Published in final edited form as:

Sleep Med. 2017 September ; 37: 130–134. doi:10.1016/j.sleep.2017.06.022.

Adherence to Continuous Positive Airway Pressure Improves Attention/Psychomotor Function and Sleepiness: A Bias-Reduction Method with Further Assessment of APPLES

Tyson H. Holmes^a and Clete A. Kushida^b

^aHuman Immune Monitoring Center (HIMC); Institute for Immunity, Transplantation and Infection; Stanford University School of Medicine; 299 Campus Drive, Fairchild Science Building, Stanford, CA, 94305, USA

^bStanford Sleep Medicine Center and Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, 450 Broadway Street, MC 5704, Pavilion C, 2nd Floor, Redwood City, CA, 94063-5704, USA

Abstract

Objective/Background—Variable adherence to prescribed therapies for sleep disorders is commonplace. This study was designed to integrate three available statistical technologies (instrumental variables, residual inclusion, and shrinkage) to allow sleep investigators to employ data on variable adherence in the estimation of the causal effect of treatment as received on clinical outcomes.

Patients/Methods—Using data from the Apnea Positive Pressure Long-term Efficacy Study (APPLES), regression adjustment for observed and unobserved confounders was applied to two primary neurocognitive outcomes, plus two measures of sleepiness. We demonstrate how to obtain estimates of reduced uncertainty for the causal effect of treatment as received for continuous positive airway pressure (CPAP) within clinical subpopulations (defined by baseline disease severity) of sleep apnea patients.

Results and Conclusions—Following six months of treatment, statistically significant improvements caused by device adherence were detected for subjective sleepiness in mild, moderate and severe disease, objective sleepiness in severe disease, and attention and psychomotor function in moderate disease. Some evidence for worsening of learning and memory due to increased adherence in moderate disease was also detected. Application to APPLES illustrates that this method can yield bias corrections for unobserved confounders that are substantial—revealing new clinical findings. Use of this fully general method throughout sleep research could sharpen understanding of the true efficacy of pharmacotherapies, medical devices, and behavioral interventions. Extensive technical appendices are provided to facilitate application of this general method. Clinicaltrials.gov identifier NCT00051363.

Corresponding Author: Tel: 650-725-2925, FAX: 650-725-0160, tholmes@stanford.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

adherence; continuous positive airway pressure; instrumental variables; shrinkage estimation; sleep apnea

1. INTRODUCTION

1.1. Conceptual Background

1.1.1. Variable adherence and problem of confounders—Variable adherence to prescribed therapies for sleep disorders is widespread [1]. In sleep research, assessment of clinical response to variable adherence is sometimes made by estimating association rather than causation, between treatment as received and treatment response [2]. Estimates of association can be misleading if interpreted as estimates of causation. Unlike an externally imposed factor, such as random treatment assignment within clinical trials, a patient's level of adherence is largely self-selected. Self-selection introduces confounding variables. Confounding variables compromise consistent estimation of the causal effect of treatment as received ("local average treatment effect" among adherers [3]). For instance, those who adhere more may be older and without adjusting for the possible confounder of age, the estimate of the causal effect of treatment as received will be inconsistent. Bias exists when the expected value of a parameter estimate differs from the true value of that parameter. Roughly, an estimator is inconsistent if bias remains even as sample size grows infinitely large.

1.1.2. Subpopulations and the need for reduced uncertainty of parameter estimates—Estimation bias is not the only obstacle to understanding the causal relationship between treatment as received and clinical outcome. Equally problematic is estimation uncertainty; uncertain estimates of the same parameter can differ widely in value among separate samples drawn from the same population. Uncertainty widens confidence intervals and inflates standard errors. As sample size diminishes, uncertainty increases, and uncertainty reduces statistical power.

Randomized trials in sleep research are designed to have sufficient statistical power for testing hypotheses regarding treatment effect on the primary outcome in the full sample of participants. Nevertheless, sleep investigators are often interested in testing hypotheses regarding treatment effects within subpopulations defined by baseline moderators, such as disease severity. Sample sizes within these subpopulations are often too small to provide adequate statistical power for testing treatment effects using standard statistical procedures. Estimates of treatment effects are uncertain in these subpopulations. The small sample sizes of these subpopulations likewise introduces uncertainty into estimates of the causal effect of treatment as received on clinical outcome. Special methods are required to reduce this uncertainty.

