Employer-Mandated Obstructive Sleep Apnea Treatment and Healthcare Cost Savings among Truckers

Stephen V. Burks[1,2,3], Jon E. Anderson[4], Bibhudutta Panda[1], Rebecca Haider[1], Tim Ginader[1], Nicole Sandback[1], Darya Pokutnaya[4], Derek Toso[1], Natalie Hughes[1], Humza S. Haider[4], Resa Brockman[4], Alice Toll[4], Nicholas Solberg[4], Jesse Eklund[1], Michael Cagle[4], Jeffery S. Hickman[5], Erin Mabry[5], Mark Berger[6], Charles A. Czeisler[7,8], Stefanos N. Kales[8-10]

1. Division of Social Science, University of Minnesota Morris (UMN Morris)
2. Institute of Labor Economics (IZA), Bonn, DE
3. Roadway Safety Institute, Region 5 University Transportation Center, Minneapolis, MN
4. Division of Science and Math, University of Minnesota Morris (UMN Morris)
5. Virginia Tech Transportation Institute, Blacksburg, VA
6. Precision Pulmonary Diagnostics, Houston, TX
7. Division of Sleep and Circadian Disorders, Departments of Medicine and Neurology, Brigham and Women's Hospital, Boston, MA
8. Division of Sleep Medicine, Harvard Medical School, Boston, MA
9. Department of Environmental & Occupational Medicine & Epidemiology, Harvard TH Chan School of Public Health, Boston, MA
10. Occupational Medicine, Cambridge Health Alliance, Harvard Medical School, Cambridge, MA

Corresponding author: Stephen V. Burks, Ph.D., University of Minnesota Morris, Division of Social Science, 600 East 4th Street Morris, MN 56267-2134; svburks@morris.umn.edu; 320-589-6191

Abstract

Objective: To evaluate the effect of an employer-mandated obstructive sleep apnea (OSA) diagnosis and treatment program on non-OSA-program trucker medical insurance claim costs.

Methods: Retrospective cohort analysis; cohorts constructed by matching (randomly, with replacement) Screen-positive Controls (drivers with insurance screened as likely to have OSA, but not yet diagnosed) with Diagnosed drivers (n=1,516; cases=1,224, OSA Negatives=292), on two factors affecting exposure to medical claims: experience level at hire and weeks of job tenure at the Diagnosed driver's polysomnogram (PSG) date (the "matching date"). All cases received auto-adjusting positive airway pressure (APAP) treatment, and were grouped by objective treatment adherence data: any "Positive Adherence" (n=932) versus "No Adherence" (n=292). Bootstrap resampling produced a difference-in-differences estimate of aggregate non-OSA-program medical insurance claim cost savings for 100 Diagnosed drivers as compared to 100 Screen-positive Controls before and after the PSG/matching date, over an eighteen-month period. A two-part multivariate statistical model was used to set exposures and demographics/-anthropometrics equal across sub-groups, and to generate a difference-in-differences comparison across periods that identified the effect of OSA treatment on per-member per-month costs of an individual driver, separately from cost differences associated with adherence choice.

Results: Eighteen-month non-OSA-program medical claim costs savings from diagnosing (and treating as required) 100 Screen-positive Controls: $153,042 (95% CI: -$5,352, $330,525). Model-estimated effect of treatment on those adhering to APAP: -$441 per-member per-month (95% CI: -$861, -$21).

Conclusions: Results suggest a carrier-based mandatory OSA program generates substantial savings in non-OSA-program medical insurance claim costs.

Keywords: OSA, OSA - PAP Therapy, commercial motor vehicle operator, healthy worker selection, medical insurance costs, truckload motor carrier, treatment adherence, mandatory OSA program

Clinical Trials: Not Applicable

Statement of Significance

Industry cost concerns contributed to the USDOT's 2017 withdrawal of a regulatory process on mandatory screening of truckers for obstructive sleep apnea (OSA). We analyze the non-OSA-program medical insurance claim cost savings that resulted from the first large-scale mandatory motor carrier program to screen, diagnose, and treat OSA. We estimate that OSA treatment lowers other medical insurance costs for those adhering to treatment by $441 per-member per-month and that diagnosing and treating 100 drivers screened as likely to have OSA saved $153,042 over the course of eighteen months in non-OSA-program medical insurance claims. One of these results is clearly statistically significant and the other is borderline; together they suggest that a mandatory OSA program can generate savings that offset a substantial portion of its costs.

1. Introduction

Obstructive sleep apnea (OSA) is associated with specific neurocognitive deficits in attention/working memory, vigilance, and executive functioning[1-3] and is one of the most common medical causes of excessive daytime sleepiness or fatigue.[4] Untreated OSA is associated with many adverse health conditions, including cardiovascular events in males,[5] systemic hypertension and pulmonary hypertension,[6,7] depression,[8-10] insomnia,[11-13] Type II Diabetes,[14-16] obesity,[16-19] and mortality.[20] Accordingly, untreated OSA has been implicated in higher healthcare utilization and higher healthcare costs.[21-24]

Among the 1.87 million U.S. commercial drivers estimated by the Bureau of Labor Statistics to be operating non-farm-based heavy trucks (gross weight ≥ 26,000 lbs.),[25] 17-28% or 318,00 to 524,000 are expected to have at least mild OSA based on prevalence studies on commercial drivers.[26-30] If the larger population of the 4.0 million drivers estimated by the Federal Motor Carrier Safety Administration to be using commercial driver's licenses in interstate and intrastate transportation[31] is considered, 0.68 to 1.1 million drivers may have OSA. The majority of these drivers are thought to be undiagnosed and untreated.[26,32] There is thus considerable scope for healthcare cost savings through treatment of OSA among commercial drivers.

