Karger
OPEN ACCESS

Respiration

Respiration , DOI: 10.1159/000553192

Received: April 9, 2026

Accepted: June 10, 2026

Published online: June 25, 2026

Residual AHI Measurement from PAP Devices Versus Home Polygraphy: A Method Comparison Study Using Real World Data

Mansell SK, Augustt S, Gowing F, Ridout D, Hilton S, Kilbride C, Main E, Schievano S, Mandal S

ISSN: 0025-7931 (Print), eISSN: 1423-0356 (Online)

https://www.karger.com/RES

Respiration

Residual AHI Measurement from PAP Devices Versus Home Polygraphy: A Method Comparison Study Using Real World Data

Stephanie K. Mansell[a,b], Sharon Augustt[b], Francesca Gowing[b], Deborah Ridout[a], Prof Stephen Hilton[a], Prof Cherry Kilbride[b, c], Prof Eleanor Main[a], Prof Silvia Schievano[a], Dr Swapna Mandal[a,b]

[a] University College London, London, UK

[b] Royal Free London NHS Foundation Trust, London, UK

[c] Brunel University of London, London, UK

Short title: Poor Agreement Between PAP AHI and Home Polygraphy in OSA Management

Corresponding Author:

Stephanie Mansell

E-mail address: stephanie.mansell.21@ucl.ac.uk

Keywords:

Obstructive Sleep Apnoea

Apnoea Hypopnoea Index

Positive Airway Pressure Therapy

Telemonitoring

Abstract

Introduction

Residual apnoea hypopnoea index (AHI) derived from positive airway pressure (PAP) device downloads is frequently used to assess treatment effectiveness in obstructive sleep apnoea (OSA). However, the accuracy of device reported AHI in real-world settings remains uncertain. This study evaluated the agreement between PAP device derived AHI and AHI measured using home based cardiorespiratory polygraphy.

Methods

This secondary analysis of the 3DPiPPIn randomised controlled trial included adults with OSA (AHI ≥15 events/hr) naïve to domiciliary PAP therapy. Participants underwent two-night home cardiorespiratory polygraphy at three and six months. Residual AHI was obtained from Löwenstein PAP devices via data download. Agreement between methods was assessed using Bland–Altman analysis, intraclass correlation coefficients (ICC), and weighted Kappa. Sensitivity and specificity for detecting inadequate control (AHI ≥7 events/hr) were calculated. The association between average leak and AHI discrepancy was examined using linear regression.

Results

Ninety-eight participants were included. PAP derived AHI values consistently underestimated polygraphy derived AHI across all time points. Agreement was poor, with ICCs of 0.02 at three months and 0.16 at six months, and low diagnostic agreement across severity categories at both three months ($_{K}$0.10, 95%CI: 0 to 0.26) and six months ($_{K}$0.09, 95%CI: 0 to 0.28). Bland–Altman plots demonstrated systematic underestimation with wide limits of agreement. Sensitivity for detecting inadequate control was very low at both time points (3 months 12% and 6 months 21%), despite high specificity (97% at 3 months and 95% at 6 months). Higher interface leak was significantly associated with greater underestimation of AHI (three months: slope = 0.45, R^2 = 0.18, $p < 0.0001$; six months: slope = 0.38, R^2 = 0.14, $p < 0.0001$).

Conclusion

PAP device downloads substantially underestimated residual AHI and demonstrated poor agreement with home cardiorespiratory polygraphy. These findings highlight important limitations of device derived AHI in real-world practice, particularly in the presence of interface leak. Where treatment adequacy is uncertain, confirmatory assessment with polygraphy remains necessary.

Introduction

Obstructive sleep apnoea (OSA) is a prevalent condition associated with significant morbidity and mortality[1]. The severity of OSA is conventionally assessed using the apnoea hypopnoea index (AHI) derived from polysomnography or cardiorespiratory polygraphy[2]. Continuous positive airway pressure (CPAP) therapy remains the gold standard treatment for moderate to severe OSA[3]. Monitoring residual AHI during therapy is essential for evaluating treatment effectiveness and guiding clinical decisions[4, 5]. Modern CPAP devices incorporate proprietary algorithms to estimate AHI from airflow signals, enabling remote monitoring and telemedicine approaches[6]. However, the validity and reliability of these device derived AHI values remain uncertain, particularly across different manufacturers and in real world conditions [7-11], much of the published data are over a decade old and relate to obsolete devices[12-22].

Previous studies have reported variable agreement between device-based AHI and AHI measured by polygraphy, with some demonstrating clinically acceptable correlation and others highlighting significant discrepancies [23-25]. Meta-analyses suggest that while mean bias may be small, limits of agreement and percentage error can be substantial, raising concerns about precision and diagnostic categorisation [26]. Furthermore, most published data focus on devices from a limited number of manufacturers (predominantly Philips Respironics, ResMed and Fisher and Paykel), leaving gaps in evidence for others. Given the increasing reliance on device downloads for clinical and research purposes, robust validation against gold standard methods is critical.

This study aims to evaluate the agreement between AHI obtained from CPAP device data downloads (Löwenstein, Germany) and AHI measured by home based cardiorespiratory polygraphy (Embletta MPR, Natus, USA) systems. By comparing continuous and categorical measures across multiple time points, we aimed to determine whether device derived AHI can be considered a reliable surrogate for polygraphy measurements in patients receiving CPAP therapy.

Methods

This was a secondary analysis of data from the 3dPiPPIn RCT (Clinical Trials registration: ISRCTN74082423)[27]. The study was approved by Hampshire B Research Ethics Committee (REC reference: 22/SC/0405), and all participants provided informed consent. The methods for the 3DPiPPIn trial have been reported previously[28]. In brief, the 3DPiPPIn RCT compared the clinical effectiveness of customised PAP therapy interfaces with conventional interfaces with the primary endpoint of AHI at six months. Patients were recruited from those under the care of the Sleep and Ventilation service at the Royal Free London NHS Foundation Trust (RFL), United Kingdom (UK).

Inclusion criteria

- Diagnosis of Sleep Disordered Breathing (SDB): AHI ≥15 events/hr.
- Patients naive to domiciliary PAP therapy.
- Age ≥18 years.
- Completed the 3DPiPPIn study

Exclusion criteria

- AHI <15 events/hr.
- Excessive facial hair which they were unwilling to shave.
- Age <18 years.

