Title

Application of inverse-probability-of-treatment weighting to estimate the effect of daytime sleepiness in obstructive sleep apnea patients

short title

Causal inference methods in sleep apnea

Author

François Bettega[1], Clémence Leyrat[2], Renaud Tamisier[1], Monique Mendelson[1],Yves Grillet[3], Marc Sapène[4], Maria R Bonsignore[5], Jean Louis Pépin[1], Michael W Kattan[6], Sébastien Bailly[1]

Affiliations

[1] University Grenoble Alpes, Inserm, Grenoble Alpes University Hospital, HP2, 38000 Grenoble, France

[2] Department of Medical Statistics, Inequalities in Cancer Outcomes Network, London School of Hygiene and Tropical Medicine, London, UK

[3] Private Practice Sleep and Respiratory Disease Centre, Nouvelle Clinique Bel Air, Bordeaux, France

[4] Private Practice Sleep and Respiratory Disease Centre, Valence, France

[5] Respiratory Medicine, PROMISE Department, University of Palermo and IRIB-CNR, Palermo, Italy

[6] The Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic, Cleveland, Ohio, United States

Corresponding author:

Sébastien Bailly Laboratoire EFCR - CHU de Grenoble Rond point de la Chantourne - CS10217 38043 Grenoble Cedex 9 France. Email: sbailly@chu-grenoble.fr

Conflict of interest:

None to declare

Fundings

JLP, SB, and FB are supported by the French National Research Agency in the framework of the "Investissements d'avenir" program (ANR-15-IDEX-02) and the "e-health and integrated care and trajectories medicine and MIAI artificial intelligence" Chairs of excellence from the Grenoble Alpes University Foundation. CL is supported by the UK Medical Research Council (MRC Skills Development Fellowship MR/T032448/1). This work has been partially supported by MIAI University Grenoble Alpes, (ANR-19-P3IA-0003)

Prior abstract publication:

None

Word count

Text word count: 3404

Abstract word count: 299

Glossaire

ADR: Adverse Drug Reactions

AHI: Apnea-Hypopnea index

ATE: Average Treatment Effect

ESS: Epworth Sleepiness Scale

IPW-RA: Inverse Propensity Weighted Regression Adjustment

BMI: Body Mass Index

CI: Confidence Interval

CPAP: Continuous Positive Airway Pressure

IPTW: Inverse Probability of Treatment Weighting

IQR: Inter Quartile Range

OSA: Obstructive Sleep Apnea

OSFP: Observatoire Sommeil de la Fédération de Pneumologie

RCT: Randomised Controlled Trial

Abstract

Introduction : Continuous positive airway pressure (CPAP), the first line therapy for obstructive sleep apnea (OSA), is considered effective in reducing daytime sleepiness. Its efficacy relies on adequate adherence, often defined as >4 hours per night. However, this binary threshold may limit our understanding of the causal effect of CPAP adherence and daytime sleepiness and multilevel approach for CPAP adherence can be more appropriate.

Objective : In this study, we show how two causal inference methods can be applied on observational data for the estimation of the effect of different ranges of CPAP adherence on daytime sleepiness measured by the Epworth sleepiness score (ESS) .

Methods : Data were collected from a large prospective observational French cohort for OSA patients. Four groups of CPAP adherence were considered (0-4; 4-6; 6-7 and 7-10 hours per night). Multivariable regression, inverse-probability-of-treatment weighting (IPTW) and IPTW with regression adjustment (IPTW-RA) were used to assess the impact of CPAP adherence level on daytime sleepiness.

Results : In this study, 9,244 OSA patients treated by CPAP were included. The mean initial ESS was 11 ($\pm$5.2) with a mean reduction of 4 points ($\pm$5.1). Overall, there was an evidence of the causal effect of CPAP adherence on daytime sleepiness which was mainly observed between the lower CPAP adherence group (0-4h) compared to the higher CPAP adherence group (7-10h). There are no differences by considering higher level of CPAP adherence (>4h).

Conclusion : We showed that IPTW and IPTW-RA can be easily implemented to answer questions regarding causal effects using observational data when randomized trials cannot be conducted. Both methods give a direct causal interpretation at the population-level and allow the assessment of the appropriate consideration of measured confounders.

Keywords

Causal inference; inverse probability weight, daytime sleepiness, sleep apnea

Introduction

Obstructive sleep apnea (OSA) is a major health concern with multi-organ consequences and significant economic cost and social burdens.[1] OSA is defined by recurrent complete or partial obstruction of the upper airway during sleep. It has been estimated that more than one billion adults aged 30–69 years (men and women) worldwide suffer from moderate to severe OSA.[2] Obstructive sleep apnea frequently co-occurs with comorbidities such as obesity, diabetes, hypertension or other cardiovascular and metabolic diseases and has a major impact on quality of life.[1,3–5] Continuous positive airway pressure (CPAP), the first-line therapy for OSA, is highly effective in terms of symptom improvement, even in minimally symptomatic patients who initially complain of fatigue and non-restorative sleep.[6] Previous studies have demonstrated that adequate adherence to CPAP treatment is the pre-requisite for reducing symptoms. CPAP also has an effect on quality of life by improving daytime sleepiness. Indeed, CPAP use is associated with a significant decrease in the Epworth Sleepiness Score (ESS)[7] and a majority of initially sleepy patients (with ESS>10) experience a significant improvement in their ESS after CPAP initiation.[8] The effect size of the response and the dose-response relationship have mainly been established by meta-analyses summarizing existing randomized clinical trials.[9,10]

