Apnoea–hypopnoea indices determined via continuous positive airway pressure (AHI-CPAP\textsubscript{flow}) versus those determined by polysomnography (AHI-PSG\textsubscript{gold}): a protocol for a systematic review and meta-analysis

Fanny Bertelli,\textsuperscript{1,2} Carey Meredith Suehs,\textsuperscript{3,4} Jean Pierre Mallet,\textsuperscript{3,5} Marie Caroline Rotty,\textsuperscript{2} Jean Louis Pepin,\textsuperscript{6} Frédéric Gagnadoux,\textsuperscript{6} Eric Matzner-Lober,\textsuperscript{8} A Bourdin,\textsuperscript{2,3,5} Nicolas Molinari,\textsuperscript{1,4} Dany Jaffuel\textsuperscript{3,5,9}

ABSTRACT

Introduction To date, continuous positive airway pressure (CPAP) remains the cornerstone of obstructive sleep apnoea treatment. CPAP data describing residual sleep-disordered breathing events (ie, the CPAP-measured apnoea–hypopnoea indices (AHI-CPAP\textsubscript{flow})) is difficult to interpret because it is an entirely different metric than the polysomnography (PSG) measured AHI gold standard (AHI-PSG\textsubscript{gold}). Moreover, manufacturer definitions for apnoea and hypopnoea are not only different from those recommended for PSG scoring, but also different between manufacturers. In the context of CPAP initiation and widespread telemedicine at home to facilitate sleep apnoea care, there is a need for concrete evidence that AHI-CPAP\textsubscript{flow} can be used as a surrogate for AHI-PSG\textsubscript{gold}.

Methods and analysis No published systematic review and meta-analysis (SRMA) has compared the accuracy of AHI-CPAP\textsubscript{flow} against AHI-PSG\textsubscript{gold} and the primary objective of this study is therefore to do so using published data. The secondary objectives are to similarly evaluate other sleep disordered breathing indices and to perform subgroup analyses focusing on the inclusion/exclusion of central apnoea patients, body mass index levels, CPAP device brands, pressure titration modes, use of a predetermined pressure, and fixed pressure level or not, and the impact of a 4% PSG desaturation criteria versus 3% PSG on accuracy. The Preferred Reporting Items for SRMA protocols statement guided study design. Randomised controlled trials and observational studies of adult patients (\geq 18 years old) treated by a CPAP device will be included. The CPAP intervention and PSG comparator must be performed synchronously. PSGs must be scored manually and follow the American Academy of Sleep Medicine guidelines (2007 AASM criteria or more recent). To assess the risk of bias in each study, the Quality Assessment of Diagnostic Accuracy Studies 2 tool will be used.

Ethics and dissemination This protocol received ethics committee approval on 16 July 2020 (IRB\_MTP\_2020\_07\_202000404) and results will be disseminated via peer-reviewed publications.

Strengths and limitations of this study

- This will be the first systematic review and meta-analysis to compare published data for apnoea–hypopnoea indices determined via continuous positive airway pressure devices (AHI-CPAP\textsubscript{flow}) versus the gold standard determined by polysomnography (AHI-PSG\textsubscript{gold}) in terms of accuracy.
- Given that AHI-CPAP\textsubscript{flow} and AHI-PSG\textsubscript{gold} are often used interchangeably, the results of this study will fill a pertinent knowledge gap particularly suited to the context of in-home CPAP initiation and the associated increase in telemedicine.
- This study will also evaluate how different parameters (central apnoea inclusion/exclusion, body mass index, oxygen saturation thresholds and devicespecific summary measurements) affect accuracy, and may therefore impact practice, study design or analyses.
- The main limitations of this study, as for most systematic reviews and meta-analyses, are likely to result from a paucity of eligible publications and their methodological quality.
- In the context of patients requiring long-term CPAP management, why a single night point estimate (like AHI-PSG\textsubscript{gold}) is considered as a gold standard rather than longitudinally repeated measurements (like AHI-CPAP\textsubscript{flow}) requires consideration.

