GoodSleepco™

& EnsoSleep

GoodSleep Test & EnsoSleep PPG Clinical Performance

If you would like more information, please contact us:

hello@goodsleepco.health

Introduction

Performance relates to the ability of the system to safely, effectively, and efficiently fulfill the role for which it is designed. The primary metric for assessing performance of EnsoSleep PPG is the Apnea Hypopnea Index (AHI). The AHI depends on the number of sleep disordered breathing events detected, normalized by a determination of total sleep time (TST), where the TST is the total amount of sleep time during the recording period. An AHI below 5 is considered normal (normative sleep breathing and respiration), more than 5 is considered Obstructive Sleep Apnea (OSA) in adult patients.

The performance testing objective was to validate EnsoSleep PPG's determination of the AHI on single-channel photoplethysmography (PPG) and oximetry signals for adult patients.

The primary endpoint was to determine whether the AHI-5 cutoff as determined by EnsoSleep PPG from a PPG recording was non-inferior to clinically acceptable performance levels defined for home sleep testing. The acceptable performance was based on AASM clinical practice guidelines for home sleep apnea testing (HSAT) sensitivity and specificity, and compared to the 2/3 majority manual scoring AHI determination from a simultaneous recording by a PSG device.

The secondary endpoints included a continuous evaluation of the AHI, which measured EnsoSleep PPG agreement performance for AHI determination across the full range of observed AHI severities. Finally, a supportive analysis included agreement metrics for Wake, Light non-rapid eye movement (NREM), Deep NREM, and REM Sleep, and performance metrics for TST, as well as Central Index (CI) at various cutoffs (CI-5, CI-10, CI-15); Sleep Efficiency (SE); Sleep Latency (SL); Wake After Sleep Onset (WASO); Oxygen Desaturation Events Index (ODI); and Central Respiratory Events Index (CI).

Dataset

Clinical Trial

In this study, physiologic signals from FDA cleared polysomnography (PSG) devices were collected, utilizing the Philips Respironics Sleepware G3 (K202142), Natus Sandman Elite (K153353), and Polysmith Sleep System (K161650) PSG systems, to establish gold-standard comparator data. Simultaneously, PPG signals were recorded utilizing the FDA cleared single-channel PPG device, Viatom Checkme O2 (K191088), to establish the primary validation endpoints to evaluate EnsoSleep PPG performance. Enrollment utilized a non-randomized, prospective, all-comers enrollment approach as part of routine attended diagnostic in-lab PSG testing.

A Majority Scoring Panel (MSP) consisted of three (3) Registered Polysomnographic Technologists (RPSGTs) holding current certification and licensure with the Board of Registered Polysomnographic Technologists (BRPT) and possessing five or more years of clinical sleep medicine experience. Each event or stage was evaluated independently by all three individuals. For an event to be officially scored or reported, a consensus of at least two-thirds among the scorers was required. This method aimed to improve the reliability and accuracy of comparator labels. Each PSG was reviewed by a board certified sleep physician to provide a clinical confirmation of scoring and technical adequacy of PSGs prior to final incorporation into MSP and use for AHI comparison to EnsoSleep PPG.

Clinical Trial Dataset Profile

The following is a profile of the adult patients in the gold standard dataset. It demonstrates the dataset is a representative sample of the adult patient population related to this device.

Age:

Age	Number of Subjects (n)	Percent of Subjects (%)
18-29	24	15.2%
30-39	30	19.0%
40-49	29	18.4%
50-59	27	17.1%
60-69	34	21.5%
70-79	12	7.6%
80-89	1	0.6%
Not Specified	1	0.6%

Sex (at birth)

Sex (at birth)	Number of Subjects (n)	Percent of Subjects (%)
Female	88	55.7%
Male	68	43.0%
Not Specified	2	1.3%

Body Mass Index (BMI)

Body Mass Index (BMI)	Number of Subjects (n)	Percent of Subjects (%)
Normal	20	12.7%
Overweight	41	25.9%
Obese	72	45.6%
Morbidly Obese	24	15.2%
Not Specified	1	0.6%

Epworth Sleepiness Scale (ESS)