Randomized Controlled Trials (RCT) are considered as the gold standard for causal inference in medical research, providing the highest level of evidence. Unfortunately, RCTs are not always feasible for ethical, logistical or financial reasons.[11,12] Furthermore, RCTs often have strict inclusion criteria, and typically include younger patients with few or no comorbidities, thus limiting the generalizability of the findings.[13] Current data emerging from RCTs might not represent the true impact of CPAP for reducing subjective sleepiness in unbiased real life populations.[14–16]

When RCTs cannot be implemented or when real-world evidence is needed, observational studies, such as cohorts or registries, contain a wealth of data for causal inference. However, unlike RCTs, observational studies are prone to confounding bias due to the absence of randomization, meaning that treatment groups might be unbalanced. Therefore, specific statistical methods have been proposed to address this issue in order to target the causal nature of the relationship between

multilevel exposures and outcomes. There are several ways to account for the effect of measured confounding factors. In medical research, the most standard approaches are multivariable regression and standardization for the estimation of marginal effects. In medical research, the most standard approaches are multivariable regression and standardization for the estimation of marginal effects. Propensity score (PS) approaches, are a first methodological way to replicate covariate balance associated with randomized trials (with the difference that PSs only achieve balance on measured variables) and minimize selection bias. This was explored by Keenan et al.[17] for balance CPAP adherence. However, PS approaches result in a decrease in the overall sample size. Finally, the inverse probability of treatment weighting (IPTW) estimator developed within the counterfactual theory has been increasingly used.[18]

When a study population is large enough, propensity score based methods, such as IPTW, and multivariable regression lead to similar results.[19] Unlike multivariable regression, IPTW allows the comparison and evaluation of covariate balance after weighting and leads to directly interpretable marginal effects (and not conditional effects). Most of the IPTW theory has been developed for binary exposures and its implementation for multilevel exposures has been given little attention, which can explain its limited use in practice.[20]

In the present study, we aim to describe two weighted methods for the estimation of causal effects – namely inverse-probability-treatment weighting (IPTW) and IPTW with regression adjustment (IPW-RA) and to illustrate their implementation and interpretation for the analysis of the causal effect of CPAP adherence on the change in ESS from baseline in a large national prospective cohort.

Material and methods

Patients with a diagnosis of sleep apnea by polygraphy or polysomnography, older than 18 years old and treated by CPAP were included from the "Observatoire Sommeil de la Fédération de Pneumologie" (OSFP) database, a National French Registry for sleep apnea. Patients with missing values for CPAP adherence or ESS either at the diagnostic visit or at the first follow up visit were excluded from our study.

We used a multiple imputation by chained equations (MICE) method to replace missing data values in the dataset under certain assumptions about the data missingness mechanism (i.e. assuming that the data are missing at random). Details on imputation are available in supplementary materials 1, number of missing value are presented in supplementary table S1.

The exposure, i.e. average objective adherence, came from CPAP devices download during the first follow-up visit by the pulmonologist. In order to evaluate adherence as a multilevel treatment, patients were divided into four equally sized adherence groups based on average nightly CPAP adherence as follows: i. CPAP use between 0 and 4 hours by night, ii. between 4 and 6 hours by night, iii. between 6 and 7 hours by night and iv. between 7 and 10 hours by night. The last group was used as the reference.

The outcome, daytime sleepiness was assessed using the self-administered ESS questionnaire which leads to a score between 0 and 24, with 24 being the maximum drowsiness. The patient's reported Epworth score was considered at two time points: 1) at the diagnostic visit and 2) at the first follow-up visit.

In order to assess the impact of CPAP adherence on the Epworth score, all potential confounders must be accounted for (i.e. all variables that can be related with Epworth score and CPAP adherence must be considered). For example, among OSA patients, younger or very elderly patients or female much more likely to present lower adherence to CPAP.[21] Thus, age, obesity and sex have an impact on the outcome (daytime sleepiness) and therefore must be considered as confounding factors, otherwise the estimate of the causal effect between the exposure and response variables would be biased. To ensure that these potential confounding factors are accounted for, the IPTW

estimator can be applied as explained in the next part of the manuscript. Moreover, the Inverse Propensity Weighting with Regression Adjustment (IPW-RA) can be used to increase the robustness of standard IPTW to a mis-specification of the weight model. This method is a double robust estimator of average treatment effect.[22] In both methods, two steps are considered: 1) a weight is computed for each patient and 2) these weights are subsequently used in a regression model to predict the average treatment effect (ATE) on the Epworth score. The ATE is defined as the average difference between the potential outcomes for every individual in the population. It is the contrast between two hypothetical worlds. When the exposure has multiple levels, there are as many ATEs as possible contrasts.

Counterfactual theory

Contrary to traditional statistics which aim to assess associations between an exposure and an outcome, causal inference refers to specific assumptions and study design to be able to draw causal conclusions from the data.[23] In the potential outcome framework (the framework developed for causal inference), the potential outcome refers to what would have happened if a patient had received a treatment.[24] In order to evaluate the effect of several treatments, it is necessary to establish the effect of each treatment on each patient. There are as many potential outcomes as there are treatments. However, the observation of each potential outcome is nearly impossible because each patient usually receives only one treatment.

