The diagnostic method has a strong influence on classification of obstructive sleep apnea

PIERRE ESCOURROU[1], LUDGER GROTE[2], THOMAS PENZEL[3], WALTER T. MCNICHOLAS[4], JOHAN VERBRAECKEN[5], ROSA TKACOVA[6], RENATA L. RIHA[7], JAN HEDNER[2] and ON BEHALF OF THE ESADA STUDY GROUP

[1]Centre de Médecine du Sommeil, Hôpital Béclère, Clamart, France; [2]Sleep Disorder Center, Pulmonary Medicine, Sahlgrenska University Hospital, Gothenburg, Sweden; [3]Charité - Universitaetsmedizin Berlin, Schlafmedizinisches Zentrum, Berlin, Germany; [4]Pulmonary and Sleep Disorders Unit, St. Vincent's University Hospital, School of Medicine and Medical Science, University College Dublin, Dublin, Ireland; [5]Pulmonary Medicine and Multidisciplinary Sleep Disorders Centre, Antwerp University Hospital and University of Antwerp, Edegem, Belgium; [6]Department of Respiratory Medicine, Faculty of Medicine, P.J. Safarik University and L. Pasteur Teaching Hospital, Kosice, Slovakia; [7]Department of Sleep Medicine, Royal Infirmary of Edinburgh, Edinburgh, UK

SUMMARY

Polygraphy (PG) and polysomnography (PSG) are used in clinical settings in Europe for diagnosing obstructive sleep apnea (OSA), but their equivalence in unselected clinical cohorts is unknown. We hypothesized that the method would affect both diagnostic outcomes and disease severity stratification. Data from 11 049 patients in the multi-centre European Sleep Apnea Cohort (ESADA) with suspected OSA (male and female, aged 18–80 years) were used in two groups of patients to compare PG ($n = 5745$) and PSG ($n = 5304$). Respiratory events were scored using the 2007 American Association of Sleep Medicine (AASM) criteria. In subjects who underwent PSG, mean apnea–hypopnea index (AHI) using sleep time (AHI_{PSG} 31.0 ± 26.1 h^{-1}) and total analysed time (TAT) (AHI_{TAT} 24.7 ± 22.0 h^{-1}) were higher than in subjects who underwent PG (AHI_{PG} 22.0 ± 23.5 h^{-1}) ($P < 0.0001$). The oxygen desaturation index (ODI) was lower in subjects investigated with PG (ODI_{PG} 18.4 ± 21.7 h^{-1}) compared to subjects investigated with PSG (ODI_{PSG} 23.0 ± 25.3 h^{-1}) but not different when the PSG was indexed by TAT (ODI_{TAT} 18.6 ± 21.4 h^{-1}, $P < 0.65$). The proportion of patients with an AHI ≥ 15 was 64% in the subjects who underwent PSG and 47% in the subjects who underwent PG ($P < 0.001$). Overall, patients investigated using PG are likely to have a 30% lower AHI on average, compared to patients investigated by PSG. This study suggests that PG interpreted using standard guidelines results in underdiagnosis and misclassification of OSA. We advocate the development of PG-specific guidelines for the management of OSA patients.

INTRODUCTION

The severity of obstructive sleep apnea (OSA) is quantified conventionally by the apnea–hypopnea index (AHI), which reflects the number of apneas and hypopneas per hour of electroencephalography (EEG)-measured sleep. Specific threshold levels are used to establish the severity of the disorder, particularly in a clinical context (AASM, 1999; Epstein *et al.*, 2009). The AHI determines reimbursement of

continuous positive airway pressure (CPAP) therapy in many health-care systems. For instance, an AHI of 15 h^{-1}, or lower if combined with relevant comorbidities and/or daytime sleepiness, is the minimum requirement for prescription of CPAP in 12 of 21 European countries (Fietze *et al.*, 2011; Fischer *et al.*, 2012).

Apnea–hypopnea index is derived after scoring data obtained using multi-channel polysomnography (PSG) performed in specialized sleep laboratories. Currently, PSG is

considered the gold standard method for diagnosing OSA (Trikalinos and Lau, 2007). Conversely, apneas and hypopneas recorded using polygraphy (PG) are reported as a total using the approximate patient-reported denominator of time in bed, rather than EEG-measured sleep (A+H/time in bed). Several studies have advocated that EEG-based studies are not necessary for the diagnosis and therapeutic monitoring of sleep-disordered breathing (Douglas *et al.*, 1992; Masa *et al.*, 2011; Mulgrew *et al.*, 2007; Rosen *et al.*, 2012).

Despite the paucity of validated outcome studies in the area of PG, many health-care systems have adopted this less resource-consuming method for the clinical assessment of OSA.

The European Sleep Apnea Database (ESADA) is a network of 26 sleep centres in Europe. The ESADA collaborative documents and evaluates multiple aspects of OSA management (Hedner *et al.*, 2011). Consecutive patients referred with suspected OSA are enrolled and studied using PSG or PG in accordance with local guidelines at the different centres. Both centre-specific and centralized quality control is applied to information entered into the database.

In controlled studies aimed largely at validating new PG equipment, results for PG and PSG are not equivalent unless 'pretest probability' of OSA is elevated, but to date there is no consensus on the evaluation of this disease probability (Ayas *et al.*, 2010; Chesson *et al.*, 2003; Collop *et al.*, 2011; Corral-Penafiel *et al.*, 2013; Trikalinos and Lau, 2007; Trikalinos *et al.*, 2007) . We hypothesized that the type of test would affect both diagnostic outcomes and disease severity stratification in patients with suspected OSA, and aimed at comparing the AHI results obtained in patients investigated by PG with a second group of patients investigated by PSG used routinely among the ESADA centres to reflect the actual clinical settings in Europe.

