



Phenotyping REM OSA by means of peripheral arterial tone-based home sleep apnea testing and polysomnography: a critical assessment the sensitivity and specificity of both methods

Journal:	<i>Journal of Sleep Research</i>
Manuscript ID	JOSR-21-409
Manuscript Type:	Research Article
Date Submitted by the Author:	21-Jun-2021
Complete List of Authors:	Massie, Frederik; University of Oxford, Department of Engineering Van Pee, Bart; University of Oxford, Department of Engineering Vits, Steven; University of Antwerp Verbraecken, Johan; Antwerp University Hospital, Pulmonary Medicine Bergmann, Jeroen; University of Oxford, Department of Engineering
Keywords:	inter-rater variability, multi-night testing

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Title Page

Title:

Phenotyping REM OSA by means of peripheral arterial tone-based home sleep apnea testing and polysomnography: a critical assessment the sensitivity and specificity of both methods

Short Title:

Assessment of the accuracy of REM OSA phenotyping

Authors:

1. Frederik Massie, MSc, Natural Interaction Lab, Thom Building,
Department of Engineering, University of Oxford, Oxford, United Kingdom
2. Bart Van Pee, MSc, Natural Interaction Lab, Thom Building,
Department of Engineering, University of Oxford, Oxford, United Kingdom
3. Steven Vits, MSc, Faculty of Medicine and Health Sciences, University of Antwerp,
Wilrijk, Belgium
4. Johan Verbraecken, MD, PhD, Multidisciplinary Sleep Disorder Center, Antwerp
University Hospital, Edegem, Antwerp, Belgium, and Faculty of Medicine and Health
Sciences, University of Antwerp, Wilrijk, Belgium
5. Jeroen Bergmann, MSc, PhD, Natural Interaction Lab, Thom Building,
Department of Engineering, University of Oxford, Oxford, United Kingdom

Corresponding Author:

Name: Frederik Massie

Address: Natural Interaction Lab, Department of Engineering, Thom Building, University of
Oxford. Parks Road, Oxford OX1 3PJ, United Kingdom

Email: frederik.massie@eng.ox.ac.uk

Telephone: +32 474942710

Word count:

4587

Reference count:

36

Disclosures:

Frederik Massie, Bart Van Pee, and Steven Vits are shareholders of and employed by
Ectosense, the manufacturer of the NightOwl device.

Author contributorship:

All authors contributed to the study conception and design. Data analysis was performed by
FM, BVP, and SV. The first draft of the manuscript was written by FM, BVP, and SV. All
authors read and approved the final manuscript.

Data availability statement:

The data that support the findings of this study are available from the corresponding author
upon reasonable request.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Abstract

The clinical relevance of rapid eye movement sleep-related obstructive sleep apnea (REM OSA) is supported by its associated adverse health outcomes and impact on optimal treatment strategies. To date, no assessment of REM OSA phenotyping performance has been conducted for any type of sleep testing technology. The objective of this study was to assess this for polysomnography (PSG) and a peripheral arterial tone-based home sleep apnea test (PAT HSAT).

In a dataset comprising 261 participants, the sensitivity and specificity of the agreement on REM OSA phenotyping was assessed for two independent scorings of PSG and synchronously administered PAT HSAT.

The sensitivity and specificity of REM OSA phenotyping were 0.87 and 0.89 respectively for the PSG inter-scorer comparison, and 0.68 and 0.97 for PAT HSAT on a single-night basis, using the conventional minimum required REM sleep time of 30 minutes.

PSG-based REM OSA phenotyping was found to be sensitive and specific even for a single-night testing protocol. PAT-based REM OSA phenotyping showed a lower sensitivity but a slightly higher specificity compared to PSG. In order to increase performance and conclusiveness of PAT-based REM OSA phenotyping, a multi-night protocol of 2 to 5 nights could be considered.

Finally, the minimum required REM sleep time could be lowered from the conventional 30 minutes to 15 minutes without significantly lowering REM OSA phenotyping sensitivity and specificity while increasing the level of phenotyping conclusiveness.

Keywords

inter-rater variability, multi-night testing,

Introduction

It has been well studied that for a significant subgroup of people with obstructive sleep apnea (OSA), the condition is more severe during rapid eye movement (REM) sleep (Varga & Mokhlesi, 2019). Indeed, during REM sleep, muscle tone decreases, which increases the propensity for the tissues surrounding the upper airway to collapse (Zhou et al., 2020). Hypoxic and hypercapnic ventilatory drive also tend to reduce during REM sleep (N. J. Douglas, White, Weil, Pickett, & Zwillich, 1982; Neil J. Douglas, 2010). These phenomena may result in more frequent, or more extended respiratory events, and deeper oxygen desaturations associated with these events (Varga & Mokhlesi, 2019). The sympathetic discharge near the end-stage of respiratory events is exacerbated during REM sleep (Somers, Dyken, Clary, & Abboud, 1995), which causes repetitive surges in blood pressure and heart rate (Schächinger, Weinbacher, Kiss, Ritz, & Langewitz, 2001). The increased severity of OSA during REM sleep likely contributes to an increased risk for and severity of the well-studied OSA consequences, such as endothelial dysfunction (Budhiraja, Parthasarathy, & Quan, 2007), systemic hypertension (Dean et al., 2015), insulin resistance (Budhiraja & Quan, 2005), atherosclerosis (Lévy et al., 2009), adverse neurocognitive outcomes (Stansbury & Strollo, 2015), and cardiovascular mortality (Budhiraja & Quan, 2005). The clinical relevance of identifying REM-related OSA (REM OSA) as a distinct OSA phenotype is supported not only by its adverse health outcomes, but also by its impact on optimal treatment strategies. It is well known that the majority of REM sleep epochs tend to occur during the last part of patient's sleep. Once treated with continuous positive airway pressure (CPAP), patients with increased severity of OSA during REM sleep could benefit from ensuring that the CPAP is also active throughout the final hours before awakening (Alzoubaidi & Mokhlesi, 2017). Since many patients cannot achieve such high levels of

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

CPAP adherence, it is imperative to explore alternative treatment strategies or even combine less effective therapeutic approaches (e.g., oral appliance plus nasal EPAP) in order to achieve clinical efficacy in lowering the REM apnea-hypopnea index (REM AHI) (Varga & Mokhlesi, 2019). Knowledge of a patient’s REM OSA phenotype and its adverse impact on comorbidities might equally influence treatment strategies of these comorbidities.