- Existing facial pressure ulcers.
- Unable to provide informed consent.
- Known allergy to silicone.
- Keloid scarring.
- Previous domiciliary PAP therapy.
- Failed to complete the 3DPiPPIn study

The PAP therapy devices used were Prisma Smart and Prisma 25S (Löwenstein Germany). Auto CPAP (Prisma Smart) or Auto S mode (Prisma 25S) were used in patients, with minimum pressure 4cmH_2O and maximum pressure of 20cmH_2O and 25cmH_2O respectively. These devices use Forced Oscillation Technique (FOT) technology to indirectly identify respiratory events and infer residual apnoea and hypopnea events via airflow and pressure signals within the breathing circuit. It is known that AHI presents with some night-to-night variability in patients with mild and severe OSA [29], but this has only been investigated during diagnostics and not in response to PAP therapy. To navigate this issue, patients undertook two-night home based cardiorespiratory polygraphy at three and six months concurrently with their PAP therapy, resulting in a total of four nights data, (Embletta MPR, Natus, USA). The sleep study system recorded multi channels: abdominal effort, thoracic effort, nasal pressure, nasal thermistor flow, snore, SpO_2, pulse rate, position and activity. Polygraphy was scored manually by one rater(SKM) according to AASM 2012 criteria for respiratory events[2]. AHI from CPAP therapy devices can be obtained from PAP therapy devices via a data download either via modem link or secure digital (SD) card into PrismaCloud (Löwenstein, Germany).

Study data were collected and managed using REDCap electronic data capture tools hosted at UCL [30, 31]. Baseline demographic and anthropometric measures were taken at baseline, three months and six months.

Analysis

Descriptive statistics were used to report patient demographics, mean ±SD was used for continuous data and frequency (proportion) for categorical data. Skewed data were described with median [interquartile range]. Bland–Altman plots were used to visually assess agreement and systematic bias between PAP device derived AHI and cardiorespiratory sleep study calculated AHI values, at each time point and night. These analyses were exploratory and descriptive; no pooled estimates across nights were calculated. A comparison of the AHI as measured from the PAP therapy devices and the AHI as measured from the cardiorespiratory sleep study was made using Intraclass Corelation Coefficient calculated using a two-way mixed effects model for absolute agreement. Furthermore, weighted quadratic Kappa Coefficient with 95% confidence intervals are reported to indicate levels of agreement across diagnostic thresholds (no SDB, mild, moderate and severe OSA). Weighted kappa coefficients were calculated using observed categories at each time point due to low frequencies in higher severity groups. A comparison of the polygraphy derived AHI and ODI was made using the same ICC and Kappa Coefficient methods . Although physiologically related, ODI can be derived from oximetry alone and is more accessible in routine clinical practice. This analysis was therefore included to explore whether ODI may act as a pragmatic surrogate for AHI when repeat cardiorespiratory sleep studies are not feasible.

Using cardiorespiratory-sleep study calculated AHI as the reference standard, sensitivity and specificity of PAP device derived AHI for identifying inadequate treatment control were calculated. Previous authors have previously found an AHI ≤7 to be indicative of adequate control of SDB on PAP therapy[23, 32]. Therefore, inadequate control was defined a priori as AHI

>7 events. Analyses were restricted to night one recordings to avoid non-independence arising from repeated measures within individuals. The statistical approaches used assume independent observations and do not readily accommodate clustered data. Inclusion of both nights without accounting for within-subject correlation would risk pseudo-replication and overestimation of precision. Night one was therefore selected as a single representative measure per participant, consistent with standard clinical assessment. The relationship between interface leak over the last seven days and levels of agreement between PAP device derived AHI and cardiorespiratory sleep study was assessed using simple linear regression for night one at three and six months to quantify the direction and magnitude of association. Linearity was evaluated through visual inspection of scatter plots. Data were analysed using Excel (Microsoft Office 365, Microsoft, USA), SPSS (version 29, IBM, USA) and R (version 4.4.1, The R Foundation for Statistical Computing, Austria) in R studio (version 2025.05.01, Posit software, USA).

Results

There were 98 participants who completed the study with 56 in the control group and 42 in the intervention group, demographics are displayed in Table 1. The primary diagnosis for all participants was obstructive sleep apnoea (OSA).

Residual AHI values obtained from PAP device downloads were consistently lower than those measured by cardiorespiratory sleep study at both three and six months (Table 2). Median AHI from cardiorespiratory sleep study was approximately twice that reported by the device, and these differences were statistically significant across all time points and nights.

Bland–Altman plots (Fig. 1) confirmed systematic bias, with PAP devices underestimating AHI by an average of one to three events/hr and wide limits of agreement, particularly at higher AHI values. Specifically. Bland–Altman analyses demonstrated mean differences (bias) of 5.6 (95% CI 3.6 to 7.6) at three months (night 1), 4.8 (95% CI 2.8 to 6.8) at three months (night 2), 3.7 (95% CI 2.5 to 4.9) at six months (night 1), and 3.6 (95% CI 1.8 to 5.5) at six months (night 2). Corresponding limits of agreement were -12.7 (95% CI -16.1 to -9.3) to 23.9 (95% CI 20.5 to 27.3) , -13.7(95% CI -17.1 to -10.2) to 23.4 (95% CI 19.9 to 26.8), -7.1 (95% CI -9.2 to -5.1) to 14.5 (95% CI 14.5 to 12.5), and -10.9 (95% CI -14.2 to -7.7) to 18.2 (95% CI 15.0 to 21.4) events/hr, respectively Agreement analysis demonstrated poor concordance between methods. Intraclass correlation coefficients (ICC) for night one were 0.02 (95% CI: –0.16 to 0.21) at three months and 0.16 (95% CI: –0.04 to 0.35) at six months, indicating negligible levels of agreement (Fig. 2).

There are diagnostic thresholds for AHI[2]; therefore, an analysis of level of agreement for AHI between PAP therapy data downloads and cardiorespiratory polygraphy with the data transformed into categorical data (No remnant OSA, Mild OSA, Moderate OSA and Severe OSA) using weighted (quadratic) Kappa Coefficient was also undertaken. Data were categorised based upon the first night cardiorespiratory polygraphy results (Table 3a and Table 3b).There were poor levels of agreement at both three months ($_{K}$0.10, 95%CI: 0 to 0.26) and six months (K0.09, 95%CI: 0 to 0.28)., reflecting frequent misclassification of severity.