METHODS

The ESADA study sample

The European Sleep Apnea Cohort (ESADA) study is a pan-European, multi-centre, prospective study involving 26 sleep clinics in 15 European countries and Israel (20 centres were university-affiliated sleep clinics). The rationale and investigative techniques underlying the establishment of ESADA have been discussed in detail elsewhere (Hedner *et al.*, 2011). The overall study objective is the prospective evaluation of a large clinically representative cohort of subjects with suspected sleep-disordered breathing in order to identify cross-sectional and longitudinal associations with cardiovascular and metabolic morbidity and mortality. ESADA employs a web-based collection platform to facilitate transfer of data from individual centres to the central database at the University of Gothenburg, Sweden.

In brief, patients suspected of OSA (male or female, aged 18–80 years) were eligible for inclusion into the study. Exclusion criteria included treated OSA and a limited life expectancy due to illness unrelated to OSA (e.g. HIV, advanced renal disease, uncontrolled malignancies), as well as documented alcohol or drug abuse within 1 year prior to inclusion in the study. The ESADA protocol states that any sleep investigations should be performed in accordance with local practice. Thus, recordings obtained using either PG or PSG are included. The data report contains standardized modules for a range of anthropometric measures, medical history, ongoing medication, patient-rated daytime symptoms and sleep data. All data are transferred electronically, stored and reviewed continuously for quality in a common database (the ESADA database). The total sample reviewed for this study was 12 548 patients from 26 centres. Data from centres reporting fewer than 50 patients ($n = 3$, total 81 patients) were discarded from the analysis. An additional 1328 incomplete data files, e.g. poor recording quality or missing data, were rejected, and the final sample identified for the analysis comprised 11 049 patients. Research ethics committee approval for the study was obtained at each of the participating centres. Informed consent was obtained from all participants.

Sleep data

The database accepts data from either PG or PSG recordings and the protocol states that PG recordings should include a minimum of four channels, not including devices with one channel EEG [level III devices in accordance with the American Sleep Disorders Association (ASDA)] (Ferber *et al.*, 1994). All sleep data were examined visually and edited manually according to prescribed protocol definitions prior to entry into the database. The protocol mandates the AASM 2007 scoring rules (Iber *et al.*, 2007), which stated 'recommended' rules for hypopnea as 'nasal pressure signal excursions drop by $\geq 30\%$ of baseline, lasting at least 10 s, with a $\geq 4\%$ desaturation from pre-event baseline' and 'alternative' rule as 'nasal pressure signal excursions drop by $\geq 50\%$ of baseline, lasting at least 10 s, with a $\geq 3\%$ desaturation from pre-event baseline or the event is associated with arousal'. As no arousals are scored in PG due to lack of EEG, respiratory events consistent with a hypopnea without desaturation were ignored. The total AHI on PSG AHI or apneas/hypopneas per time in bed (PG) were recorded in the database. Indices of respiratory events were calculated using the denominator of EEG-recorded total sleep time (TST) in the PSG recordings or of time in bed between lights-out and lights-on in PG recordings. The latter is labelled total analysis time (TAT) for the purposes of this study (periods of upright position or non-interpretable data were not excluded). TAT was derived from the PSG recordings to allow for comparison with PG derived indices. Minimum PG reporting requirements were: total analysis time, subjective sleep time, apnea + hypopnea per time in bed, oxygen desaturation index (ODI 4% or ODI 3% according to the scoring rule used), mean peripheral capillary oxygen saturation (SpO_2) and lowest SpO_2. The following additional indices were

reported for PSG recordings: total sleep time, sleep efficiency, percentage of different sleep stages, periodic limb movement (PLM) index, PLM arousal index, respiratory-related arousal index and spontaneous arousal index. Each recording was graded in terms of technical quality using a four-level rating scale (excellent: no missing data; good: partial or complete loss of data in one to two channels; acceptable: three to four channels involved; poor: more than four channels involved). The quality of recordings was classified as 'excellent' or 'good' in 99 and 98.5% of PSGs and PGs included in the study, respectively.

Statistics

Data analysis was undertaken using IBM SPSS version 20. Between groups, comparison was performed using the χ^2 test (or Fisher's exact test where any expected cell value was less than 10) and paired and unpaired Student's t-tests. Results are reported as total number, percentage and mean $\pm$ standard deviation (SD) unless stated otherwise.

Two-way analysis of variance (ANOVA) was used to compare AHI between PG and PSG according to the hypopnea rule used for diagnosis (recommended or alternative). Forward multiple linear regression analyses were performed using AHI as the dependent variable with 'method used' as the confounding factor in the unadjusted model, with further adjustment for age, body mass index (BMI), gender and cardiovascular-, pulmonary- and sleep-related comorbidities. All tests were two-tailed and statistical significance was taken at $P = 0.05$.

RESULTS

Clinical characteristics of the cohort according to type of sleep study are shown in Table 1. Age and BMI distributions were similar among patients investigated using either method, but the proportion of males was higher in the PSG group ($P < 0.0001$). There was also a high proportion of patients with comorbidities; for example, cardiovascular diseases were reported in 52.7 and 46% of the PSG and PG groups, respectively. Pulmonary disease [asthma, chronic obstructive pulmonary disease (COPD), restrictive pulmonary disease] was present in 14.0 and 11.7% of PSG and PG groups, respectively. COPD was the most prevalent disorder (4.1 and 5.6% in the PSG and PG groups, respectively). Metabolic diseases were also common.