To date, no consensus has been established on the clinical criterion to phenotype REM OSA (Mokhlesi & Punjabi, 2012). Nevertheless, most papers use the criterion set out below, which comprises three independent requirements. The first requirement is that the REM AHI has to be at least two times higher than the AHI during non-REM sleep stages (NREM AHI) (Oksenberg, Arons, Nasser, Vander, & Radwan, 2010). Second, to ensure REM OSA is only labeled when there is significant OSA, and to ensure that the results are numerically stable, the overall AHI has to be larger than 5 (Alzoubaidi & Mokhlesi, 2017; Gabryelska & Białasiewicz, 2020; Haba-Rubio, Janssens, Rochat, & Sforza, 2005; Su, Liu, Panjapornpon, Andrews, & Foldvary-Schaefer, 2012). Third, the phenotyping is rendered inconclusive for participants for which the total REM sleep time is less than 30 minutes (Duce, Kulkas, Langton, Töyräs, & Hukins, 2018; Nisha Aurora, Crainiceanu, Gottlieb, Kim, & Punjabi, 2018) since short REM sleep time might lead to imprecise results (Mokhlesi & Punjabi, 2012) or an overestimation of the clinical effects of REM OSA (Mokhlesi et al., 2014). The three criteria combined can be formulated as:

$$\text{REM AHI} / \text{NREM AHI} \geq 2 \text{ and AHI} > 5 \text{ and REM sleep time} > 30 \text{ minutes, else}$$

phenotyping is inconclusive.

Duce et al. (2018) criticized the minimum REM sleep time requirement, finding that employing a minimum required REM sleep time risks underdiagnosing REM OSA. The first-night effect might cause an underestimation of REM sleep time during a single-night sleep

test, further adding to this risk (Duce et al., 2018). Regardless, fixing this cutoff at exactly 30 minutes appears to be a rule of thumb without support of a statistical analysis. In this study we therefore assessed what minimum REM sleep time is required to obtain an adequately sensitive and specific REM OSA phenotyping.

The value of OSA phenotyping for the development of personalized OSA management has been well recognized (Zinchuk & Yaggi, 2020), and the body of research on PSG-based OSA phenotyping is rapidly growing (Finnsson et al., 2020). Coinciding with the prominent role of the PSG for such phenotyping, there is an increased utilization of reduced-channel home sleep apnea testing (HSAT) (Chiao & Durr, 2017) – a trend fueled by the impact of COVID-19 on the admittance of patients in the sleep lab (Johnson, Sullivan, Nti, Rastegar, & Indira Gurubhagavatula, 2020). These reduced channel HSATs have more restrictive capabilities in determining OSA phenotypes, such as REM OSA, primarily due to their lack of an electroencephalogram (EEG) for sleep recording. Despite the absence of EEG, the American Academy of Sleep Medicine (AASM) Scoring Manual reports one particular category of HSAT devices that is required to demonstrate REM sleep, namely the peripheral arterial tonometry (PAT) -based devices (Berry et al., 2020). These devices monitor changes in peripheral arterial tone related to sympathetic surges, alongside surges in pulse rate and oxygen desaturations (Berry et al., 2020), and do so through the registration of a conditioned photoplethysmography (PPG) signal. As sympathetic tone increases during REM sleep, its manifestations on the peripheral arterial tone and the pulse rate can be used to identify episodes of REM sleep. To date, two PAT-based HSAT devices have been cleared by the FDA: the WatchPAT (Itamar Medical, Israel) (Ioachimescu et al., 2020) and the NightOwl (Ectosense, Belgium) (Massie et al., 2018), the latter of which was used in this study. For both devices there currently exist no studies reporting on their ability to accurately determine

the REM OSA according to any conventional definition, hence this was evaluated in this study.

To fairly assess the ability of a HSAT to estimate a clinical parameter, it is important to compare its performance against the inter-scorer agreement of the PSG for the same parameter. While significant inter-scorer variability has been demonstrated for REM sleep time (Malhotra et al., 2013), no similar exercise has been performed to date for REM OSA phenotyping.

Therefore, the primary aim of our study was to compare the REM OSA phenotyping performance of PAT HSAT to PSG.

The results of the study may help to provide insight into whether the trend towards lower complexity sleep testing is reconcilable with an increased understanding of the importance of REM OSA phenotyping.

Methods

Dataset

The dataset analyzed in this manuscript comprises 261 participants with suspicion of OSA, which were prospectively recruited in a consecutive cohort across four different centers, namely three in the USA (where all centers were part of the United Health Services Group in Miami, Florida), and one located in Belgium (Ziekenhuis Oost Limburg, ZOL, Genk). All participants were scheduled for one overnight in-lab PSG. The USA branch of the study was approved by Aspire IRB, part of the WIRB-Copernicus Group. The Belgian branch of the study was approved by the Ethics Committee of ZOL. Underaged or mentally disabled participants were excluded from participating in the study. For all participants, gender, age, and BMI were recorded. Persons of diverse racial and ethnic backgrounds were included.

Protocol and devices

A graphical representation of the study setup is provided in Figure 1. Routine PSG was performed for all study participants. Qualified lab technicians at each participating study center were responsible for setting up the equipment and capturing PSG data. During setup of PSG, the PAT HSAT (NightOwl) was attached to the middle finger of the hand to which the pulse oximeter of PSG was applied. All PSG data was double-scored by two independent centers which were blinded from one-another's analysis.

[SUGGESTED INSERTION FIGURE 1]

Polysomnography

For the European and USA center respectively, the Alice 6 PSG (Philips Respironics, USA) and Cadwell Easy PSG (Cadwell, USA) were used.

Each PSG was scored by two independent scoring centers. The first scoring was performed by the team of sleep technicians of the center where the patient was admitted (further referred to as **Local Analysis**).

Another independent scoring was performed by scorers of Cerebra Medical (Canada). This company provides computer-aided sleep scoring services to support PSG scoring for clinical centers and clinical trials. The studies were first analyzed by their proprietary Michele Sleep Scoring System (MSSS) and were subsequently complemented with complete manual rescored by an expert technologist.

Malhotra et al. (2013) confirmed in a multi-centric trial that the MSSS, complemented with manual editing by an expert scorer, is more robust than the results of a single expert scorer.