PROSPERO/Trial registration numbers CRD42020159914/NCT04526366; Pre-results INTRODUCTION

Description of the condition

Nearly 1 billion adults aged 30–69 years are affected by obstructive sleep apnoea (OSA), including 425 million with moderate to severe OSA requiring treatment according to
current recommendations. In 2020, continuous positive airway pressure (CPAP) remains the cornerstone of OSA treatment. Several studies have shown that CPAP therapy can effectively reduce upper airway obstruction with subsequent improvements in daytime sleepiness, sleep quality and quality of life. OSA treatment with CPAP has been demonstrated to be cost-effective in various countries with different healthcare systems. The treatment and monitoring of OSA is therefore gaining recognition as an increasingly important public health issue. Considering the number of patients to be diagnosed and monitored on a long-term basis, there is a need for simplified diagnostic and monitoring methods. Because initiation of CPAP at home demonstrated equivalent effects on patient outcomes when compared with an in-laboratory titration approach, the 2019 American Academy of Sleep Medicine (AASM) guidelines recommended that CPAP therapy be initiated using either auto-CPAP at home or in-laboratory CPAP titration in adults with OSA and no significant comorbidities. In addition, it was suggested that clinicians use telemonitoring-guided interventions during the initial period of CPAP therapy in adults with OSA.

**Description of the interventions**

In clinical practice, to summarise data and conclude with an OSA diagnosis, the high-dimensional data contained within a polysomnography (PSG) is reduced down to the number of apnoea or hypopnoea events occurring per hour, that is the ‘apnoea–hypopnoea index (AHI)’. As underlined by the 2013 American Thoracic Society Statement, CPAP data describing residual sleep-disordered breathing events are difficult to interpret. CPAP devices rely only on a reduction in airflow for determining AHI and a recording time (corresponding to the length of time the device is turned on associated with a measurable breathing signal irrespective of the sleep/awake patient status), whereas PSG includes more data such as respiratory flow patterns, electroencephalogram (EEG) arousal, total sleep time, respiratory effort and oxyhaemoglobin desaturation measures. AHI determined by CPAP data is a different metric than the gold standard AHI determined by PSG (hereafter termed ‘AHI-CPAP\_flow’, corresponding to the ratio between an airflow reduction and a recording time) is a different metric than the gold standard AHI determined by PSG (hereafter termed ‘AHI-PSG\_gold’, corresponding to an airflow reduction associated with EEG arousal and/or oxyhaemoglobin desaturation and total sleep time). The extent to which the AHI-CPAP\_flow can be used interchangeably with or as a surrogate for AHI-PSG\_gold is unclear. Moreover, the manufacturer definitions for apnoea, hypopnoea and flow limitations are not only different from those recommended for PSG scoring, but also different among manufacturers (online supplementary appendix 1). Depending on how a given manufacturer defines an event, the PSG percentage of desaturation and event considered, the differences in total sleep time versus recording time, AHI-CPAP\_flow can theoretically overestimate or underestimate AHI-PSG\_gold. Why it is important to do this review?

Considering the increasing use of in-home auto-CPAP titration and associated telemedicine initiatives (both for titration and long-term monitoring), the issue of whether or not AHI-CPAP\_flow is a valid surrogate for AHI-PSG\_gold should be resolved.

**Objectives**

To date, no published systematic review and meta-analysis (SRMA) has compared AHI-CPAP\_flow and AHI-PSG\_gold. Therefore, the primary objective of this study is to compare published data for AHI-PSG\_gold and AHI-CPAP\_flow in patients treated by CPAP. The secondary objectives are to evaluate, in a manner similar to the primary objective, data for apnoea index (AI), hypopnoea index (HI), respiratory disturbance index (RDI) and respiratory effort-related arousals (RERAs) and to perform subgroup analyses focusing on the inclusion/exclusion of central apnoea patients, body mass index (BMI) levels, CPAP device brands, pressure titration modes, use of a predetermined and fixed pressure level or not, and the impact of a 4% PSG desaturation criteria versus 3% PSG on index accuracy.