As shown in Fig. 1, individual centres contributed between 0.6 and 14.3% of the total sample. PG recordings accounted for 5745 patients (52%) in the total sample compared with 5304 patients (48%) with PSG recordings. Diagnostic practice varied considerably between centres, with some centres using PG or PSG exclusively. The use of the recommended and alternative hypopnea scoring criteria varied among centres. The alternative rule was used in scoring respiratory hypopneas in 87.7% of PSGs, while the recommended rule was applied in the scoring of 83.4% of PGs.

Table 1 Patient characteristics and comorbidities in the groups investigated with polygraphy and polysomnography

Test	Polysomnography	Polygraphy	P
n	5304 (48%)	5745 (52%)	
Gender			
Male	73%	70%	<0.0001
Female	27%	30%	
Age (years)	51.7 ± 12.4	51.9 ± 13.1	NS
BMI (kg m²)	31.0 ± 6.4	31.2 ± 6.8	NS
ESS (/24)	9.8 ± 5.3	9.9 ± 5.2	NS
CV disease total %	52.7	46.0	<0.0001
Cardiac failure %	2.5	1.9	0.02
Ischaemic heart disease %	9.7	6.9	0.051
Stroke and TIA %	2.3	2.8	0.23
Hypertension %	45.5	38.8	<0.0001
Pulmonary disease %	14	11.7	0.0002
COPD %	4.1	5.6	0.0005
Respiratory failure %	1.4	0.3	<0.0001
Metabolic disease %	34.1	24.9	<0.0001
Diabetes NID %	9.7	9.2	0.55
Diabetes ID %	2.1	3.4	<0.0001
Any sleep disorder other than SDB %	13.6	17.5	<0.001
Insomnia %	4.6	3.6	0.008
RLS %	2.1	2.4	0.82

RLS, restless legs syndrome; SDB, sleep-disordered breathing; NID, non-insulin-dependent; COPD, chronic obstructive pulmonary disease; TIA, transient ischaemic attack; CV, cardiovascular; BMI, body mass index; NS, not significant; ESS, Epworth Sleepiness Scale; SDB, sleep-disordered breathing.

The reported AHI distribution differed between the PSG and PG groups: 50% of patients undergoing PSG had an AHI_{PSG} above 23.4 h⁻¹, whereas the median AHI_{PG} was 13.6 h⁻¹ in patients investigated using PG (Fig. 2a). The proportion of patients with an $AHI_{PSG} \geq 15$ was 65% in the patients who underwent PSG and 46% in those who underwent PG ($P < 0.001$). When the AASM-defined OSA severity ranges were applied (0–4.9, 5–14.9, 15–29.9 and ≥ 30 h⁻¹), prevalence values obtained in patients recorded by PSG were significantly lower than in patients recorded by PG for AHI ranges below 15 ($P < 0.0001$) and higher for AHI above 30 ($P < 0.001$) (Fig. 2b). When the AHI distributions using the two methods were compared for the same percentage of the population, AHI_{PSG} of 5 h⁻¹ was found to correspond to an AHI_{PG} of 2 h⁻¹, an AHI_{PSG} of 15 h⁻¹ to an AHI_{PG} of 10 h⁻¹ and an AHI_{PSG} of 30 h⁻¹ to an AHI_{PG} of 20 h⁻¹.

Mean AHI_{PSG} was higher compared with AHI_{PG} (31.0 ± 26.1 versus 22.0 ± 23.5 h⁻¹, respectively ($P < 0.0001$). ODI_{PSG} was also higher (23.0 ± 25.3 h⁻¹) than ODI_{PG} 18.4 ± 21.7 h⁻¹ ($P < 0.0001$). However, the difference between AHI and ODI scores was greater in the patients
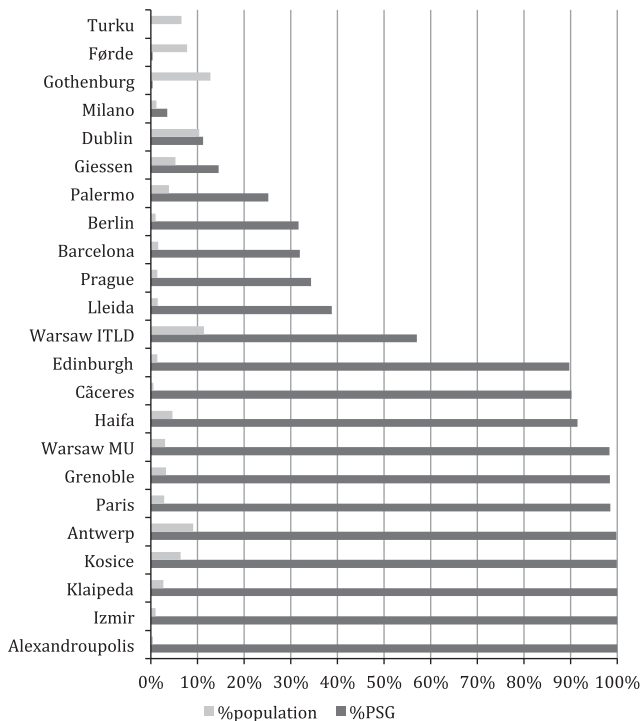
Figure 1. Percentage of total recruited patients (open bars) and percentage of patients investigated with polysomnography (closed bars) at the 23 European Sleep Apnea Cohort (ESADA) centres providing data for the analysis.

who underwent PSG than in the patients who underwent PG (7.7 ± 15.0 versus 2.0 ± 8.9 h^{-1}, $P < 0.0001$) (Fig. 3a).