Bias and uncertainty are related. This is the statistical phenomenon of bias-variance trade-off. Namely, statistical methods that reduce bias, including correctives for confounders, can increase uncertainty of parameter estimates. In this paper, we show sleep investigators how

to apply techniques that reduce uncertainty in confounder-adjusted estimates of the causal effect of treatment as received on clinical outcome within subpopulations of interest.

1.1.3 “Adherence dose response”—Throughout the remainder of this paper, we operationalize the causal effect of treatment as received as “adherence dose response.” Specifically, we define “adherence dose response” as the amount and direction of change in the clinical outcome of interest caused by each unit increment of adherence.

1.2. Why Reanalyze APPLES?

Here we reanalyze a dataset from the NHLBI-sponsored Apnea Positive Pressure Long-term Efficacy Study (APPLES). Our original analysis of APPLES data assessed the impact of adherence on three neurocognitive outcomes by adjusting for 102 possible confounders observed at baseline in the comparison of continuous positive airway pressure (CPAP) and sham devices [4]. However, not all confounders are observed. Consistent estimates cannot necessarily be obtained by simply adjusting for measured confounders. In application to the APPLES dataset, Holmes et al. [5] extended [4] in two ways. They demonstrated a method for analysis of longitudinal adherence and outcome data and that method adjusted estimates of adherence dose response for observed and unobserved confounders. Holmes et al. [5] limited analysis to a single neurocognitive outcome—Pathfinder Number Test total time (PFNT-TT). They detected improvement in PFNT-TT due to increased adherence to CPAP but not due to increased adherence to the sham device.

The analysis presented here extends [4] and [5] in two important respects. Estimation of adherence dose response with adjustment for observed and unobserved confounders is expanded to now include two measures of sleepiness. We have included sleepiness to demonstrate clinical breadth of application, and sleepiness is clearly a domain of interest throughout sleep research. Further, we demonstrate how to obtain estimates of adherence dose response of reduced uncertainty within clinical subpopulations of sleep-apnea patients. With these extensions from our prior work, the goal of the present study is to integrate currently available statistical technologies to provide a fully general and accessible method for estimation of adherence dose response to sleep therapies in clinical subpopulations.

1.3. Organization of Paper

The remainder of this paper is organized as follows. Section 2.1 describes the APPLES dataset. Section 2.2 provides an overview of the study’s statistical methods. These methods, though built from existing, proven statistical technologies, have not yet seen widespread use in sleep research. To facilitate their use, for clinical investigators who wish to apply these methods, the Online Supplement provides extensive technical guidelines and details for their collaborating statisticians. The study’s results are summarized in Section 3. The paper concludes with discussion of clinical and research implications in Section 4.

2. MATERIALS AND METHODS

2.1. Materials

2.1.1. Data from APPLES—We begin by describing the dataset analyzed for the present study. In APPLES, double-blind randomization assigned participants to either CPAP or sham device [6]. The present study examined two neurocognitive measures, PFNT-TT and Buschke Selective Reminding Test sum recall (BSRT-SR), plus objective (Maintenance of Wakefulness Test mean sleep latency, MWT-MSL) and subjective (Epworth Sleepiness Scale total score, ESS-TS) sleepiness measures from APPLES. The third primary neurocognitive outcome from APPLES, Sustained Working Memory Test (SWMT) overall midday score, was excluded from analysis here for technical reasons outlined in the Online Supplement. Participants were assessed on the above four outcomes at baseline, two and six months post-randomization [4]. Analyses presented here are for the six-month visit only ($n = 443$ and $n = 403$ participants remaining of originally randomized to CPAP and sham devices, respectively, for 77% follow-up). This dataset contains nearly the same sample of participants as in [5] but does not include data from the two-month visit. The present study's goal was to obtain minimal-uncertainty, minimal-bias estimates of the adherence dose response on each of these four outcomes at six months. Estimates were made in each of three subpopulations defined on baseline apnea-hypopnea index (AHI, number of abnormal sleep-related breathing events per hour of sleep) values of 10 to 15 ($n = 113$), > 15 to 30 ($n = 249$), and > 30 ($n = 484$). Nightly hours of device usage (active CPAP or sham) were captured on an Encore® Pro SmartCard© monitor (Phillips Respironics® Inc., Murrysville, Pennsylvania, USA). Current analysis was on completely de-identified data. All APPLES participants provided written informed consent. The APPLES study protocol was approved by the institutional review board at each participating center. Full details for APPLES have been previously published [4, 6].

