position; Neck circ., neck circumference; NREM, non-rapid eye movement; REMOV, sleep time spent in increased respiratory effort; SE, sleep efficiency; SOL, sleep onset latency; WASO, wake time after sleep onset.

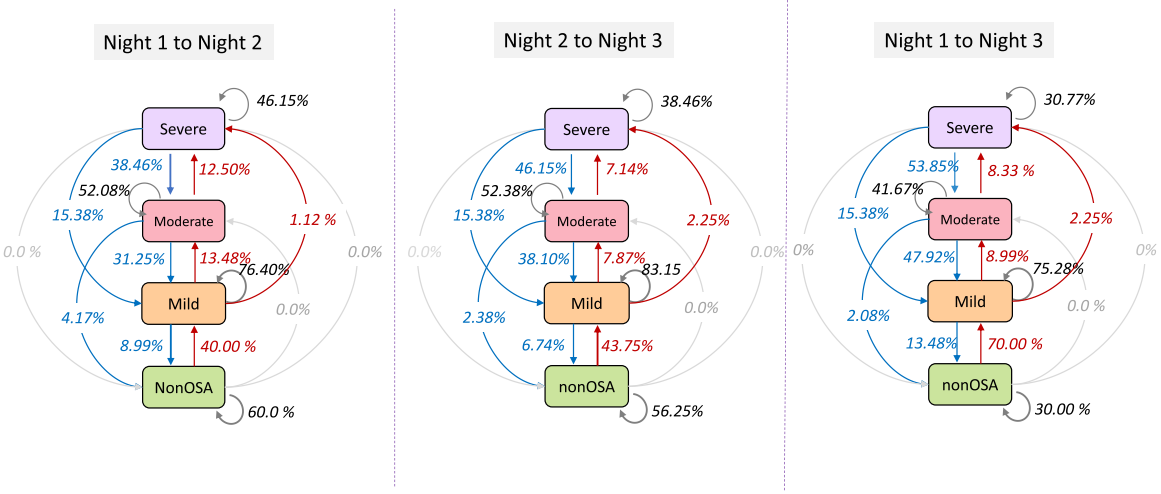


Fig. 4. Night-to-night transition patterns of the apnea-hypopnea index (AHI) across different levels of obstructive sleep apnea (OSA) severity. Each of the three networks describes the probabilities of every possible transition of OSA severity category from night 1 to night 2 (left), night 2 to night 3 (middle), and night 1 to night 3 (right). Four OSA severity categories were used: non-OSA with AHI <5 events/hour (green); mild OSA with AHI 5 to <15 events/hour (orange); moderate OSA with AHI 15 to <30 events/hour (red); and severe OSA with AHI ≥30 events/hour (purple). Probabilities were estimated from a 4 × 4 transition matrix under assumption of multinomial distribution. The arrows indicate different patterns of transition: moving to a higher (red), lower (blue) or unchanged (black) OSA category. Grey arrows indicate a transition of null probability. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 2
Impact of adding a second night of assessment on the accuracy of obstructive sleep apnea (OSA) severity grading.

Metrics	Detecting AHI ≥5 events/h		Detecting AHI ≥15 events/h		Detecting AHI ≥30 events/h	
	One night	Two nights	One night	Two nights	One night	Two nights
BAC	0.85	0.86	0.86	0.93	0.89	0.89
F1 score	0.97	0.98	0.82	0.92	0.64	0.80
Sensitivity	0.95	0.99	0.81	0.88	0.82	0.89
Specificity	0.75	*0.75**	0.93	0.97	0.97	0.98
FPR	0.25	0.29	0.07	0.03	0.03	0.03
FNR	0.05	0.01	0.19	0.12	0.18	0.11
LR+	3.79	3.80	11.06	31.36	22.37	51.33
LR-	0.06	0.02	0.21	0.12	0.19	0.12
PPV	0.99	*0.99**	0.86	0.95	0.54	0.73
NPV	0.43	0.67	0.90	0.94	0.99	0.99

Diagnostic performance metrics were estimated by comparing the apnea-hypopnea index (AHI) on any night (1, 2 or 3) and the average AHI from sleep tests on two different nights (1 and 2, 1 and 3, or 2 and 3) against the expected "true outcome" (i.e., mean AHI from the 3 nights). Statistically significant differences in the median AHI between single-night and double-night evaluations were determined using permutation testing (* indicates not statistically significant at a p-value threshold of 0.005).
AHI, apnea-hypopnea index; BAC, balanced accuracy; F1 score, harmonic mean of sensitivity and positive predictive value; FNR, false negative rate; FPR, false positive rate; LR-, negative likelihood ratio; LR+, positive likelihood ratio; NPV, negative predictive value; PPV, positive predictive value.

head in supine position is independently associated with greater AHI variability, whereas sleep quality (based on the proportion of sleep time spent in deep NREM sleep) is an independent contributor to less variability in AHI.