The mean TAT used for the calculation of AHI in patients recorded using PG exceeded the mean TST used in the PSG analysis by 47 min (Table 2). Indeed, sleep efficiency measured in patients by PSG was 83 ± 13.6%. In the PSG group, when the index of apneas and hypopneas was recalculated in each patient as the number of recorded events (product of reported AHI and TST) divided by TAT, the result produced a significantly lower AHI (AHI_TAT = 24.7 ± 22.0 h^{-1}; $P < 0.0001$) (Table 2). However, it was still higher than the AHI$_{PG}$ ($P < 0.0001$) (Fig. 3b). Apneas/hypopneas may have been scored in periods that were classified as awake during the PSG recordings. When computing the AHI_TAT in our study we did not account for such events. Therefore, we have to assume that the difference between AHI$_{PG}$ and AHI_TAT represents a conservative estimate which could be higher if 'wake' events had been included in the analysis. Similarly, the ODI was recalculated in the patients who underwent PSG from the number of oxygen desaturation events (product of ODI times TST) and indexed over TAT resulting in a lower ODI_TAT (18.6 ± 21.4 h^{-1}). This result was significantly lower than the original ODI$_{PSG}$ ($P < 0.0001$), but did not differ significantly from the ODI$_{PG}$ (Fig. 3b).

The contribution of the diagnostic method used (PSG or PG) is given in Table 3 for the unadjusted model, as well as

for several models accounting for confounders such as anthropometrics and comorbidities. The diagnostic method used contributed consistently and significantly to the AHI index, and its contribution varied only marginally after control for confounders (from AHI$_{PSG}$ +8.60 ± 0.47 events h^{-1} in the unadjusted model to AHI$_{PSG}$ +7.61 ± 0.43 events h^{-1} in the fully adjusted model, Table 3).

DISCUSSION

The data from this large-scale, multi-centre European cohort show that AHI and ODI values obtained in patients who underwent PSG polygraphic recordings are, on average, lower than those derived in patients using polysomnography. Although this may seem obvious at first to all those who score or use both systems regularly, we have demonstrated that the differences are unrelated to the anthropometric characteristics or comorbid conditions of the patients. We have shown, using collated data acquired from the recordings of many different systems, that the differences lie in scoring definitions of respiratory events and in the overestimation of actual sleep time in polygraphic recordings. Thus, the AASM criteria for grading OSA severity are not pertinent to PG recordings and our findings suggest that decision-making may thus become more arbitrary if attempts are made to apply them in this setting.

Methods used to diagnose sleep apnea vary considerably across European health-care systems (Fietze *et al.*, 2011). The current study, which reflects the practice in sleep centres with or without a university affiliation, confirms that a large proportion of European patients are diagnosed using PG. The high prevalence of sleep-disordered breathing within most westernized countries has led to the use through necessity of simplified recording techniques (levels III and IV). Several European health-care systems apply strict AHI-related rules for reimbursement of OSA therapy. Consequently, classification of OSA by different diagnostic techniques needs to be based on uniform criteria that should take the diagnostic method into account.

Our study has shown that patients diagnosed using PG are likely to have a 30% lower AHI (A + H) compared to patients investigated by PSG. The median AHI in patients investigated using PG was approximately 10 units lower than the median in the PSG group. If clinically relevant cutoff values for classification of OSA are applied, it may be estimated that a PSG-based diagnosis of OSA would be ruled out in 14% of the patients if a diagnostic PG method is used. Conversely, approximately 14% of patients would escape a diagnosis of severe OSA (AHI ≥ 30) with a PG diagnostic study. Considering that most epidemiological and outcome data are based on cutoff values derived from PSG standards, our results may have considerable clinical implications. There is an evident need for outcome data based on PG classification of OSA, and potentially rectified cutoff levels for AHI need to be applied in the clinical management of OSA in order to align with the high frequency of PG-based studies. It may be
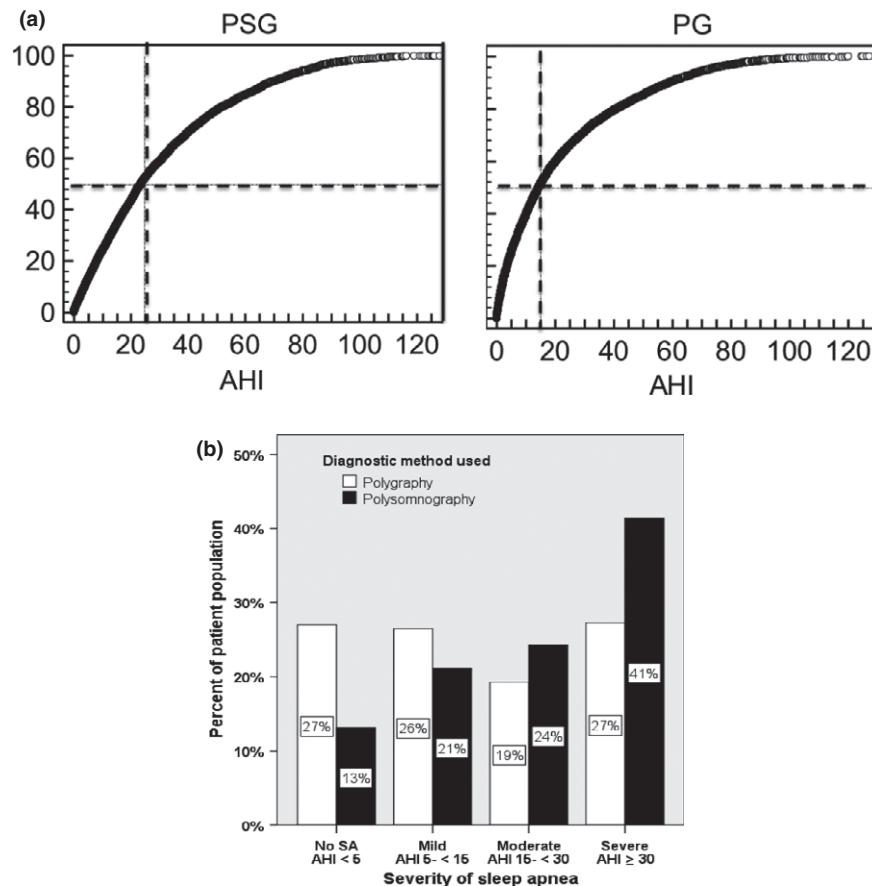
Figure 2. (a) Cumulative distribution of apnea–hypopnea index (AHI) values obtained in patients studied with polysomnography (PSG, left) and polygraphy (PG, right) and (b) proportion of patients with no sleep apnea or sleep apnea in various severity groups by either a polysomnographic (filled bars) or polygraphic (open bars) study. The horizontal dashlines are the medians of the distribution and the vertical dashlines the corresponding AHI.

