## SCIENTIFIC INVESTIGATIONS

# Validation of the Nox-T3 Portable Monitor for Diagnosis of Obstructive Sleep Apnea in Chinese Adults

Liyue Xu, MD[1]; Fang Han, MD[2]; Brendan T. Keenan, MS[3]; Elizabeth Kneeland-Szanto, MBA[3]; Han Yan, MD[2]; Xiaosong Dong, MD[2]; Yuan Chang, MD[1]; Long Zhao, BS[2]; Xueli Zhang, BS[2]; Jing Li, BS[2]; Allan I. Pack, MBChB, PhD[3]; Samuel T. Kuna, MD[3,4]

[1]PKU-UPenn Sleep Center, Peking University International Hospital, Beijing, China; [2]Department of Respiratory Medicine, Peking University People's Hospital, Beijing, China; [3]Department of Medicine and Center for Sleep and Circadian Neurobiology, University of Pennsylvania, Philadelphia, Pennsylvania; [4]Department of Medicine, Corporal Michael J. Crescenz Veterans Affairs Medical Center, Philadelphia, Pennsylvania

**Study Objectives:** To evaluate the performance of a portable monitor (Nox-T3, Nox Medical Inc. Reykjavik, Iceland) used to diagnose obstructive sleep apnea in Chinese adults.

**Methods:** Eighty Chinese adults (mean ± standard deviation age 47.6 ± 14.0 years, 77.5% males, body mass index 27.5 ± 5.4 kg/m$^2$) underwent overnight, unattended home sleep apnea testing (HSAT) with the Nox-T3 portable monitor followed by an overnight in-laboratory polysomnogram (PSG) with simultaneous portable monitor recording. The portable monitor recordings were scored using automated analysis and then manually edited using different criteria for scoring hypopneas. Polysomnography was scored based on recommended guidelines.

**Results:** When scoring of hypopneas required a ≥ 4% oxygen desaturation event, the mean ± standard deviation apnea-hypopnea index (AHI) was 24.4 ± 20.8 events/h on HSAT, 28.0 ± 22.9 events/h on in-laboratory portable monitor recording, and 28.6 ± 23.9 events/h on PSG ($P$ < .0001). Bland-Altman analysis of AHI on PSG versus HSAT showed a mean difference (95% confidence interval) of −4.64 (−7.15, −2.13); limits of agreement (equal to ± 2 standard deviations) was −26.62 to 17.35 events/h. Based on a threshold of AHI ≥ 5 events/h, HSAT had 95% sensitivity, 69% specificity, 94% positive predictive value, and 75% negative predictive value compared to PSG. Using an AHI ≥ 15 events/h, HSAT had 93% sensitivity, 85% specificity, 89% positive predictive value, and 91% negative predictive value. Closer agreements were present when comparing the simultaneous recordings. Similar results were obtained using different scoring criteria for hypopneas.

**Conclusions:** Despite known differences between HSAT and PSG, the results show close agreement between the two diagnostic tests in Chinese adults, especially when controlling for night-to-night variability and changes in sleeping environment.

**Keywords:** apnea-hypopnea index, home sleep apnea testing, polysomnography

**Citation:** Xu L, Han F, Keenan BT, Kneeland-Szanto E, Yan H, Dong X, Chang Y, Zhao L, Zhang X, Li J, Pack AI, Kuna ST. 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.

## INTRODUCTION

Although a polysomnogram (PSG) performed in a sleep laboratory and attended by a technologist remains the reference standard for the diagnosis of obstructive sleep apnea (OSA), unattended home sleep apnea testing (HSAT) with portable monitors is increasingly being used to diagnose OSA. This is of particular clinical importance in countries where sleep medicine is rapidly developing. For example, in China, polysomnography is currently used to diagnose almost all instances of sleep apnea. However, using this paradigm, it is estimated that only 0.6% of patients with OSA in China have received a diagnosis (Han F, personal communication). Validation of HSAT monitors will allow HSAT to gain wider acceptance in China and increase patient access to testing. Validation of HSAT monitors is also of importance given their utility in the ambulatory telemedicine-based clinical pathways that are being developed for the diagnosis of OSA and treatment of patients with the condition. Moreover, HSAT is less costly than PSG and allows assessment in the home environment rather than the unfamiliar sleep laboratory.

### BRIEF SUMMARY

**Current Knowledge/Study Rationale:** The purpose of this study was to evaluate the performance of a portable monitor (Nox-T3) used for home sleep testing to diagnose obstructive sleep apnea. In particular, we wanted to validate its use in Chinese adults, a less obese patient population than that evaluated in most previous studies. **Study Impact:** The results show that home sleep testing using the Nox-T3 monitor has close agreement with the results of in-laboratory polysomnography. The ability to use home sleep testing rather than polysomnography to diagnose sleep apnea will improve access to care.

However, HSAT is less comprehensive than PSG, and differences between the two procedures need to be considered when using HSAT results in clinical management. For example, electroencephalogram, electrooculogram, and chin muscle activity signals are usually not recorded during HSAT. The absence of these signals prevents detection of when the patient is sleeping and the stages of sleep. Consequently, the apnea-hypopnea index (AHI) on HSAT is generally calculated as the

mean number of respiratory events per hour of recording time rather than, as on PSG, events per hour of sleep. This increases the likelihood that the HSAT will underestimate the severity of the patient's OSA. The lack of electroencephalogram signals on HSAT also prevents detection of arousals. This may also result in underestimation of OSA severity in contrast to PSG when hypopneas on PSG are scored based on an associated 3% or greater oxygen desaturation event and/or an arousal.

Given these differences between the two diagnostic testing methods, it is important to thoroughly understand the portable monitor used for HSAT as compared to the information provided by PSG. Validation studies are therefore performed prior to introduction of a portable monitor into clinical practice.[1] These studies are usually designed to take into consideration the known night-to-night variability in PSG results and the known effects of environment on sleep quality. One of the most widely used HSAT portable monitors is the Nox-T3 monitor (Nox Medical, Inc., Reykjavik, Iceland) that records nasal pressure, snoring, rib cage and abdominal movement, pulse oximetry, activity, and body position. Bipolar channels are also available, but not commonly used. A pilot study in the United States by Cairns et al.[2] compared a simultaneous in-laboratory Nox-T3 recording and PSG in 32 adults (mean ± standard deviation [SD] body mass index [BMI] $32.8 \pm 6.8$ kg/m$^2$). On manually edited scoring, the mean ± SD AHI was $16.3 \pm 19$ events/h on PSG and $18.6 \pm 19$ events/h on the portable monitor recording. Comparing manually edited scoring at an AHI cutoff of $\geq 5$ events/h, the portable monitor recording had 100% sensitivity, 70% specificity, a positive predictive value of 88%, and a negative predictive value of 100%. That pilot study also reported a close agreement between AHI and oxygen desaturation index between the Noxturnal software's (Nox Medical, Inc., Reykjavik, Iceland) automatic scoring without manual editing and manually scored PSG. HSAT was not performed in that study. One of the purposes of the current study was to validate the Nox-T3 monitor in a larger number of subjects and evaluate its performance during HSAT, the monitor's intended purpose.