2.2. Statistical Methods

2.2.1. Inconsistent estimator—Each clinical outcome was regressed on average CPAP adherence over four to six months post-randomization, age, race (white vs. non-white) and gender. All participants randomized to sham were assigned zero hours for CPAP adherence except for fifteen participants. These fifteen participants had switched to active CPAP by four months. For the present analysis, they were assigned their recorded CPAP adherence. All variables were centered and scaled prior to analysis. Centering and scaling entailed subtraction of total sample mean and division by total sample standard deviation. These multivariate regression models were fit using ordinary least squares.

2.2.2. Consistent Estimator—Consistent estimation of adherence dose response employed a two-stage regression procedure. In the first stage, average CPAP adherence over four to six months post-randomization was regressed on randomly assigned treatment condition, age, race and gender, all coded as described in Section 2.2.1. Randomly assigned treatment condition served as an instrumental variable. An instrumental variable is an external factor that drives variation in clinical outcome but solely through its effects on adherence, a restriction that is facilitated by the double-blind design [4–6]. The second-stage regression was limited to those who received CPAP. This consisted of those who were

randomly assigned to CPAP (none of which switched to sham device) or were randomly assigned to sham device but switched to CPAP by four months. In the second-stage regression, clinical outcome was regressed on age, race, gender, average CPAP adherence over four to six months post-randomization, and the residual from first-stage regression. The residual is that portion of an observation that is unexplained by the fit of the regression model. This “residual inclusion” is the technique of using the unexplained variation from the first-stage regression (here, variation in adherence not explained by condition, age, race, and gender) as an explanatory variable in the second-stage regression. Residual inclusion thereby permits adjustment for unobserved confounders (eg, not age, race and gender) to yield a more consistent estimate [7] of adherence dose response. The regression model in each stage was fit using ordinary least squares.

2.2.3. Shrinkage—Uncertainty compounds through both stages of the two-stage consistent estimator. As such, some means of uncertainty reduction is essential, especially since sample sizes are reduced by subpopulation analyses. Statistical shrinkage shifts an original, unconstrained estimate $\hat{\theta}$ of a parameter of interest θ , such as an estimate of adherence dose response, toward a shrinkage target t to yield a shrunken estimate $\dot{\theta}$.

$$\dot{\theta} \stackrel{\text{def}}{=} \lambda t + (1 - \lambda)\hat{\theta}, \quad 0 \leq \lambda \leq 1. \quad \text{Eq. 1}$$

Degree of shrinkage λ and target t are selected to minimize total error (squared bias + variance), where variance quantifies the uncertainty of the estimate. A useful target t for this purpose has 1) lower variance than the original estimate(s) and 2) close proximity to θ . For the current analysis, we shrank consistent subpopulation estimates toward their grand average across all three subpopulations. Shrinkage toward the grand average reduces uncertainty by drawing upon information from the entire sample. We estimated that optimal degree of shrinkage λ^* toward the consistent grand average that minimized total error. See Online Supplement for technical details.

3. RESULTS

Section 3.1 summarizes statistical findings. Section 3.2 interprets the statistical findings in clinical terms.

Section 3.1. Statistical Findings

Inconsistent, consistent, and shrunken estimators' estimates of adherence dose response are compared by outcome and baseline severity in Fig. 1. Estimates from the inconsistent estimator are generally markedly different from estimates based on the consistent estimator. That strongly suggests that unobserved confounders are present and that this correction may be important for obtaining a more accurate quantification of adherence dose response in these subpopulations. The shrunken estimator generally produced estimates that are less uncertain (narrower confidence intervals) than the consistent estimator. This is particularly evident for all outcomes within AHI 10 – 15 and AHI > 15 – 30, which are those subpopulations where sample sizes were smallest. These improvements in reduced

uncertainty come at the cost of shrunken estimates being somewhat biased toward the shrinkage target (consistent grand average represented by gray horizontal line, Fig. 1). This added bias is modest compared to the bias reduction from adjustment for unobserved confounders (inconsistent vs. consistent estimators, Fig. 1).

Section 3.2. Clinical Findings

Several clinical findings emerge (Fig. 1). The grand average across all three subpopulations of the consistent estimate (horizontal gray line, Fig. 1) of adherence dose response is in the direction of improvement with increased CPAP adherence for each outcome. The null hypothesis of adherence dose response = 0 is rejected for any consistent estimator's confidence interval that excludes zero. On this basis, adherence dose response is statistically significant for ESS-TS for mild, moderate and severe AHI, MWT-MSL for severe AHI, PFNT-TT for moderate AHI, and BSRT-SR for moderate AHI. Unlike ESS-TS, MWT-MSL, and PFNT-TT, the consistent-estimator finding for BSRT-SR for moderate AHI, with confidence interval of $(-3.068, -0.018)$, is in direction of worsening with increased adherence. For MWT-MSL, bias correction of adherence dose response shifted identification of therapeutic benefit from mild/moderate disease (inconsistent estimator) to severe disease (consistent and shrunken estimators).