Epworth Sleepiness Scale (ESS)	Number of Subjects (n)	Percent of Subjects (%)
Lower normal sleepiness (ESS under 5)	38	24.1%
Higher normal sleepiness (ESS 6-10)	58	36.7%
Mild excessive sleepiness (ESS 11-12)	15	9.5%
Moderate excessive sleepiness (ESS 13-15)	20	12.7%
Severe excessive sleepiness (ESS ≥16)	21	13.3%
Not Specified	6	3.8%

Sleep Apnea Disease State Severity

Sleep Apnea Disease State Severity	Min AHI cutoff (inclusive)	Max AHI cutoff (exclusive)	Number of Subjects (n)	Percentage of Subjects (%)
Normative	None	5	49	31.0%
Mild	5	15	53	33.5%
Moderate	15	30	29	18.4%
Severe	30	None	27	17.1%

Medical Condition Groups

Medical Condition Groups	Number of Subjects (n)	Percent of Subjects (%)
Sleep Disorders	69	43.7%
Psychiatric Disorders	58	36.7%
Neurologic Disorders	18	11.4%
Neurodevelopmental Disorders	11	7.0%
Cardiac Disorders	11	7.0%
Pulmonary Disorders	32	20.3%
Metabolic Disorders and Other	26	16.5%

Medication Groups

Medication Groups	Number of Subjects (n)	Percent of Subjects (%)
Benzodiazepines	11	7.0%
Antidepressants	43	27.2%
Stimulants	5	3.2%
Opiates	3	1.9%
Sleep Aids	8	5.1%

Study Sample

A participant was considered to have completed the study if they completed a single-night in-lab sleep test with ≥4 hours of technically adequate PPG signals required for performance validation. The end of the study was defined as completion of the last sleep test to go through the schedule of activities in the study globally.

Overall, 158 subjects were included in the final analysis, where 109 subjects received a positive reference standard AHI determination that indicated a positive sleep apnea diagnosis, and 49 received a negative reference standard AHI determination that reflected normative sleep breathing (i.e. clinical control subjects). Both the positive population and negative population in the final study sample met the required sample size and representativeness criteria reported in the clinical trial protocol.

Desaturation Rules

The American Academy of Sleep Medicine (AASM) for the Scoring of Sleep and Associated Events recommendations Version 3 (February 2023) recommends utilizing the 3% sleep disordered breathing event scoring rule as the current standard of care (SOC), with the 4% rule as optional for clinical reporting. Based on the procedures recommended by the AASM scoring manual for the SOC, the calibration of this protocol's sample size statistical power analysis was based on the AASM guideline recommended parameters for out-of-center testing which uses the AASM recommended 3% scoring criteria. Therefore, the primary endpoint AHI calculation utilized the recommended 3% rule.

However, all performance measures were also analyzed and reported based on the 4% rule. Construction of the panel clinician scoring reference standard for sleep disordered breathing based on the 4% rule was interpolated from the primary 3% sleep disordered event and AHI measures that were collected by applying a software-based procedure for the exclusion of sleep disordered breathing events from the 2/3 MSP defined in the study protocol, whereby each event that was associated with a desaturation of less than 4% was removed.

For reference, we provide here the verbatim guidelines from the AASM manual for the 3% rule (i.e., AASM 1A) vs. the 4% rule (i.e., AASM 1B).

1A. Score a respiratory event as a hypopnea if ALL the following criteria are met:

1. The peak signal excursions drop by ≥30% of pre-event baseline using nasal pressure (diagnostic study), PAP device flow (titration study), or an alternative hypopnea sensor (diagnostic study).
2. The duration of the ≥30% drop in signal excursion is ≥10 seconds.
3. There is a ≥3% oxygen desaturation from the pre-event baseline or the event is associated with an arousal.

1B. Score a respiratory event as a hypopnea if ALL the following criteria are met:

1. The peak signal excursions drop by ≥30% of pre-event baseline using nasal pressure (diagnostic study), PAP device flow (titration study), or an alternative hypopnea sensor (diagnostic study).
2. The duration of the ≥30% drop in signal excursion is ≥10 seconds.
3. *There is a ≥4% oxygen desaturation from the pre-event baseline.*

Performance Metrics

Agreement Analysis

In order to analyze the performance at the defined AHI severity threshold, each patient is associated with a negative or positive diagnosis based on their AHI. The table specifies the definitions for a positive and negative diagnosis for the primary endpoint AHI severity threshold of AHI ≥ 5.