One of the main problems with observational data is the fact that the exposure is not independent of the other variables. In order to address this issue, the IPTW is based on the creation of a pseudo-population in which the exposure variable (i.e. CPAP adherence) becomes independent of the potential outcomes given the covariates.

The pseudo-population is the result of assigning a weight to each participant that is, informally, proportional to the participant's probability of receiving his/her own exposure.

Assumptions

In order to use a pseudo-population to measure the effect of CPAP adherence on daytime sleepiness without bias, we need to verify four assumptions to identify the causal effect: 1) consistency 2) non-interference, 3) conditional exchangeability, 4) positivity, and one assumption about the estimation of the causal effect: no model mis-specification. The four assumptions are summarized in Table 1.

The consistency assumption is often stated such that an individual's potential outcome under his/her observed exposure is exactly the same outcome as it would have been if they received his/her observed intervention via the hypothetical intervention.[18,25] This assumes that observing is the same as intervening. The treatment needs to be precisely defined to ensure that observed treatment and hypothetical treatment use in causal framework lead to the same outcomes for a given patient.

Non-interference assumption states that an individual's treatment has no influence on other individuals potential outcomes An example of a violation of this assumption is to consider vaccines because vaccinating one individual may affect the disease status of other individuals.

Conditional exchangeability refers to the assumption of no unmeasured confounders. In causal inference, all joint predictors of exposure and outcomes must be accounted for. Thus, all variables related with treatments and outcomes, (i.e. in this case all variables linked with CPAP adherence and daytime sleepiness variations) have to be included. In our study, this is illustrated by a causal directed acyclic graphs (Figure 1).[26] Information on the design of the DAG and the links between the variables are provided in supplementary Material 2. This assumption is empirically untestable and can only be verified through expert knowledge.[27] However, we can assume that most important confounders have been properly included in the OSFP database due to expert medical knowledge and the selection of variables.

Positivity states that given their own characteristics, every individual has a non-zero probability of receiving any exposure level.[27] For example, the existence of formal contraindications to one of the treatments evaluated among the observed population is a violation of the positivity assump-

tion because the patients with the contraindication could not be exposed to the contraindicated treatment.

In addition to these fours assumptions, we need a correct model specification i.e. the unknown probabilities for a patient to belong to each treatment group knowing all confounders is modelled through a correctly specified model. For instance, modelling an exponential relationship using a linear model is a violation of this assumption. This assumption also states that all confounding variables and their real functional forms are used to fit the model. Double robust estimators such as IPW-RA can help address this assumption.

Weight estimation and balancing properties

In our study, we predicted for the patient i the probability of belonging to the adherence group noted A given their confounding factors. We used age and BMI for illustrative purpose in this example.

$$P(A_i = a_i | AGE_i = age_i, BMI_i = bmi_i, ...)$$

Then, each individual was weighted according to the inverse of the probability of receiving the treatment they actually received (i.e. their adherence group). The probability of each individual of belonging to their treatment group were computed using a multinomial logistic regression:

$$IPW_i = \frac{1}{P(A_i = a_i | AGE_i = age_i, BMI_i = bmi_i, ...)}$$

An example of weight assessment is illustrated in Figure 2. In order to minimise the bias-variance compromise, a weight truncation was performed. All weights that exceed a specified threshold were each set to that threshold. Several thresholds for weight truncation were investigated from 1-99th to 25-75th percentiles and the threshold which offer the best bias-variance ratio was chosen.

To create a model able to estimate the causal effect, we needed to select a set of variables. We will use the recommendations of Lefebvre et al. on model specification.[28] These authors recommended the inclusion of all risk factors (confounders or not) and avoid including pure predictors of exposure, also known as instrumental variables, in the treatment model. According to these rules,

for this study, candidate variables are all variables which are not instrumental variables, selected by using univariable linear regressions for the outcome, with less than 60% of missing values, without colinearity. To control for the type I error rate, selection is carried out on a subgroup consisting of 20 % patients stratified by adherence group. These patients are used only to choose the variables and were removed for weight and final models. To ensure that no major confounding factors have been overlooked by our procedure, the choice of variables to be included and their relationship to each other was made in collaboration with OSA clinical experts (RT, JLP and MB).

Four approaches were implemented and compared for the estimation of the outcome: 1) mean comparison using a t-test , 2) a multivariable regression, 3) using the IPTW estimator and 4) using the double-robust IPW-RA. The simple t-test does not account for confounding, and therefore is not appropriate for the analysis of non-randomised studies. We report these results as a benchmark for adjusted methods. Final results are expressed as Average Treatment Effect (ATE)(95% confidence interval) from the reference group (7-10 hours).

For IPTW the final weighted model was adjusted for the exposure. For the IPW-RA, the final weighted model was adjusted for the exposure and all the confounders included in the weight model, to account for potential remaining imbalances in confounders between groups. IPW-RA combines the strengths of IPTW and multivariable regression: confounders are adjusted for in a multiple regression model, also weighted by the inverse of the propensity score. By doing so, the causal effect estimate will be unbiased if either the weight model or the outcome model is correctly specified. The weights are the IPTW weights, and the covariates in the outcome model can be any covariate still unbalanced despite the weighting. A simple IPW-RA allows the estimation of a conditional causal effect, but it is possible to marginalize on the distribution of the covariates to estimate a marginal effect (the ATE).