proposed that such classification documents and standardizes recording times between lights-out and lights-on times and provides corrected cutoff values. It has been suggested that portable studies would be economically efficient when the pretest probability of OSA is high, but no clinical decision model has been confirmed to date.

The different AHI values obtained by PSG and PG could not be explained by anthropometric differences as mean age and BMI were similar in both groups, whereas the contribution of male patients was slightly higher in the PSG group. Could patients referred for PSG have had more comorbid conditions, suggesting that they may be 'sicker' patients? A separate multivariate analysis addressing the influence of comorbid disease on AHI suggested only marginal influence of cardiovascular, metabolic, pulmonary and sleep disorders other than sleep apnea on the difference in AHI between both methods (Table 3). Importantly, complaints of sleepiness or insomnia as well as other relevant sleep disorders were recorded to a similar extent in both groups. These results are at variance with current guidelines indicating PSG as the preferred diagnostic test for OSA in the presence of comorbid conditions (Collop *et al.*, 2007; Epstein *et al.*, 2009; Ferber *et al.*, 1994).

The continued analysis addressed the possibility that the lower AHI in the PG group could be accounted for by overestimation of sleeping time, which would yield a lower index despite a similar number of events. In addition, rules applied to score hypopneic events may have accounted for a higher AHI in PSG, as concomitant arousals are taken into account.

A previously published health technology assessment reviewing 14 studies (879 patients) addressing PSG versus level III devices confirmed the underestimation of AHI values by 3–11 events h^{-1} when PG was used (Trikalinos and Lau, 2007; Trikalinos *et al.*, 2007). Masa *et al.* also found a mean bias of seven events h^{-1} between PSG performed in the hospital and home respiratory polygraphy, increasing with higher AHI (Masa *et al.*, 2011), as confirmed by a recent meta-analysis on 19 studies on diagnostic accuracy of PG versus PSG showing decreasing sensitivity as disease severity increased (El Shayeb *et al.*, 2014). The discrepancy between the total recording time and the total sleep time increases with the severity of obstructive sleep apnea–hypopnea syndrome (OSAHS) (patients with more severe OSAHS have longer cumulative arousals, and thus more wake after sleep onset). Quality and length of sleep may also
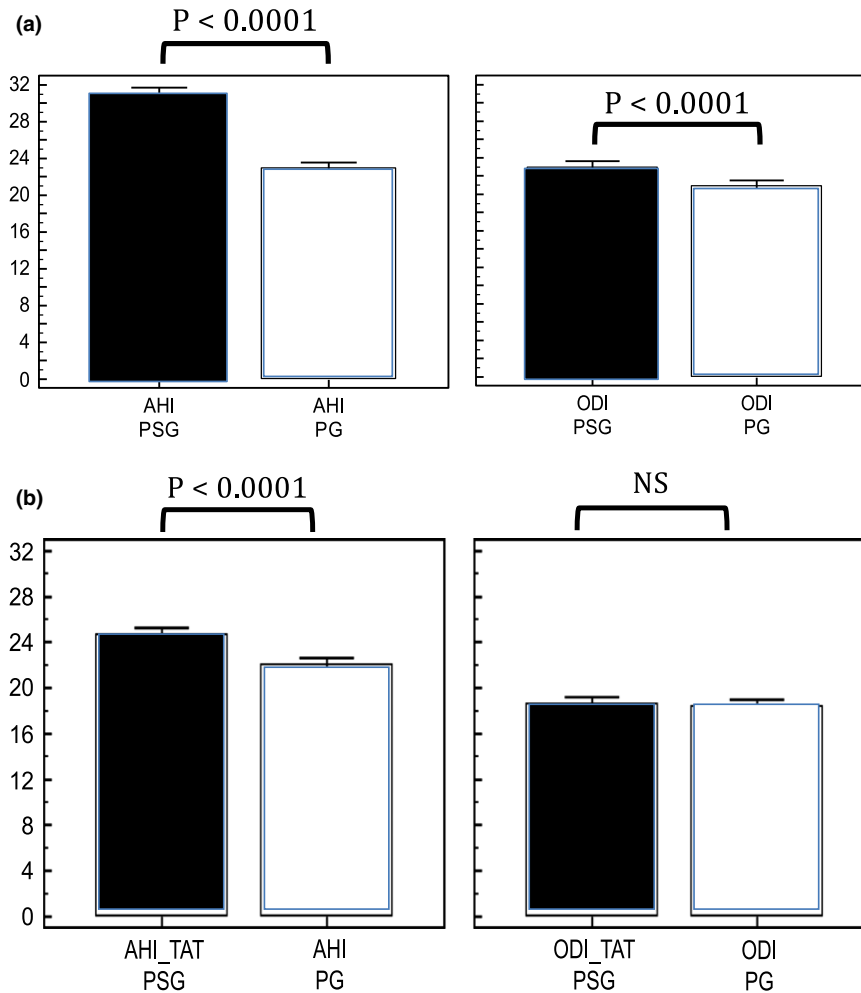
Figure 3. (a) Apnea–hypopnea index (AHI) and oxygen desaturation index (ODI) in patients undergoing polysomnography (PSG) and polygraphy (PG) (mean and 95% confidence interval). (b) AHI and ODI calculated over total analysed time [AHI_total analysed time (TAT) and ODI_TAT] in patients undergoing polysomnography and AHI and ODI in patients undergoing polygraphy.