Indeed, in a 2010 retrospective analysis of health care costs of OSA in commercial motor vehicle (CMV) drivers at a large U.S. fleet,[33] estimated savings from OSA treatment were on the order of $6,000 per driver over 24 months (or $250 per-member per-month (PMPM)). A 2013 study examined the cost savings of an education campaign on the diagnosis and treatment of OSA in the medical plan of a large railroad,[34] and found that positive airway pressure (PAP) therapy was associated with a cost decrease on the order of $150 to $200 PMPM. However, both studies were small and were based on voluntary participation, limiting their generalizability to larger

populations and mandated settings. And although both analyzed the costs of individuals with OSA receiving therapy in comparison to those with OSA not receiving therapy, neither study deployed the differences-in-differences analytic framework that has since become generally accepted.[35]

Among non-CMV drivers, untreated OSA increases the risk of motor vehicle crashes by 1.2- to 4.9-fold, while adherence to treatment with PAP significantly reduces this excess crash risk; similar results have been found for CMV operators.[36] Because of the potential excess crash risk associated with untreated OSA[36-39] the Federal Motor Carrier Safety Administration's (FMCSA) Medical Expert Panel (in 2007) and Medical Review Board (in 2008 and 2011), and the National Transportation Safety Board (in 2009), all recommended that safety regulations should require comprehensive screening and diagnosis of commercial drivers for OSA during their required biennial commercial vehicle operator's medical examination.[40,41] That recommendation was renewed by the Medical Review Board and FMCSA's Motor Carrier Safety Advisory Committee in November, 2016.[42]

However, the financial costs of OSA screening, diagnosis, and treatment have been a significant concern among elements of the trucking industry, and have generated resistance to the potential for mandated OSA screening standards. Indeed, industry lobbying is credited with the passage by Congress of a law signed in 2013 that prevented the FMCSA from issuing guidance on OSA screening to CMV medical examiners in the absence of a full rulemaking on the topic.[43] More recently, the US Department of Transportation (USDOT) issued an Advance Notice of Proposed Rulemaking on March 8, 2016 on the topic of sleep apnea screening,[44] but withdrew it on July 27, 2017 after negative comments from industry, on the grounds of "insufficient information."[45,46]

The present study analyzes medical insurance claim costs incurred by CMV drivers in the context of an employer-mandated OSA program, the first and largest internal OSA program to date operated by a motor carrier, and the largest employer-based program to date in the US, as far as the authors are aware. The program included screening, diagnosis, auto-adjusting positive airway pressure (APAP) treatment for those with OSA, and APAP treatment adherence monitoring.[26,47] Two hypotheses are addressed. Hypothesis 1: diagnosing and treating the subset of drivers screened as likely to have OSA lowered their aggregate medical insurance claim costs, as compared to not doing so. Hypothesis 2: OSA treatment lowered medical insurance claim costs of those individuals with OSA who adhered to treatment. Hypothesis 2 is addressed from two perspectives, on the basis of "those who adhered to treatment," and also on the basis of "those to whom treatment was offered" (or an "intention to treat") basis.

2. Methods

a. Carrier OSA Program Protocol

The OSA screening, diagnosis, and treatment program was implemented by Schneider National, Inc., a major North American trucking firm.[48,49] A pilot in 2005 was followed by full implementation beginning in April, 2006.[26,47] The Somni-Sage® screening questionnaire was used, assigning drivers to one of four classes ranging from "High Priority" to "Low Priority," for receiving polysomnogram (PSG) diagnostic testing.[26] Due to the startup process in the presence of turnover, about one-half (n=17,098) of the drivers employed from 2006 to the study end date (December 31, 2009) were screened. The carrier selected who to refer from those screened as High Priority for an overnight, multi-channel, laboratory, technician-attended PSG ("Type 1" PSG as defined by the American Academy of Sleep Medicine)[50] at a national network of sleep

laboratories. Referral was based on factors such as the driver's schedule, route, continuing employment, and sleep lab availability. PSG records show 5 conducted in 2005, 493 in 2006, 370 in 2007, 632 in 2008, and 662 in 2009. Diagnosis and treatment were covered without co-pays as preventive medicine for drivers enrolled in the firm's voluntary medical insurance plan.

PSGs were interpreted immediately using standard criteria with diagnostic clinical evaluations the morning after the overnight tests.[50] Drivers who were diagnosed as "positive" for OSA by board-certified sleep physicians (generally with an apnea-hypopnea index (AHI) ≥5) were given first-line treatment: an APAP machine, heated humidifier, and mask interface, usable both in the truck sleeper berth while on the road and at home. For the first 14-90 days, and longer if necessary, drivers' adherence to APAP therapy was monitored using wireless data transmission. After APAP adherence was initially demonstrated, periodic batch downloads from the APAP machine's internal adherence memory maintained monitoring. Adherence trouble-shooting, education, and monitoring used follow-up phone and face-to-face contacts. Drivers with OSA who remained non-adherent as demonstrated by objective APAP monitoring, despite this multifaceted process of remediation, were eventually terminated after the remediation process failed.[47]

b. **Retrospective Matching of Diagnosed Drivers to Screen-positive Controls**

To address Hypothesis 1, Diagnosed drivers (both those with OSA and those without OSA) must be compared to otherwise similar control drivers who were not diagnosed. This required considering how drivers entered the OSA program. The study firm is engaged in long distance for-hire trucking in which driver turnover rates at large firms are very high (averaging 94% per year between 1996 and 2018).[48,51] Though the study firm had many long-term drivers and turnover rates lower than industry averages, many drivers joined and departed the firm in a

continuous process during the study period.[48,52] The screening and diagnosis process, plus a significant group of high-priority drivers already employed at the study start date (January 1, 2005), created a time lag between initial employment and the PSG for those drivers sent to be diagnosed. This created the potential for "healthy worker survivor" selection effects—drivers who became sick enough to be unable to work did not enter the study, and the extent of this selection process was correlated with the time drivers had been able to file a medical insurance claim.[53,54] In addition, prior work in this setting has shown the existence of a substantial safety-selection effect—untreated OSA is associated with the risk of having a serious, preventable crash (a 4-to 5-fold increase compared to drivers who don't have OSA),[36] which in turn is associated with the risk of discharge (specifically, a 30-fold increase in the hazard of discharge after a serious preventable crash).[55] Thus, the study implemented a retrospective cohort approach through the process of matching each Diagnosed driver (a potential case) with a driver designated as a "Screen-positive Control," who had a similar length of exposure for a medical insurance claim, and also equal exposure to selection into the participant pool of currently employed drivers.

The potential for having a medical insurance claim required enrollment in the study firm's employee insurance program. Since the data only included medical insurance enrollment during the study period (January, 2005-December, 2009), and because a significant fraction of the study firm's workforce had longer tenure than the entire study period, enrollment was not a statistically appropriate exposure measure for the matching process. (That is, two drivers with insurance observed to start in January of 2005 could have been on the job for very different lengths of time prior to 2005, and therefore could have had different lengths of exposure to selection.) However, job tenure could be identified for all drivers. In addition, while some employed drivers were not enrolled in the firm's medical insurance program, employment for a specific minimum period

(generally, 3 months for inexperienced-at-hire and 1 month for experienced-at-hire drivers) was a necessary condition for enrollment. As a result, job tenure and experience-at-hire were appropriate proxies for the length of exposure to the potential for a medical claim for the purpose of matching Screen-positive Controls to Diagnosed drivers, in order to ensure both had been subjected to equal amounts of safety and healthy worker selection. (Length of insurance enrollment was still the appropriate measure of exposure to the chance of costs during the 2005-2009 study period, when medical claims can be observed, and is used thus in the statistical model described below.)