The performance characteristics of a portable monitor may be influenced by the characteristics of the patient population in which it is used. For example, oxygen desaturation events are more likely to occur secondary to a respiratory event in an obese patient than a nonobese patient, because of the lower oxygen tension commonly present in obese patients. It is not known how well the monitor performs in Chinese adults, a less obese population with craniofacial features that are associated with increased risk of OSA.[3] Therefore, another purpose of the current study was to validate HSAT using the Nox-T3 monitor in Chinese adults.

## METHODS

### Protocol

Eighty adults referred to the sleep center at the Peoples' Hospital (Peking University, Beijing, China) for evaluation of OSA volunteered to participate. Participants were between the ages of 18 to 80 years and had no previous sleep testing or treatment for OSA. Individuals were excluded for the following reasons: no telephone access or inability to return for follow-up; prior diagnosis of central sleep apnea/Cheyne-Stokes respiration, obesity hypoventilation syndrome, narcolepsy, rapid eye movement behavior disorder, chronic obstructive pulmonary disease, or heart failure; shift work, regular jet lag or irregular work schedules by history over the past 3 months; supplemental oxygen therapy (daytime or nocturnal); or a clinically unstable medical condition as defined by a change in medications in the previous month, or a new medical diagnosis in the previous 2 months (eg, myocardial infarction, active infection, thyroid disease, depression or psychosis, cirrhosis, surgery, or cancer). This project was conducted to evaluate HSAT prior to its use in the sleep center's routine clinical practice. The Institutional Review Board at Peking University People's Hospital approved the project. Written informed consent was not required, but all participants were informed about the purpose of the project and all activities in this study conformed to the principles outlined by the Declaration of Helsinki.

All participants were asked to initially perform an overnight HSAT using the Nox-T3 portable monitor followed, within 1 week, by an in-laboratory PSG (Alice6, Philips Respironics, Inc, Murrysville, Pennsylvania, United States) with a simultaneous Nox-T3 portable monitor recording. One subject declined to perform the HSAT. The order of home and in-laboratory testing was fixed to assess the ability of individuals with no previous experience with sleep testing to successfully perform the HSAT. For both the home and in-laboratory testing, participants were instructed to sleep in whatever position was comfortable for them and to take their regular medications.

### Portable Monitor Recordings

Nox Medical, Inc., which supplied the monitors and software, had no other involvement in the study. The following signals were recorded during the portable monitor recordings: nasal pressure, rib cage and abdominal movement by inductance plethysmography, snoring, body position, activity, and heart rate and oxygen saturation by pulse oximetry. For the HSAT, the participants came to the sleep center to receive instructions on how to perform the recording. During the session, a trained sleep technologist demonstrated how to apply the sensors and the participant was then asked to apply the sensors. After the technician confirmed proper placement, the sensors were removed and the participant reapplied the sensors at home just prior to bedtime. The morning after the HSAT, the participant completed an after-study questionnaire to report events during the recording. During the in-laboratory sleep testing, the sleep technologist applied the portable monitor sensors and initiated the recording. Separate sensors were used for the simultaneous portable monitor and PSG recordings. Therefore, during the PSG with simultaneous portable monitor recording, the participant wore two sets of nasal cannula, two sets of rib cage and abdominal belts, and two pulse oximeters.

The ability of subjects to perform HSAT was assessed as the percentage of individuals with a successful initial HSAT and the quality of the HSAT used for analysis. A successful HSAT required at least 3 hours of recording containing the oxygen saturation and at least one of the respiratory signals (airflow,

rib cage movement, abdominal movement). If the initial HSAT was unsuccessful, the participant took a portable monitor home after the PSG and performed another HSAT. If the second attempt was unsuccessful, the HSAT was not repeated.

The quality of the HSAT was assessed by automated analysis of signal quality for oxygen saturation, airflow, abdominal movement, and thoracic movement. The automated analysis scores artifacts when the signal is absent or deemed to be invalid. For example, if the oxygen saturation values are outside an acceptable range, an artifact event is marked in that area on the oxygen saturation signal. Signal quality reported by the automated analysis is calculated as a percentage of the total duration of scoreable signal within the analysis period divided by the total duration of the signal within the analysis period. The signal quality results were not used to include or exclude Nox-T3 recordings.

Analysis start time and stop time on the portable monitor recordings was manually determined based on the participant's responses on a morning questionnaire and the activity signal on the recording. The scorer was blinded to whether the portable monitor recording was performed at home or in the laboratory and to a particular participant's PSG results. The portable monitor recordings were initially scored automatically using Noxturnal software. The software program defined apneas as ≥ 90% reduction in airflow from baseline for at least 10 seconds. Obstructive apneas were defined as an apnea associated with respiratory effort and central apneas were defined as an apnea during which respiratory effort was absent. Mixed apneas were defined as an apnea during which respiratory effort was initially absent but appeared during the latter part of the event. Hypopneas were defined as a ≥ 30% reduction in a respiratory signal for ≥ 10 seconds associated with a ≥ 4% reduction in oxygen saturation. The recordings were then manually edited by an experienced PSG technologist with the aid of the software program using 2012 American Academy of Sleep Medicine scoring criteria.[4] The same start and stop time selected for the automatic scoring was used for the manually edited scoring. Two separate manually edited scorings were performed using different definitions for hypopnea: (1) the same criteria used for automatic scoring, and (2) hypopneas defined by a ≥ 30% reduction in a respiratory signal for at least 10 seconds associated with a ≥ 3% reduction in oxygen saturation. When the portable monitor recording's nasal pressure signal was absent or not able to be scored throughout the recording or during portions of the recording, the flow signal derived from the rib cage and abdominal respiratory inductance plethysmography signals was used for scoring.[5] The AHI on the Nox-T3 recordings was calculated as the average number of apneas and hypopneas per hour of analysis time.