The within-subject variability in AHI was caused by a variety of factors related to natural changes in sleep position and sleep quality [5,9]. A strength of our study was to use an unobtrusive tool using mandibular jaw movement measurements coupled with an automated scoring algorithm across three separate nights of recording. This approach reduced the classical "first night effect" observed with in-laboratory PSG [8] and eliminated biases related to interscorer variability in manual scoring [8].

The GEE model showed that internight AHI variability was significantly associated with variations in the ratio of light versus deep NREM sleep and in the proportion of TST spent with the head in supine position. Previous studies have also reported that OSA severity with the trunk in supine position could be significantly mitigated by rotating the head from a supine to lateral position [33–37]. A critical influence of head position on AHI, independently of trunk position and sleep stage, was also reported in a recent study by Zhu et al. [37]. Our findings quantitatively confirm the relationship between the supine head position during sleep and natural internight AHI variability.

Diagnostic and therapeutic decision-making could be inaccurate when based on a single night of sleep testing, even when sleep testing is performed at home with an unobtrusive technology (as used in this study). These findings highlight the importance of repeating simplified and automated home sleep testing to better support clinical diagnosis and personalized management of OSA.

In this study we also separated the potential effects of factors related to sleep duration/quantity (including SOL, WASO, and TST) on internight AHI variability. Given that those effects were small, the use of mandibular jaw movement-based technology and automated scoring could be beneficial for reducing AHI variation due to bias related to the time of measurement. The mandibular jaw movement-based technology has been shown to outperform actigraphy-based signals for the sleep/wake classification task and provides therefore better accuracy in TST, WASO and SOL calculation [22], a frequent pitfall for other type 3 or 4 recording devices in multi-night variability measurements.

The conventional OSA severity grading system (ICSD-3) [25] was used to classify OSA severity after each night of sleep testing (non-OSA, mild OSA, moderate OSA or severe OSA). This study showed that there is a high risk of diagnosis misclassification based on the results of a single night of assessment because for instance the probability of being diagnosed as having non-OSA on one night and mild OSA on a later night was 40–70% (Fig. 4).

It is widely recognized that although AHI has been extensively used to gauge the severity of sleep apnea, taking solely AHI into consideration to predict the clinical consequences of obstructive sleep apnea (OSA) or to make treatment decisions comes, however, with significant limitations [38,39]. Nevertheless, estimates of OSA prevalence are based on a variety of different AHI thresholds, including those as low as 5 events/h.

Current clinical guidelines (ICSD-3) indicate that an AHI of ≥15 events/h even in the absence of symptoms is sufficient for the diagnosis of OSA and initiation of therapy [25,30]. Community-based cohort studies also indicate that this AHI cut-off value is associated with adverse cardiometabolic outcomes [40]. Thus, like other authors, the analyses of this study primarily focused on an AHI cutoff value of ≥15 events/h. If the decision to initiate OSA treatment was solely based on an AHI exceeding 15 events per hour regardless of symptoms, night-to-night transitional probabilities showed that, based on a single-night test, around 15% would have been undertreated (i.e., patients diagnosed with no or mild OSA but suffering from moderate to severe OSA) and 50% overtreated (i.e., patients diagnosed with moderate to severe OSA but suffering from no or mild OSA) (Fig. 4). Similar changes in the OSA severity category from one night to another have also been reported in previous studies [5,9]. The greatest risk for misclassification in this study was observed in individuals with mild to moderate OSA, which is consistent with recent data published by Lechat et al. [8].