The positive predictive value of a "High Priority" designation from Somni-Sage® is 80% for mild OSA (AHI ≥5).[26] The study firm's OSA program was designed to diagnose all drivers screened at High Priority, but due to driver turnover, the study end date, and the fact that the data include the startup period of the OSA program, when diagnostic examinations were not necessarily available in large numbers quickly, there were sufficient potential Screen-positive Controls (n=1,573).

Each driver diagnosed with OSA who had medical insurance enrollment data (potential cases, n=2,186) was matched with a Screen-positive Control driver who also had medical insurance enrollment (a factor which further restricted potential matches). Screen-positive Controls were drawn randomly under two conditions: that the control had the same experience-at-hire as the Diagnosed driver, and that the control's job tenure in weeks was the same (±1 week) as the Diagnosed driver's on the week of the PSG (the "matching date" for the control). Drivers used as controls were replaced in the pool of potential Screen-positive Controls; some were thus used more than once (775 unique drivers), but in this event were matched to a new Diagnosed driver on that driver's PSG date. The following sub-groups were thus created.

i. Screen-positive Controls: drivers screened as likely to have OSA but who had not yet had a PSG (matched n=1,516).

ii. Diagnosed: drivers who received a PSG (matched n=1,516; this group was compared as a whole to Screen-positive Controls to address Hypothesis 1).

   a. OSA Negative: drivers whose PSG showed AHI <5 (matched n=292).

   b. Cases: OSA-diagnosed drivers who were clinically judged to have the disease (PSG showed AHI ≥ 5 and treatment recommended by a board certified sleep physician interpreting the sleep study, matched n= 1,224), and who were then provided with APAP and instructions on usage. Treatment adherence was a condition of continued employment.

iii. For use in addressing Hypothesis 2, cases were further subdivided based upon treatment adherence, which was not assigned randomly, but chosen by drivers.

   a. Positive Adherence: cases who recorded some level of APAP treatment adherence (matched n= 932) in two additional sub-groups (distinguished here but combined in the presentation of results as their costs were never statistically different):

      i. Full Adherence: cases who always met or exceeded the minimum consensus standard of 4 hours/night mean APAP usage[56] (matched n=510).

      ii. Partial Adherence: cases who recorded treatment, but did not meet the minimum requirements for full adherence (matched n=422).

   b. No Adherence: cases who never recorded any adherence with APAP treatment (matched n=292).

c. Medical Cost Identification and Data Synthesis

Data from the study firm's human resource system provided driver demographics, such as age, gender, racial or ethnic category, hiring date, and separation date and type (if applicable). These were merged with operational data that provided information such as home terminal, weekly miles, and job type; then records from the sleep medicine services provider were added, including the results of the Somni-Sage® screening questionnaire and, when applicable, PSG results and APAP adherence data. Finally, records for all periods of medical insurance enrollment and of all claims made through the medical insurance manager, United Health Care (UHC), by the study carrier's employee drivers during the study period were assembled and added. It should be noted that these data reflect the general medical expenses incurred by relevant employee drivers through the study carrier's voluntary medical insurance program, but do not include the direct costs of diagnosing and treating OSA (and also do not include pharmaceutical claims). Only some of the OSA expenses flowed through UHC as a medical insurance manager, and the proportion of costs not included could not be determined. Thus, since the total costs of the program could not be accumulated, those costs that were identified were removed (see Online Supplement, Section II). All costs have been adjusted to 2018 dollars using the Personal Consumption Expenditures index for medical services,[57] which includes medical expenses paid on behalf of the consumer, such as payments from a medical insurance program.

3. Statistical Analysis

The primary response variable, calculated separately for each study participant, was the medical insurance claim costs (excepting the costs of the OSA program itself) paid by the study firm per member (PM), separately for the periods before and after the PSG (each Diagnosed

driver) or matching date (each Screen-positive Control). Week-by-week data on participants' operational characteristics and claim-by-claim information on participants' medical insurance claims were cumulated appropriately to provide relevant summary information about driver characteristics, claims costs, and the diagnoses involved in claims, in one observation per participant in each period. To address Hypothesis 1, aggregate results over the 18 months before and after were considered. To address Hypothesis 2, model estimated cost savings over 18 months of enrollment were divided by 18 to generate per-member per-month (PMPM) cost savings for an average treatment-adherent driver in each period.

a. **Evaluation of Hypothesis 1: Bootstrapped Difference-in-Differences**

Because the study retrospectively analyzes observational data, drivers were not randomly assigned into the subgroups whose results were to be compared to test the two hypotheses. As a result, there may be differences in costs between the relevant study sub-groups that arose from reasons other than the hypotheses that are addressed by the study's statistical tests.

More specifically, testing Hypothesis 1 requires an estimate of the aggregate medical claim cost differences between Screen-positive Controls and all Diagnosed drivers. Although the selection of Screen-positive Controls to be diagnosed was primarily based on operational characteristics, it cannot be ruled out that the way drivers were selected for diagnosis led to medical cost differences between undiagnosed Screen-positive Controls and Diagnosed drivers that were due to reasons that were not related to Hypothesis 1.[47]

The appropriate correction for such a potential bias is a difference-in-differences approach. Specifically, the costs of Diagnosed drivers were subtracted from those of Screen-positive Controls in the period before the PSG/matching date. This provides a measure of initial differences between the two sub-groups, before the drivers selected for diagnosis received their

PSG. Then the medical costs of Diagnosed drivers were subtracted from those of Screen-positive Controls in the period after the PSG/matching date, providing a measure of the effect of the OSA program combined with that of any initial differences. Finally, the difference in the before-period was subtracted from that in the after-period. Conditional on the assumption that the initial cost differences were persistent through both periods (also called the "parallel trends" assumption),[35] this procedure provides an estimate of the effect of the OSA program on medical costs for Diagnosed drivers as compared to undiagnosed Screen-positive Controls, isolated from any initial differences between the two sub-groups.[35]

The estimated cost savings and associated 95% empirical confidence intervals are formed using the bootstrap method.[58] This method permits the calculation of realistic estimates and confidence intervals without making any assumptions about the nature of the underlying data distribution. The bootstrap method resamples the existing matched case-control pairs repeatedly to produce an empirical probability distribution for the statistics of interest. The statistics that were calculated were the aggregate costs of Screen-positive Control and Diagnosed drivers in the periods before and after the PSG/matching date, along with the appropriate cost differences between the sub-groups, and the final difference-in-differences estimate. The process was repeated 10,000 times to produce the empirical distributions of the statistics. This method produces estimate distributions that closely reflect the variation in the underlying population for samples that adequately represent that population, and the study sample size is large enough to provide confidence that population characteristics are well-represented in the sample, and thus are also well-represented in the bootstrap distribution.