As model based standard errors are incorrect because they do not account for the uncertainty in PS estimation, we chose to bootstrap weight and treatment effect estimations. This allowed us to estimate confidence intervals based on percentiles for the treatment effect without making assumptions about the distribution of the parameters. We evaluated the possibility of integrating missing value imputation into the bootstrap, but this procedure was time consuming and increased

the time needed far too much. To keep a reasonable execution time, we therefore chose to impute the missing values before the bootstrap.[29]

We performed a complementary analysis using the same model and method with morning fatigue as outcome. This analysis is presented in the supplementary material 3.

All statistical analyses were performed using R Statistical Software (version 4.0.2).The tests were performed at a 5% significant level.

Results

Population

From the OSFP database, 9,244 patients were included in the study.

The included patients were mainly men (n = 6,492, 70.2 %) with a mean age of 57 years (standard deviation $\pm$12.4) and mean body mass index (BMI) was 32 kg/m^2 ($\pm$6.9). The mean apnea hypopnea index (AHI) was 41 events/hour ($\pm$20.4) and 6,510 (70 %) of patients had severe OSA.

The mean observance was 05 hours 35 min by night and patients were divided in four groups according to their average CPAP use by night as follows: 1) CPAP use between 00 hours and 04 hours, n = 1,977 (21.4 %), 2) CPAP use between 04 hours and 06 hours, n = 3,519 (38.1 %) CPAP use between 06 hours and 07 hours, n = 2,023 (21.9%) and 4) CPAP use between 07 hours and 10 hours, n = 1,725 (18.7 %). For more information, differences across all variables and subgroups are presented in Table 2.

Overall, the unadjusted mean initial ESS was 11 ($\pm$5.2). There was a mean reduction in the Epworth score of 4 ($\pm$5.1) at the follow-up visit under CPAP treatment. This reduction was different according to the CPAP adherence groups: the smallest difference was observed in the 0-4 hours adherence group with a mean Epworth score which varied from 11 ($\pm$5.3) to 8 ($\pm$4.7) resulting in a difference of 3 ($\pm$5) which was lower compared to the three other adherence groups. In the 4-6 h adherence group, the mean reduction was 5 ($\pm$5.2) and was similar to the higher level of CPAP adherence groups (6-7 hours and 7-10 hours adherence groups) with a mean reduction of 5 ($\pm$5.2) (Figure 3).

Comparison of statistical approaches

First, we performed a weighted regression analysis using the IPTW estimator. The positivity assumption was verified by making sure that the mean of the maximum IPTW weights obtained by bootstrap was reasonable. In order to investigate the presence of outliers, we verified the distribution of the weights by adherence group (supplementary table S2). The mean of truncated weights was

4 (4; 4) (supplementary table S3).

After weighting, the standardized mean differences of all variables for each adherence group showed no imbalance on confounders (Figure 4). Adjustment in the final model was performed to correct for potential remaining imbalance in confounders.

The coefficients of the multinomial logistic regression are available in supplementary table S4.

The second modelling approach was based on the double robust approach, IPW-RA. Coefficients of the IPW-RA are available in supplementary table S5.

Confounders selected are the following variables at diagnosis : age (years), sotwobody mass index (kg/m^2), neck circumference (cm), sleepiness at the wheel, morning tiredness, morning headaches,libido disorder, dysfunction, night sweating, daytime sleepiness measured by ESS scale, fatigue measured by Pichot's scale, depression measured by Pichot's depression scale, apnea hypopnea index, sex, , restless legs syndrome, morning tiredness, morning headaches, night sweating and fatigue measured by Pichot's scale.

When the adherence groups are compared with an unweighted mean difference, patients in 0-4 hours adherence group, have an average Epworth score of 2 (95% bootstrap confidence intervals based on percentiles 1.9; 2.1) points higher than the reference group. Patients in 4-6 hours adherence group, have an average Epworth score of 0.8 (0.7; 0.9) points higher than patients in 7-10 hours adherence group. Patients with a 6-7 hours adherence group, have an average Epworth score of 0.2 (0; 0.3) point higher than patients in 7-10 hours adherence group, there is an evidence of a difference between groups (overall CPAP adherence effect, $p<0.001$). By using multivariable regression, IPTW or IPW-RA estimators, the results are attenuated as compared to the unadjusted analysis but similar to each other: patients in 0-4 hours adherence group, have an average Epworth score of 1.1 points (0.8; 1.3) higher than patients in 7-10 hours adherence group with IPTW. Patients in 4-6 hours adherence group, have an average Epworth score of 0.5 points (0.3; 0.7) higher than patients in 7-10 hours adherence group. Patients with a 6-7 hours adherence group, have an average Epworth score of 0.2 points (0; 0.5) higher than patients in 7-10 hours adherence group. (results for the four methods are presented in Figure 5). IPTW-RA was the most efficient approach, leader to narrower 95% confidence intervals.

Discussion

In this study, we illustrated the application of an inverse probability of treatment weighting estimator by assessing the impact of CPAP adherence levels on ESS in a large population (n = 9,244) of CPAP-treated patients. Our results suggested evidence of a difference in the Epworth score at the follow-up visit between low adherence groups and the high adherence group (>7h). However, there was no evidence of a difference for patients with high CPAP adherence level (6-7h vs. 7-10h groups). This result is consistent with other studies.[30,31]Further applications should be performed to investigate the use of such methods on other symptoms and signs of daytime sleepiness.