OSA Severity	AHI Range	OSA Negative	OSA Positive
No OSA (-)	[0,5]	X	
OSA (+)	[5,∞]		X

By comparing the reference standard diagnosis with the EnsoSleep PPGdiagnosis at the defined AHI threshold, each patient was tagged as a True Positive (TP) - both the reference standard and subject device produced a condition-positive result, True Negative (TN) - both the reference standard and subject device produced a condition-negative result, False Positive (FP) - the reference standard produced a condition-negative result while EnsoSleep PPG produced a positive result, and False Negative (FN) - the reference standard produced a condition-positive result while EnsoSleep PPG produced a negative result. The definition for each condition-positive and negative AHI results associated OSA diagnostic outcome agreement can be seen in the following table, i.e. confusion matrix:

	Positive Reference Standard Outcome	Negative Reference Standard Outcome
Positive Test Outcome	*TP*	*FP*
Negative Test Outcome	*FN*	*TN*

By tabulating the number of patients associated with each diagnostic outcome agreement category by AHI values, the sensitivity, and specificity can be calculated for the defined AHI severity threshold of AHI ≥ 5 utilizing the following equations:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \qquad \text{Specificity} = \frac{TN}{TN + FP}$$

*Note that this same procedure, pooled on an epoch-by-epoch basis, can be used to calculate the sensitivity and specificity of EnsoSleep PPG sleep staging.

Two-sided 95% bootstrap percentile method confidence intervals (R=2,000) were calculated for sensitivity and specificity. The final Apnea-Hypopnea Index (eAHI) sensitivity and specificity for the primary endpoint sleep apnea diagnostic threshold of AHI>5 was reported and found to meet the predefined objective performance criteria.

Performance Results

Primary Endpoint:

EnsoSleep PPG Sleep Apnea diagnostic agreement validation - Apnea-Hypopnea Index performance (eAHI ≥ 5)

The Primary Endpoint validated the performance of EnsoSleep PPG's Apnea-Hypopnea Index (eAHI) in comparison to gold-standard PSG by evaluating the PPG-based AI software's ability to meet and exceed clinically acceptable performance levels established for home sleep apnea testing (HSATs) when compared to PSGs.

Utilizing the AASM recommended 3% hypopnea scoring criteria 1.A. and a sleep apnea diagnostic threshold of AHI ≥ 5, EnoSleep PPG demonstrated a sensitivity of 92.6% (PPG eAHI results agreed with PSG AHI results in over 9 in 10 subjects with sleep apnea (AHI ≥ 5)) and a specificity of 71.6% (PPG eAHI results agreed with PSG AHI results in over 7 in 10 subjects with normative sleep breathing (AHI ≤ 5)).

Results were also reported for the alternative 4% hypopnea scoring criteria 1.B., demonstrating 89.4% sensitivity and 76.8% specificity between the EnsoSleep PPG eAHI and gold-standard PSG AHI findings the AHI ≥ 5 sleep apnea diagnostic threshold. The EnsoSleep PPG prospective clinical validation performance results for sleep apnea diagnostic agreement with simultaneous PSG AHI findings were confirmed to exceed the predefined clinical performance criteria.

Sleep Apnea: Apnea Hypopnea Index (eAHI) > 5 (3% and 4% Criteria performance)

Desaturation	Sample Size (n)	Sensitivity	Specificity
eAHI ≥ 5 (3% Hypopnea Criteria 1.A.)	158 (n=109 AHI≥5, n=49 AHI≤5))	92.6% (87.2%, 97.2%)	71.6% (59.2%, 83.7%)
eAHI ≥ 5 (4% Hypopnea Criteria 1.B.)	158 (n=76 AHI≥5, n=82 AHI≤5)	89.4% (81.6%, 96.1%)	76.8% (67.1%, 85.4%)

Secondary Endpoints:

EnsoSleep PPG Sleep Apnea diagnostic severity validation - Apnea-Hypopnea Index continuous performance evaluation measures

The Secondary Endpoints evaluated continuous performance measures for AHI.
To validate performance in terms of agreement and correlation between EnsoSleep PPG's eAHI determination and gold-standard PSG AHI results across the full spectrum of Sleep Apnea diagnostic severities, secondary endpoints were predefined for mean difference (MD) and limits of agreement (LOA) produced by the Bland-Altman (BA) analysis. Secondary endpoints were also predefined for slope (ß1) and intercept (ß0) parameters produced by a Deming Regression analysis as continuous performance validation measures and acceptance for AHI performance across the full range of sleep apnea severities.