## Polysomnograms

Polysomnography was performed according to the recommendations of the American Academy of Sleep Medicine.[6] The following signals were recorded: electroencephalogram (F3M2, F4M1, C3M2, C4M1, O1M2, O2M1), bilateral electro-oculogram, chin muscle electromyogram, oronasal thermistor, nasal pressure, rib cage and abdominal movement, electrocardiogram (lead 1), snoring, body position, bilateral anterior

tibialis electromyograms, and heart rate and oxygen saturation by pulse oximetry.

Using American Academy of Sleep Medicine 2012 scoring criteria,[4] PSG was scored manually with the aid of computer software by an experienced sleep technologist without knowledge of the results of the portable monitor recordings. Apneas were scored when there was ≥ 90% reduction in airflow from baseline for ≥ 10 seconds on the oronasal thermistor signal. The same criteria used to identify obstructive, central, and mixed apneas on the portable monitor recordings were used to score those events on PSG. Two separate PSG scorings were performed using different definitions for hypopnea: (1) events with ≥ 30% reduction in airflow from baseline for ≥ 10 seconds accompanied by ≥ 4% oxygen desaturation and (2) events with ≥ 30% reduction in airflow from baseline for ≥ 10 seconds associated with ≥ 3% reduction in oxygen saturation and/or an arousal. AHI on PSG was calculated as the average number of apneas and hypopneas per hour of sleep.

## Statistical Analysis

Continuous variables are summarized using means and SD and categorical variables using counts and percentages. Comparisons of respiratory parameters across the three monitoring methods (in-home Nox-T3, in-laboratory Nox-T3, and in-laboratory PSG) were compared using a repeated-measures analysis of variance, accounting for multiple observations per subject. To assess the level of agreement between the monitoring methods, we utilized paired $t$ tests and methods described by Bland and Altman.[7,8] Specifically, for a given metric, we first calculated the subject-specific difference for each pair of methods and tested whether this was significantly different from zero using paired $t$ tests. Next, for each pair of techniques, we examined the relationship between the subject-specific difference and the subject-specific average value using the two techniques. This relationship was evaluated graphically and statistically for bias, including examining the average subject-specific difference and associated limits of agreement (equal to the mean difference ± 2 SDs) and testing for significant correlation between the subject-specific difference and mean (eg, whether differences between techniques are larger/smaller for higher/lower average values). Primary agreement analyses compared the ≥ 4% AHI on PSG to that obtained from in-home and in-laboratory portable monitoring, separately. Secondary analyses repeated this comparison using the ≥ 3% rule. Using similar methods, we also examined the agreement between manual and automated AHI scoring within the in-home and in-laboratory monitors, separately.

Finally, we examined the diagnostic characteristics of home- or laboratory-based portable testing by calculating the sensitivity, specificity, positive predictive value, and negative predictive value at AHI thresholds of ≥ 5, ≥ 10, ≥ 15 and ≥ 30 events/h, using the results of the in-laboratory PSG as the reference standard. These analyses help in the understanding of the ability of home-based testing to accurately diagnose OSA when compared to PSG. Statistical analyses were performed using Stata/SE Version 14.1 (StataCorp LP, College Station, Texas, United States) and SAS Version 9.4 (SAS Institute Inc.,

**Table 1**—Comparison of respiratory parameters observed in PSG, simultaneous in-laboratory portable monitor recording and HSAT.

| | PSG | | Nox-T3_lab | | Nox-T3_home | | |
|---|---|---|---|---|---|---|---|
| | n | mean ± SD | n | mean ± SD | n | mean ± SD | *P* ‡ |
| **Total Time, min †** | 80 | 400.7 ± 65.8 | 76 | 467.3 ± 54.1 | 77 | 463.3 ± 58.4 | < .0001 |
| **Respiratory Indices (events/h)** | | | | | | | |
| AHI 3%* | 80 | 33.5 ± 23.2 | 76 | 32.6 ± 22.3 | 77 | 28.8 ± 20.4 | < .001 |
| AHI 4%** | 79 | 28.6 ± 23.9 | 76 | 28.0 ± 22.9 | 77 | 24.4 ± 20.8 | < .0001 |
| OAI | 80 | 19.8 ± 20.6 | 76 | 19.2 ± 19.2 | 77 | 17.3 ± 18.4 | .162 |
| CAI | 80 | 1.0 ± 1.6 | 76 | 1.2 ± 2.8 | 77 | 0.7 ± 1.4 | .173 |
| MAI | 80 | 2.7 ± 5.8 | 76 | 2.0 ± 4.2 | 77 | 1.6 ± 3.6 | .103 |
| HI 3%* | 80 | 9.9 ± 11.3 | 76 | 10.2 ± 7.8 | 77 | 9.7 ± 6.1 | .919 |
| HI 4%** | 79 | 4.7 ± 6.4 | 76 | 5.6 ± 5.6 | 77 | 4.9 ± 4.1 | .250 |
| **No. of Respiratory Events** | | | | | | | |
| Total apneas | 80 | 158.2 ± 163.7 | 76 | 176.2 ± 179.2 | 77 | 151.1 ± 165.3 | .146 |
| Obstructive apneas | 80 | 132.6 ± 141.4 | 76 | 151.7 ± 155.3 | 77 | 133.8 ± 147.0 | .203 |
| Central apneas | 80 | 7.0 ± 11.3 | 76 | 9.0 ± 21.5 | 77 | 5.3 ± 11.4 | .194 |
| Mixed apneas | 80 | 18.6 ± 39.4 | 76 | 15.6 ± 32.3 | 77 | 12.1 ± 29.0 | .341 |
| Hypopneas (3%)* | 80 | 67.4 ± 80.8 | 76 | 79.5 ± 60.2 | 77 | 74.0 ± 46.9 | .405 |
| Hypopneas (4%)** | 79 | 31.6 ± 44.0 | 76 | 43.6 ± 42.7 | 77 | 37.5 ± 31.6 | .018 |
| **Oxygen Saturation** | | | | | | | |
| Mean SpO_2, % | 78 | 94.2 ± 2.3 | 76 | 93.2 ± 2.2 | 77 | 93.4 ± 2.9 | < .0001 |
| Nadir SpO_2, % | 79 | 78.1 ± 12.8 | 76 | 77.1 ± 10.6 | 77 | 78.3 ± 11.1 | .0504 |
| % Time SpO_2 < 90% | 80 | 11.7 ± 16.8 | 76 | 10.5 ± 15.0 | 77 | 9.3 ± 16.6 | .009 |
| ODI 3%, events/h | 80 | 29.5 ± 24.6 | 76 | 31.6 ± 22.0 | 77 | 27.7 ± 20.2 | .006 |
| ODI 4%, events/h | 80 | 24.9 ± 24.1 | 76 | 24.3 ± 21.2 | 77 | 20.9 ± 19.9 | < .001 |

† = total sleep time for PSG or total analysis time for Nox-T3_lab and Nox-T3_home. ‡ = *P* value from mixed model analysis of variance comparing among methods. * = scoring of hypopneas on PSG required an associated oxygen desaturation event of ≥ 3% and/or an arousal and scoring of hypopneas on the portable monitor recordings required an associated oxygen desaturation event ≥ 3%. ** = scoring of hypopneas on all three types of sleep test required an associated oxygen desaturation event ≥ 4%. AHI = apnea-hypopnea index, CAI = central apnea index, HI = hypopnea index, HSAT = home sleep apnea testing, MAI = mixed apnea index, Nox-T3_home = home sleep apnea testing using the Nox-T3 device, Nox-T3_lab = in-laboratory portable monitor recording using the Nox-T3 device, OAI = obstructive apnea index, ODI = oxygen desaturation index, PSG = polysomnogram, SpO_2 = saturation of oxygen.