Because of these conclusions, Cerebra Medical's analysis served as the expert benchmark (**Expert Analysis**) to which the Local Analysis and PAT HSAT's analysis were compared.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

All PSG data were scored according to the latest AASM scoring rules using the recommended 1A (3%) rule for hypopnea scoring (Berry et al., 2020).

PAT HSAT and its principle of operation

NightOwl, similar to WatchPAT, is built around a fingertip mounted PPG probe. PPG operates on simple optical technology to detect blood volume changes in the tissue microvascular bed. From the PPG measurement, the arterial blood oxygen saturation (SpO₂), pulse rate (PR), and peripheral arterial tone (PAT) are derived. The latter signal reflects changes in the blood volume due to peripheral artery constriction following an increase in sympathetic tone at the end-stage of a respiratory event. In line with the AASM’s definition of PAT technology, the combination of a decrease in PAT, a fall in SpO₂, and an increase in PR is used to detect respiratory events (Berry et al., 2020). The probe also contains an accelerometer for the detection of limb movement to determine sleep/wake based on actigraphy.

REM sleep time estimation based on PAT

Bursts of rapid eye movements have been related to phasic vasoconstrictions throughout REM episodes (Herscovici, Pe’er, Papyan, & Lavie, 2007). The temporal pattern of these vasoconstrictions as detected by PAT were found to be different from that observed in NREM sleep (Dvir, Adler, Freimark, & Lavie, 2002). Furthermore, PR low frequency power has been shown to increase (Chouchou & Desseilles, 2014), and oxygen desaturation dynamics are different in REM sleep than NREM sleep (Choi, Park, Yu, Ryu, & Ha, 2016). NightOwl automatically analyzes the abovementioned phenomena to identify whether or not a sleep epoch comprises REM sleep.

Statistical analysis

General

Statistical analysis was performed using MATLAB (version 2019a, MathWorks, USA). For all sensitivity and specificity endpoint parameters, 95% confidence intervals were computed. Sleep apnea severity was defined as no clinical sleep apnea if the AHI was less than 5, mild sleep apnea if the AHI was between 5 and 15, moderate sleep apnea if the AHI was between 15 and 30, and severe sleep apnea if the AHI was larger than 30. Significance levels were determined for an alpha (p-value) of 0.05.

Data synchronization

PSG and PAT HSAT data were algorithmically synchronized by matching the instantaneous heart rate traces derived from the electrocardiogram trace of PSG and the PR trace of PAT HSAT. Data epochs that were of insufficient quality to be interpreted by the sleep technician or PAT HSAT were rejected from PSG and the PAT HSAT traces.

Data adequacy

All 261 participants included in the dataset met the following data adequacy criteria: all PSG channels could be interpreted by the technicians (e.g. there were no detachments of the nasal cannula or pulse oximeter), and at least 4 hours of analyzable signal could be obtained for the PAT HSAT, as recommended by the AASM (Kapur et al., 2017).

Participants with missing patient characteristics, such as age and gender data, were omitted from the analysis of population demographic statistics. These participants were still included for the calculation of endpoint parameters.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Endpoint analysis

Primary study aim: assessing REM OSA phenotyping agreement

The criterion used to determine REM OSA was:

***REM AHI / NREM AHI ≥ 2 and AHI > 5 and REM sleep time > 30 minutes, else
phenotyping is inconclusive.***

For both the PAT HSAT analysis and the Local Analysis, the sensitivity and specificity of the agreement with the Expert Analysis were calculated. To describe the agreement of REM OSA phenotyping by just a single parameter, the sum of the sensitivity and specificity was also calculated (Bianchi, 2017).

Auxiliary study aim: assessing minimum required REM sleep time

To better understand the impact of the minimum required REM sleep time (further referred to as the REM exclusion criterion, or REC), it was investigated how the REM OSA phenotyping performance changed as the REC varied from 1 minute to the conventional 30 minutes. Participants for which the REM sleep time was less than the REC according to either the Expert Analysis or the comparator analysis were labeled as inconclusive. Statistical significance of changes in sensitivity by varying the REC was determined by means of a Test of Proportions.

Results

Demographic information

As listed in Table 1, participants were predominantly male (60%), of middle age (mean 54 yrs, STD 14), and overweight (mean BMI 30.0 kg/m² and STD 5.9). The mean AHI was 31.9

(STD 25.6). 29 participants had no clinical sleep apnea, 59 participants had mild sleep apnea, 60 participants had moderate sleep apnea and 113 participants had severe sleep apnea.

[SUGGESTED INSERTION TABLE 1]

REM OSA phenotyping performance and influence of minimum required REM sleep time

The prevalence of REM OSA in this dataset, calculated at a REC of 30 minutes, was 26.1%.

The sensitivity and the specificity of the REM OSA phenotyping by the Local Analysis and PAT HSAT is visualized in Figure 2 for REC values ranging from 1 to 30 minutes. The red segment of the sensitivity line of the PAT HSAT highlights those REC values for which the sensitivity was statistically significantly lower than at the conventional REC of 30 minutes.

An increase in sensitivity can be observed as REC increases from 1 minute to 30 minutes. For PAT HSAT, as REC values exceeded 10 minutes, sensitivities were no longer significantly lower than the sensitivity at REC 30. For the Local Analysis, no statistically significant reduction in sensitivity was observed when reducing the REC from 30 minutes to 1 minute. There was no REC value in the investigated range for which the specificity of either comparator was statistically significantly lower than that at a REC of 30 minutes.

As stated in the Methods section, only assessments for which both the Expert Analysis and the comparator determined a REM sleep above the REC were labeled as conclusive. The percentage of participants for which there was conclusive phenotyping is visualized in Figure 2 for each REC value. As the REC increases, the percentage of conclusiveness drops. A decline from more than 85% conclusiveness at a REC of 1 minute to approximately 45% conclusiveness at a REC of 30 minutes was observed. The numeric results presented in Figure 2 are listed in Table 2 for RECs of 15 and 30 minutes. For both REC values, the PAT HSAT sensitivity is substantially lower than the Local Analysis, while its specificity was slightly higher. For both comparators, a marked difference in conclusiveness from

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

approximately 45% at a REC of 30 minutes to more than 65% at a REC of 15 minutes was found.