Based on these findings, AHI values obtained from a single-night test may result in uncertainty and misdiagnosis of OSA, and a closer approximation of an individual's true OSA status is likely to be obtained by averaging AHI values from recordings taken on three different nights. Having an accurate indication of OSA severity is key as it allows appropriate treatment decisions to be made, such as whether or not to treat a patient, and which treatment is most appropriate (e.g., positive airway pressure therapy or an oral appliance) [9]. Interestingly, the within-subject variability in SDB studies is higher than that in other fields of medicine such as lung function testing, where respiratory measurements are repeated over time and are therefore less prone to inaccuracies related to technical or natural variability.

The key strengths of this study are its novelty, the use of a validated non-invasive HSAT based on automated analysis of mandibular jaw movements, and multiple recordings over a short period of time. These findings reinforce the recent literature reporting about multi-night variability in SDB recordings. Future studies performed by other researcher teams in the field should reproduce these findings.

5. Conclusion

In conclusion, three nights of sleep evaluation with a non-invasive tool using automated analysis of mandibular jaw movement biosignals provide useful information about the natural short-term variability in AHI and highlight the risk of inaccurate determination of OSA severity based on a single-night recording. This data reinforces the value of multi-night sleep recordings to reduce potential OSA misdiagnosis and severity misclassification. Furthermore, the AHI variability over three nights that has the potential to alter therapeutic decision-making at an AHI threshold of 15 events/h is associated with the proportion of sleep time spent in deep non-REM sleep and with the head in supine position.

Funding and support

JLP is supported by the French National Research Agency in the framework of the "Investissements d'avenir" program (ANR-15-IDEX-02) and the "e-health and integrated care and trajectories medicine and MIAI artificial intelligence" chairs of excellence from the Grenoble Alpes University Foundation. This work was partially supported by MIAI @ Grenoble Alpes, (ANR-19-P3IA-0003).

CRediT authorship contribution statement

Jean-Benoît Martinot: Formal analysis, Conceptualization, Data curation, Investigation, Methodology, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Nhat-Nam Le-Dong:** Formal analysis, Data curation, Methodology, Visualization, Writing – original draft. **Renaud Tamisier:** Conceptualization, Writing – review & editing. **Sébastien Bailly:** Data curation, Methodology, Writing – review & editing. **Jean-Louis Pépin:** Conceptualization, Formal analysis, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

Jean-Benoît Martinot is a scientific advisor to Sunrise and has been an investigator in pharmaceutical trials for Jazz Pharmaceuticals and Theranexus.

Nhat-Nam Le-Dong is an employee of Sunrise.

Jean-Louis Pépin is a scientific advisor to Sunrise, has received grants and/or personal fees from ResMed, Philips, Fisher & Paykel, Sefam, AstraZeneca, AGIR à dom, Elevie, VitalAire, Boehringer Ingelheim, Jazz Pharmaceuticals and Itamar Medical, and has received research support for clinical studies from Mutualia and Air Liquide Foundation.

There was no funding or other financial support for this research from Sunrise.

Renaud Tamisier and **Sébastien Bailly** have no financial disclosures.

Acknowledgements

The authors wish to thank Ms. Liesbeth Ory and Ms. Ravzat Ashurlaeva for their secretarial assistance and support. Manuscript formatting assistance was provided by Nicola Ryan (New Zealand), independent medical writer.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.sleep.2023.09.002.