The results presented in Figure 5 showed that the effect size (the mean difference in Epworth score between each adherence group and the reference) is of greater magnitude when confounding is not accounted for (unadjusted analysis) compared with adjusted multivariable regression and weighted methods. This is due to the presence of positive confounding when comparing groups, and therefore these estimates are strongly biased. To avoid inducing bias, multivariable regression and weighted methods accounting for all confounding factors should be preferred. In summary, unadjusted methods lead to bias estimates if groups being compared differ in terms of characteristics also associated with the outcome, which is almost always the case in non-randomised studies. Multivariable regression models, under a correct specification of the model and the validity of the four identifiability assumptions, lead to the estimation of unbiased conditional causal effects. Although marginal and conditional effects are identical in linear models, it is not the case in non-linear model and conditional effects are harder to interpreting in a public health context. Using weighted approaches, an investigation of the weight distribution allows us to identify violations of the positivity assumption and potential lack of overlap. When these problems exist, standard regression methods rely on extrapolation, but an incorrect extrapolation will be more difficult to identify . Furthermore, the treatment allocation mechanism is sometimes easier to model than the outcome mechanism, thus reducing the risk of model misspecification using weighted estimators. In addition, weighted estimators lead to the estimation of marginal effects, which are often more easily interpretable and useful for policy making, as compared with conditional estimates from

multivariable regression. Finally, when using multivariable regression, the model is built to estimate the causal effect of the intervention, but the other coefficients of the model do not have a causal interpretation. However, in practice, the other coefficients are reported and interpreted causally, whereas this is prevented with the use of weighted estimators, which provide a single estimate. In addition, IPW-RA has a double robust property, which minimizes the risk of bias due to model misspecification.

By comparing results of IPTW and IPW-RA, conclusions are similar, however, the use of a double-robust IPW-RA estimator allows an increase in confidence in the consideration of possible risk of model misspecification.

However, as in any observational study, unmeasured confounders cannot be ruled out, but it cannot be checked from the data.

From a methodological point of view, the present study highlights the benefits of applying IPW methods to estimate the effect of a multilevel exposure in assessing marginal causal effects. Under a set of assumptions, it is possible to estimate the causal effect of an exposure on an outcome with well-designed observational studies, which can be used as an alternative to randomized clinical trials.[32] In this study the ATE is the most relevant estimand to understand the potential benefits if all the patients were adherent. However, the weights can be easily modified for the estimation of the Average Treatment effect on the Treated. IPTW and IPW-RA are examples of modern statistical methods developed over the past decades which have improved health research by moving the interpretation from associational to causal.[33]

In order to limit the risk of bias when using IPTW, it is important to assess the distribution of the weights and to consider whether the functional form of the weight model is correctly specified. In addition, a careful investigation of the plausibility of the assumptions required for causal inference is needed. This implies extensive discussion between the statistician and clinician to be vigilant with regards to variable selection and model validation. Moreover, if IPW estimators are now well-known and extensively used, mainly for binary exposure, application of such methods need to be carefully performed and the reporting of these methods in clinical research should be improved.[34]

Unlike for binary exposures, a few published studies[35,36] have applied IPW to multilevel expo-

sure, in order to assess the marginal causal effect on an outcome. However, multilevel exposure are of great relevance in the medical field as many treatments have several levels or need to be compared to other treatments or combination of treatments.

We proposed a method to consider CPAP adherence as a multilevel variable, in contrast to a binary one and we have illustrated the application of IPW method to reduce confounding bias. Further methodological research for causal inference applied to multivalued exposures could be proposed. Indeed it could be of interest to compare these approaches to those based on machine learning algorithms to calculate weights, such as the Gradient boosting algorithms that have already been used for this purpose.[37] Beyond the trimming we used, which relies on the choice of an arbitrary cutoff, other methods, such as the overlap weighting methods,[38] have been developed to minimize the risk of extreme propensity score.

In conclusion, IPW for multi-level exposures is a promising approach. This study showed that patients with different levels of CPAP adherence experienced a different reduction in their daytime sleepiness measured by the Epworth score. We also showed that patients who had high CPAP adherence experienced a greater reduction in daytime sleepiness than non-adherent patients at their first follow-up visit.

Acknowledgments

Guarantor

Bettega François

Author contributions

F.B., C.L., S.B. contributed to the study design, analysis and interpretation of the data, F.B., C.L., S.B., M.M., R.T., M.R.B., J.L.P, M.K contributed substantially to the study design, data interpretation, and the writing of the manuscript, R. T., Y. G., and M. S. were responsible for acquisition of data, contributed to the discussion, and reviewed the manuscript. S.B., C.L. and J.-L. P. were responsible for study concept and design, supervised the study, and critically revised the manuscript. S.B. is the guarantor of this work and, as such, had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Financial/nonfinancial disclosures

None declared.

Role of the sponsors

The sponsor had no role in the design of the study, the collection and analysis of the data, or the preparation of the manuscript

References

[1] Lévy P, Kohler M, McNicholas WT, Barbé F, McEvoy RD, Somers VK, et al. Obstructive sleep apnoea syndrome. *Nature Reviews Disease Primers* . 2015. 1(1):1–21.