Table 4 shows the descriptive results comparing polygraphy derived AHI and ODI at three and six months for both night one and night two.

ICC comparing AHI and ODI for night one were 0.89 (95% CI: 0.85 to 0.93) at three months and 0.84 (95% CI: 0.75 to 0.90 at six months), indicating excellent levels of agreement. For the categorical classification, there were excellent levels of agreement at three months ($_{K}$0.85, 95%CI: 0.71 to 0.92) and good levels of agreement at six months ($_{K}$0.65, 95%CI: 0.44, to 0.86).

Diagnostic performance was explored using an AHI threshold of >7 events/hr to define inadequate control. At three months, sensitivity was 12%, while specificity was 97%, indicating that the device rarely overestimated AHI but frequently failed to detect inadequate control.

Positive predictive value was 60% and negative predictive value 73%. At six months, sensitivity improved slightly to 21%, with specificity remaining high (95%), but overall diagnostic accuracy remained limited. These findings suggest that, when using an AHI threshold of ≥7 events/hr, the PAP device is highly specific but has limited sensitivity for detecting inadequate control of AHI and tends to underestimate residual AHI.

Leak analysis revealed an association between interface leak and AHI discrepancy. Scatter plots (Fig. 3) showed that higher average leak over the preceding seven days correlated with greater underestimation of AHI by the device. Linear regression demonstrated a significant relationship at both time points (three months: slope = 0.45, R^2 = 0.18, $p < 0.0001$; six months: slope = 0.38, R^2 = 0.14, $p < 0.0001$). However, the low R^2 values indicate that leak explains only a small proportion of the variance in AHI discrepancy, suggesting that additional factors contribute to measurement error.

Discussion

In this study, PAP therapy data downloads were found to be unreliable for measuring AHI. Poor levels of agreement were observed in both continuous (ICC) and categorical (Kappa Coefficient) data. indicating that device derived AHI values did not consistently reflect those obtained from the reference standard. However, the unbalanced nature of the categorical data should be taken into consideration during interpretation as this may influence the magnitude of estimates. These discrepancies were evident at both three and six months, suggesting that the lack of accuracy is not simply a transient initial treatment phenomenon but persists over time. . Previous authors have suggested a minimally clinically important difference of 5 events/hr [3]. In the present study, mean differences from the Bland–Altman analyses ranged from 3.6 to 5.6 events/hr, suggesting that, at a group level, the average bias between device-derived and polygraphy-derived AHI may not reach a clinically meaningful threshold. However, the limits of agreement were wide, indicating substantial variability at individual level. This suggests that, despite acceptable mean agreement, device-derived AHI may not reliably reflect residual disease in individual patients. Agreement deteriorated with increasing AHI, indicating a directionally dependent bias whereby the device increasingly underestimated AHI at higher levels of residual disease. This suggests reduced reliability in patients with more severe residual AHI. Agreement between device-derived and polygraphy-derived AHI was assessed using multiple complementary metrics, as no single method fully characterises agreement. The combined use of these approaches therefore provides a more robust and clinically meaningful assessment than any single metric. Our findings are consistent with wider evidence demonstrating marked variability in PAP device performance. A recent meta-analysis using Bland–Altman methods reported substantial variability in percentage error across CPAP manufacturers[25], highlighting that algorithmic differences contribute significantly to measurement inconsistency. Iftikhar et al.[26] report broad good levels of agreement in ResMed and Philips Respironics devices but since Iftikhar et al. meta-analysis did not include Löwenstein devices its findings may not be directly comparable to our results. This is an important omission because device algorithms differ in how they detect and classify respiratory events, and the performance of one manufacturer cannot be assumed to generalise to others.

Richter et al [25] specifically evaluated Löwenstein devices under controlled laboratory conditions and reported excellent agreement with simultaneously recorded polysomnography (mean difference of 0.9 events/hr (±0.9, 95% CI: ±1.2, p =0.108, ICC 0.961). In contrast, our results, conducted under real world conditions, demonstrated weak agreement (ICC 0.16, 95%CI: -0.04 to 0.35). This divergence may reflect methodological differences between studies. Unlike highly controlled laboratory titration, real world home PAP use introduces day-to--day variability, patient movement, environmental noise, and fluctuations in interface leak, all of which can interfere with device algorithms. These contextual factors are important, because in bench studies several authors have shown that device accuracy declines significantly in the

presence of high or unstable interface leak[33-38]. Interface leak may contribute to measurement error. Excessive leak can distort the airflow signal within the PAP circuit, reducing the accuracy of event detection and potentially leading to underestimation of residual events. In particular, leak may attenuate flow reductions or obscure changes in airflow that are used by device algorithms to identify hypopnoeas. This may explain the observed association between higher leak and greater discrepancy in AHI. Within the 3DPiPPIn study[27], high interface leak was observed in the intervention group, which might explain some of the reduced reliability between device derived AHI and AHI calculated through cardiorespiratory sleep study. Our own analysis confirmed that leak was strongly associated with greater underestimation of AHI (Fig. 3). Importantly, the specific Löwenstein devices used in this study have not been validated under high leak conditions, which may partly explain the poorer agreement -observed.

The observed discrepancies between device-derived and polygraphy-derived AHI may reflect fundamental differences in measurement principles and signal acquisition. Cardiorespiratory polygraphy estimates AHI using direct physiological signals, including airflow, respiratory effort, and oxygen desaturation. In contrast, PAP devices infer respiratory events indirectly from airflow and pressure signals within the circuit, using proprietary algorithms to detect flow limitation, apnoea, and hypopnoea.

Previous authors have compared AHI and ODI in the diagnostic phase of the OSA pathway[39-41]. In the United Kingdom, the National Institute for Health and Care Excellence has recently issued guidance on home testing devices incorporating newer technological developments[42]. To our knowledge, comparison of AHI and ODI during PAP therapy has not previously been reported. In this study, ODI demonstrated closer agreement with polygraphy-derived AHI than PAP device–derived AHI. However, it is important to note that polygraphy-derived AHI is not a gold standard and may underestimate events, particularly hypopnoeas associated with arousals without oxygen desaturation. As such, these findings should not be interpreted as evidence that ODI is a superior surrogate for AHI, but rather that it may provide a pragmatic and accessible measure that aligns more closely with polygraphy-derived estimates in this context. While oximetry alone may be useful in straightforward cases, more comprehensive assessment with polygraphy remains appropriate where more complex sleep-disordered breathing is suspected. These findings have potential implications for the practical monitoring of patients receiving PAP therapy.