**METHODS AND ANALYSIS**

The Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols statement guided the design of this protocol (see online supplementary appendix 2 for the checklist). The study is sponsored by the University Hospitals of Montpellier, Montpellier, France. Registration was submitted to PROSPERO (https://www.crd.york.ac.uk/PROSPERO/) on 6 February 2020, and first published on 28 April 2020, with updates on 21 August 2020 and 23 March 2021. Any changes or amendments made to the protocol will be tracked and dated on PROSPERO.

**Study eligibility criteria**

Detailed eligibility criteria were designed to take into account the relevant patient population, interventions, comparators, outcomes and statistical analysis/study types.

**Patient population**

Studies reporting results for adult patients (≥18 years old) treated by a CPAP device will be included.

**Intervention/comparator**

The intervention of interest is the measure of residual sleep-disordered breathing events by a CPAP device during a given night. The comparator of interest is the simultaneous measure of the residual sleep-disordered breathing events by a PSG during the same night of CPAP treatment. The CPAP intervention and PSG comparator
must be performed synchronously, and the CPAP device name and series must be mentioned in the study.

**Outcome measures**

The main outcome is the mean difference between AHI-CPAP\(_{\text{low}}\) and AHI-PSG\(_{\text{gold}}\) measures (and the associated SD for the mean difference). Secondary outcomes include analogous results for other frequently co-occurring indices: AI, HI, RDI or RERA.

PSG scoring must be manually performed by a qualified physician or technician using 2007 AASM recommended/alternative criteria or more recent AASM criteria. The scoring criteria used in the study must be detailed. In particular, oxyhaemoglobin desaturation level (3% or 4%), and apnoea and hypopnoea scoring criteria must be mentioned.

**Statistical analyses/study types**

The primary analysis will synthesise estimates for the average of individual differences between measures made via PSG minus those made via CPAP. Such individual differences are most conveniently provided by Bland–Altman test results. However, any of the following analysis types, comparing the aforementioned PSG versus CPAP data, are also of interest: scatter plots, correlation coefficients, differences in central tendency or any type of conformity test. We will include randomised controlled trials and observational studies published in English. Case series will be included but not case reports. Meta-analyses will be excluded. In general, because we are focused on a within-patient comparison, we expect more observational studies than randomised controlled trials. In addition, each arm of a randomised trial can be considered an independent estimate of PSG–CPAP differences.

**Information sources**

We will perform a search in MEDLINE (PubMed), Embase, Web of Science and the Google Scholar database. Additional studies will be sought by manually checking the references of included studies and relevant reviews. Searches will be restricted to publications appearing from 2007 onwards (to the day of search results, with an update just before publication). A supplementary search for ongoing/unpublished trials will be made using the https://www.clinicaltrials.gov/ website.

**Search strategy**

Our search strategy was developed using the following key concepts: AHI, CPAP, PSG and adults. For each key concept, we will also use acronyms such as PSG or current terms used by specialists as search terms. These search terms will be combined using Boolean operators “AND” and “OR”. The full electronic strategy is presented in the online supplementary appendix 3.

**Study selection**

Figure 1 summarises the study selection process. At least two authors will screen the titles and abstracts yielded from the literature searches, independently and in duplicate. The exclusion criteria listed in box 1 will be used and sequentially deployed so as to help populate the future study flowchart (based on figure 1). Articles, which appear to meet our inclusion criteria, will be downloaded in full. Disagreements will be resolved by consensus or a third review author. We will identify and exclude duplicates, and collate multiple reports of the same study.

**Data collection**

Two reviewers will use pretested data collection forms to collect data independently and in duplicate. Disagreements will be resolved by consensus or by a third reviewer/author.

The data of interest include those for describing the articles analysed, and those pertaining to outcomes, and will be managed in spreadsheets (Microsoft Excel). In the first group, the following will be tabulated: author(s), journal, year of publication, study type (eg, retrospective cohort, randomised controlled trial), population size(s), mean BMI for the population(s), a short description of the OSA population, the AASM criteria applied during the study (eg, 2007 or 2012), the oxyhaemoglobin desaturation level used (3% or 4%), major exclusion criteria, device information (brand, model and mode if pertinent), the pressure mode used (eg, fixed, automatic, manual) and information concerning predetermined pressure levels used (eg, physician-determined vs CPAP-determined). Outcomes are a minima those issuing from Bland–Altman analyses (mean difference (PSG\(_{\text{gold}}\) minus CPAP\(_{\text{low}}\)) and the associated SD, SE, lower limit of agreement and upper limit of agreement for each of the following: AHI, AI, HI, RDI and RERA).