One iteration of the bootstrap was conducted by obtaining a simple random sample of 1,516 matched pairs drawn (with replacement) from among the original 1,516 pairs constructed for the

study. (Since sampling was with replacement, almost all the bootstrap samples contain many duplicated pairs, and omit some of the original pairs, and hence have different costs than those of the original data set.) For every selected pair, the actual PM cost in the 18 month period before, and the 18 month period after, the PSG/matching date was calculated for both drivers. While not all drivers were observed for a full period of 18 months, observed costs were accrued for drivers who entered or exited within the 18 month observation windows, reflecting the fact that driver turnover affected the medical insurance costs incurred by the study firm.

However, some drivers were observed for fewer than 18 months before or after not because of entry or exit, but because the driver's observation was truncated by the beginning or end of the study period. In this case the actual costs the driver incurred during the period of observation were scaled up to reflect the expected period of observation for his/her study subgroup. The expected observation period was 18 months or longer in the before-period for all sub-groups, and 18 months for all drivers in the after-period except No Adherence drivers, for whom the expected observation period was 6.23 months (27 weeks). The expected insurance enrollment duration in the after-period was based on the Kaplan-Meier survival function for each sub-group, which correctly accounts for exits and censoring (see Figure 1; details in Section XIV of the online supplement).

The total costs were then divided by 1,516 and multiplied by 100 to produce cost and cost difference estimates based on 100 Screen-positive Control and 100 Diagnosed drivers. The mean of each distribution for a cost, or for a cost difference, is reported as the point estimate, and the 2.5 and 97.5 percentiles of each distribution are identified to form an empirical 95% confidence interval.

b. **Evaluation of Hypothesis 2: Model-based Difference-in-Differences Treatment Effect**

Evaluating Hypothesis 2, which addresses the existence of an individual treatment effect, also faces a challenge from non-random subgroup assignments. Drivers diagnosed with OSA were divided into those who exhibited no treatment adherence and those showing positive treatment adherence. However, these subgroup memberships were self-selected, and so it is likely there were also differences in medical costs between the two groups that were independent of treatment. For instance, it might be conjectured that drivers who refuse mandated treatment were generally higher in risk taking and lower in "self-care" than those who attempted treatment.

This issue can be addressed with the same difference-in-differences approach used for Hypothesis 1, on the same assumption that the differences between subgroups are maintained over time (the "parallel trends" assumption).[35] This method can be applied using observed PMPM costs computed for the two subgroups in each period (see Supplemental Online Material, Section IV). However, while the difference-in-differences approach adjusts for persistent differences between the No and Positive Adherence subgroups, an analysis done with observed PMPM costs fails to address other important potential biases: those due to unequal average values of important demographic characteristics, and to unequal periods of observation, across the two subgroups. Testing Hypothesis 2 is best addressed by creating an analysis based upon the average driver, and on equal periods of observation across all drivers. This calls for the use of a multivariate regression model, and the two-part model used is described in detail in the following section. (It may be noted that the concerns just noted are not directly relevant to Hypothesis 1, which is about aggregate savings for the Diagnosed subgroup, as compared to the Screen-positive Controls; there it is sufficient that the study subgroups are typical of those who would screen positive in the driver population as a whole.)

It is important to note how the model, once selected, is deployed to provide a difference-in-differences estimate of the treatment effect. An underlying idea is that once the model has been estimated on the entire data set, the coefficient estimates capture, conditional on the specification of the model, the systematic relationship between the covariates (i.e. the independent variables denoting such features as age or body mass index (BMI), as well as time period and study subgroup membership) and the dependent variable, medical costs. (This creates the averaging of covariate effects across subgroups that would have been the result of randomization in a prospective trial.) Once these estimates are in hand, they can be used to compute predicted values of costs using any set of covariate values selected, not just the ones that individual subjects actually possess. A typical use is to compute the effect on predicted costs of a change of one unit in the value of any one covariate (or related set of covariates when one covariate interacts with another): just compute the prediction twice, once each for the initial and the changed values of the selected covariate, holding all else equal. It also permits the computation of counterfactual cost predictions, that is, predictions based on covariate values no subjects actually possess.

In the predictions presented here, the exposure variable was first set to 18 months, thus making the period of exposure equal for all participants. Second, the values of all the covariates except exposure, study subgroup, and time period, were set to the mean values observed in the Positive Adherence subgroup, to create an estimated cost for the typical driver that accepted treatment. Finally, time period was set to "after" and subgroup identity was set first to Positive Adherence (giving predicted costs in the after-period for a Positive Adherence driver who accepted treatment), and then subgroup identity is set to No Adherence (giving the *counterfactual* predicted cost for a Positive Adherence driver *who did not accept treatment*). The difference between the two predictions is the predicted effect of treatment, in the after-period, for a driver

with characteristics typical of the Positive Adherence subgroup. When this is repeated for the before-period, the final estimate of the treatment effect (on those who accepted treatment) is the difference between the two within-period differences. As a variation, to create an "intention to treat" estimate of the treatment effect, this entire process is repeated, but using mean covariate values (other than exposure, subgroup, and time period) derived from all drivers offered treatment (i.e. both Positive and No Adherence drivers, instead of just the Positive Adherence drivers). (For further details, see Online Supplement, Section V.2).

c. **Hypothesis 2: Structure of the Two-part Multivariate Statistical Model**

The model selected for the estimation of Hypothesis 2 has a complex structure, which reflects several specific features of the study data. First, the response variable is a pair of per-member costs, one cost value for the before PSG/matching date and the other for the after-period. Because these represent two measurements for the same individual over time, they are likely to be correlated. Second, as discussed above, there is variation across study subgroups in demographic and work characteristics that may be associated with differences in health care expenditures, such as age, BMI, type of job, and length of time observed. And third, health care cost data has some particular distributional features. Specifically, there is a substantial portion of zero cost observations (which vary across subgroups), and the distribution of positive costs is right-skewed with many small costs, and a tail of less frequent but much larger costs.