[2] Benjafield AV, Ayas NT, Eastwood PR, Heinzer R, Ip MSM, Morrell MJ, et al. Estimation of the global prevalence and burden of obstructive sleep apnoea: A literature-based analysis. *The Lancet Respiratory Medicine* . 2019. 7(8):687–698.

[3] Pépin JL, Timsit JF, Tamisier R, Borel JC, Lévy P, Jaber S. Prevention and care of respiratory failure in obese patients. *The Lancet Respiratory Medicine* . 2016. 4(5):407–418.

[4] Zinchuk AV, Jeon S, Koo BB, Yan X, Bravata DM, Qin L, et al. Polysomnographic phenotypes and their cardiovascular implications in obstructive sleep apnoea. *Thorax* . 2018. 73(5):472–480.

[5] Pépin JL, Bailly S, Tamisier R. Incorporating polysomnography into obstructive sleep apnoea phenotyping: Moving towards personalised medicine for OSA. *Thorax* . 2018. 73(5):409–411.

[6] McEvoy RD, Antic NA, Heeley E, Luo Y, Ou Q, Zhang X, et al. CPAP for Prevention of Cardiovascular Events in Obstructive Sleep Apnea. *New England Journal of Medicine* . 2016. 375(10):919–931.

[7] McDaid C, Durée KH, Griffin SC, Weatherly HL, Stradling JR, Davies RJ, et al. A systematic review of continuous positive airway pressure for obstructive sleep apnoea–hypopnoea syndrome. *Sleep Medicine Reviews* . 2009. 13(6):427–436.

[8] Gasa M, Tamisier R, Launois SH, Sapene M, Martin F, Stach B, et al. Residual sleepiness in sleep apnea patients treated by continuous positive airway pressure. *J Sleep Res* . 2013. 22(4):389–397.

[9] Crook S, Sievi NA, Bloch KE, Stradling JR, Frei A, Puhan MA, et al. Minimum important difference of the Epworth Sleepiness Scale in obstructive sleep apnoea: Estimation from three randomised controlled trials. *Thorax* . 2019. 74(4):390.

[10] Patel S, Kon SSC, Nolan CM, Barker RE, Simonds AK, Morrell MJ, et al. The Epworth Sleepiness Scale: Minimum Clinically Important Difference in Obstructive Sleep Apnea. *Am J Respir Crit Care Med* . 2018. 197(7):961–963.

[11] Ware JH, Hamel MB. Pragmatic Trials — Guides to Better Patient Care? *N Engl J Med* . 2011. 364(18):1685–1687.

[12] Benson K. A Comparison of Observational Studies and Randomized, Controlled Trials. *The New England Journal of Medicine* . 2000. 9.

[13] Kennedy-Martin T, Curtis S, Faries D, Robinson S, Johnston J. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. *Trials* . 2015. 16(1):495.

[14] Pack AI, Magalang UJ, Singh B, Kuna ST, Keenan BT, Maislin G. To RCT or not to RCT? Depends on the question. A response to McEvoy et al. *Sleep* . 2021. 44(zsab042).

[15] Pack AI, Magalang UJ, Singh B, Kuna ST, Keenan BT, Maislin G. Randomized clinical trials of cardiovascular disease in obstructive sleep apnea: Understanding and overcoming bias. *Sleep* . 2021. 44(zsaa229).

[16] McEvoy RD, Sánchez-de-la-Torre M, Peker Y, Anderson CS, Redline S, Barbe F. Randomized clinical trials of cardiovascular disease in obstructive sleep apnea: Understanding and overcoming bias. *Sleep* . 2021. 44(4).

[17] Keenan BT, Maislin G, Sunwoo BY, Arnardottir ES, Jackson N, Olafsson I, et al. Obstructive sleep apnoea treatment and fasting lipids: A comparative effectiveness study. *European Respiratory Journal* . 2014. 44(2):405–414.

[18] Hernan MA. A definition of causal effect for epidemiological research. *Journal of Epidemiology & Community Health* . 2004. 58(4):265–271.

[19] Stuart EA. Matching Methods for Causal Inference: A Review and a Look Forward. *Statistical Science* . 2010. 25(1):1–21.

[20] Yoshida K, Solomon DH, Haneuse S, Kim SC, Patorno E, Tedeschi SK, et al. Multinomial Extension of Propensity Score Trimming Methods: A Simulation Study. *Am J Epidemiol* . 2019. 188(3):609–616.

[21] Pépin JL, Bailly S, Rinder P, Adler D, Szeftel D, Malhotra A, et al. CPAP Therapy Termination Rates by OSA Phenotype: A French Nationwide Database Analysis. *JCM* . 2021. 10(5):936.

[22] Glynn AN, Quinn KM. An Introduction to the Augmented Inverse Propensity Weighted Estimator. *Polit. anal.* . 2010. 18(1):36–56.

[23] Pearl J. An Introduction to Causal Inference. *The International Journal of Biostatistics* . 2010. 6(2).

[24] Vandenbroucke JP, Broadbent A, Pearce N. Causality and causal inference in epidemiology: The need for a pluralistic approach. *International Journal of Epidemiology* . 2016. 45(6):1776–1786.

[25] Robins JM, Hernán MÁ, Brumback B. Marginal Structural Models and Causal Inference in Epidemiology:. *Epidemiology* . 2000. 11(5):550–560.