4. DISCUSSION AND CONCLUSIONS

In this section, we discuss the clinical implications of our results. We then further explore the advantages of statistical shrinkage when estimating adherence dose response in clinical subpopulations. The generality of the proposed method within sleep research is also discussed.

4.1. Clinical Implications

In our original report [4], the comparison of interest, as required by that randomized clinical trial's design, was between CPAP and sham devices at each of nine levels of adherence. Those comparisons were made with adjustment for over one hundred possible baseline confounders. In contrast, in further analyses of the sample from APPLES in [5] and here, emphasis instead has been across levels of adherence within a randomly assigned treatment condition with comparisons adjusted for observed as well as unobserved confounders. The longitudinal method of Holmes et al. [5] found improvements in PFNT-TT attributable to increased adherence on average over two-month and six-month visits and all baseline severity levels combined. In the present study, using cross-sectional data at the six-month visit, we extend findings from [4] and [5] by localizing an attributable, beneficial dose-response effect on PFNT-TT to the subpopulation of moderate baseline disease. Results for sleepiness outcomes are broadly concordant with the original report of therapeutic benefit from CPAP [4]. Moreover, with bias correction applied here, an attributable therapeutic benefit of CPAP adherence on subjective sleepiness at six months has been identified within the subpopulation of mild baseline disease (Fig. 1). The above findings are largely consistent with published clinical observations as reviewed by our group [8] and others [9]. We also present some evidence here that CPAP adherence worsens learning and memory; however, we consider this finding with some skepticism because 1) of the confidence interval's very

close proximity to zero, 2) it stands as the only finding that disappears with shrinkage estimation, and 3) the grand average of consistent estimates (horizontal gray line, Figure 1) is in the direction of therapeutic benefit for BSRT-SR.

4.2. The Advantage of Shrinkage

Though adjustment for unobserved confounders can result in large corrections in estimates of mean dose response (gray vertical bars, Fig. 1), consistent estimation can greatly inflate uncertainty, as evidenced by the much wider confidence intervals for consistent estimates compared to inconsistent estimates (Fig. 1). This is where shrinkage estimators have an important role by reducing inflated uncertainty. As observed in Section 3.2, shrunk estimators yielded essentially the same clinical findings as consistent estimators but with generally narrower confidence intervals (reduced uncertainty). Reduced uncertainty is not trivial. More stable estimates tend to be more closely reproduced in future studies. That focuses and accelerates progress in research. Reduced uncertainty may also improve statistical power with the caveat that these gains in statistical power may be somewhat constrained by the minor bias that optimal shrinkage introduces by shrinkage toward the grand average consistent estimate. Reducing inflated uncertainty may prove especially important in the setting of subpopulation analyses where the certainty of estimates is already compromised by small sample sizes. Indeed, notice that confidence intervals narrow least from consistent to shrinkage estimator estimates where sample size is largest ($AHI > 30$ of Fig. 1). Gains in statistical power could be greatest when adherence dose response is moderate to strong, the subpopulation's sample size is small, but total sample size is large. A collaborating statistician can assess the conditions under which shrinkage can be most beneficial for a particular study. Improved statistical power, even incremental, is important because it can change decisions about efficacy of a proposed medical therapy.

4.3. Generality of Method

The method demonstrated here in application to APPLES is entirely general in that it can be applied across a wide range of clinical trials in sleep research. Application to randomized trials is especially straightforward because randomization can serve as the instrumental variable in the first-stage of the consistent estimator (Section 2.2.2). That said, any mechanism that is truly external to the patient might serve as an instrument, such as reduced adherence strictly caused by medical device malfunction. Moreover, outcomes need not be measured on continuous scales (eg, nightly hours of adherence) because residual inclusion (Section 2.2.2) was specifically designed to accommodate a diversity of measurement scales [7]. This breadth of scales encompasses, for instance, adherence measured by pill counts. To help broaden application, the particular shrinkage method presented here (Section 2.2.3) is distribution-free [10]. Namely, it does not make any strong statistical assumptions about the distribution of the clinical outcome (eg, outcome need not be normally distributed; Online Supplement).