In addition to poor agreement, diagnostic performance analyses revealed a consistent pattern of high specificity but very low sensitivity for detecting inadequate treatment control. This means that while the device rarely overestimated AHI, it frequently failed to identify patients with clinically significant residual SDB defined as an AHI >7events/hr. From a clinical perspective, this has important implications. Low device reported AHI may provide false reassurance and delay the identification and management of suboptimal therapy. As telemonitoring based models of sleep care become more common, understanding the limitations of device derived AHI will become increasingly important, particularly in settings where formal sleep studies are less readily available. In routine clinical practice, device derived AHI should not be interpreted in isolation. Instead, it should be considered alongside symptoms, concordance, interface leak data, and formal reassessment with our data supporting the use of oximetry as being adequate rather than more time consuming cardiorespiratory polygraphy.

These findings carry clear implications for clinical trials. When AHI is used as a primary or key secondary endpoint, gold-standard diagnostic tools such as polygraphy or polysomnography should remain the preferred methods of outcome assessment. Substituting device derived values risks outcome misclassification, reduced internal validity, and incorrect inference regarding treatment effect.

Further research is needed to determine whether the discrepancies observed reflect intrinsic limitations of the device algorithms, patient specific factors, or interactions between the two. Event level analyses may help identify which types of events are most frequently misclassified. In addition, studies evaluating device accuracy under controlled, experimentally varied leak conditions could provide important insights into the robustness of device algorithms across real world scenarios. Although one strength of the present analysis is that all participants used devices from the same manufacturer, thereby eliminating between manufacturer variability, this also limits generalisability.

Limitations

This study has several limitations. First, cardiorespiratory polygraphy rather than polysomnography was used as the reference standard, and while appropriate for assessing respiratory events, it does not measure sleep stages or arousals, which may influence AHI calculation. Second, although leak was examined, other factors such as pressure variability and device specific auto adjustment behaviour were not directly evaluated. Third, we did not conduct breath-by-breath or event level matching, which may have provided additional insight into the mechanisms of misclassification. Additionally, the post hoc analysis nature of this study add inherent bias that should be considered. Furthermore, the leak data was limited to an average over the last seven nights, night specific data for the nights of the polygraphy could have added further insights into the relationship between interface leak and AHI, future work should prospectively collect these data. Finally, because the study population was drawn from a single clinical trial cohort, the results may not generalise to broader OSA populations.

In conclusion, this study demonstrates that PAP device derived AHI underestimates residual sleep disordered breathing when compared with cardiorespiratory polygraphy. These findings emphasise that device algorithms, particularly under real world conditions influenced by leak and other sources of signal distortion, cannot currently substitute for gold-standard diagnostic tools when accurate assessment of treatment response is required. As remote monitoring and telemedicine become increasingly embedded within sleep pathways, careful interpretation of device reported AHI, supported by clinical judgement and confirmatory testing where necessary, will remain essential.

Statements

Acknowledgement

The authors would like to acknowledge the support of the Clinical Research Facility at the Royal Free Hospital for providing space for data collection. The authors would like to acknowledge the support of the Associate Principal Investigators: Lynne McDonnell, Nneoma Okoli and Lauren Geddes. The authors would like to acknowledge the support of pre-registration physiotherapy students on research placements in recruitment and data collection: Amelia Simcock Brown, Laurie Parks, Sonia Waika, Ciaran Brennan and Adam Turrin. The authors would like to acknowledge and thank Tolu Williams for supporting trial governance.

Statement of Ethics

Study approval statement:

This protocol was reviewed by the Hampshire B Research Ethics Committee (REC reference: 22/SC/0405), who granted a favourable ethical opinion.

Consent to participate statement:

Participants who were unable to provide informed consent were excluded from the study. The research team obtained written informed consent before the study baseline measurements and after the participant's information sheet review. Potential participants were given a minimum of

24 hours to consider if they wished to participate in the study. Consent was obtained from participants who were non-English-speaking using a translator.

Conflict of Interest Statement

SKM: Educational grants from Dolby Vivisol.SG: Nil to declare FG: Nil to declare. DR: Nil to declare. CK: Nil to declare. SH: Nil to declare. EM: Nil to declare. SS: Nil to declare. SM: consultancy for Fisher and Paykel, educational and small research grants from Dolby Vivisol.

Funding Sources

This project is funded by the NIHR Doctoral Fellowship (NIHR award: 302337). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. The funder had no role in the design, data collection, data analysis, and reporting of this study. The study was sponsored by University College London Joint Research Office.

Author Contributions

Conceptualisation: SKM Methodology: SKM, SM, CK, EM, SS, EM, SH, DR Resources: SKM, FG, SA Data Curation: SKM, FG, SA, DR Writing - Original Draft: SKM, Writing - Review and Editing: All authors Supervision: SM, EM, SS, CK, SH, DR Project administration: SA, FG, SKM Funding acquisition: SKM, FG, EM, SH, SM, SS Guarantor: Swapna Mandal

Data Availability Statement

Data supporting this study cannot be made available due to ethical, legal and commercial reasons. Further enquires can be directed to the corresponding author.

References

1. Lyons MM, Bhatt NY, Pack AI, Magalang UJ. Global burden of sleep-disordered breathing and its implications. Respirology. 2020;25(7):690-702.

2. Berry RB, Budhiraja R, Gottlieb DJ, Gozal D, Iber C, Kapur VK, et al. Rules for scoring respiratory events in sleep: update of the 2007 AASM Manual for the Scoring of Sleep and Associated Events. Deliberations of the Sleep Apnea Definitions Task Force of the American Academy of Sleep Medicine. Journal of clinical sleep medicine : JCSM : official publication of the American Academy of Sleep Medicine. 2012 Oct 15;8(5):597-619.

3. National Institute for Clinical Excellence. Obstructive sleep apnoea/hypopnoea syndrome and obesity hypoventilation syndrome in over 16s. NICE guideline [NG202]. https://www.nice.org.uk/guidance/ng202: NICE; 2021.