**Missing data**

Where necessary, we will contact the authors of studies to obtain missing data/information. In cases where authors are unreachable or for some reason missing data cannot be provided, but a way to glean such data is available (for eg, from the pixels of a scatter plot), such will be tolerated and indicated. In the discussion section of the review, the potential impact of any missing data will be discussed.

The Quality Assessment of Diagnostic Accuracy Studies 2 tool\(^{17}\) will be used to assess full article quality and risk of bias. The latter is a specialised tool for diagnostic accuracy studies that addresses four domains (patient selection, index test, reference standard, and flow and timing), assessing the risk of bias for each. Two independent reviewers will assess the methodological quality of selected articles and disagreements about scoring will be resolved by a third reviewer.

If a sufficient number of studies are available (at least ten is suggested), a funnel plot will be used to graphically summarise the extent of publication bias. Additionally, the Egger test\(^{18}\) will be used to measure funnel plot asymmetry with statistical significance set at p<0.05.

The overall quality of findings will be assessed using the Grading of Recommendations Assessment, Development and Evaluation system.
Data synthesis
If we identify a sufficient number of studies with homogeneous populations and characteristics, we will carry out meta-analyses of primary (AHI-CPAP_{flow} vs AHI-PSG_{gold}) and secondary outcomes using parametric analyses. In case of a low number of studies or studies with small sample sizes, we will adjust the methods and estimators used (regardless of the number and quality of studies found, a systematic narrative review will be written). Differences in means (PSG_{gold} minus CPAP_{flow}) will be evaluated both as directional differences and as absolute values. Meta-analysis will be performed using a random-effects model to avoid homogeneity problems between studies results. The heterogeneity among studies included in each meta-analysis will be assessed with the Q-test statistic.

Analyses will be performed in the R statistical programming environment. Meta-analysis will be conducted using the meta package.

Subgroup analysis and investigation of heterogeneity
Subgroup analyses will be performed as described in the secondary objectives. If a sufficient number of studies are included in this review, we will perform sensitivity analyses to assess the consistency and robustness of our results.

Patient and public involvement
The present work is based on a review of relevant studies and does not include original patient data. Therefore, patients or public are not involved in this review protocol.

Protocol amendments
Any amendments to this protocol will be reported in PROSPERO with the justification and date of modification.

ETHICS AND DISSEMINATION
This protocol was approved by the Internal Review Board at the Montpellier University Hospitals on 16 July 2020.
Box 1  Sequential exclusion process

All (full scientific) study designs and publication types written in English and reporting Bland-Altman test results (or other tests suggesting the presence of the required data) between simultaneous, paired, polysomnography (PSG)-derived and continuous positive airway pressure (CPAP)-derived data describing sleep disordered breathing. To achieve this, the following exclusion criteria will be sequentially applied to search results:

- Duplicates.
- Does not refer to a full scientific article (eg, case reports are excluded).
- Meta-analysis.
- Paediatric populations (populations <18 years are excluded).
- Absence of appropriate paired results: synchronised PSG-derived and CPAP-derived data must be reported for ≥1 of the following measures: apnoea-hypopnoea indices, apnoea index, hypopnoea index, respiratory disturbance index and respiratory effort related arousals.
- Absence of appropriate test comparing paired results: the study must use ≥1 of the following tests to compare the previously mentioned PSG-derived versus CPAP-derived data: a Bland-Altman or correlation test, other conformity tests or tests of differences in central tendency.
- Inappropriate PSG scoring 1: PSG-derived variables must be coded following the American Academy of Sleep Medicine 2007 (or more recent) guidelines.
- Inappropriate PSG scoring 2: PSG-derived variables must be scored manually.
- Inappropriate PSG scoring 3: oxygen desaturation level used for scoring must be mentioned and at 3% or 4%.
- Inappropriate CPAP description: brand/device names must be mentioned.