Cary, North Carolina, United States). A value of *P* < .05 was used to determine statistical significance.

## RESULTS

### Sample Characteristics
Eighty patients were evaluated. Participants were middle-aged (47.6 ± 14.0 years), overweight (BMI 27.5 ± 5.4 kg/m²) and predominantly male (77.5%). Patients were moderately sleepy, with an average Epworth Sleepiness Scale score of 10.1 ± 4.9. During the PSG, participants slept 6.7 ± 1.1 hours on average with a sleep efficiency of 82.4 ± 15.7% and arousal index of 27.8 ± 17.4 events/h. The mean percentages of stage N1, N2, N3, and rapid eye movement sleep were 19.2 ± 13.6, 52.5 ± 13.6, 10.1 ± 8.4 and 16.4 ± 6.7, respectively.

### Success Rate and Quality of HSAT
The initial HSAT fulfilled criteria for an acceptable study in 74 of the 79 participants (94.7%) who attempted the HSAT. The HSAT was unsuccessful in 5 of the 79 of the participants due to the absence or loss of the oximetry signal. Two

subjects refused to repeat the study, whereas the other three individuals successfully performed the HSAT on the second attempt. Based on the Noxturnal software rating of signal quality on the HSAT used in the analysis, a scoreable signal was present in 97.2 ± 7.9% of the recording for the nasal pressure signal, 99.7 ± 1.5% of the recording for the respiratory inductance plethysmography signals, and 99.0 ± 4.7% of the recording for the oximetry signal. The in-laboratory portable monitor recording was unsuccessful in 4 subjects due to the oximetry signal being absent or not able to be interpreted. The flow signal derived from the rib cage and abdominal respiratory inductance plethysmography signals was used to score portions of the recording or the entire recording in 25 of the home studies and 14 of the in-laboratory studies.

### Comparison of Respiratory Parameters Across Techniques
**Table 1** compares the results of manually edited scoring of PSG, in-laboratory portable monitor recording, and HSAT. As expected given the lack of sleep staging on portable monitors, total analysis time was more than 60 minutes longer on both the in-laboratory and in-home portable monitor

**Figure 1**—Comparison of manually edited AHI on PSG to HSAT and in-laboratory portable monitor recording.
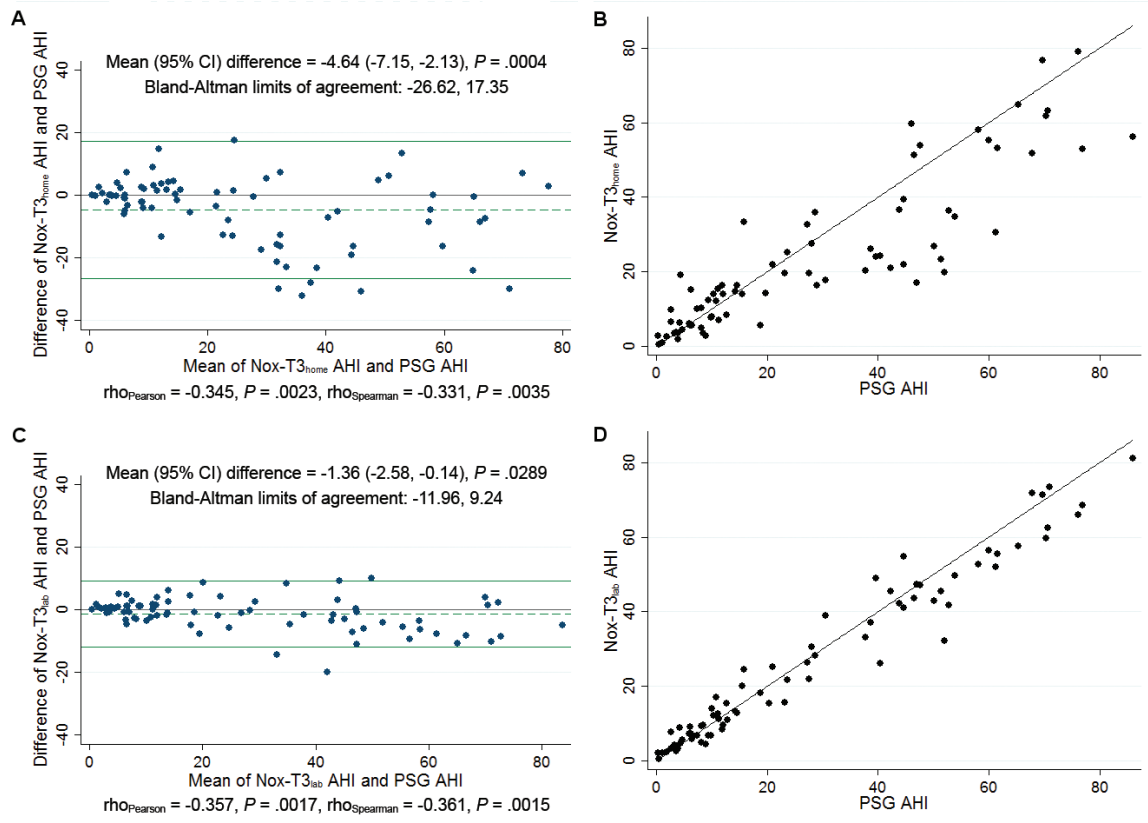


(A) Bland-Altman plot of manually edited AHI on PSG compared to HSAT (Nox-T3home). (B) Identity plot of manually edited AHI on PSG compared to HSAT (Nox-T3home). (C) Bland Altman plot of manually edited AHI on PSG compared to in-laboratory portable monitor recording (Nox-T3lab). (D) Identity plot of manually edited AHI on PSG compared to (Nox-T3lab). Scoring of hypopneas in all recordings required a ≥ 4% oxygen desaturation event. AHI = apnea-hypopnea index, CI = confidence interval, HSAT = home sleep apnea testing, Nox-T3home = home sleep apnea testing using the Nox-T3 device, Nox-T3lab = in-laboratory portable monitor recording using the Nox-T3 device, PSG = polysomnogram.

recordings when compared to total sleep time on the PSG ($P < .0001$).