References

[1] Benjafield AV, Ayas NT, Eastwood PR, Heinzer R, Ip MSM, Morrell MJ, Nunez CM, Patel SR, Penzel T, Pépin JL, Peppard PE, Sinha S, Tufik S, Valentine K, Malhotra A. Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. Lancet Respir Med 2019;7:687–98.
[2] Kapur VK, Auckley DH, Chowdhuri S, Kuhlmann DC, Mehra R, Ramar K, Harrod CG. 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;13:479–504.
[3] Lechat B, Catcheside P, Reynolds A, Adams RJ, McEvoy RD, Eckert DJ. Reply to Martinez-Garcia et al. and to Abreu and Punjabi. Am J Respir Crit Care Med 2022;206:126–9.
[4] Lechat B, Nguyen DP, Reynolds A, Loffler K, Escourrou P, McEvoy RD, Adams R, Catcheside PG, Eckert DJ. Single-night diagnosis of sleep apnea contributes to inconsistent cardiovascular outcome findings. Chest 2023;(Jan 28):S0012–3692(0023)00157-00155 (online ahead of print).
[5] Roeder M, Sievi NA, Bradicich M, Grewe FA, Siegfried S, Gaisl T, Kohler M. The accuracy of repeated sleep studies in OSA: a longitudinal observational study with 14 nights of oxygen saturation monitoring. Chest 2021;159:1222–31.
[6] Sforza E, Roche F, Chapelle C, Pichot V. Internight variability of apnea-hypopnea index in obstructive sleep apnea using ambulatory polysomnography. Front Physiol 2019;10:849.
[7] Aarab G, Lobbezoo F, Hamburger HL, Naeije M. Variability in the apnea-hypopnea index and its consequences for diagnosis and therapy evaluation. Respiration 2009;77:32–7.
[8] Lechat B, Naik G, Reynolds A, Aishah A, Scott H, Loffler KA, Vakulin A, Escourrou P, McEvoy RD, Adams RJ, Catcheside PG, Eckert DJ. Multinight prevalence, variability, and diagnostic misclassification of obstructive sleep apnea. Am J Respir Crit Care Med 2022;205:563–9.
[9] Punjabi NM, Patil S, Crainiceanu C, Aurora RN. Variability and misclassification of sleep apnea severity based on multi-night testing. Chest 2020;158:365–73.
[10] Bliwise DL, Benkert RE, Ingham RH. Factors associated with nightly variability in sleep-disordered breathing in the elderly. Chest 1991;100:973–6.
[11] Lorenzo JL, Barbanoj MJ. Variability of sleep parameters across multiple laboratory sessions in healthy young subjects: the "very first night effect". Psychophysiology 2002;39:409–13.
[12] Stöberl AS, Schwarz EI, Haile SR, Turnbull CD, Rossi VA, Stradling JR, Kohler M. Night-to-night variability of obstructive sleep apnea. J Sleep Res 2017;26:782–8.