[26] Etminan M, Collins GS, Mansournia MA. Using Causal Diagrams to Improve the Design and Interpretation of Medical Research. *Chest* . 2020. 158(1S):S21–S28.

[27] Cole SR, Hernan MA. Constructing Inverse Probability Weights for Marginal Structural Models. *American Journal of Epidemiology* . 2008. 168(6):656–664.

[28] Lefebvre G, Delaney JAC, Platt RW. Impact of mis-specification of the treatment model on estimates from a marginal structural model. *Statist. Med.* . 2008. 27(18):3629–3642.

[29] Schomaker M, Heumann C. Bootstrap inference when using multiple imputation: Bootstrap Inference When Using Multiple Imputation. *Statistics in Medicine* . 2018. 37(14):2252–2266.

[30] Weaver TE, Maislin G, Dinges DF, Bloxham T, George CFP, Greenberg H, et al. Relationship between hours of CPAP use and achieving normal levels of sleepiness and daily functioning. *Sleep* . 2007. 30(6):711–719.

[31] Stepnowsky CJ, Dimsdale JE. Dose–response relationship between CPAP compliance and measures of sleep apnea severity. *Sleep Medicine* . 2002. 3(4):329–334.

[32] Frieden TR. Evidence for Health Decision Making — Beyond Randomized, Controlled Trials. *N Engl J Med* . 2017. 377(5):465–475.

[33] Glass TA, Goodman SN, Hernán MA, Samet JM. Causal Inference in Public Health. *Annu. Rev. Public Health* . 2013. 34(1):61–75.

[34] Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Statist. Med.* . 2015. 34(28):3661–3679.

[35] Ahn HJ, Lee SR, Choi EK, Han KD, Jung JH, Lim JH, et al. Association between exercise habits and stroke, heart failure, and mortality in Korean patients with incident atrial fibrillation: A nationwide population-based cohort study. *PLoS Med* . 2021. 18(6):e1003.659.

[36] Ali AK, Hartzema AG, Winterstein AG, Segal R, Lu X, Hendeles L. Application of multicategory exposure marginal structural models to investigate the association between long-acting beta-agonists and prescribing of oral corticosteroids for asthma exacerbations in the Clinical Practice Research Datalink. *Value Health* . 2015. 18(2):260–270.

[37] Olmos A, Govindasamy P. A Practical Guide for Using Propensity Score Weighting in R. *Practical Assessment, Research and Evaluation* . 2015. 20.

[38] Li F, Thomas LE, Li F. Addressing Extreme Propensity Scores via the Overlap Weights. *American Journal of Epidemiology* . 2019. 188(1):250–257.

List of tables

Table 1: Table of causality assumptions

Assumptions	Definitions	Can it be tested from the data?
Consistency	The outcome of an individual under their observed exposure is the same as their potential outcome had they received their observed intervention via the hypothetical intervention	No
Non-interference	The treatment received by an individual has no influence on the potential outcomes of the other individuals	No
Conditional exchangeability	Given the measured variables, the exposure and potential outcomes are independent, i.e. all joint predictors of exposure and outcomes are accounted for	No, but the investigation of the balance between exposure groups after weighting may give an indication of the plausibility of this assumption for the measured variables (but not the unmeasured variables)
Positivity	Given their own characteristics, every individual has a non-zero probability of receiving any exposure level.	Yes, by investigating the range of the estimated propensity score values

Table 2: Patient characteristics according to the adherence group

	All groups 9,244	0-4 h (1) 1,977 (21.4%)	4-6 h (2) 3,519 (38.1%)	6-7 h (3) 2,023 (21.9%)	7-10 h (4) 1,725 (18.7%)
Variables at diagnosis					
ESS score	10.9 (5.2)	11.0 (5.3)	10.9 (5.2)	10.8 (5.2)	10.8 (5.3)
Gender (male)	6,485 (70.2%)	1,337 (67.6%)$_{3}$	2,470 (70.2%)	1,457 (72%)$_{1}$	1,221 (70.8%)
Age (years)	57.3 (12.4)$_{1,4}$	55.5 (12.8)$_{2,3,4}$	56.7 (12.1)$_{1,3,4}$	57.9 (11.9)$_{1,2,4}$	59.7 (12.4)$_{1,2,3}$
Body mass index (kg/m2)	32.0 (7.1)	31.9 (6.8)	31.9 (7.4)	32.0 (7.2)	32.4 (6.8)
Apnea hypopnea index (event/h)	40.7 (20.5)$_{1,3,4}$	37.8 (19.6)$_{2,3,4}$	39.7 (19.7)$_{1,3,4}$	42.7 (21.5)$_{1,2}$	43.6 (21.2)$_{1,2}$
Tobacco status	0.6 (0.7)	0.6 (0.8)$_{4}$	0.6 (0.7)$_{4}$	0.6 (0.7)	0.5 (0.7)$_{1,2}$
Depression scale	4.0 (3.8)	4.3 (3.9)$_{2}$	3.9 (3.8)$_{1}$	3.9 (3.7)	4.1 (3.8)
Pichot's fatigue scale	13.3 (8.2)	13.8 (8.6)$_{2}$	12.9 (8.1)$_{1,4}$	13.1 (8.1)	13.6 (8.2)$_{2}$
Morning headaches	3,697 (40%)	820 (41.5%)	1,423 (40.4%)	813 (40.2%)	641 (37.2%)
Morning tiredness	7,178 (77.7%)	1,546 (78.2%)	2,743 (77.9%)	1,560 (77.1%)	1,329 (77%)
Diabetes	2,492 (27%)	568 (28.7%)	906 (25.7%)	518 (25.6%)	500 (29%)
Variables at follow-up					
ESS score	6.3 (4.2)$_{1,3,4}$	7.5 (4.8)$_{2,3,4}$	6.3 (4.1)$_{1,3,4}$	5.7 (3.9)$_{1,2}$	5.5 (3.9)$_{1,2}$
Residual apnea hypopnea index under CPAP	4.1 (4.7)$_{1}$	4.5 (5.5)$_{2,3}$	3.8 (4.2)$_{1}$	3.9 (4.5)$_{1}$	4.2 (4.9)
Pichot's fatigue scale	7.7 (7.2)$_{1,3}$	9.4 (7.7)$_{2,3,4}$	7.4 (6.9)$_{1,3}$	6.8 (6.7)$_{1,2,4}$	7.4 (7.3)$_{1,3}$
Morning headaches	2,479 (26.8%)$_{4}$	579 (29.3%)$_{4}$	982 (27.9%)$_{4}$	533 (26.3%)$_{4}$	385 (22.3%)$_{1,2,3}$
Morning tiredness	4,954 (53.6%)$_{1,4}$	1,166 (59%)$_{2,3,4}$	1,906 (54.2%)$_{1,4}$	1,055 (52.2%)$_{1}$	827 (47.9%)$_{1,2}$
Number of ADR types under CPAP	0.7 (1.1)$_{1,3,4}$	1.0 (1.3)$_{2,3,4}$	0.6 (1.0)$_{1,3,4}$	0.5 (1.0)$_{1,2}$	0.5 (1.0)$_{1,2}$
Duration since diagnosis (year)	0.9 (1.5)$_{1,4}$	0.8 (1.2)$_{3,4}$	0.9 (1.4)$_{4}$	1.0 (1.7)$_{1}$	1.1 (1.9)$_{1,2}$