[SUGGESTED INSERTION FIGURE 2]

[SUGGESTED INSERTION TABLE 2]

Discussion

Minimum required REM sleep time when phenotyping REM

OSA: the current consensus might be too conservative

The findings presented in this manuscript support that a reduction of the minimum required REM sleep time for REM OSA phenotyping from 30 minutes to 10 minutes can be permitted before the reduction in phenotyping sensitivity becomes statistically significant.

By lowering the REC from 30 to 15 minutes, the conclusiveness percentage increases from 44% to 67% for the Local PSG Analysis, and from 46% to 70% for the PAT HSAT.

While Mokhlesi et al. (2014) argue that a minimum REM sleep time of 30 minutes would be required to reduce the possibility of overestimating the effects of REM OSA in individuals with short REM duration, we hold the opinion that in light of significant inconclusiveness, it would rather be more prudent to avoid false negative or inconclusive REM OSA phenotyping. In view of steep gains in conclusiveness while experiencing only a non-significant reduction in sensitivity and specificity resulting from a reduction of the minimum required REM sleep time from 30 minutes to 15 minutes, the authors therefore propose the use of a minimum required REM sleep time of 15 minutes instead of 30 when electing to phenotype REM OSA. In contrast to Mokhlesi et al.(2014), it is the authors’ opinion that whenever REM OSA was phenotyped based on a short REM sleep time, a multi-night

assessment either by PSG or PAT HSAT would be desirable over the altogether abstention of phenotyping.

Phenotyping REM OSA: PAT-HSAT may fall short

Since this work presents the first reporting on REM OSA phenotyping performance of PSG and PAT HSAT according to a conventional REM OSA criterion, no consensus exists on an adequate phenotyping sensitivity and specificity.

In clinical practice, a certain margin of error for the diagnosis of OSA has been accepted, as implied from the inherent residual inter-scorer variability of PSG-derived AHI (Malhotra et al., 2013). It can be argued that the accepted clinical performance of REM OSA phenotyping ought not to be expected to exceed the accepted clinical performance of the OSA diagnosis itself, as the latter has generally more impactful treatment implications than the former.

Therefore, a conservative performance target for REM OSA sensitivity and specificity would be to take the inter-rater agreement level of the diagnosis of OSA.

As highlighted by Bianchi (2017), the agreement between a comparator and a reference can be characterized by the sum of its sensitivity and specificity at a binary cutoff. To assess the clinical performance of OSA diagnosis on the same patient demographic for which the REM OSA performance was determined in this study, the sensitivity and specificity of diagnosing OSA by the Local PSG Analysis were calculated from the same dataset of 261 double scored PSGs, while the Expert PSG Analysis was again designated as the reference.

The AHI-cutoffs most important to the diagnosis of OSA are 5 and 15, depending on the presence of OSA-related symptoms (Epstein, David, Strollo, & Friedman, 2009). As displayed in Table 3, the sum of the sensitivity and specificity was 1.70 for both AHI cutoffs. To support the generalizability of these results, it was investigated whether the inter-rater agreement of this dataset was in line with those reported in the literature. Since inter-scorer variability literature reports mostly on a correlation coefficient, despite its obvious

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

shortcomings as a parameter for disease severity accuracy, the ICC(A,1) between the two AHI estimates was calculated. The correlation coefficient between the two AHIs was 0.91, which is close to the across-site ICC of 0.95 reported by Malhotra et al.(2013).

As displayed in Table 2 of the result section, the sum of the sensitivity and specificity for determining REM OSA when respectively employing REC values of 15 and 30 minutes were 1.69 and 1.76, which was close to, or larger than, the target of 1.70. Therefore, it can be concluded that the inter-rater agreement of phenotyping REM OSA using PSG is similar to the inter-rater agreement of diagnosing OSA. For PAT HSAT, the same sums calculated for REC values of 15 and 30 minutes were respectively 1.58 and 1.64, which were lower than the target of 1.70. Hence, though it has not yet been formally explored, this type of test might benefit from multi-night testing in order to increase its sensitivity and specificity.

[SUGGESTED INSERTION TABLE 3]

A theoretical model of multi-night REM OSA phenotyping: it may move the needle

It could be concluded that, contrary to PSG, REM OSA phenotyping performance of PAT HSAT did not meet the proposed, although very demanding, target sum of sensitivity and specificity on a single night basis. It was preliminarily investigated whether multi-night testing could improve the performance above the target level. To this end, multi-night testing was modeled as a binomial process in which each single-night test outputs a REM OSA phenotype assessment with a sensitivity and specificity equal to those introduced in Table 2. This model can be interpreted as a voting model in which a single vote is cast after every single-night test, judging whether REM OSA was present during that particular night. The global vote on the presence of REM OSA is cast when at least k out of n votes were positive. Model parameter n is to be interpreted as the number of single-night tests performed for which the REM OSA vote was conclusive. In the event of one or more inconclusive single-

night tests, repeat testing continues until n conclusive single-night phenotype determinations are obtained. Therefore, the actual number of single-night tests would be significantly dependent on the choice of the REC value, since that parameter was demonstrated to significantly influence the conclusiveness of phenotyping. Model parameter k is to be interpreted as the minimum number of positive single-night REM OSA votes required to render the global REM OSA phenotype determination positive.

It can be shown by simple mathematical derivation that the sensitivity and specificity of the global vote under the assumptions laid out above can be estimated as set out in equations 1 (EQ1) and 2 (EQ2).

$$se(k,n) = 1 - \sum_{i=0}^{k-1} \frac{n!}{i!(n-i)!} se(1,1)^i (1 - se(1,1))^{n-i} \quad \text{EQ1}$$

$$sp(k,n) = 1 - \sum_{i=0}^{n-k} \frac{n!}{i!(n-i)!} sp(1,1)^i (1 - sp(1,1))^{n-i} \quad \text{EQ2}$$

The resulting $se(k,n)$ and $sp(k,n)$ parameters are respectively the sensitivity and specificity of the global REM OSA phenotype determination (i.e., the global vote) for any given choice of model parameters k and n . The resulting sum of $se(k,n)$ and $sp(k,n)$ for different values of k and n are displayed in Figure 3.