[13] Tamaki M, Nittono H, Hayashi M, Hori T. Examination of the first-night effect during the sleep-onset period. Sleep 2005;28:195–202.
[14] White LH, Lyons OD, Yadollahi A, Ryan CM, Bradley TD. Night-to-night variability in obstructive sleep apnea severity: relationship to overnight rostral fluid shift. J Clin Sleep Med 2015;11:149–56.
[15] Ahmadi N, Shapiro GK, Chung SA, Shapiro CM. Clinical diagnosis of sleep apnea based on single night of polysomnography vs. two nights of polysomnography. Sleep Breath 2009;13:221–6.
[16] Tschopp S, Wimmer W, Caversaccio M, Borner U, Tschopp K. Night-to-night variability in obstructive sleep apnea using peripheral arterial tonometry: a case for multiple night testing. J Clin Sleep Med 2021;17:1751–8.
[17] Kelly JL, Ben Messaoud R, Joyeux-Faure M, Terrail R, Tamisier R, Martinot JB, Le-Dong NN, Morrell MJ, Pépin JL. Diagnosis of sleep apnoea using a mandibular monitor and machine learning analysis: one-night agreement compared to in-home polysomnography. Front Neurosci 2022;16:726880.
[18] Martinot JB, Pépin JL, Malhotra A, Le-Dong NN. Near-boundary double-labeling-based classification: the new standard when evaluating performances of new sleep apnea diagnostic solutions against polysomnography? Sleep 2022;45.
[19] Pépin JL, Letesson C, Le-Dong NN, Dedave A, Denison S, Cuthbert V, Martinot JB, Gozal D. Assessment of mandibular movement monitoring with machine learning analysis for the diagnosis of obstructive sleep apnea. JAMA Netw Open 2020;3:e1919657.
[20] Pepin JL, Le-Dong NN, Cuthbert V, Coumans N, Tamisier R, Malhotra A, Martinot JB. Mandibular movements are a reliable noninvasive alternative to esophageal pressure for measuring respiratory effort in patients with sleep apnea syndrome. Nat Sci Sleep 2022;14:635–44.
[21] Martinot Jean-Benoit, Le-Dong Nhat-Nam, Malhotra Atul, Pépin Jean-Louis. Respiratory effort during sleep and prevalent hypertension in obstructive sleep apnoea. Eur Respir J 2023;61(3):2201486. https://doi.org/10.1183/13993003.01486-2022.
[22] Le-Dong NN, Martinot JB, Coumans N, Cuthbert V, Tamisier R, Bailly S, Pépin JL. Machine learning-based sleep staging in patients with sleep apnea using a single mandibular movement signal. Am J Respir Crit Care Med 2021;204:1227–31.
[23] Le Bon O, Hoffmann G, Tecco J, Staner L, Noseda A, Pelc I, Linkowski P. Mild to moderate sleep respiratory events: one negative night may not be enough. Chest 2000;118:353–9.
[24] Prasad B, Usmani S, Steffen AD, Van Dongen HP, Pack FM, Strakovsky I, Staley B, Dinges D, Maislin G, Pack AI, Weaver TE. Short-term variability in apnea-hypopnea index during extended home portable monitoring. J Clin Sleep Med 2016;12:855–63.
[25] American Academy of Sleep Medicine. International classification of sleep disorders – third ed. (ICSD-3)..
[26] Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, Kern R, Picus M, Hoyer S, van Kerkwijk MH, Brett M, Haldane A, Del Río JF, Wiebe M, Peterson P, Gérard-Marchant P, Sheppard K, Reddy T, Weckesser W, Abbasi H, Gohlke C, Oliphant TE. Array programming with NumPy. Nature 2020;585:357–62.
[27] Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, Polat İ, Feng Y, Moore EW, VanderPlas J, Laxalde D, Perktold J, Cimrman R, Henriksen I, Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F, van Mulbregt P. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 2020;17:261–72.
[28] Lê S, Josse J, Husson F. FactoMineR: an R package for multivariate analysis. J Stat Software 2008;25:1–18.
[29] Husson F, Le S, Pagès J. Exploratory multivariate analysis by example using R. Chapman and Hall/CRC; 2017.
[30] Epstein LJ, Kristo D, Strollo Jr PJ, Friedman N, Malhotra A, Patil SP, Ramar K, Rogers R, Schwab RJ, Weaver EM, Weinstein MD. Clinical guideline for the evaluation, management and long-term care of obstructive sleep apnea in adults. J Clin Sleep Med 2009;5:263–76.
[31] Hardin J, Hilbe J. Generalized estimating equations. London: Chapman and Hall/CRC; 2003.
[32] Malhotra A, Ayappa I, Ayas N, Collop N, Kirsch D, McArdle N, Mehra R, Pack AI, Punjabi N, White DP, Gottlieb DJ. Metrics of sleep apnea severity: beyond the apnea-hypopnea index. Sleep 2021;44.
[33] Joosten SA, O'Donoghue FJ, Rochford PD, Barnes M, Hamza K, Churchward TJ, Berger PJ, Hamilton GS. Night-to-night repeatability of supine-related obstructive sleep apnea. Ann Am Thorac Soc 2014;11:761–9.
[34] Levendowski DJ, Oksenberg A, Vicini C, Penzel T, Levi M, Westbrook PR. A systematic comparison of factors that could impact treatment recommendations for patients with Positional Obstructive Sleep Apnea (POSA). Sleep Med 2018;50:145–51.
[35] Martinot J-B, Le-Dong N, Cuthbert V, Denison S, Tamisier R, Borel JC, Pepin JL. Head and Body positions as risk-factors for sleep obstructive breathing. ERJ Open Res 2019;5:P116.
[36] Safiruddin F, Koutsourelakis I, de Vries N. Upper airway collapse during drug induced sleep endoscopy: head rotation in supine position compared with lateral head and trunk position. Eur Arch Oto-Rhino-Laryngol 2015;272:485–8.
[37] Zhu K, Bradley TD, Patel M, Alshaer H. Influence of head position on obstructive sleep apnea severity. Sleep Breath 2017;21:821–8.
[38] Dirk A, Pevernagie DA, Schiza S, Verbraecken J, Arnardottir ES. On the rise and fall of the apnea−hypopnea index: a historical review and critical appraisal. J Sleep Res 2020;29:e13066. https://doi.org/10.1111/jsr.13066.
[39] Lévy P, Tamisier R, Pépin JL. Assessment of sleep-disordered-breathing: quest for a metric or search for meaning? J Sleep Res 2020:e13143. https://doi.org/10.1111/jsr.13143. 00.
[40] Punjabi NM, Sorkin JD, Katzel LI, Goldberg AP, Schwartz AR, Smith PL. Sleep-disordered breathing and insulin resistance in middle-aged and overweight men. Am J Respir Crit Care Med 2002;165:677–82.