4. Patil SP, Ayappa IA, Caples SM, Kimoff RJ, Patel SR, Harrod CG. Treatment of Adult Obstructive Sleep Apnea With Positive Airway Pressure: An American Academy of Sleep Medicine Systematic Review, Meta-Analysis, and GRADE Assessment. Journal of Clinical Sleep Medicine. 2019;15(02):301-34.

5. Read N, Anderson KN, Craig S, Kendrick A, Quinnell T, J. S, et al. The Optimal Sleep Pathway: Towards better care for patients with sleep conditions. NHS England; 2025.

6. Verbraecken J, Amodio E, Basoglu OK, Bellazzi R, Bradicich M, Bruyneel M, et al. European Respiratory Society statement on advanced telemedicine for obstructive sleep apnoea (e-Sleep). European Respiratory Journal. 2025;66(5):2500557.

7. Agrawal R, Wang JA, Ko AG, Getsy JE. A real-world comparison of apnea–hypopnea indices of positive airway pressure device and polysomnography. PloS one. 2017;12(4):e0174458.

8. Gagnadoux F, Pevernagie D, Jennum P, Lon N, Loiodice C, Tamisier R, et al. Validation of the System One RemStar Auto A-Flex for Obstructive Sleep Apnea Treatment and Detection of Residual Apnea-Hypopnea Index: A European Randomized Trial. Journal of Clinical Sleep Medicine. 2017;13(02):283-90.

9. Khirani S, Delord V, Olmo Arroyo J, De Sanctis L, Frapin A, Amaddeo A, et al. Can the analysis of built-in software of CPAP devices replace polygraphy in children? Sleep Medicine. 2017 2017/09/01/;37:46-53.

10. Mihai R, Ellis K, Davey MJ, Nixon GM. Interpreting CPAP device respiratory indices in children. Journal of Clinical Sleep Medicine. 2020;16(10):1655-61.

11. Ni Y-N, Thomas RJ. A longitudinal study of the accuracy of positive airway pressure therapy machine-detected apnea-hypopnea events. Journal of Clinical Sleep Medicine. 2022;18(4):1121-34.

12. Prasad B, Carley DW, Herdegen JJ. Continuous positive airway pressure device-based automated detection of obstructive sleep apnea compared to standard laboratory polysomnography. Sleep and Breathing. 2010 2010/06/01;14(2):101-07.

13. Ueno K, Kasai T, Brewer G, Takaya H, Maeno K-i, Kasagi S, et al. Evaluation of the Apnea-Hypopnea Index Determined by the S8 Auto-CPAP, a Continuous Positive Airway Pressure Device, in Patients with Obstructive Sleep Apnea-Hypopnea Syndrome. Journal of Clinical Sleep Medicine. 2010;06(02):146-51.

14. Denotti AL, Wong KKH, Dungan GC, Gilholme JW, Marshall NS, Grunstein RR. Residual sleep-disordered breathing during autotitrating continuous positive airway pressure therapy. European Respiratory Journal. 2011;39(6):1391-97.

15. Berry RB, Kushida CA, Kryger MH, Soto-Calderon H, Staley B, Kuna ST. Respiratory Event Detection by a Positive Airway Pressure Device. Sleep. 2012;35(3):361-67.

16. Ikeda Y, Kasai T, Kawana F, Kasagi S, Takaya H, Ishiwata S, et al. Comparison between the Apnea-Hypopnea Indices Determined by the REMstar Auto M Series and Those Determined by Standard In-

Laboratory Polysomnography in Patients with Obstructive Sleep Apnea. Internal Medicine. 2012;51(20):2877-85.

17. Cilli A, Uzun R, Bilge U. The accuracy of autotitrating CPAP-determined residual apnea–hypopnea index. Sleep and Breathing. 2013 2013/03/01;17(1):189-93.

18. Kim D-E, Hwangbo Y, Bae JH, Yang KI. Accuracy of residual apnea-hypopnea index obtained using the continuous positive airway pressure device: application of new version 2.0 scoring rules for respiratory events during sleep. Sleep and Breathing. 2015 2015/12/01;19(4):1335-41.

19. Li QY, Berry RB, Goetting MG, Staley B, Soto-Calderon H, Tsai SC, et al. Detection of Upper Airway Status and Respiratory Events by a Current Generation Positive Airway Pressure Device. Sleep. 2015;38(4):597-605.

20. Nigro CA, González S, Arce A, Aragone MR, Nigro L. Accuracy of a novel auto-CPAP device to evaluate the residual apnea-hypopnea index in patients with obstructive sleep apnea. Sleep and Breathing. 2015 2015/05/01;19(2):569-78.

21. Baek JH, Jeon J-Y, Lee S-A. Accuracy of the Auto Scoring by the S9 CPAP in Patients with Obstructive Sleep Apnea. Sleep Med Res. 2016 6;7(1):26-32.

22. Reiter J, Zleik B, Bazalakova M, Mehta P, Thomas RJ. Residual Events during Use of CPAP: Prevalence, Predictors, and Detection Accuracy. Journal of Clinical Sleep Medicine. 2016;12(08):1153-58.

23. Fernandez Alvarez R, Rabec C, Rubinos Cuadrado G, Cascon Hernandez J, Rodriguez P, Georges M, et al. Monitoring Noninvasive Ventilation in Patients with Obesity Hypoventilation Syndrome: Comparison between Ventilator Built-in Software and Respiratory Polygraphy. Respiration; international review of thoracic diseases. 2017;93(3):162-69.

24. Iftikhar IH, Finch CE, Shah AS, Augunstein CA, Ioachimescu OC. A meta-analysis of diagnostic test performance of peripheral arterial tonometry studies. Journal of Clinical Sleep Medicine. 2022;18(4):1093-102.

25. Richter M, Schroeder M, Domanski U, Schwaibold M, Nilius G. Reliability of respiratory event detection with continuous positive airway pressure in moderate to severe obstructive sleep apnea — comparison of polysomnography with a device-based analysis. Sleep and Breathing. 2023 2023/08/01;27(4):1639-50.

26. Iftikhar IH, BaHammam A, Jahrami H, Ioachimescu O. Accuracy of residual respiratory event detection by CPAPs: a meta-analysis. Sleep & breathing = Schlaf & Atmung. 2023 Oct;27(5):1759-68.

27. Mansell SK, Mandal S, Gowing F, Ridout D, Kilbride C, Olsen O, et al. Clinical impact of customised positive airway pressure (PAP) therapy interfaces (3DPiPPIn): a single site randomised controlled trial. Thorax. 2026:thorax-2025-224135.