The figure illustrates how extending the number of nights markedly improves the performance, with a second night of testing crossing the 1.70 performance target for k equal to 1. As n exceeds 3 nights of testing for k equal to 1, the performance starts to decrease, and it drops below the 1.70 performance target as n exceeds 4 nights of testing. This is to be expected, since the requirement of only one positive REM OSA vote out (with $k = 1$) of three or more single-night tests would likely result in a high false positive rate of the global vote. Extending the number of test nights to 4 results in marked performance surges for $k = 2$. By increasing the number of test nights to 5, a performance increase from 1.58 to 1.91 was simulated. Increasing the number of test nights beyond 5 shows diminishing incremental performance gains. For a test duration exceeding 5 nights, one might argue that the

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

incremental burden for the patient to test for an additional night starts to outweigh the limited incremental performance gain.

Aside from the above discussion, one should note that multi-night testing has the additional benefit of further mitigating inconclusiveness from which both PAG HSAT and PSG still suffer even after reducing the REC from 30 to 15 minutes.

This model has several significant limitations. First, the model assumes that the REM OSA phenotype is stable throughout the multi-night testing protocol. The validity of this assumption requires further research. Though it has been well established that the AHI is subject to significant night-to-night variability (Punjabi, Patil, Crainiceanu, & Aurora, 2020), it stands to reason that the REM OSA phenotype itself might remain mostly stable over time. Indeed, whereas the AHI might be influenced by the sleeping position and the REM sleep time, among others, which may vary significantly from night to night, the underlying pathophysiology driving the REM predominance of the OSA is likely to remain stable. Second, the sensitivity and specificity values cannot be determined with perfect accuracy since we have shown that PSG itself suffers from significant inter-rater variability when phenotyping REM OSA. A more stable sensitivity and specificity estimate could be obtained by averaging PSG data scored by a large number of expert sleep technicians. Third, a binomial model assumes statistical independence between the repeat measurements. However, a degree of dependence between individual single-night REM OSA phenotyping errors within a multi-night protocol is expected. For example, when a patient’s REM episodes are difficult to estimate from the PAT measurement because of sustained and abnormally depressed sympathetic activity, a systematic underestimation of REM sleep time may occur which may result in repeated false negative REM OSA phenotyping.

Nevertheless, statistical dependence of the phenotyping performance of repeat-measures would result in an underestimation of the number of nights required to increase the REM

OSA phenotyping performance beyond the target value. Hence, if this assumption doesn't hold, the conclusion would stand that a multi-night protocol would be desirable to increase REM OSA phenotyping performance of PAT HSAT.

[SUGGESTED INSERTION FIGURE 3]

Strengths and weaknesses of the study

The strength of this study lies in its analysis of a large and independently double-scored dataset, which includes both PSG data and synchronously acquired PAT HSAT data.

A number of weaknesses of the study should be noted. Firstly, only one embodiment of PAT HSAT was included. Investigating the generalizability of the results to the WatchPAT device will be the subject of future research. Finally, the simulation used to estimate the potential of multi-night testing by PAT HSAT relies on challengeable assumptions. Multi-night clinical trial data would be required to test these assumptions more conclusively.

Conclusion

This work presents the first assessment of the REM OSA phenotyping performance of either PSG or PAT HSAT. Firstly, it was found that PSG-based REM OSA phenotyping was sensitive and specific even for a single-night test. The PAT-based REM OSA phenotyping showed a lower sensitivity but a slightly higher specificity to PSG. In order to increase the performance of PAT-based REM OSA phenotyping to a level comparable to the single-night PSG, a multi-night testing protocol should be considered. A multi-night test duration of 2 to 5 nights was simulated to yield the most substantial incremental performance gains. Secondly, it was suggested that in order to increase the level of conclusiveness of REM OSA phenotyping, the minimum required REM sleep time can be lowered from the conventional 30 minutes to 15 minutes without significantly impacting the sensitivity and specificity of the

phenotyping. A multi-night testing protocol would further increase conclusiveness for both PSG and PAT HSAT.

Abbreviations

AASM	American Academy of Sleep Medicine
BMI	body mass index
COVID-19	Coronavirus disease of 2019
CPAP	continuous positive airway pressure
EPAP	expiratory positive airway pressure
EEG	electroencephalography
HSAT	home sleep apnea test
ICC	intraclass correlation coefficient
ICC(A,1)	ICC of type <i>two-way fixed model with single measures of absolute agreement</i>
MSSS	Michele Sleep Scoring System
NREM AHI	AHI during non-REM sleep
OSA	obstructive sleep apnea
PAT	peripheral arterial tonometry
PAT HSAT	PAT-based HSAT
PPG	photoplethysmography
PR	pulse rate
PSG	polysomnography
REC	REM exclusion criterion
REM	rapid eye movement
REM AHI	AHI during REM sleep

REM OSA	REM-related OSA
SpO ₂	blood oxygen saturation
STD	standard deviation
ZOL	Ziekenhuis Oost-Limburg

Acknowledgements

/

References

- Alzoubaidi, M., & Mokhlesi, B. (2017). Obstructive sleep apnea during REM sleep: Clinical relevance and therapeutic implications, *22*(6), 545–554.
<https://doi.org/10.1097/MCP.0000000000000319>. Obstructive
- Berry, R. B., Quan, S. F., Abreu, A., Brooks, R., Gamaldo, C., Harding, S. M., ... Vaughn, B. V. (2020). The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications. Version 2.6. *Darien, IL: American Academy of Sleep Medicine*, *13*(5), 665–666. <https://doi.org/10.5664/jcsm.6576>
- Bianchi, M. T. (2017). Bayes' theorem and the rule of 100: a commentary on 'Validity of administrative data for identification of obstructive sleep apnea.' *Journal of Sleep Research*, *26*(3), 401. <https://doi.org/10.1111/jsr.12534>
- Budhiraja, R., Parthasarathy, S., & Quan, S. F. (2007). Endothelial dysfunction in obstructive sleep apnea. *Journal of Clinical Sleep Medicine*, *3*(4), 409–415.
<https://doi.org/10.5664/jcsm.26864>
- Budhiraja, R., & Quan, S. F. (2005). Sleep-disordered breathing and cardiovascular health.

Current Opinion in Pulmonary Medicine, 11(6), 501–506.