The approach taken here to address these issues is a standard one: a multivariate model with two parts that are jointly estimated.[59-61] The first part of the model used a logistic specification for the binary response of zero versus positive costs, and is appropriate for modeling the probability of positive costs. The second part used a gamma generalized linear specification for the total value of costs over the observation period, when this total was positive, which is appropriate for

the skewed distribution of such costs. Both parts of the model were estimated simultaneously, using the same covariate list. While the matching of Diagnosed drivers to Screen-positive Controls was based on the length of tenure and experience at hire, the months of observed insurance enrollment were used as an "offset" variable in both parts of the two-part model; this feature adjusted model parameter estimates for the observed length of exposure to the risk of medical insurance claims that each driver actually incurred during the study period.[59]

The initial set of covariates was specified based upon extensive prior work with similar data, and a resulting background understanding of important aspects of the data generating process.[36] An examination of residuals, likelihood ratio tests, and the Akaike Information Criterion (AIC) were used to examine fit and determine the model. "Testing down" by dropping covariates that were not statistically significant was not employed for two reasons: the fact that there were prior empirical grounds for the set of initial covariates, and to avoid the enhanced risk that such a procedure carries of overfitting or "modeling sample noise," i.e. of finding spurious effects that are due to random sampling variation.[62]

d. **Covariate (Independent Variable) Specifications for Two-part Model**

i. Time Period, Demographic, Anthropometric, and Work Covariates

- Time Period: in the form of indicator variables for "Before PSG/matching date," "After PSG/matching date". "Before PSG/matching date" is the base (omitted) category

Personal characteristics that are used in all models and may be associated with variation in medical costs include:

- Sex: One indicator variable for "Female"; the base (omitted) category is "Male."
- Age at PSG/matching: Age at PSG (Diagnosed drivers) or at comparison date (Screen-positive Controls) specified as indicators for the ranges of "40 ≤ Age < 50", and "Age ≥

50"; the lowest range "21 ≤ Age < 40" is the base (omitted) category. The minimum age for a commercial driver's license in interstate transportation is 21.

- Race: Race/ethnicity is in the form of indicator variables for "African-American", and "Other." "Other" subjects are neither African-American nor White. The base (omitted) category is "White".
- Marital Status: Marital status over the study period is a binary indicator variable for "Married" if the status of the driver was "married" during the majority of the observation period. Drivers not classified as married were classified as single, the base (omitted) category.
- Geographic Location: The geographic location of the driver's home is in the form of indicator variables for "Northeast", "South", and "West." The base (omitted) category is "Midwest".
- Body Mass Index: Body Mass Index (recorded at the PSG date for Diagnosed drivers, and as a screening result for Screen-positive Controls) specified as a set of indicator variables: "25 ≤ BMI < 35", and "BMI ≥ 35"; the base (omitted) category was "BMI < 25".
- Type of Work: in the form of indicator variables for "System," "Local," "Regional," "Team," and "Other." "Dedicated" is the base (omitted) category.

ii. Exposure, Sub-group Identifiers, and Interactions

- Exposure to medical costs was measured in months of enrollment in the health insurance program before or after the PSG date, whichever was relevant.
- Study Sub-groups: As defined above, in the form of indicator variables for OSA Negatives, Positive Adherence, and Screen-positive Controls. No Adherence drivers are the base (omitted) category.

- Time interacted with Type of Work (five additional indicator variables).
- Time interacted with Race (two additional indicator variables).
- Time interacted with Geographic Region (three additional indicator variables).
- Age interacted with Study Sub-groups and Time Period: This formulation produces six indicator variables for the combinations of Age category (above) and Study Sub-group, two indicator variables for the combinations of Time and Age category, three indicator variables for the combinations of Time and Study Sub-group, and six indicator variables for combinations of Time, Study Sub-group, and Age category.

All indicator variables are binary, coded as "1" if the driver had the characteristic, and "0" otherwise. Three discrete categories were used for Age and BMI to allow for non-linear effects without the complexity of a polynomial specification.

Note that the size and statistical significance of each specific coefficient estimate from each part of the model enters into the calculation of the size and statistical significance of every cost and cost difference estimate. However, the coefficient estimates are intermediate, and not final, results. That is, neither the size nor the statistical significance of any of the individual coefficient estimates are directly informative about either the size or the statistical significance of any specific cost or cost difference estimate. Thus the estimated coefficients from both portions of the two-part model are presented only in the Online Supplement, and the focus herein is on the model-adjusted predictions.

As previously noted, the coefficient estimates capture the systematic part of the relationship between each covariate and costs, simultaneously adjusting for the other covariates. Conditional on the structure of the model, the model-adjusted predictions thus account for the following aspects of the underlying data: a) variations across study sub-groups with respect to the propensity

to have positive (versus zero) costs, b) variations across study sub-groups in the level of costs incurred when costs were positive,[63] c) variation in costs across time periods, d) variations across study sub-groups in the risk of exposure to a medical claim (the number of medical insurance enrollment months observed in each period), and e) variations across study sub-groups in the distributions of potentially confounding demographic, anthropometric, and work type covariates.

e. **Supporting Analyses**

Several supporting analyses were undertaken. Some relevant results are presented below, and all details are reported in the Online Supplement.