28. Mansell SK, Mandal S, Ridout D, Olsen O, Gowing F, Kilbride C, et al. Clinical impact of customised positive airway pressure (PAP) therapy interfaces versus usual care in the treatment of patients with sleep-disordered breathing (3DPiPPIn): a randomised controlled trial protocol. BMJ Open. 2024;14(11):e087234.

29. Sforza E, Roche F, Chapelle C, Pichot V. Internight variability of apnea-hypopnea index in obstructive sleep apnea using ambulatory polysomnography. Frontiers in physiology. 2019;10.

30. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. Journal of biomedical informatics. 2009;42(2):377-81.

31. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: building an international community of software platform partners. Journal of biomedical informatics. 2019;95:103208.

32. Aarrestad S, Qvarfort M, Kleiven AL, Tollefsen E, Skjonsberg OH, Janssens JP. Diagnostic accuracy of simple tools in monitoring patients with chronic hypoventilation treated with non-invasive ventilation; a prospective cross-sectional study. Respiratory medicine. 2018 Nov;144:30-35.

33. Luján M, Sogo A, Pomares X, Monsó E, Sales B, Blanch L. Effect of Leak and Breathing Pattern on the Accuracy of Tidal Volume Estimation by Commercial Home Ventilators: A Bench Study. Respiratory care. 2013;58(5):770-77.

34. Sogo A, Montanyà J, Monsó E, Blanch L, Pomares X, Lujàn M. Effect of dynamic random leaks on the monitoring accuracy of home mechanical ventilators: a bench study. BMC pulmonary medicine. 2013 2013/12/10;13(1):75.

35. Luján M, Sogo A, Grimau C, Pomares X, Blanch L, Monsó E. Influence of Dynamic Leaks in Volume-Targeted Pressure Support Noninvasive Ventilation: A Bench Study. Respiratory care. 2015;60(2):191-200.

36. Zhu K, Rabec C, Gonzalez-Bermejo J, Hardy S, Aouf S, Escourrou P, et al. Combined effects of leaks, respiratory system properties and upper airway patency on the performance of home ventilators: a bench study. BMC pulmonary medicine. 2017 2017/11/21;17(1):145.

37. Lebret M, Fresnel E, Prouvez N, Zhu K, Kerfourn A, Richard JC, et al. Responses of Bilevel Ventilators to Unintentional Leak: A Bench Study. Healthcare (Basel). 2022 Nov 30;10(12).

38. Richard M, Fresnel E, Mallet JP, Gilson R, Kerfourn A, Patout M, et al. Performances of Auto-CPAP Devices Under Real-Life Leak Patterns: A High-Fidelity Modeling Approach. Arch Bronconeumol. 2025 Jan 18.

39. Rashid NH, Zaghi S, Scapuccin M, Camacho M, Certal V, Capasso R. The Value of Oxygen Desaturation Index for Diagnosing Obstructive Sleep Apnea: A Systematic Review. The Laryngoscope. 2021;131(2):440-47.

40. Incerti Parenti S, Bartolucci ML, Fiordelli A, Gigola P, Paganelli C, Alessandri-Bonetti G. The Diagnostic Accuracy of Overnight Oximetry for Pediatric Obstructive Sleep Apnea: A Systematic Review and Meta-Analysis. Applied Sciences. 2024;14(22):10208.

41. da Silva Dantas EL, Stelzer FG, Bernardo WM, Eckeli AL. Oximetry-based devices in diagnosis of obstructive sleep apnea: A systematic review and meta-analysis. Sleep Medicine Reviews. 2025 2025/10/01/;83:102139.

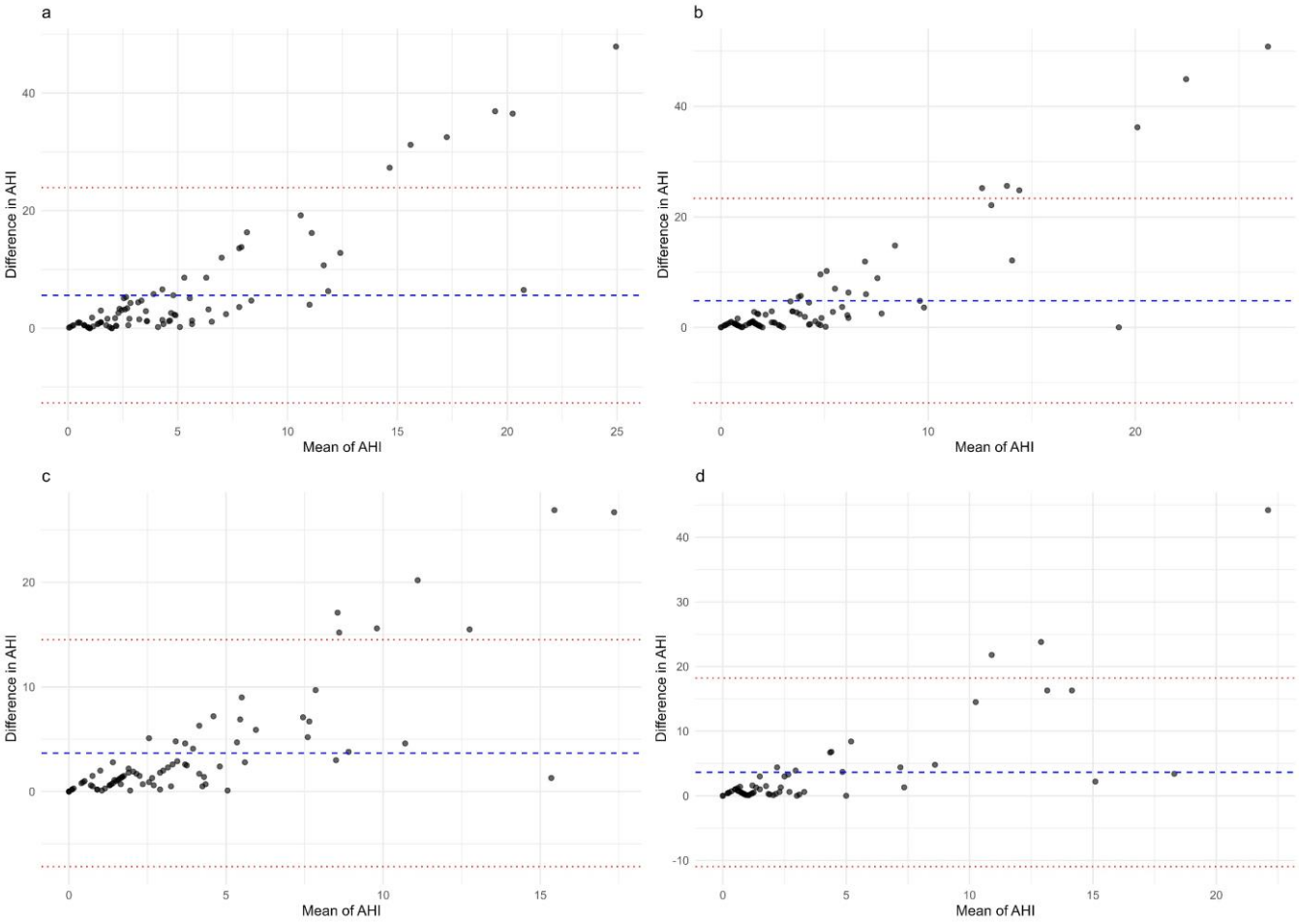
42. National Institute for Health and Care Excellence. Home-testing devices for diagnosing obstructive sleep apnoea hypopnoea syndrome. In: National Institute for Health and Care Excellence, editor. https://www.nice.org.uk/guidance/htg735: National Institute for Health and Care Excellence; 2024.