<https://doi.org/10.1097/01.mcp.0000183058.52924.70>

Chiao, W., & Durr, M. L. (2017). Trends in sleep studies performed for Medicare beneficiaries. *Laryngoscope*, 127(12), 2891–2896. <https://doi.org/10.1002/lary.26736>

Choi, E., Park, D. H., Yu, J. H., Ryu, S. H., & Ha, J. H. (2016). The severity of sleep disordered breathing induces different decrease in the oxygen saturation during rapid eye movement and non-rapid eye movement sleep. *Psychiatry Investigation*, 13(6), 652–658. <https://doi.org/10.4306/pi.2016.13.6.652>

Chouchou, F., & Desseilles, M. (2014). Heart rate variability: A tool to explore the sleeping brain? *Frontiers in Neuroscience*, 8(DEC), 1–9. <https://doi.org/10.3389/fnins.2014.00402>

Dean, D. A., Wang, R., Jacobs, D. R., Duprez, D., Punjabi, N. M., Zee, P. C., ... Redline, S. (2015). A systematic assessment of the association of polysomnographic indices with blood pressure: The multi-ethnic study of atherosclerosis (MESA). *Sleep*, 38(4), 587–596. <https://doi.org/10.5665/sleep.4576>

Douglas, N. J., White, D. P., Weil, J. V., Pickett, C. K., & Zwillich, C. W. (1982). Hypercapnic ventilatory response in sleeping adults. *American Review of Respiratory Disease*, 126(5), 758–762. <https://doi.org/10.1164/arrd.1982.126.5.758>

Douglas, Neil J. (2010). *Respiratory Physiology: Understanding the Control of Ventilation. Principles and Practice of Sleep Medicine: Fifth Edition* (Fifth Edit). Elsevier Inc. <https://doi.org/10.1016/B978-1-4160-6645-3.00022-0>

Duce, B., Kulkas, A., Langton, C., Töyräs, J., & Hukins, C. (2018). The prevalence of REM-related obstructive sleep apnoea is reduced by the AASM 2012 hypopnoea criteria. *Sleep and Breathing*, 22(1), 57–64. <https://doi.org/10.1007/s11325-017-1526-1>

- Dvir, I., Adler, Y., Freimark, D., & Lavie, P. (2002). Evidence for fractal correlation properties in variations of peripheral arterial tone during REM sleep. *American Journal of Physiology - Heart and Circulatory Physiology*, 283(1 52-1), 434–439. <https://doi.org/10.1152/ajpheart.00336.2001>
- Epstein, L. J., David, K., Strollo, P. J. J., & Friedman, N. (2009). Clinical Guideline for the Evaluation, Management and Long-term Care of Obstructive Sleep Apnea in Adults. *Journal of Clinical Sleep Medicine*.
- Finnsson, E., Ólafsdóttir, G. H., Loftsdóttir, D. L., Jónsson, S. Æ., Helgadóttir, H., Ágústsson, J. S., ... Wellman, A. (2020). A scalable method of determining physiological endotypes of sleep apnea from a polysomnographic sleep study. *Sleep*, (September), 1–10. <https://doi.org/10.1093/sleep/zsaa168>
- Gabryelska, A., & Białasiewicz, P. (2020). Association between excessive daytime sleepiness, REM phenotype and severity of obstructive sleep apnea. *Scientific Reports*, 10(1), 1–6. <https://doi.org/10.1038/s41598-019-56478-9>
- Haba-Rubio, J., Janssens, J. P., Rochat, T., & Sforza, E. (2005). Rapid eye movement-related disordered breathing: Clinical and polysomnographic features. *Chest*, 128(5), 3350–3357. <https://doi.org/10.1378/chest.128.5.3350>
- Herscovici, S., Pe'er, A., Pappan, S., & Lavie, P. (2007). Detecting REM sleep from the finger: An automatic REM sleep algorithm based on peripheral arterial tone (PAT) and actigraphy. *Physiological Measurement*, 28(2), 129–140. <https://doi.org/10.1088/0967-3334/28/2/002>
- Ioachimescu, O. C., Shirine Allam, J., Samarghandi, A., Anand, N., Fields, B. G., Dholakia, S. A., ... Collop, N. A. (2020). Performance of peripheral arterial tonometry–based testing for the diagnosis of obstructive sleep apnea in a large sleep clinic cohort. *Journal*

- of *Clinical Sleep Medicine*, 6(10), 1663–1674. <https://doi.org/10.5664/jcsm.8620>
- Johnson, K. G., Sullivan, S. S., Nti, A., Rastegar, V., & Indira Gurubhagavatula. (2020). The impact of the COVID-19 pandemic on sleep medicine practices. *Journal of Clinical Sleep Medicine*. Retrieved from <http://dx.doi.org/10.1016/j.cplett.2015.08.035>
- Kapur, V. K., Auckley, D. H., Chowdhuri, S., Kuhlmann, D. C., Mehra, R., Ramar, K., & Harrod, C. G. (2017). Clinical Practice Guideline for Diagnostic Testing for Adult Obstructive Sleep Apnea: An American Academy of Sleep Medicine Clinical Practice Guideline. *J Clin Sleep Med*, 13(3), 479–504.
- Lévy, P., Pépin, J. L., Arnaud, C., Baguet, J. P., Dematteis, M., & Mach, F. (2009). Obstructive Sleep Apnea and Atherosclerosis. *Progress in Cardiovascular Diseases*, 51(5), 400–410. <https://doi.org/10.1016/j.pcad.2008.03.001>
- Malhotra, A., Younes, M., Kuna, S. T., Benca, R., Kushida, C. A., Walsh, J., ... Pien, G. W. (2013). Performance of an automated polysomnography scoring system versus computer-assisted manual scoring. *Sleep*, 36(4), 573–582. <https://doi.org/10.5665/sleep.2548>
- Massie, F., De Almeida, D. M., Dreesen, P., Thijs, I., Vranken, J., & Klerkx, S. (2018). An evaluation of the Night Owl home sleep apnea testing system. *Journal of Clinical Sleep Medicine*, 14(10), 1791–1796. <https://doi.org/10.5664/jcsm.7398>
- Mokhlesi, B., Finn, L. A., Hagen, E. W., Young, T., Hla, K. M., Van Cauter, E., & Peppard, P. E. (2014). Obstructive sleep apnea during REM sleep and hypertension: Results of the Wisconsin sleep cohort. *American Journal of Respiratory and Critical Care Medicine*, 190(10), 1158–1167. <https://doi.org/10.1164/rccm.201406-1136OC>
- Mokhlesi, B., & Punjabi, N. M. (2012). “REM-related” obstructive sleep apnea: An epiphenomenon or a clinically important entity? *Sleep*, 35(1), 5–7.