First, additional details about the removal of the partial OSA program costs from the medical insurance claim costs are presented (Online Supplement, Section II). Second, additional details are presented about the characteristics of the medical insurance cost distributions (differences across sub-groups in zero costs, and an outlier description; Online Supplement, Section III). Third, PMPM cost estimates from the observed data are presented for main study sub-groups in both time periods (Online Supplement, Section IV). Fourth, details about model development and estimating predicted PMPM costs for Hypothesis 2 are presented (Online Supplement, Section V). Fifth, the individual coefficient estimates for both parts of the two-part model are presented (Online Supplement, Section VI). Sixth, to complement the cost differences highlighted in the evaluation of Hypothesis 2, the underlying PMPM cost estimates derived from the two-part model are presented for main study sub-groups in both periods (Online Supplement, Section VII). Seventh, PMPM cost estimates derived from the model are presented for smaller subsets of the main study sub-groups, broken out by age for both periods (Online Supplement, Section VIII). Eighth, two robustness checks on the use of the multivariate two-part model are presented: the main results are re-done with AHI ≥15 as the threshold for a positive OSA diagnosis (Online

Supplement Section IX), and without a large outlier cost value (Online Supplement, Section X). Ninth, evidence is offered on the reason two adherence sub-groups have been collapsed into one (Online Supplement, Section XI). Tenth, evidence is provided on how costs break out by major diagnostic category (MDC) and by whether a claim was associated with some selected comorbidities of OSA or not (Online Supplement Section XII). Eleventh, details on the pattern of exits from employment across study sub-groups are presented (Online Supplement, Section XIII). Twelfth, details on the process of estimating the aggregate cost savings are presented (Online Supplement Section XIV). Finally, a robustness check is presented for the estimate of the aggregate cost savings: removing a large outlier (Online Supplement Section XV).

The data synthesis and analysis were performed by the University of Minnesota, Morris Truckers & Turnover Project (S.V. Burks, Principal Investigator, J.E. Anderson and B. Panda, Co-Investigators). Retrospective analysis of individually identified protected health information was approved by the Institutional Review Board of the University of Minnesota. All analyses were conducted with Stata Version 14.2 software.

4. Results

a. Participant Characteristics

Two statistical profiles of the study participants are presented in Tables 1 and 2. Table 1 breaks out participants into Screen-positive Controls versus all Diagnosed drivers. As is to be expected from the method of matching of controls to Diagnosed drivers, mean job tenure before the PSG/matching date is the same, as is the mean number of observed months of enrollment. Screen-positive control drivers have, overall, slightly higher BMI than Diagnosed drivers (considering the BMI category variable; Fisher's exact test, p=.04). Table 2 breaks out

participants into Screen-positive Controls versus the three Diagnosed driver sub-groups: OSA Negative, any Positive Adherence, and No Adherence. In particular, demographic factors (age, marital status, geographic location, and BMI), and exposure characteristics (months of medical insurance enrollment and of job tenure) vary across study sub-groups, which made accounting for such differences with a multivariate model appropriate for assessing Hypothesis 2.

The participant characteristics reflect that participants were all screened, from among employee drivers of the study firm, as at High Priority for an OSA diagnosis (by the Somni-Sage® self-report questionnaire). As a result, they were older and more obese than the overall driver population. For comparison, among a stratified random sample of drivers interviewed by the National Institute of Occupational Safety and Health at truck stops across the US (a population likely to be similar to the study firm's drivers), 69% were found to be obese (BMI ≥ 30),[64] while among study participants 83.6% of Screen-positive Controls, 61.6% of negatives, 82.8% of Positive Adherence drivers, and 82.5% of No Adherence drivers fell in this category. Since the full driver population at the study carrier can be expected to reflect the characteristics of the long-haul driver population in the US,[64,65] the selection of drivers from that population for OSA diagnosis through any method analogous to that of the Somni-Sage® screening tool is likely to identify a group with similar age and obesity characteristics.[26]

b. Hypothesis 1: Aggregate Medical Cost Savings from Diagnosis and Treatment of Screen-positive Drivers

Table 3 shows cost estimates derived from 10,000 bootstrap samples rescaled to show non-OSA-program medical insurance claim costs per 100 drivers over a period of 18 months before and 18 months after the PSG/matching date. Subtracting the difference between Screen-positive Controls and Diagnosed drivers before the PSG/matching date from the same difference

calculated in the after-period produces the values in the lower right cell, a difference-in-differences estimate of the aggregate cost savings in non-OSA-program medical insurance claims: $153,042. A 95% bootstrap confidence interval formed from the 2.5 and 97.5 percentiles of the cost difference distribution is (-$5,352, $330,525). This range for the aggregate cost savings contains a negative lower bound, showing that it is not statistically significant at the conventional standard of p=.05. However, the range includes primarily positive values, and the estimate is statistically significant at the p=.06 level, providing suggestive evidence for aggregate cost savings in non-OSA-program medical insurance costs from the OSA program.

The costs of the OSA program itself are not estimated due to data limitations. But for 100 drivers to be diagnosed, and treated as required, they would consist of: (1) the costs of screening enough drivers to find 100 at high priority for an OSA diagnosis (approximately 333 drivers),[26] (2) the cost of diagnosing 100 drivers, and (3) the costs of providing 81 drivers found to have OSA with APAP, along with any needed follow-up and maintenance.

c. **Hypothesis 2: Effect of OSA Treatment on the Medical Costs of a Driver with OSA**

Table 4 presents the results of the estimation process described in Methods using the two-part multivariate model. The top rows display the findings when the process is applied using mean covariate values computed only from the drivers who showed positive evidence of APAP treatment adherence. The predicted differences in PMPM non-OSA-program medical insurance claim costs associated with accepting treatment, as opposed to refusing it, are shown for both periods, and the treatment effect (the difference in the predicted differences) is in the rightmost column: -$441 PMPM 95% CI: (-$861, -$21). The bottom rows provide the same display but using mean covariate values calculated from all drivers offered treatment (i.e. including both the Positive and the No Adherence subgroups, or an “intention to treat” analysis). In both cases, the

estimated PMPM costs savings associated with accepting APAP treatment are substantial, and in both cases the results are statistically significant by conventional standards.

d. **Supplementary Results: Robustness Check on the Aggregate Cost Savings**

As described in Methods, a robustness check was performed by repeating the analysis without the matched pair (Diagnosed driver and Screen-positive Control) containing an outlier, a Screen-positive Control participant with an observed PM cost over $250,000. The main results include this outlier, as high variance and outliers are characteristics of medical insurance data. The impact of removing the high-cost Control is noticeable; estimated mean savings become $107,106 (a $45,936 reduction; 95% CI is -$29,697, $252,323; see Online Supplement, Section XV).

e. **Supplementary Results: Two Robustness Checks on the Two-part Model**

As described in Methods, two robustness checks were performed. First, the analysis of Table 4 is repeated, but under the new assumption that the threshold for a positive OSA diagnosis is increased from AHI ≥5 to AHI ≥15. The reclassification causes changes in sub-group sizes, and in particular, the Positive Adherence subgroup is 27% smaller, and the No Adherence subgroup is 37% smaller, since the number of OSA positives is smaller. Under this modification, the difference-in-differences estimate of the OSA treatment effect on medical insurance costs is lower, and is now not statistically significant: -$318 PMPM (95% CI: -$934, $298, for drivers who accepted treatment), as compared to the value of -$441 in Table 4. However, the qualitative pattern of the results is similar to that in Table 4 (see Online Supplement Section IX).