Table 2 Sleep data in the groups investigated with polygraphy and polysomnography			
Test	*Polysomnography*	*Polygraphy*	*P*
Analysed time min	470.9 ± 67.0	428.8 ± 59.0	<0.0001
Subjective length min	409.44 ± 97.5	421.2 ± 89.4	<0.0001
TST_{PSG} and TAT_{PG} min	381.7 ± 78.8	428.8 ± 59.0	<0.0001
AHI h^{-1} in PSG (AHI_{PSG}) and PG (AHI_{PG})	31.0 ± 26.1	22.0 ± 23.5	<0.0001
ODI h^{-1} in PSG (ODI_{PSG}) and PG (ODI_{PG})	23.0 ± 25.3	18.4 ± 21.7	<0.0001
Time below $SaO_2 = 90\%$ min	30.6 ± 64.0	40.6 ± 80.5	<0.02
AHI_TAT h^{-1}	24.7 + 22.0	22.0+ 23.5	0.0001
ODI_TAT h^{-1}	18.6 + 21.4	18.4 + 21.7	0.65

AHI, apnea–hypopnea index; TAT, total analysed time; TST, total sleep time; PSG, polysomnigraphy; ODI, oxygen desaturation index.

be affected adversely during a subject's first sleep study, irrespective of whether it is performed using PG or PSG ('first night effect') (Toussaint *et al.*, 1995).

The differences between ODI scored in patients tested by PG and in others tested by PSG were considerably smaller than observed differences in AHI (Fig. 3a), whereas ODI

Table 3 Results of six different multiple linear regression analyses with apnea–hypopnea index (AHI) as the dependent variable

Model	Parameters included	AHI difference between PSG and PG (mean and SD)	P-value for the variable 'diagnostic method used'	P-value for anthropometric variables and comorbidities	R^2 of the model	Degrees of freedom
1	Unadjusted diagnostic method (M)	+8.60 ± 0.47	< 0.001		0.030	10 942
2	M plus anthropometrics (A) (age, BMI, gender)	+8.45 ± 0.42	< 0.001	Age, B = 0.3 ± 0.02, $P < 0.001$ BMI, B = 1.5 ± 0.03, $P < 0.001$ Male gender, B = 11.3 ± 0.46, $P < 0.001$	0.230	10 886
3	M, A and any CVD	+8.06 ± 0.42	< 0.001	CVD B = 2.4 ± 0.48, $P < 0.001$	0.228	10 572
4	M, A and any pulmonary disease	+8.27 ± 0.43	< 0.001	Pulmonary disease, B = −1.6 ± 0.65, $P = 0.02$	0.227	10 618
5	M, A and any metabolic disease	+8.08 ± 0.43	< 0.001	Any metabolic disease, B = 0.7 ± 0.22, $P = 0.002$	0.228	10 618
6	M, A and any sleep disorder	+7.89 ± 0.42	< 0.001	Sleep disorder B = −9.9 ± 0.59, $P < 0.001$	0.245	10 612
7	Fully adjusted model including all factors mentioned above	+7.61 ± 0.43	< 0.001	Age, B = 0.2 ± 0.02, $P < 0.001$ BMI, B = 1.3 ± 0.34, $P < 0.001$ Male, B = 10.0 ± 0.47, $P < 0.001$ CVD 1.8 ± 0.49, $P < 0.001$ Pulmonary disease, B = −1.4 ± 0.65, $P = 0.027$ Sleep disorder, B = −9.7 ± 0.60, $P < 0.001$ Metabolic disease, B = 1.4 ± 0.49, $P = 0.003$	0.248	10 572

BMI, body mass index; CVD, cardiovascular disease; SD, standard deviation.
The contribution of the diagnostic method used [polysomnography (PSG) or polygraphy (PG)] is given for the unadjusted model, as well as for several models accounting for potential confounders (anthropometrics, comorbidities). The diagnostic method used consistently and contributed significantly to the AHI index, which varied only marginally between the models.

computed in apnea PSG over TAT was not different compared to patients measured in PG (Fig. 3b). As oxygen desaturations in both PG and PSG recordings are conventionally analysed automatically in the ESADA and are unlikely to happen during wakefulness in patients with a low incidence of cardiac failure (2.5–1.9%), this favours the hypothesis of a time dilution effect to explain the difference between AHI in PSG and PG. AHI on PSG is a more complex measure of sleep-disordered breathing, as it takes several flow and desaturation criteria into account, including the scoring of arousals on EEG in relation to respiratory events.