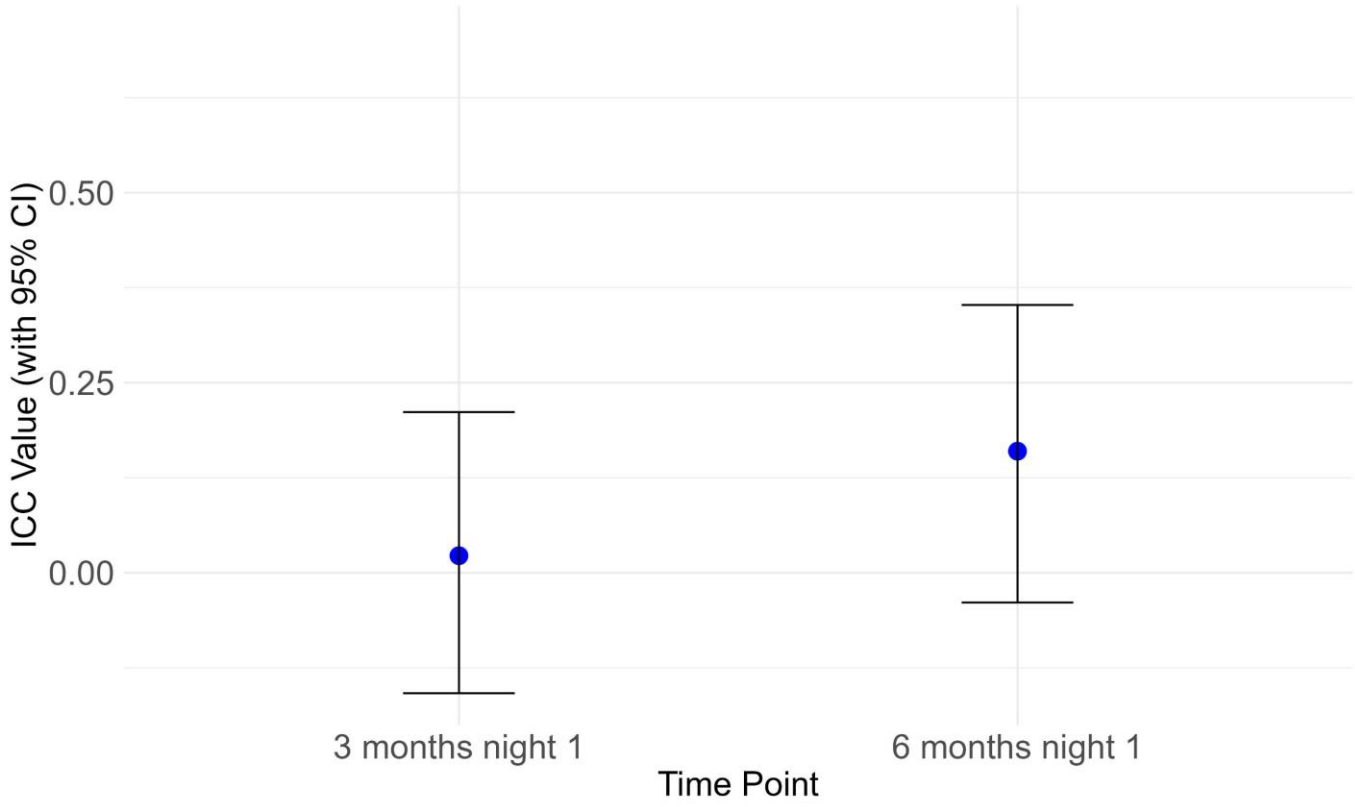
Figure Legends

Fig. 1. Bland–Altman plots comparing PAP derived AHI with cardiorespiratory sleep study AHI

Fig. 2. Intraclass Correlation Coefficient for Apnoea Hypopnea Index between PAP therapy downloads and cardiorespiratory polygraphy

Fig. 3. Scatter plot of average leak during the last 7 days against the difference in AHI derived from PAP therapy devices and the AHI from cardiorespiratory sleep study scoring.