When examining the number of events and per-hour indices, there were no differences among PSG, in-laboratory portable monitor recording, and HSAT in the total number of apneas and the number of specific types of apnea (ie, obstructive, central, mixed, and their respective indices). For hypopnea-specific measures, there was a difference in the total number of hypopneas using the 4% definition ($P = .018$), with more hypopneas on the in-laboratory Nox-T3 compared to PSG. However, there were no differences among PSG, in-laboratory portable monitor recording, and HSAT in the number of 3% hypopneas or either of the hypopnea indices.

Irrespective of the method used to score hypopneas and despite the different methods for calculating AHI on PSG versus type 3 portable monitor recording (total sleep time versus analysis time), the AHI on PSG and in-laboratory portable monitor recording was greater than that on HSAT (all pairwise values of $P \leq .002$ for PSG or in-laboratory Nox-T3 compared to HSAT). Positional differences may have accounted for this finding. However, no difference in percentage of time in the supine position was found on HSAT versus PSG, restricting analysis to those subjects with valid PSG positional data (n = 65; $P = .416$). The mean analysis time on the HSAT

and in-laboratory portable monitor recordings was similar (**Table 1**) and therefore did not explain the difference in AHI between those two recordings.

Although statistically significant differences in oxygen desaturation severity measures were observed among the three methods (**Table 1**), the magnitudes of the differences were relatively small and unlikely to be of clinical significance. Mean oxygen saturation was lower on both the in-laboratory portable monitor recording ($P < .001$) and HSAT ($P < .001$) compared to PSG. There was also a difference in percent time oxygen saturation < 90%, with significantly lower values on HSAT compared to PSG ($P = .002$). The oxygen desaturation index for events ≥ 4% was lower on HSAT compared to both the PSG ($P < .001$) and in-laboratory portable monitor ($P = .001$). The oxygen desaturation index with events ≥ 3% was significantly lower on HSAT than in-laboratory Nox-T3 ($P = .001$), although neither HSAT ($P = .121$) nor in-laboratory Nox-T3 ($P = .091$) were statistically different from the PSG.

## Agreement Between Monitoring Methods
Bland-Altman and identity plots comparing AHI on PSG with that on HSAT and in-laboratory portable recording are shown in **Figure 1**. The Bland-Altman analysis of AHI on PSG versus HSAT showed a mean difference of −4.6 (95% confidence

**Table 2**—Prevalence, sensitivity, specificity, and positive and negative predictive value for different cutoffs of manually edited AHI* from the HSAT and in-laboratory portable monitor recording versus the PSG.

| | Prevalence | Sensitivity | Exact 95% CI LB | Exact 95% CI UB | Specificity | Exact 95% CI LB | Exact 95% CI UB | PPV | NPV |
|---|---|---|---|---|---|---|---|---|---|
| | | | **Nox-T3_home versus PSG** | | | | | | |
| **AHI, events/h** | | | | | | | | | |
| ≥ 5 | 0.83 | 0.95 | 0.87 | 0.99 | 0.69 | 0.39 | 0.91 | 0.94 | 0.75 |
| ≥ 10 | 0.68 | 0.92 | 0.81 | 0.98 | 0.79 | 0.58 | 0.93 | 0.91 | 0.83 |
| ≥ 15 | 0.55 | 0.93 | 0.81 | 0.99 | 0.85 | 0.69 | 0.95 | 0.89 | 0.91 |
| ≥ 30 | 0.39 | 0.63 | 0.44 | 0.80 | 0.93 | 0.82 | 0.99 | 0.86 | 0.80 |
| | | | **Nox-T3_lab versus PSG** | | | | | | |
| **AHI, events/h** | | | | | | | | | |
| ≥ 5 | 0.84 | 0.97 | 0.89 | 1.00 | 0.75 | 0.43 | 0.95 | 0.95 | 0.82 |
| ≥ 10 | 0.69 | 0.96 | 0.87 | 1.00 | 1.00 | 0.85 | 1.00 | 1.00 | 0.92 |
| ≥ 15 | 0.55 | 1.00 | 0.91 | 1.00 | 0.94 | 0.80 | 0.99 | 0.95 | 1.00 |
| ≥ 30 | 0.40 | 0.97 | 0.83 | 1.00 | 0.98 | 0.88 | 1.00 | 0.97 | 0.98 |

* = scoring of hypopneas on the Nox-T3 recordings required a ≥ 4% oxygen desaturation event; hypopneas on the PSG required a ≥ 4% oxygen desaturation event and/or an arousal. AHI = apnea-hypopnea index, CI = confidence interval, HSAT = home sleep apnea testing, LB = lower bound, Nox-T3_home = home sleep apnea testing using the Nox-T3 device, Nox-T3_lab = in-laboratory portable monitor recording using the Nox-T3 device, NPV = negative predictive value, PPV = positive predictive value, PSG = polysomnogram, UB = upper bound.