In our computation, we evaluated the effect of using the TOff/TOn duration (total analysed time) to derive the AHI in PSG which was lower than in patients tested by PSG, but still higher than in patients tested by PG (Fig. 3a, b). This leads to the conclusion that event classification was also certainly contributing to the observed difference. Most probably, the difference was accounted for by the scoring of hypopneas; however, this could not be ascertained, as hypopneas were not reported separately in the ESADA. Such a difference has already been outlined in several studies scoring hypopneas by PSG visual analysis of arousals associated with respiratory events (Ayappa *et al.*, 2008; Collop, 2014; Ruehland *et al.*, 2009), whereas adding surrogate arousals to respira-

tory polygraphy does not lead to substantially different results from hypopnea analysis based only on desaturation (Masa *et al.*, 2013). Indeed, the use of the recommended rule in 83.4% of PG in our study was associated with a significantly lower AHI (21.2 ± 0.33 h^{-1}) than applying the alternative rule in the remaining patients (32.2 ± 0.8 h^{-1}) ($P < 0.001$) probably due mainly to the lower desaturation threshold in the alternative rule. In PSG, the majority of studies (87.7%) used the alternative rule, which resulted in AHI values closer to those obtained by the recommended scoring (31.2 ± 0.4 versus 30.9 ± 1.1 h^{-1}, respectively) ($P < 0.001$). This sub-analysis clearly underlines the importance of standardized scoring rules in order to obtain uniform clinical classification of patients. In the context of the ESADA cohort, we were able to document a strong impact of the scoring rules applied for PG.

Therefore, our results also support the need for further standardization of the scoring procedures for both PG and PSG. As the difference in ODI between PSG and PG was, on average, 4.6 h^{-1}, it may be extrapolated that time dilution effect could explain approximately half the difference of nine events h^{-1} in AHI between PSG and PG, while the other half could be explained by hypopnea events classification. The study by Masa *et al.* (2011) in 348 patients without heart

disease in eight Spanish centres determined, in the same subjects, the agreement between home respiratory polygraphy (HRP) and in-hospital PSG for therapeutic decision-making, according to a variant of the 2007 AASM recommended criteria for definition of hypopnea (oxygen desaturation greater or equal to 3%, instead of 4%). Their findings suggested that the difference in AHI between the two methods primarily explained the discrepant therapeutic decisions made. However, differences in the interpretation of clinical variables and comorbidity also contributed. Moreover, the study did not provide detailed information about relevant AHI thresholds obtained by HRP and PSG for therapeutic decision-making. Nevertheless, the authors also underlined the importance of a time dilution effect and the impact of arousal scoring to explain the difference in AHI obtained by HRP or PSG.

This study has both strengths and weaknesses. To our knowledge, this is the largest evaluation of different diagnostic methods for OSA evaluation in the literature. The study includes a wide range of clinical recording devices in different sleep laboratory settings and data are collected under controlled conditions. The ESADA uses the clinical routines applied at clinical sleep centres across Europe. This study was purposely not designed as a randomized comparison of PG and PSG in the same subjects, but rather as a study of clinical routines, which reflect a mixture of routines in real life. It shows that current practice in Europe for the diagnosis of OSA is heterogeneous, and creates different diagnostic outcomes.

There is a possibility of recruitment bias to either PG or PSG, although this is less likely, as there was no systematic difference in the findings related to proportion of patients investigated by either method. Our comparisons are not based on data obtained in the same subjects; however, clinical characteristics and comorbidities did not differ substantially between the two groups. Differences in sleep architecture between home and laboratory studies are also possible, but it has been shown that reliable studies can be made in the laboratory and at the patient's home. Furthermore, normative data do not consider the location of sleep recording as an important confounder.

Methods for sleep scoring may have differed to some extent between participating centres; however, scoring techniques were standard following the AASM 2007 criteria and routine for scoring were stated in the protocol. The use of different scoring rules is an additional potential cause of data variation in the analysis, but emphasizes the importance of standardized rules for the correct quantification of SDB in PG assessments. It is highly likely that interscorer and between-centre variability would have affected both PG and PSG scoring results. The mean AHI difference of 7.6 h^{-1} we reported may therefore be considered as a rather conservative estimate due to the inconsistency of the analysis method.

The latest update of the 2007 AASM *Manual for the Scoring of Sleep and Associated Events* (Berry *et al.*, 2012) would not change our conclusions, as it does not take into account the actual sleeping time in PG, and the 'recommended' rule still allows for scoring hypopnea when 'there is a $\geq$ 3% oxygen desaturation from pre-event baseline or the event is associated with an arousal'.

ESADA protocol states that prevailing clinical routines should be employed at participating centres; this means that equipment from different manufacturers is used in the study, and this could have influenced the data reported in the study. Although this factor was beyond our control, the fact remains that the size of study provides a relevant reflection of actual clinical practice in Europe.

In conclusion, evaluation of data from the ESADA cohort has highlighted the difference in clinical results obtained between using PG and PSG recording techniques. These differences need to be taken into account, considering that PG techniques are used increasingly in clinical sleep medicine. As therapeutic decisions and guidelines are based on the results provided by various types of recording devices, there is an obvious need for standards of relevant cutoff values to define outcome in OSA using specific PG thresholds different than those for PSG AHI.

REFERENCES

Anonymous. Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. Report of an American Academy of Sleep Medicine Task Force. *Sleep*, 1999, 22: 667–689.

Ayappa, I., Norman, R. G., Seelall, V. and Rapoport, D. M. Validation of a self-applied unattended monitor for sleep disordered breathing. *J. Clin. Sleep Med.*, 2008, 4: 26–37.