4.4. Conclusions

This study successfully integrated a set of available statistical technologies to provide a fully general method for sleep investigators and their collaborating statisticians for estimation of adherence dose response within the smaller samples of clinical subpopulations. Application

to APPLES illustrates that this method can yield bias corrections for unobserved confounders that are substantial to the extent of revealing new clinical findings. Use of this method throughout sleep research could sharpen understanding of the true efficacy of medical devices, pharmacotherapies, and behavioral interventions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Authors were partially supported by National Heart, Lung, and Blood Institute [5U01-HL-068060] and Patient-Centered Outcomes Research Institute (PCORI) [CE-12-11-4137] (per Study Protocol Section 7.2.5. Special Analyses), each awarded to C.A.K. The statements presented in this article are solely the responsibility of the authors and do not necessarily represent the views of PCORI, its Board of Governors or Methodology Committee. For methodological study reported here, sponsors had no role in design, analysis and interpretation, writing of report, or decision to submit article for publication. APPLES full acknowledgements are in Online Supplement. This paper was improved substantially as a result of comments from three anonymous referees.

FUNDING

Authors were partially supported by National Heart, Lung and Blood Institute [contract 5U01-HL-068060] and Patient-Centered Outcomes Research Institute [per Section 8.2 of study proposal, grant CE-12-11-4137], each awarded to C.A.K. For methodological study reported here, sponsors had no role in design, analysis and interpretation, writing of report, or decision to submit article for publication.

References

1. Weaver TE, Grunstein RR. Adherence to continuous positive airway pressure therapy; The challenge to effective treatment. *Proc Am Thorac Soc*. 2008; 5:173–178. [PubMed: 18250209]
2. Weaver TE, Maislin G, Dinges DF, et al. Relationship between hours of CPAP use and achieving normal levels of sleepiness and daily functioning. *SLEEP*. 2007; 30:711–719. [PubMed: 17580592]
3. Huber M. Sensitivity checks for the local average treatment effect. *Econ Lett*. 2014; 123:220–223.
4. Kushida CA, Nichols DA, Holmes TH, et al. Effects of continuous positive airway pressure on neurocognitive function in obstructive sleep apnea patients: The apnea positive pressure long-term efficacy study (APPLES). *SLEEP*. 2012; 35:1593–1602. [PubMed: 23204602]
5. Holmes TH, Zulman DM, Kushida CA. Adjustment for variable adherence under hierarchical structure: instrumental variable modeling via compound residual inclusion. *Med Care*. 2016 in press.
6. Kushida CA, Nichols DA, Quan SF, et al. The apnea positive pressure long-term efficacy study (APPLES): Rationale, design, methods, and procedures. *J Clin Sleep Med*. 2006; 2:288–300. [PubMed: 17561541]
7. Terza JV, Basu A, Rathouz PJ. Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. *J Health Econ*. 2008; 27:531–543. [PubMed: 18192044]
8. Zhou J, Camacho M, Tang X, Kushida CA. A review of neurocognitive function and obstructive sleep apnea with or without daytime sleepiness. *Sleep Med*. 2016; 23:99–108. [PubMed: 27288049]
9. Weaver TE, Chasens ER. Continuous positive airway pressure treatment for sleep apnea in older adults. *Sleep Med Rev*. 2007; 11:99–111. [PubMed: 17275370]
10. Daniel, WW. Applied nonparametric statistics. Boston: PWS-KENT; 1990.

- Variable adherence to sleep therapies is commonplace.
- Study integrates three statistical methods to estimate adherence dose-response.
- Dose-response to continuous positive airway pressure examined on four outcomes.
- Adjustment for unobserved confounders can alter dose-response findings.

Figure 1.

Inconsistent (Inconsist.), consistent (Consist.), and shrunk estimator estimates of adherence dose response by outcome and subpopulation. By design, consistent estimators provide estimates of reduced bias and shrunk estimators provide estimates of reduced uncertainty (ie, narrow confidence intervals) of consistent estimates. Horizontal gray line is shrinkage target (grand average consistent estimate). Horizontal black line is zero. Height of each gray vertical bar demarcates adherence dose response estimate (95% confidence-interval error bars). Null hypothesis (adherence dose response = 0) is rejected where 95% confidence interval excludes zero. Vertical units are original scale for each outcome (PFNT-TT = Pathfinder Number Test total time; BSRT-SR = Buschke Selective Reminding Test sum recall; MWT-MSL = Maintenance of Wakefulness Test mean sleep latency; ESS-TS = Epworth Sleepiness Scale total score).