**Figure 2**—Percentage of patients falling into clinical OSA groupings.
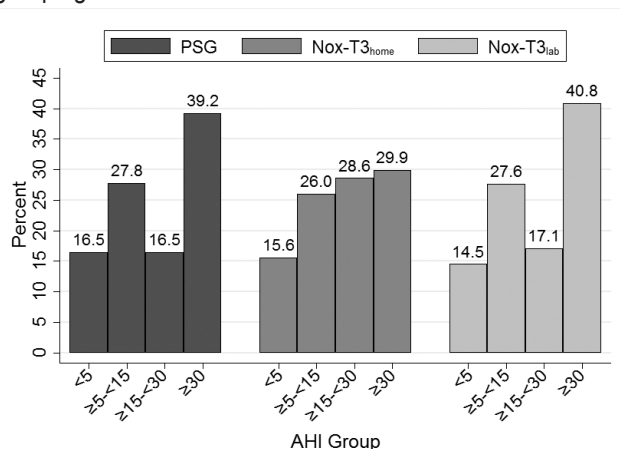


This graph illustrates the percentage of patients falling into clinical OSA groupings of none (AHI < 5 events/h), mild (5 ≤ AHI < 15), moderate (15 ≤ AHI < 30), and severe (AHI ≥ 30) based on PSG, HSAT (Nox-T3_home), and in-laboratory portable monitor recording (Nox-T3_lab). Scoring of hypopneas in all recordings required a ≥ 4% oxygen desaturation event. AHI = apnea-hypopnea index, HSAT = home sleep apnea testing, Nox-T3_home = home sleep apnea testing using the Nox-T3 device, Nox-T3_lab = in-laboratory portable monitor recording using the Nox-T3 device, OSA = obstructive sleep apnea, PSG = polysomnogram.
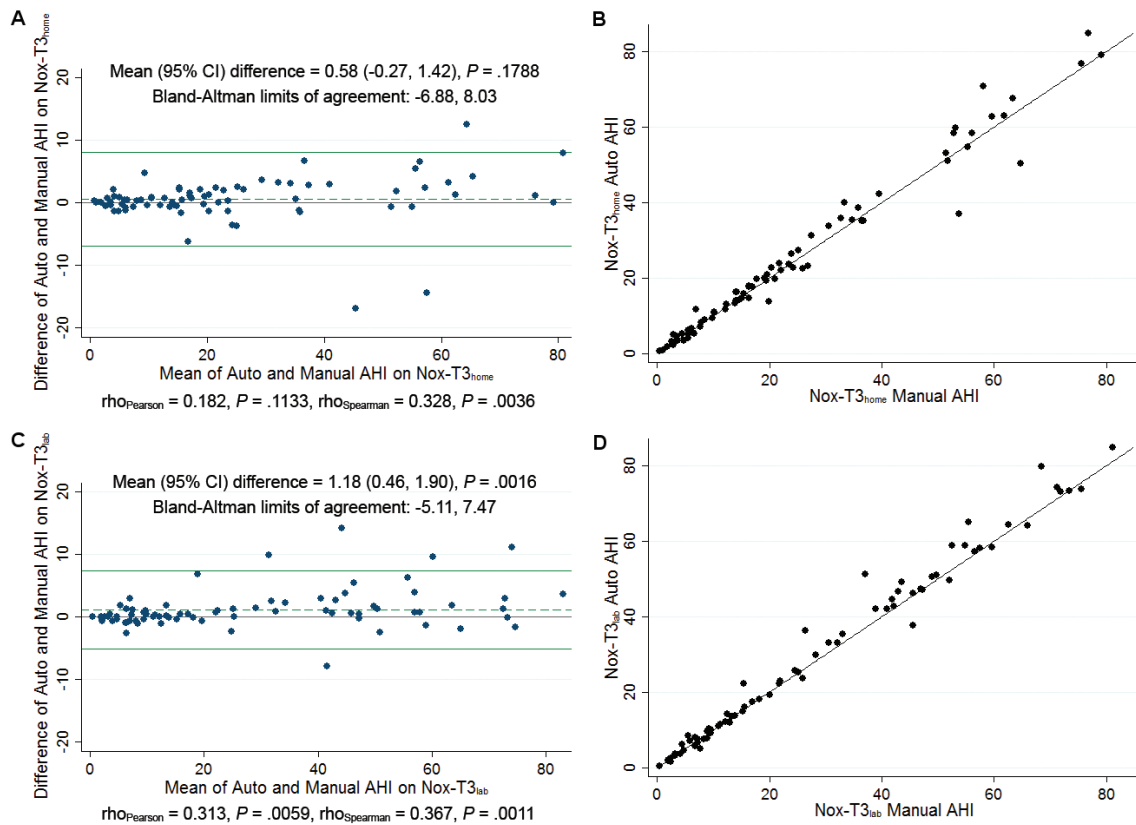
interval [CI]: −7.2, −2.1; $P < .001$) with limits of agreement ranging from −26.6 to 17.4 events/h. In contrast, in the Bland-Altman plot of AHI on PSG versus in-laboratory portable monitor recording, the mean difference was only −1.4 (95% CI: −2.6, −0.1; $P = .029$) with narrower limits of agreement of −12.0 to 9.2 events/h. We note that in both cases, there was evidence for a significant negative correlation between the difference and mean, suggesting that at higher AHI values the

portable monitors result in larger underestimates of the PSG AHI. The squared correlation coefficient ($R^2$) for AHI was 0.79 on the PSG versus HSAT identity plot and 0.96 on the PSG versus in-laboratory portable monitor recording identity plot, suggesting an extremely high amount of shared variability. Similar results were observed when manually edited scoring of hypopneas on portable monitor required a ≥ 3% oxygen desaturation event and hypopneas on PSG required a ≥ 3% oxygen desaturation event and/or an arousal (**Figure S1**, supplemental material). The closer relationship between PSG and simultaneous in-laboratory portable monitor recording than between PSG and HSAT supports the importance of differences in environment and night-to-night variability on sleep test results.

**Table 2** compares the diagnostic characteristics for different cutoffs of manually edited AHI from the HSAT and in-laboratory portable monitor recording compared to reference-standard PSG when hypopneas in all recordings required ≥ 4% oxygen desaturation. Using a threshold of AHI ≥ 5 events/h, the HSAT had 95% sensitivity, 69% specificity, 94% positive predictive value, and 75% negative predictive value. Similar results for HSAT were observed at AHI cutoffs of ≥ 10 and ≥ 15 events/h, with specificity increasing to 85% with only a small decrease in sensitivity (93%) at an AHI cutoff ≥ 15 events/h. The in-laboratory portable monitor recording had similar or better sensitivities, specificities, and predictive values at these thresholds. Similar results for both HSAT and in-laboratory portable monitor recording were also observed when hypopneas on portable monitor required a ≥ 3% oxygen desaturation event and hypopneas on PSG required a ≥ 3% oxygen desaturation event and/or an arousal (**Table S1**, supplemental material).

**Figure 2** illustrates the percentage of subjects with no OSA and mild, moderate, and severe OSA based on the AHI on PSG, HSAT, and in-laboratory portable monitor recording. As

**Figure 3**—Comparison of automatically scored AHI to manually edited AHI on HSAT and in-laboratory portable monitor recording.



**(A)** Bland-Altman plot comparing automatically scored AHI to manually edited AHI on HSAT (Nox-T3$_{home}$). **(B)** Identity plot comparing automatically scored AHI to manually edited AHI on HSAT (Nox-T3$_{home}$). **(C)** Bland Altman plot comparing automatically scored AHI to in-laboratory portable monitor recording (Nox-T3$_{lab}$). **(D)** Identity plot comparing automatically scored AHI to in-laboratory portable monitor recording (Nox-T3$_{lab}$). Scoring of hypopneas in all recordings required a ≥ 4% oxygen desaturation event. AHI = apnea-hypopnea index, CI = confidence interval, HSAT = home sleep apnea testing, Nox-T3$_{home}$ = home sleep apnea testing using the Nox-T3 device, Nox-T3$_{lab}$ = in-laboratory portable monitor recording using the Nox-T3 device.

suggested given the strong agreement between the PSG and simultaneous in-laboratory Nox-T3 for the AHI using the 4% hypopnea criteria, the proportions within each of the clinical groupings were similar between these two techniques. There were similar percentages in the non-OSA and mild OSA groups across the three recordings, but a higher proportion of participants fell within the moderate group and a lower proportion in the severe group on the HSAT compared to PSG or in-laboratory Nox-T3.