Ayas, N. T., Fox, J., Epstein, L., Ryan, C. F. and Fleetham, J. A. Initial use of portable monitoring versus polysomnography to confirm obstructive sleep apnea in symptomatic patients: an economic decision model. *Sleep Med.*, 2010, 11: 320–324.

Berry, R. B., Budhiraja, R., Gottlieb, D. J. *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. *J. Clin. Sleep Med.*, 2012, 8: 597–619.

Chesson, A. L. Jr, Berry, R. B. and Pack, A. Practice parameters for the use of portable monitoring devices in the investigation of suspected obstructive sleep apnea in adults. *Sleep*, 2003, 26: 907–913.

Collop, N. Breathing related arousals: call them what you want, but please count them. *J. Clin. Sleep Med.*, 2014, 10: 125–126.

Collop, N. A., Anderson, W. M., Boehlecke, B. *et al.* Clinical guidelines for the use of unattended portable monitors in the diagnosis of obstructive sleep apnea in adult patients. Portable Monitoring Task Force of the American Academy of Sleep Medicine. *J. Clin. Sleep Med.*, 2007, 3: 737–747.

Collop, N. A., Tracy, S. L., Kapur, V. *et al.* Obstructive sleep apnea devices for out-of-center (OOC) testing: technology evaluation. *J. Clin. Sleep Med.*, 2011, 7: 531–548.

Corral-Penafiel, J., Pepin, J. L. and Barbe, F. Ambulatory monitoring in the diagnosis and management of obstructive sleep apnoea syndrome. *Eur. Respir. Rev.*, 2013, 22: 312–324.

Douglas, N. J., Thomas, S. and Jan, M. A. Clinical value of polysomnography. *Lancet*, 1992, 339: 347–350.

El Shayeb, M., Topfer, L. A., Stafinski, T., Pawluk, L. and Menon, D. Diagnostic accuracy of level 3 portable sleep tests versus level 1 polysomnography for sleep-disordered breathing: a systematic review and meta-analysis. *Can. Med. Assoc. J.*, 2014, 186: E25–E51.

Epstein, L. J., Kristo, D., Strollo, P. J. Jr *et al.* Clinical guideline for the evaluation, management and long-term care of obstructive sleep apnea in adults. *J. Clin. Sleep Med.*, 2009, 5: 263–276.

Ferber, R., Millman, R., Coppola, M. *et al.* Portable recording in the assessment of obstructive sleep apnea. ASDA standards of practice. *Sleep*, 1994, 17: 378–392.

Fietze, I., Penzel, T., Alonderis, A. *et al.* Management of obstructive sleep apnea in Europe. *Sleep Med.*, 2011, 12: 190–197.

Fischer, J., Dogas, Z., Bassetti, C. L. *et al.* Standard procedures for adults in accredited sleep medicine centres in Europe. *J. Sleep Res.*, 2012, 21: 357–368.

Hedner, J., Grote, L., Bonsignore, M. *et al.* The European Sleep Apnoea Database (ESADA): report from 22 European sleep laboratories. *Eur. Respir. J.*, 2011, 38: 635–642.

Iber, C., Ancoli-Israel, S., Chesson, A. and Quan, S. F. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*, 1st edn. American Academy of Sleep Medicine, Westchester, IL, 2007.

Masa, J. F., Corral, J., Pereira, R. *et al.* Therapeutic decision-making for sleep apnea and hypopnea syndrome using home respiratory polygraphy: a large multicentric study. *Am. J. Respir. Crit. Care Med.*, 2011, 184: 964–971.

Masa, J. F., Corral, J., Gomez De Terreros, J. *et al.* Significance of including a surrogate arousal for sleep apnea–hypopnea syndrome diagnosis by respiratory polygraphy. *Sleep*, 2013, 36: 249–257.

Mulgrew, A. T., Fox, N., Ayas, N. T. and Ryan, C. F. Diagnosis and initial management of obstructive sleep apnea without polysomnography: a randomized validation study. *Ann. Intern. Med.*, 2007, 146: 157–166.

Rosen, C. L., Auckley, D., Benca, R. *et al.* A multisite randomized trial of portable sleep studies and positive airway pressure autotitration versus laboratory-based polysomnography for the diagnosis and treatment of obstructive sleep apnea: the HomePAP study. *Sleep*, 2012, 35: 757–767.

Ruehland, W. R., Rochford, P. D., O'Donoghue, F. J., Pierce, R. J., Singh, P. and Thornton, A. T. The new AASM criteria for scoring hypopneas: impact on the apnea hypopnea index. *Sleep*, 2009, 32: 150–157.

Toussaint, M., Luthringer, R., Schaltenbrand, N. *et al.* First-night effect in normal subjects and psychiatric inpatients. *Sleep*, 1995, 18: 463–469.

Trikalinos, T. A. and Lau, J. *Obstructive Sleep Apnea–Hypopnea Syndrome: Modeling Different Diagnostic Strategies*. Agency for Healthcare Research and Quality of the US Department of Health and Human Services, Tufts–New England Medical Center EPC, Rockville, MD, 2007: 1–116.

Trikalinos, T. A., Ip, S., Raman, G, Cepeda, MS, Balk, EM, D'Ambrosio, C and Lau, J. *Home Diagnosis of Obstructive Sleep Apnea–Hypopnea Syndrome*. Agency for Health Care Research and Quality of the US Department of Health and Human Services, Tufts–New England Medical Center, Rockville, MD, 2007: 1–127.