The Bland-Altman analysis of agreement demonstrated an mean difference (MD) between EnsoSleep PPG's eAHI and the gold-standard PSG AHI of 1.00 for the 3% Criteria (1.A.) and -1.039 for the 4% Criteria (1.B.), reflecting an average difference of approximately 1 AHI point in both the 3% and 4% result comparisons (e.g., eAHI of 9 vs. AHI of 10, eAHI of 20 vs. AHI of 19, etc.). The Bland-Altman analysis 95th percentile standard deviation (SD) upper (ULOA) and lower (LLOA) statistical limits of agreement results were also found within the objective performance criteria defined for clinically acceptable levels of AHI agreement.

Additional continuous AHI performance measures demonstrated the targeted near-unity slope (3%: 0.936, 4%: 0.982) and near-zero intercept (3%: 0.023, 4%: 1.219) Deming Regression parameter values within the targeted ranges.

Sleep Profile and Oxygen Saturation

Category	Deming Regression		Bland-Altman		
	Slope ß1	Intercept ß0	Mean Difference	Upper Limit (ULOA)	Lower Limit (LLOA)
eAHI (3%) [events/hour]	0.936 (0.853, 1.033)	0.023 (-1.185, 1.122)	1.000 (0.630, 1.367)	14.575 (13.779, 15.363)	-12.574 (-13.371, -11.786)
eAHI (4%) [events/hour]	0.982 (0.903, 1.130)	1.219 (0.116, 1.985)	-1.039 (-1.326, -0.749)	9.307 (8.692, 9.931)	-11.386 (-12.001, -10.763)
Total Sleep Time [hours]	1.159 (1.035, 1.318)	-0.695 (-1.576, -0.005)	-0.093 (-0.132, -0.059)	1.145 (1.060, 1.216)	-1.330 (-1.414, -1.259)
Sleep Efficiency [hours/hours]	1.154 (1.031, 1.317)	-0.088 (-0.205, 0.003)	-0.011 (-0.017, -0.007)	0.163 (0.151, 0.173)	-0.185 (-0.198, -0.176)
Sleep Latency [hours]	1.114 (0.997, 1.290)	-0.023 (-0.185, 0.090)	-0.129 (-0.154, -0.089)	0.884 (0.831, 0.970)	-1.143 (-1.196, -1.057)
Wake After Sleep Onset [hours]	1.073 (0.938, 1.219)	-0.271 (-0.436, -0.121)	0.167 (0.140, 0.196)	1.131 (1.073, 1.193)	-0.797 (-0.855, -0.735)
Oxygen Desaturation Index [events/hours]	0.962 (0.896, 1.056)	1.667 (0.330, 2.847)	-1.046 (-1.417, -0.677)	13.223 (12.426, 14.015)	-15.315 (-16.111, -14.522)

Sleep Staging: Epoch-by-Epoch Sleep Staging Performance Validation

EnsoSleep PPG sleep staging was trained utilizing a transfer learning inspired approach involving machine learning and statistical signal processing methods, including multiple deep neural network models. The algorithm was trained on a database of over 1,000,000 diagnostic PSGs that included concurrently recorded PPG and PSG signals from the same subjects which enabled sleep-stage specific PPG patterns to be recognized and translated into automated PPG-signal sleep stage detection. Clinical performance validation was conducted on the PPG-based AI sleep staging by comparison to the gold-standard PSG sleep staging of the physician reviewed, 2/3 MSP.

PPG-based AI sleep staging demonstrated epoch-by-epoch sleep stage agreement with a sensitivity (e.g., Positive Agreement (PPA)) and specificity (e.g., Negative Agreement NPA)) of 84.2% PPA, 97.5% NPA for Stage REM, 80.7% PPA, 86.7% NPA for Light Non-REM (Stages N1 and N2), 67.9% PPA, NPA 95.5% for Deep Non-REM (Stage N3), and 87.8% PPA, 93.7% NPA for Wake (Stage Wake) compared to the gold-standard PSG's 2/3 MSP consensus sleep staging.