AHI on automatic scoring of the portable monitor recording had good agreement with manually edited scoring (**Figure 3**). On the HSAT, the automatic scoring resulted in slightly higher AHI estimates compared to manual scoring, with a nonsignificant mean difference of 0.6 (95% CI: −0.3, 1.4; $P = .179$) and associated limits of agreement from −6.9 to 8.0 events/h. For the in-laboratory portable monitoring, there was a slightly larger difference between the automatic manual scoring [mean difference 1.2 (0.5, 1.9); $P = .002$]; although statistically significant, this difference was not clinically meaningful. The limits of agreement ranged from −5.1 to 7.5 events/h, which is narrower than that observed for HSAT. For both monitoring methods, there was evidence of a positive correlation between

the difference and the mean, suggesting larger overestimates of AHI on automatic scoring compared to manual scoring for higher AHI values. The $R^2$ between automatic and manually scored AHI in identity plots was 0.97 for HSAT and 0.98 for in-laboratory portable monitoring, reflecting near-perfect correlation. In addition to the AHI, there were no clinically meaningful differences in the number of scored apneas between the automatic and manual scoring (**Table S2**, supplemental material). On HSAT, the automatic scoring underestimated total apneas by 6.7 events on average ($P = .053$) and tended to underestimate the number of mixed apneas by 3.9 events ($P = .051$). For the in-laboratory portable monitoring, the automatic scoring resulted in 2.7 more estimated central apneas ($P = .051$) and 2.9 fewer mixed apneas ($P = .002$) compared to manual scoring, but there was no difference in the total number of apneas scored.

## DISCUSSION

The results demonstrate a close agreement between PSG and Nox-T3 portable monitor recording in Chinese adults. When

comparing the simultaneous portable monitor recordings and PSGs, we found a high sensitivity and specificity, with limits of agreement on Bland-Altman analysis of −11.96 and 9.24. In contrast, the same comparison in the smaller study of Cairns et al.[2] reported limits of agreement of −19.6 and 15.0. The current study further extends the findings of Cairns et al.[2] by comparing the Nox-T3 portable monitor recording at home with the in-laboratory PSG. The relatively wide limits of agreement on the Bland-Altman analysis of AHI on PSG versus HSAT (−26.6 to 17.4 events/h) were due to differences between the two testing methods at higher levels of AHI. Despite the known variability of sleep test results on night-to-night testing, the HSAT and PSG results were in good agreement in the range of AHI that would be used to diagnose OSA. Using a cutoff AHI $\geq 5$ events/h, 84.4% of subjects received a diagnosis of OSA on HSAT and 83.5% on PSG. At a cutoff AHI $\geq 15$ events/h, 58.5% of subjects received a diagnosis of OSA on HSAT and 55.7% on PSG. The close agreements in our study are of particular note, because the respiratory indices on both the in-laboratory portable monitor recording and HSAT were calculated using the edited recording time rather than the total sleep time used for the PSG indices.

The current study is also of importance in validating the use of HSAT in Chinese adults. Based on epidemiologic evidence, it is estimated that 60 million Chinese have OSA with symptoms, ie, AHI $\geq 5$ and excessive daytime sleepiness.[9] Many of these individuals have undiagnosed OSA and therefore are untreated. The high prevalence in China is attributed in part to the craniofacial characteristics of Asians and is likely to increase due to the increasing obesity of Chinese children and adults, particularly in urban areas of the country.[3,10] Currently, PSG is used almost exclusively in China to diagnose OSA. Greater reliance on HSAT would help to address the unmet need. Our results should lead to greater acceptance of HSAT by Chinese physicians, thereby improving patient access to diagnosis and treatment.

To our knowledge, this is the first study to validate the use of HSAT in Chinese adults. Previous studies compared in-laboratory portable monitor recordings with PSG in Chinese adults,[11–14] but HSAT was not performed in those studies. Three of the studies performed in-laboratory portable monitor testing concurrent with PSG[12–14] and the other study performed in-laboratory testing on different days. Levels of agreement were variable and may in part have been due to the different portable monitors being evaluated. The study by Ng et al.[13] used the Embletta portable monitor (Natus Medical Incorporated, Pleasanton, California, United States), which records the same signals as the Nox-T3. That study reported sensitivities, specificities, positive predictive values, and negative predictive values at different AHI cutoffs on the in-laboratory recordings that are similar to our current findings using the Nox-T3 monitor. Limits of agreement on the Bland-Altman analyses were also comparable. The current study extends the findings of these previous studies in Chinese adults by evaluating portable monitor during testing at home, the condition of its intended use. During home testing, the patient is required to attach the sensors and perform the recording without supervision. The high level of agreement of our HSAT results with in-laboratory PSG

and the low failure rate of 6.3% on initial home testing, which decreased to 3.8% on repeat testing, demonstrate the ability to use the Nox-T3 portable monitor for HSAT to diagnose OSA. It is our impression that having the participants demonstrate their ability to self-apply the sensors prior to taking the monitor home was a major factor in our excellent success rate.

Based on associations of BMI with risk factors for cardiovascular disease, the working Group in Obesity in China has recommended that Chinese with a BMI $\geq 28$ kg/m² should be considered obese.[15,16] The mean BMI of our subjects was $27.5 \pm 5.4$ kg/m². Most previous studies validating type 3 portable monitors were performed in subjects with higher BMI. This is of particular importance when the criteria used to score hypopneas require the presence of an oxygen desaturation event and/or arousal. Although many portable sleep monitors, like the Nox-T3, have the capability to record an electroencephalogram signal that would allow detection of arousals, bipolar signals are generally not recorded during HSAT due to the difficulty of self-application. The inability to detect arousals on HSAT could result in a lower AHI on HSAT than on PSG. The likelihood of this discrepancy would be reduced in obese patients, because a respiratory event associated with an arousal is more likely to also be associated with an oxygen desaturation event due to the lower arterial oxygen tension associated with obesity. Our study demonstrates that HSAT is also of utility when used to diagnose OSA in less obese patients. In our Chinese adults without major comorbidities, good agreement between HSAT and in-laboratory PSG were also observed when hypopneas on portable monitor required a $\geq 3\%$ oxygen desaturation event and hypopneas on PSG required a $\geq 3\%$ oxygen desaturation event and/or an arousal (**Table S1**, supplemental material).