(IRB_MTP_2020_07_202000404). The results of this study will be disseminated via peer-reviewed publications.

DISCUSSION

Because AHI-CPAP_{flow} is a different metric than the AHI-PSG_{gold}, the extent to which the AHI-CPAP_{flow} can be used interchangeably with or as a surrogate for AHI-PSG_{gold} is unclear. Furthermore, differences in apnoea/hypopnoea definitions between manufacturers are a potential additional limitation to the accuracy of the AHI-CPAP_{flow} measures. To date, no review and meta-analysis has addressed this subject. Considering the number of patients requiring CPAP-monitoring, an auto-CPAP home titration is recommended as an alternative option to in-laboratory PAP titration for patients without comorbidities and an initial telemedicine monitoring is proposed. In addition, for most of the recent randomised clinical trials, CPAP efficiency was based on the AHI-CPAP_{flow} rather than on the AHI-PSG_{gold}. \[21\] \[22\]

If we demonstrate the existence of significant differences between AHI-CPAP_{flow} and AHI-PSG_{gold} our conclusions may have major consequences not only in daily practice but also for the design of future studies, with a crucial need for increased PSG evaluation of CPAP effectiveness. In this regard, although subgroup analyses on central apnoea patients and BMI are of interest to the clinician, the potentially small number of studies may not allow us to perform these statistical analyses.

In the context of CPAP-treated patients requiring long-term management, whether a one night point estimate (like AHI-PSG_{gold}) could be considered as a gold standard to evaluate CPAP efficiency rather than a night-by-night serial measurement (like AHI-CPAP_{flow}) is also questionable. \[23\] Night-by-night AHI variation has been demonstrated not only for the PSG measures in stable patients \[25\] but also for the unstable patients with underlying cardiovascular diseases for example. \[24\] For certain patients, there is no doubt that significant night-to-night AHI variability exists, which can limit the diagnostic value of a single night measurement regardless of the nature of said measurement.

Aside from AHI-CPAP_{flow} accuracy, the question that also arises is the existence of an AHI-CPAP_{flow} threshold above which a more interventionist attitude should be adopted to avoid clinical consequences such as loss of CPAP-adherence and/or symptom recurrence. To date, this question remains debated. Whereas the AASM 2019 statement suggests that clinicians use telemonitoring-guided interventions during the initial period of CPAP therapy, no guiding AHI-CPAP_{flow} threshold was proposed. \[2\] The 2013 American Thoracic Society statement has speculated that an AHI-CPAP_{flow} over 10/h might be associated with a risk of CPAP-non-adherence (≤4 hours/day, 70% of the days), \[11\] but this was not the case when this threshold was tested for 650 long-term CPAP-treated patients. \[25\] On the other hand, in a 12285 patient cohort, an AHI-CPAP_{flow} >5/h (22% of the studied population) was associated with a statistically lower CPAP usage (mean of 5.75 vs 6.00 hours/night). \[26\]

To interpret individual AHI-CPAP_{flow} scores, care must also be taken to properly account for patient symptoms (ie, symptoms not only in terms of CPAP efficiency but also in terms of patient-reported side effects like patient-reported-leaks). In this regard, for telemedicine care, patient questionnaires are likely to play a complementary role in addition to CPAP-reported data. \[27\]

Regardless of our review and meta-analysis results, there is an urgent need to standardise the respiratory event definitions used by device manufacturers. The current variability limits the usefulness of AHI-CPAP_{flow} in clinical practice. Eight years after the American Thoracic Society statement and despite a clear recommendation on this issue, \[11\] no progress in this direction has been made. The results of AHI-CPAP_{flow} versus AHI-PSG_{gold} SRMA will be interpreted and discussed in this context.

Author affiliations

1IMAG, CNRS, Univ Montpellier, CHU Montpellier, Montpellier, France
2Association pour l’assistance et la réhabilitation à domicile (Apard) groupe Adène, Montpellier, France
3Department of Respiratory Diseases, Univ Montpellier, CHU Montpellier, Montpellier, France
REFERENCES