The trial findings demonstrated reliable agreement and high concordance between the PPG-based sleep staging with gold-standard PSG sleep stages on an epoch-by-epoch basis both compared to the reference standard and published clinical and scientific interscorer reliability (ISR) findings. EnsoSleep PPG also demonstrated epoch-by-epoch sleep stage agreement within ranges comparable to those observed in 3,296,905 individual scoring decisions by Rosenberg, et. al. (2013), in The American Academy of Sleep Medicine Inter-scorer Reliability Program: Sleep Stage Scoring, which reported similar Positive Agreement levels (PPA) of 90.5% for REM (vs. 83.6% PPG), 63.0% for N1 and 85.2% for N2 (vs. 80.9% in PPG), 67.4% for N3 (vs. 63.4% in PPG), and 84.1% for Wake (vs. 86.7% PPG.

Category	Sample Size (n)	Sensitivity	Specificity
Wake	52,622	86.7%, (86.5%, 87.0%)	93.5%, (93.4%, 93.7%)
Light Non-REM	69,438	80.9%, (80.6%, 81.2%)	85.5%, (85.2%, 85.7%)
Deep Non-REM	10,195	63.4%, (62.4%, 64.3%)	95.9%, (95.7%, 96.0%)
REM	14,459	83.6%, (83.0%, 84.2%)	97.5%, (97.4%, 97.5%)

Sleep Staging: Full-Night Sleep Quality & Architectural Parameters & Indices

In addition to comparisons of epoch-by-epoch agreement of sleep staging, the prospective trial also reported EnsoSleep PPG's performance for measuring key sleep quality and architectural parameters used routinely in evaluation and management of sleep disorders. Bland-Altman and Deming Regression analyses compared EnsoSleep PPG's output sleep indices to the same sleep parameters from gold-standard PSG.

EnsoSleep PPG's performance for total sleep time (TST), a critical parameter used in directly in the calculation of the AHI) showed high agreement with PSG, demonstrating average difference of -5.58 minutes (95% CI: -7.92, -3.54 minutes) in the total sleep time (TST). Beyond measures characterizing the length of sleep, further measures of sleep architecture and quality relying on underlying sleep staging performance also demonstrated high agreement with the same sleep quality measures in PSG, showing an average difference of -1.1% (-1.7%, -0.7%) in sleep efficiency (SE), -7.74 minutes (-9.42, -5.34 minutes) in sleep latency (SL), and 10.02 minutes (8.4, 11.76 minutes) in wake after sleep onset (WASO) index measures compared to the 2/3 MSP results obtained with simultaneously recorded in-lab PSG. See the chart on page 12 for the complete data.

Conclusion

Overall, the prospective clinical performance validation trial findings suggest reliable agreement and high concordance between EnsoSleep PPG results and simultaneously recorded, gold-standard in-lab PSG results across key diagnostic thresholds for sleep apnea, including (AASM recommended 1.A. criteria results summarized in parenthesis);

- Primary endpoint results demonstrating sensitive (>90%) and specific (>70%) performance for eAHI > 5 diagnostic agreement between sleep apnea positive versus normative trial subjects,
- Secondary endpoint results demonstrating continuous measures of diagnostic performance including Bland-Altman and Deming Regression analyzes evaluating AHI across the full spectrum of sleep apnea disorder severities (average difference of <1.0 AHI point),
- Supporting endpoint results demonstrating sleep staging and macro-sleep architectural parameter performance in epoch-by-epoch sleep staging and global sleep quality measure comparisons including REM (83%), Light Sleep (N1/N2) (80%), Deep Sleep (N3) (63%), Wake (86%), the total sleep time (TST) (average difference <2.6 mins), as well as sleep efficiency (SE), sleep latency (SL), wake after sleep onset (WASO), oxygen desaturation indices (3% and 4%), and other sleep quality and respiratory parameters routine to sleep study reporting.

In conclusion, EnsoSleep PPG's performance was validated for use by qualified healthcare professionals as a tool to aid physicians and clinicians in the diagnosis and management of sleep disorders including sleep apnea, the evaluation of sleep and sleep quality, and support clinician use of EnsoSleep PPG's interoperable, AI-based analysis with an FDA-cleared, compatible, medical-purpose Pulse Oximeter (Pulse Ox) device to enable the providers and patients to benefit from PPG-based sleep testing based on the ease of use and wearability of Pulse Ox devices, reliable and validated diagnostic data and results, and accessibility of sleep disorder assessments that can be clinically administered to patients both remotely and through in-clinic settings.