Second, the as noted in the description of the Hypothesis 1 robustness check (above), the data contained one notable cost outlier in the control group in the after-period. The Table 4 analysis is re-run omitting the matched pair containing this driver (Online Supplement Section X). The pattern of results is very similar to that in Table 4 – the estimated treatment effect drops by only

$9, and the estimate remains statistically significant by conventional standards.

f. **Supplementary Model Results: Age Breakout**

Model specification selection methods revealed a significant interaction of age with study sub-groups across time period. Breakout tables of model-predicted PMPM costs by age for the both periods show that while the pattern of cost estimates across sub-groups is similar overall, the oldest age category has the largest cost point estimate for every sub-group in both periods except for drivers with Positive Adherence in the after-period (Online Supplement Section VIII).

g. **Supplementary Results: Breakouts by MDC and OSA Comorbidity Status**

Findings by MDC are limited by the fact that positive costs for non-OSA-program medical claims are sparse when broken out into smaller categories. Online Supplement Section XII shows (model-unadjusted) per-member per-month costs within each of the 18 MDCs available in the clinical classification software system, by period and study sub-group; most costs are small.[66]

In addition, versions of Table 4 (estimates from the two-part model by sub-group and period using mean values of individual characteristics computed from relevant subgroups and 18 months exposure) are displayed for all costs associated with a selected list of commonly identified OSA comorbidities, and also for its complement, all costs which are not in this list (Online Supplement Section XII). Because each estimate utilizes (only) a subset of all non-OSA-program medical costs, statistical significance is less likely. The difference-in-differences point estimate of the OSA treatment effect for selected OSA comorbidities is higher than the estimate for the group of conditions not in the list of selected comorbidities; while it is not statistically significant by conventional standards, it is close (-$313, 95% CI: -$687, $60, when calculated for drivers that accepted treatment). The estimate of the treatment effect for diseases not in the selected comorbidities list is statistically significant by conventional standards when calculated for the

drivers that accepted treatment: -$184 (95% CI: -$366, -$3).

5. Discussion

a. Two Primary Findings are Substantive and together are Significant

This study examines the first large-scale employer-mandated OSA program undertaken in the US trucking industry, and the largest employer-based program in the US as far as the authors are aware. It analyzes a unique data set, the medical insurance and operational records of 3,032 truck driver participants, and provides tests of two hypotheses. Hypothesis 1 is that the non-OSA-program medical costs for Screen-positive Controls who went through the OSA program were smaller than those of Screen-positive Controls who did not, given the characteristics of the actual driver participants. The savings for 100 drivers over an 18-month period is estimated to be $153,042 (95% CI: -$5,352, $330,525). Hypothesis 2 is that OSA treatment lowered the non-OSA-program PMPM medical costs of drivers with OSA who adhered to APAP treatment, as compared to drivers with OSA who did not accept treatment. The estimated PMPM saving for those who accepted treatment is -$441 (95% CI: -$861, -$21). The analysis result for Hypothesis 1 is near statistical significance at the conventional standard (p=.06), and reaches the conventional standard (p=.035) for Hypothesis 2. Taken together, this is substantial evidence that OSA treatment is associated with savings in non-OSA-program medical insurance claim costs.

It may be asked why the estimated effect of APAP treatment on the medical insurance costs of a typical driver does not translate into a larger aggregate savings level (as, for example, a simple multiplication of the treatment effect by 100 drivers and 18 months might suggest). The main reason is the effect on the estimated aggregate savings of the continuing flow of entries into and exits from employment experienced by the study firm, a process that is typical of firms in its

business segment.[51] Specifically, the benchmark reference against which the aggregate cost of 100 Diagnosed drivers over 18 months was measured was that of 100 Screen-positive Controls over the same time period. But this cost is substantially reduced by the significant number of Screen-positive Controls who exited employment before the observation period ended; these exits lowered the benchmark cost, and thus lowered the estimated cost savings (see Figure 1).

Consideration of the turnover process does, however, lead to two considerations which strongly suggest that the estimated value of aggregate savings is a lower bound on the actual value in a continuing OSA program. First, the most relevant control group is drivers who would have been screen-positive in the absence of an OSA program, as these drivers would be accumulating continuing medical insurance costs paid by the firm employing them. Screen-positive Controls who were aware of the OSA program, but who had not yet been sent for a diagnosis, are a natural proxy. However, it is reasonable to speculate that knowledge of the OSA program caused some Screen-positive Control drivers in the study to accelerate their exit from employment, in order to avoid being sent for a PSG diagnostic test, with the consequent likelihood of being diagnosed with OSA.[47] To the extent this occurred, the aggregate costs of the study's Screen-positive Controls would have been lower than those of the most relevant control group, a similar set of drivers who did not expect to be sent for diagnosis if they stayed on the job. Holding the cost of Diagnosed drivers constant, raising the reference benchmark would in turn raise the estimated aggregate cost savings. The finding that Screen-positive Controls had statistically higher BMI than Diagnosed drivers is consistent with this speculation.

Second, the aggregate cost of the Diagnosed drivers includes the costs of the No Adherence subgroup. These drivers were very expensive, and while they exited quickly compared to other subgroups, their mean observation length in the after-period was 27 weeks, and a majority quit,

because discharging those who were not adherent to treatment took time. Thus, their aggregate costs are a substantial contribution to the aggregate costs of the Diagnosed drivers. It was a goal of the study firm to shorten the remediation process for No Adherence drivers. To the extent this occurred (and this began to occur during the study period), it would lower the costs of No Adherence subgroup, which would lower the aggregate cost of the Diagnosed drivers. Holding the costs of Screen-positive Controls fixed, this would increase the estimated aggregate cost reduction for the Diagnosed as compared to Controls. Both of these considerations suggest the true savings from an ongoing OSA program may be higher than the estimated value.

It must be noted that a standard randomized prospective controlled trial for the evaluation of these two hypotheses was not feasible (and never will be), because it was neither ethical nor legal to randomly assign some drivers with OSA either to no treatment or to a sham treatment, since an effective treatment exists, and since untreated OSA is associated with higher commercial vehicle crash risk.[36,55] As a result, a retrospective analysis of the data from an OSA program designed by managers for a business purpose (improving fleet safety performance) was required. A retrospective analysis in this context has both advantages and challenges, and these are discussed in the following sections.

b. **Advantages of this Study**

This study has several advantages over prior work.