As in the study by Cairns et al.,[2] we found close agreement between the automatic and manually edited scoring of the Nox-T3 recording. The mean difference between automatic versus manual edited scoring on both home and in-laboratory portable monitor recordings was less than 2 events/h (**Figure 3**). This close agreement increases testing efficiency by reducing the amount of time needed to edit the automatic score. Although current guidelines for HSAT strongly recommend manually edited scoring,[4] the close agreement between automatic and manually edited scoring could potentially allow practitioners without sufficient resources for manual editing to rely on the automatic score for clinical management.

The strengths of our study are the evaluation of the portable monitor during HSAT as well as during in-laboratory PSG, and assessment of the monitor's performance in Chinese adults. However, our study was limited to the direct comparison of portable monitor testing and PSG. Given its limited objective, our study did not evaluate the use of HSAT within a complete ambulatory pathway for the diagnosis and treatment of patients with OSA. This study is the first step in that process and demonstrates the feasibility of HSAT to diagnose OSA in Chinese adults. Additional studies are needed that build on these findings. We need to assess how to select patients for out-of-laboratory management, the differences in diagnostic rates beyond consideration of AHI, and the effect of this process on treatment outcomes and adherence. Finally, the cost effectiveness

of a complete ambulatory management pathway in Chinese adults needs to be evaluated. In this study, HSAT was followed by portable monitor recording during simultaneous in-laboratory PSG. The simultaneous comparison helped to control for differences in environment and known night-to-night variability in sleep study results. The study design would have been enhanced if two PSG and two HSAT had been performed to assess night-to-night variability of results.

In summary, this study validates the use of the Nox-T3 monitor on HSAT to diagnose OSA in Chinese adults referred to a sleep center. Close agreement was present between PSG and HSAT, and even closer agreement was observed on the simultaneous in-laboratory PSG and portable monitor recording. Close agreement was also present between automatic software scoring and manually edited scoring. The results should lead to greater acceptance and use of HSAT by Chinese physicians. Additional studies are needed to assess the performance of the monitor in a community-based Chinese population and the use of HSAT in ambulatory clinical pathways that allow diagnosis and treatment in Chinese patients without in-laboratory testing.

## ABBREVIATIONS

AHI, apnea-hypopnea index
BMI, body mass index
CAI, central apnea index
CI, confidence interval
HI, hypopnea index
HSAT, home sleep apnea testing
LB, lower bound
Nox-T3$_{home}$, home sleep apnea testing using the Nox-T3 device
Nox-T3$_{lab}$, in-laboratory portable monitor recording using the Nox-T3 device
NPV, negative predictive value
OAI, obstructive apnea index
ODI, oxygen desaturation index
OSA, obstructive sleep apnea
PPV, positive predictive value
PSG, polysomnogram
SD, standard deviation
SpO$_2$, saturation of oxygen
UB, upper bound

## REFERENCES

1. Collop NA, Anderson WM, Boehlecke B, et al. Clinical guidelines for the use of unattended portable monitors in the diagnosis of obstructive sleep apnea in adult patients. Portable Monitoring Task Force of the American Academy of Sleep Medicine. *J Clin Sleep Med*. 2007;3(7):737–747.

2. Cairns A, Wickwire E, Schaefer E, Nyanjom D. A pilot validation study for the NOX T3(TM) portable monitor for the detection of OSA. *Sleep Breath*. 2014;18(3):609–614.

3. Li KK, Kushida C, Powell NB, Riley RW, Guilleminault C. Obstructive sleep apnea syndrome: a comparison between Far-East Asian and white men. *Laryngoscope*. 2000;110(10 Pt 1):1689–1693.

4. Berry RB, Brooks R, Gamaldo CE, 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 2.0. Darien, IL: American Academy of Sleep Medicine; 2012.

5. Eberhard A, Calabrese P, Baconnier P, Benchetrit G. Comparison between the respiratory inductance plethysmography signal derivative and the airflow signal. *Adv Exp Med Biol*. 2001;499:489–494.

6. Kushida CA, Littner MR, Morgenthaler T, et al. Practice parameters for indications for polysomnography and related procedures: an update for 2005. *Sleep*. 2005;28(4):499–521.

7. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1(8476):307–310.

8. Bland JM, Altman DG. Comparing methods of measurement: why plotting difference against standard method is misleading. *Lancet*. 1995;346(8982):1085–1087.

9. Ip MS, Lam B, Lauder IJ, et al. A community study of sleep-disordered breathing in middle-aged Chinese men in Hong Kong. *Chest*. 2001;119(1):62–69.

10. Wu Y. Overweight and obesity in China. *BMJ*. 2006;333(7564):362–363.

11. Choi JH, Kim EJ, Kim YS, et al. Validation study of portable device for the diagnosis of obstructive sleep apnea according to the new AASM scoring criteria: Watch-PAT 100. *Acta Oto-laryngol*. 2010;130(7):838–843.

12. Gantner D, Ge JY, Li LH, et al. Diagnostic accuracy of a questionnaire and simple home monitoring device in detecting obstructive sleep apnoea in a Chinese population at high cardiovascular risk. *Respirology*. 2010;15(6):952–960.

13. Ng SS, Chan TO, To KW, et al. Validation of Embletta portable diagnostic system for identifying patients with suspected obstructive sleep apnoea syndrome (OSAS). *Respirology*. 2010;15(2):336–342.

14. To KW, Chan WC, Chan TO, et al. Validation study of a portable monitoring device for identifying OSA in a symptomatic patient population. *Respirology*. 2009;14(2):270–275.

15. Zhou BF; for the Cooperative meta-analysis group of the Working Group on Obesity in China. Effect of body mass index on all-cause mortality and incidence of cardiovascular diseases-report for meta-analysis of prospective studies on optimal cut-off points of body mass index in Chinese adults. *Biomed Environ Sci*. 2002;15(3):245–252.

16. Zhou BF; for the Cooperative meta-analysis group of the Working Group on Obesity in China. Predictive value of body mass index and waist circumference for risk factors of certain related diseases in Chinese adults - study on optimal cut-off points of body mass index and waist circumference in Chinese adults. *Biomed Environ Sci*. 2002;15(3):83–95.

## SUBMISSION & CORRESPONDENCE INFORMATION

## DISCLOSURE STATEMENT