ESS : Epworth Sleepiness Scale; CPAP : Continuous Positive Airway Pressure; ADR : adverse drug reaction

* Quantitative variables are presented as mean (standard deviation). Qualitative variables are expressed in number of individuals (% of individuals)

† t-test was performed for the quantitative variables and a Pearson's Chi squared test for the categorical variables after application of a Bonferroni correction for multiple testing.

‡ 1,2,3,4 numbers in subscript refers to columns statistically different at the 5% threshold. e.g. 1 means that there is a statistically significant difference between the adherence group of that column and the 0-4 adherence group (1) for the variable in question.

Figure Legends

Figure 1

Figure 1: Causal directed acyclic graph

Causal directed acyclic graph for the relation between multilevel CPAP adherence and residual daytime sleepiness under CPAP. Dotted straight arrow indicates causal relation under investigation; solid arrows indicate known relations. CPAP: Continuous Positive Airway Pressure; OSA: Obstructive Sleep Apnea; SES: SocioEconomic Status. Personal situation regroups: lifestyle, marital status, children
Grey background: unobserved confounders
Dotted frame: exposure; Solid frame: outcome
Symptoms at baseline : sleepiness at the wheel, morning tiredness, morning headaches, libido disorder, night sweating, fatigue measured by Pichot's scale and mean nocturnal SaO2
Comorbidities : depression measured by Pichot's depression scale and restless legs syndrome.

Figure 2

Figure 2: Illustration of the estimation the inverse probability of treatment weight with a categorical exposure

[1]The population may be divided into two groups according to sex (12 females and 8 males).
[2] Within each group, patients are divided according to their adherence to CPAP. For each individual in each subgroup, the probability of belonging to their actual adherence group ($P_{CPAP|gender}$; i.e., probability of treatment given the sex) may be estimated from empirical proportions. [3] From this probability we compute the inverse probability of treatment weight (IPTW) : $\frac{1}{P_{CPAP|gender}}$.
[4] We use this weighting to create a pseudo-population. In this pseudo population, individuals with a high probability of belonging to a treatment group are down-weighted, and in contrast individuals with a low probability of belonging to a treatment group are up-weighted. The pseudo-population encompasses both factual and counterfactual observations. In this pseudo-population, all adherence groups are exchangeable and it is possible to compute directly the difference for a specific outcome.
[5] A common issue with pseudo-population is that individuals with a very low propensity score (very close to 0) will end up with a huge weight resulting in extremely large pseudo-population and potentially making the weighted estimation unstable. A common way to address this issue is the stabilized weight, which uses the marginal probability of treatment instead of 1 in the weight numerator resulting for a patient belonging in first adherence group in $\frac{P_{CPAP=1}}{P_{CPAP|sex}}$.

Figure 3

Figure 3: Box plot of raw epworth sleepiness score accoding to the visit by adherence group

CPAP : Continuous Positive Airway Pressure

Figure 4

Figure 4: Standardized mean difference before and after weighting

SMD: Standardized Mean Difference; CPAP : Continuous Positive Airway Pressure; ADR: adverse drug reaction; SaO2: arterial oxygen saturation

Figure 5

Figure 5: Difference mean in Epworth score between each adherence group and the reference group using different methods

Each point represents the difference mean in Epworth score between each adherence group and the reference group (7-10h). The vertical bars represent the 95% confidence intervals of these estimates.