<https://doi.org/10.5665/sleep.1570>

Nisha Aurora, R., Crainiceanu, C., Gottlieb, D. J., Kim, J. S., & Punjabi, N. M. (2018).

Obstructive sleep apnea during REM sleep and cardiovascular disease. *American Journal of Respiratory and Critical Care Medicine*, 197(5), 653–660.

<https://doi.org/10.1164/rccm.201706-1112OC>

Oksenberg, A., Arons, E., Nasser, K., Vander, T., & Radwan, H. (2010). REM-related

obstructive sleep apnea: The effect of body position. *Journal of Clinical Sleep Medicine*, 6(4), 343–348. <https://doi.org/10.5664/jcsm.27875>

Punjabi, N. M., Patil, S., Crainiceanu, C., & Aurora, R. N. (2020). Variability and

Misclassification of Sleep Apnea Severity Based on Multi-Night Testing. *Chest*, 158(1), 365–373. <https://doi.org/10.1016/j.chest.2020.01.039>

Schächinger, H., Weinbacher, M., Kiss, A., Ritz, R., & Langewitz, W. (2001).

Cardiovascular indices of peripheral and central sympathetic activation. *Psychosomatic Medicine*, 63(5), 788–796. Retrieved from

<http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L32911008>

Somers, V. K., Dyken, M. E., Clary, M. P., & Abboud, F. M. (1995). Sympathetic neural

mechanisms in obstructive sleep apnea. *Journal of Clinical Investigation*, 96(4), 1897–1904. <https://doi.org/10.1172/JCI118235>

Stansbury, R. C., & Strollo, P. J. (2015). Clinical manifestations of sleep apnea. *Journal of*

Thoracic Disease, 7(9), E298–E310. <https://doi.org/10.3978/j.issn.2072-1439.2015.09.13>

Su, C. S., Liu, K. T., Panjapornpon, K., Andrews, N., & Foldvary-Schaefer, N. (2012).

Functional outcomes in patients with REM-related obstructive sleep apnea treated with

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

positive airway pressure therapy. *Journal of Clinical Sleep Medicine*, 8(3), 243–247.
<https://doi.org/10.5664/jcsm.1902>

Varga, A. W., & Mokhlesi, B. (2019). REM obstructive sleep apnea: risk for adverse health outcomes and novel treatments. *Sleep and Breathing*, 23(2), 413–423.
<https://doi.org/10.1007/s11325-018-1727-2>

Zhou, Y., Zhao, D., Yin, G., Li, J., Cao, X., Zhang, Y., & Ye, J. (2020). Analysis of the myoelectric characteristics of genioglossus in REM sleep and its improvement by CPAP treatment in OSA patients. *Sleep and Breathing*, 24(2), 471–482.
<https://doi.org/10.1007/s11325-019-01875-7>

Zinchuk, A., & Yaggi, H. K. (2020). Phenotypic Subtypes of OSA: A Challenge and Opportunity for Precision Medicine. *Chest*, 157(2), 403–420.
<https://doi.org/10.1016/j.chest.2019.09.002>

Figures

Figure 1

Title: **Figure 1** — Diagram of the data acquisition setup.

Caption: This diagram depicts that for each study participant, PSG and PAT HSAT were concurrently administered. PSG was analyzed by two independent scoring centra (Expert and Local) and PAT HSAT was analyzed automatically. Both the Local Analysis and PAT HSAT Analysis were compared to Expert Analysis (illustrated by the dashed line). PAT HSAT = peripheral arterial tonometry-based home sleep apnea testing, PSG = polysomnography.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 2

Title: **Figure 2** — Illustration of REM OSA phenotyping sensitivity, specificity and conclusiveness as a function of REM exclusion criterion.

Caption: The figure depicts how the REM OSA phenotyping sensitivity and specificity increase and phenotyping conclusiveness decrease with increasing minimum required REM time (REM exclusion criterion) in minutes. The red line segments highlight zones where the phenotyping sensitivity was significantly lower than that at a REM exclusion criterion of 30 minutes. OSA = obstructive sleep apnea, PAT HSAT = peripheral arterial tonometry-based home sleep apnea testing, REM = rapid eye movement.

Figure 3

Title: **Figure 3** — Simulation of multi-night REM OSA phenotyping model.

Caption: The figure depicts the theoretical sum of REM OSA phenotyping sensitivity and specificity of PAT HSAT for multi-night HSAT of n nights, with k the minimum number of single-night positive REM OSA phenotype determinations to render the global phenotype determination positive. The green line highlights the target REM OSA phenotyping performance expressed as the sum of phenotyping sensitivity and specificity. k = minimum required single-night positive REM OSA phenotype determinations to render global phenotype determination positive, n = number of HSAT nights OSA = obstructive sleep apnea, PAT HSAT = peripheral arterial tonometry-based home sleep apnea testing, REM = rapid eye movement, se = sensitivity, sp = specificity.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Tables

Table 1

	Mean	STD	Min	Max
Age	54	14	21	84
AHI	31.9	25.6	0.0	118.7
BMI	30.0	5.9	18.2	53.8

Title: **Table 1** — Demographic and clinical characteristics of participants in dataset.

Caption: AHI = apnea-hypopnea index, BMI = body mass index, Min = minimum, Max = maximum, STD = standard deviation.

Table 2

REC	Comparator	Sensitivity	Specificity	Sum	Conclusiveness
15 min	<i>PAT HSAT</i>	0.63 (± 0.14)	0.95 (± 0.04)	1.58 (± 0.10)	0.70 (± 0.06)
	<i>Local PSG</i>	0.81 (± 0.12)	0.88 (± 0.06)	1.69 (± 0.12)	0.67 (± 0.06)
30 min	<i>PAT HSAT</i>	0.68 (± 0.16)	0.97 (± 0.04)	1.64 (± 0.12)	0.46 (± 0.06)
	<i>Local PSG</i>	0.87 (± 0.12)	0.89 (± 0.07)	1.76 (± 0.13)	0.44 (± 0.06)

Title: **Table 2** — Tabulation of REM OSA sensitivity, specificity, their sum and conclusiveness for REM exclusion criterion (REC) of 15 and 30 minutes.