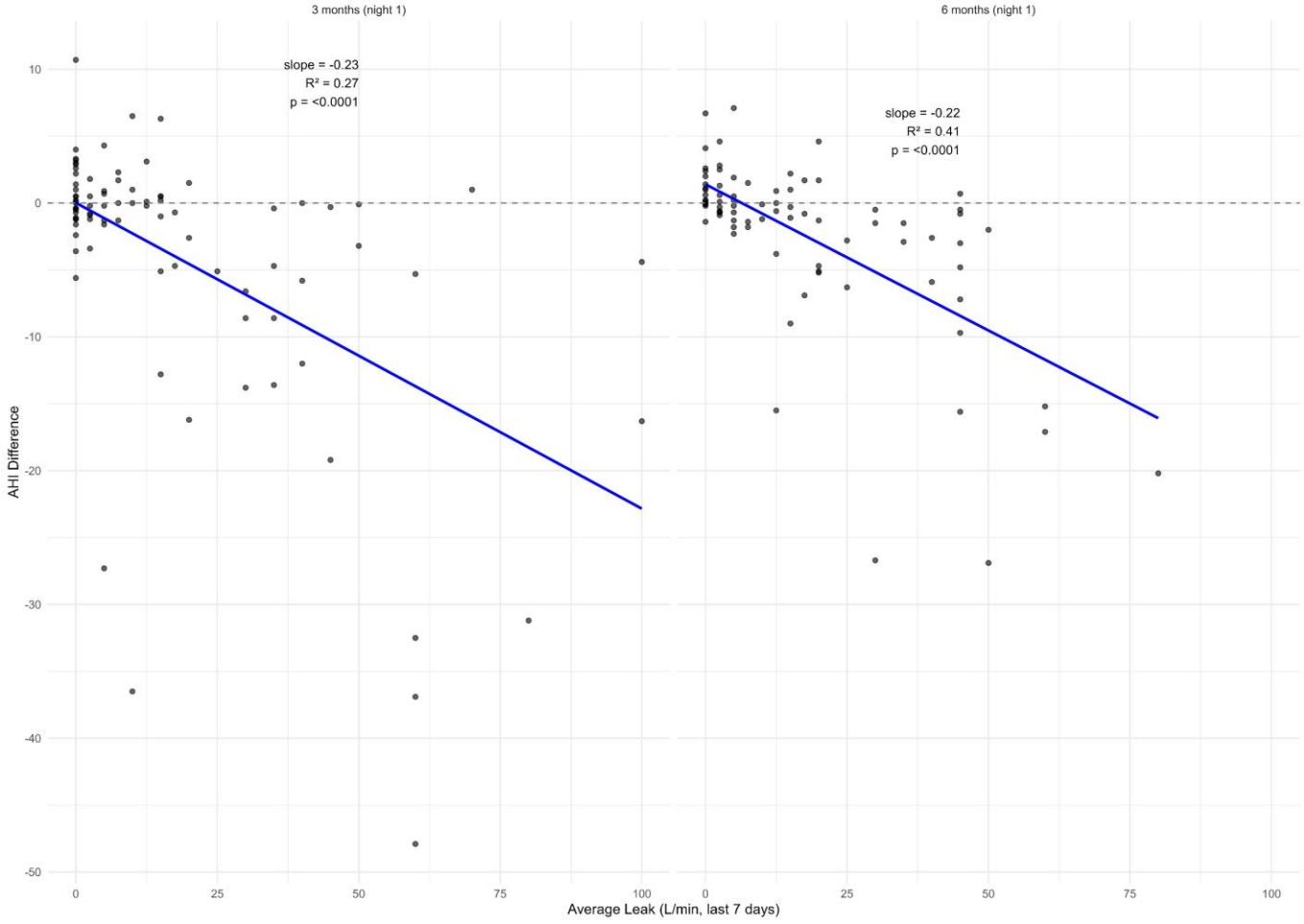


Table 1. Demographics of the participants

Demographic	
Age (years) Mean ±SD	54.7 ±11.8
Sex n (%)	
Male	76(78)
Female	22(22)
Ethnicity n (%)	
Asian	22(23)
Black	10(10)
Caucasian	66(67)
SpO_2 Mean ±SD	96.4 ±1.6
BMI Kg/m^2 Mean ±SD	35 ±6.0
Smoking Status n (%)	
Current smoker	6(6)
Ex-smoker	46(47)
Never smoker	46(47)
Baseline sleep study Mean ±SD	
AHI (events/hr)	41.2 ±18.9
ODI (events/hr)	37.5 ± 19.8
Mean SpO_2	91.9 ± 2.5
Total Sleep Time SpO_2 <90% (%)	25.2 ± 21.9
Epworth Sleepiness Score Median [IQR]	7.7±5.5
Sleepiness-Wakefulness Inability and Fatigue Score Median [IQR]	11.5 ± 8.6

Table 2 AHI from PAP therapy data download versus cardiorespiratory polygraphy

	Night 1			Night 2		
Timepoint	**AHI from cardiorespiratory polygraphy**	**AHI from PAP therapy device**	**Median diff. [95% CI] p value**	**AHI from cardiorespiratory polygraphy**	**AHI from PAP therapy device**	**Median diff. [95%CI] p value**
Three months	4.2 [1.3 to 8.1]	2.0 [1.0 to 4.0]	-1.5 [-3.1 to -0.4] *0.002*	2.7 [1.3 to 6.5]	2.0 [1.0 to 3.3]	-1.1 [-2.4 to -0.3] *0.003*
Six months	2.5 [1.4 to 5.8]	2.0 [1.0 to 3.5]	-0.9 [-2.0 to -0.2] *0.007*	2.0 [0.7 to 7.0]	1.0 [1.0 to 3.0]	-0.7 [-2.1 to -0.1] *0.020*

Table 3a

3months night 1 confusion matrix

Polygraphy derived (rows)/PAP derived(columns)	No remnant OSA	Mild OSA	Moderate OSA	Severe OSA
No remnant OSA	37	6	0	0
Mild OSA	13	8	2	0
Moderate OSA	4	1	1	0
Severe OSA				

Table 3b

6months night 1 confusion matrix

Polygraphy derived (rows)/PAP derived(columns)	No remnant OSA	Mild OSA	Moderate OSA	Severe OSA
No remnant OSA	45	8	0	0
Mild OSA	9	5	0	0
Moderate OSA	5	1	0	0
Severe OSA	1	0	0	0

Table 4 AHI versus ODI derived from cardiorespiratory polygraphy

	Night 1			Night 2		
Timepoint	**AHI from limited cardiorespiratory polygraphy**	**ODI from cardiorespiratory polygraphy**	**p value**	**AHI from cardiorespiratory polygraphy**	**ODI from cardiorespiratory polygraphy**	**p value**
Three months	4.2 [1.3 to 8.1]	3.2 [1.2 to 8.9]	0.244	2.7 [1.3 to 6.5]	2.9 [0.9 to 9.4]	*0.032*
Six months	2.5 [1.4 to 5.8]	3.4 [0.9 to 10.4]	*0.003*	2 [0.7 to 7.0]	2.7 [1.0 to 8.0]	*0.020*