Caption: 95% confidence intervals around the parameters are provided within the brackets.

PAT HSAT = peripheral arterial tonometry-based home sleep apnea testing, REC = REM exclusion criterion, REM = rapid eye movement.

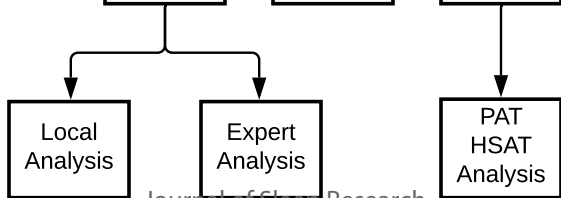
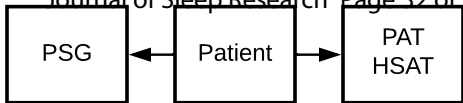
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

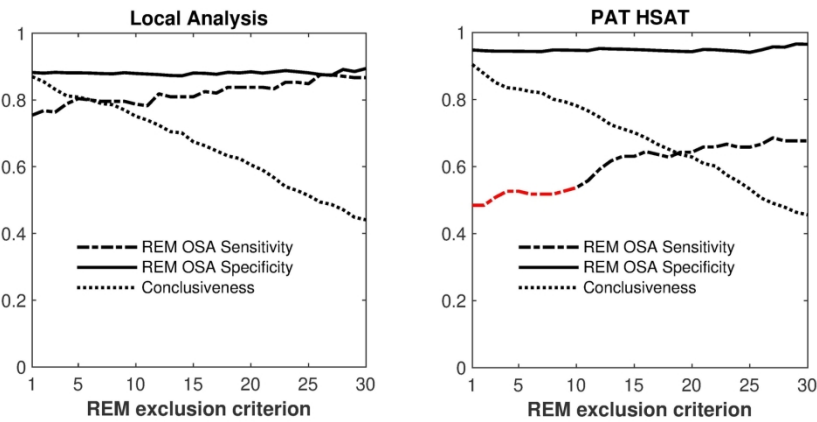
Table 3

N	AHI Cutoff	Sensitivity	Specificity	Sum
261	5	0.94 (±0.03)	0.76 (±0.16)	1.70 (±0.10)
	15	0.93 (±0.04)	0.77 (±0.09)	1.70 (±0.08)

Title: **Table 3** — Tabulation of Local Analysis’ OSA diagnosis sensitivity, specificity and their sum for AHI cutoffs 5 and 15.

Caption: The table illustrates the sensitivity and specificity of the Local Analysis’ OSA diagnosis at AHI cutoffs 5 and 15 when taking the Expert Analysis as reference for the dataset of 261 participants. 95% confidence intervals around the parameters are provided within the brackets. AHI = apnea-hypopnea index, N = sample size.

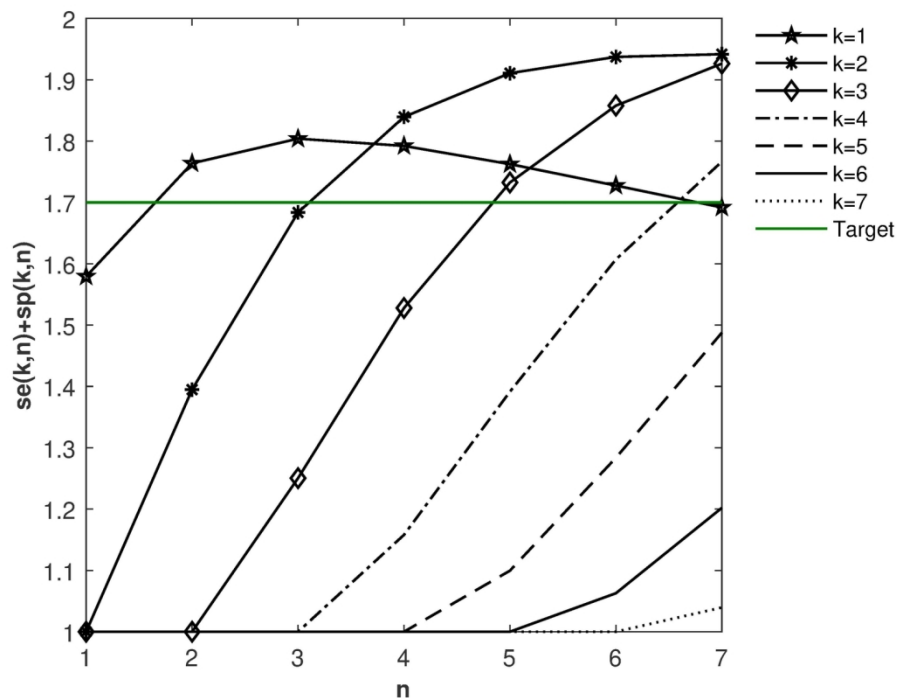




Title: Figure 2 — Illustration of REM OSA phenotyping sensitivity, specificity, and conclusiveness as a function of REM exclusion criterion.

Caption: The figure depicts how the REM OSA phenotyping sensitivity and specificity increase and phenotyping conclusiveness decrease with increasing minimum required REM time (REM exclusion criterion) in minutes. The red line segments highlight zones where the phenotyping sensitivity was significantly lower than that at a REM exclusion criterion of 30 minutes. OSA = obstructive sleep apnea, PAT HSAT = peripheral arterial tonometry-based home sleep apnea testing, REM = rapid eye movement.

176x83mm (300 x 300 DPI)



Title: Figure 3 — Simulation of multi-night REM OSA phenotyping model.

Caption: The figure depicts the theoretical sum of REM OSA phenotyping sensitivity and specificity of PAT HSAT for multi-night HSAT of n nights, with k the minimum number of single-night positive REM OSA phenotype determinations to render the global phenotype determination positive. The green line highlights the target REM OSA phenotyping performance expressed as the sum of phenotyping sensitivity and specificity. k = minimum required single-night positive REM OSA phenotype determinations to render global phenotype determination positive, n = number of HSAT nights OSA = obstructive sleep apnea, PAT HSAT = peripheral arterial tonometry-based home sleep apnea testing, REM = rapid eye movement, se = sensitivity, sp = specificity.

148x111mm (300 x 300 DPI)