1) It is a retrospective analysis of the actual non-OSA-program medical insurance claim expenses incurred by 3,032 employees in an employer-based program to screen, diagnose, and treat employees for OSA, covered with no out-of-pocket cost for employees enrolled in the study firm's voluntary medical insurance.

2) Because employee participation (and treatment adherence for those with OSA) was

mandated (on vehicle safety grounds) by the study firm as a condition of continued employment, it provides a unique estimation of the medical cost benefits of an OSA program when failure to adhere to treatment was minimized by comparison with the magnitude of this issue in a general patient population.

3) Because prescribed treatment was by APAP, objective treatment adherence data recorded by APAP units was available.

4) OSA diagnosis was by an in-lab overnight polysomnogram, a benchmark for diagnostic quality.

5) The well-established pattern of driver turnover in the part of for-hire motor freight in which the study firm operates,[51] together with the potential for healthy worker selection and safety selection,[36,53,55] implied that many drivers with poor health and high medical costs, both of which were likely to have been due to OSA, were prevented from entering the participant pool, because they left employment before the study began. A consequence was the potential for different levels of such selection between Diagnosed drivers and Screen-positive Controls. This issue was addressed effectively by matching each Diagnosed drivers with a Screen-positive Control on job tenure and experience-at-hire, to ensure both subgroups had the same exposure both to the selection processes as well as to the chance of filing a medical insurance claim.

6) The potential for bias resulting from non-random selection of which Screen-positive Controls were sent for an OSA diagnosis (an issue relevant to Hypothesis 1) was addressed effectively with a difference-in-differences approach. The cost difference from the period before the PSG/matching date was subtracted from that observed in the period after the PSG/matching date, to remove the effects of any initial differences.[35]

7) A bootstrap approach was used to generate the difference-in-differences Hypothesis 1

aggregate cost savings estimate and its confidence interval, which effectively utilized the information contained in the data about the actual costs incurred by the actual drivers in the study sample.[67]

8) As noted above, all drivers diagnosed with OSA were necessarily offered effective treatment, which implies they then self-selected into treatment adherence sub-groups. This created the challenge in that drivers choosing to avoid mandated APAP treatment may have had other behavioral differences from those accepting treatment that affected their costs. This was also effectively addressed with a difference-in-differences approach.[35]

9) The two-part model used for Hypothesis 2 not only applied the difference-in-differences methodology through the use of model-predicted values, it also effectively addressed two other features of the data. One was the distributional properties of medical insurance expenses, which contained different proportions of zero costs and of a smaller number of high costs across the study sub-groups. The second was differences across study sub-groups in anthropometric and work/related driver characteristics, along with differences in exposure in the after-period (which occurred because No Adherence drivers departed quickly compared to other subgroups).

c. **Limitations of this Study**

1) As noted above, medical and safety selection differentially removed high cost drivers before they could enter the study, and is likely to have especially affected drivers with serious OSA and/or serious complications. The potential effect of the return of these drivers to the study data set cannot be definitively determined, though it can be speculated that costs associated with OSA comorbidities would be particularly affected, and that the aggregate cost savings of Diagnosed drivers as compared to Screen-positive Controls might cross the threshold of statistical significance.

2) The evidence suggests that medical costs for all drivers who had untreated OSA rose over time, as might be expected due to the progression of the disease. This does not pose a challenge for the test of Hypothesis 1, which used actual costs during the period of observation, with a conservative approach to cost inflation for drivers whose observations were censored. However, while using the two-part model to address Hypothesis 2 has multiple advantages, and while it correctly accounts for differences in exposure, it also has the limitation that it implicitly treats PMPM costs as constant for all months when used to compute the predicted values used in the difference-in-differences estimates.

3) The inability to randomly assign participants to a sham versus real treatment means that placebo effects, if present, cannot be separately identified from the effect of OSA treatment. The difference-in-differences approach used can separate out only the "total" effect of OSA treatment (i.e. including any placebo effect) as distinct from all unmeasured effects on medical costs associated with driver self-selection into treatment groups. This is remediated to a degree by the fact that even though the estimated result is not the treatment effect most of interest to medical science, it is nonetheless a treatment effect of central interest to motor carrier managers and safety policy makers, who need to know the total effect of OSA treatment on medical insurance costs.[35]

4) The difference-in-differences methodology gives an accurate answer conditional on the assumption that the differences between study subgroups observed in the first period are maintained in the second period, an assumption which cannot be directly tested.[35] This is partly remediated by the fact that the model used to estimate the treatment effect accounts directly for any changes in the covariates, and more generally by the fact that difference-in-differences is the best feasible statistical methodology, given the prohibition of a randomized prospective trial due to ethical and legal liability issues.

5) Due to data limitations, the medical insurance claims costs presented here do not include those for pharmaceutical claims, and the costs of the OSA program itself are not included in the analysis.

6) Limitations in the way the APAP adherence data was recorded required the categorization of drivers into adherence sub-groups with limited differentiation, except for that between those showing no adherence versus any positive level of adherence. This is reflected in the fact that statistical differences in medical insurance costs between drivers with different levels of positive adherence were not found.

d. Medical and Policy Implications

The study firm, like many large motor carriers, was self-insured for most medical claims, and so benefited directly from these savings: a study firm executive was quoted in 2016, "We can fund the expense of OSA diagnosis and treatment just by generating savings on the medical side."[68] Medical insurance providers to smaller carriers may expect similar benefits to those of the study firm, if they organize and manage an OSA program for their trucking firm customers that is similar to that of the study firm. The evidence presented suggests that medical insurance firms might lower their net costs by offering such customers an OSA program for their driver employees with screening, diagnosis, and treatment as preventive care at low or no charge, as long as the motor carrier makes the program mandatory (which is separately justifiable on safety grounds).[36]

Further, the estimates of aggregate cost savings omit some parts of the cost savings to the firm, to drivers, and to society, that are associated with a mandatory OSA program. The potential savings in pharmaceutical insurance costs are not included here, nor is the value of injuries, lost work time, or disability days associated with untreated OSA,[69-71] nor the savings from avoided

preventable crashes. It is also likely that the Positive Adherence drivers, who had lower medical insurance expenses as a result of the program, experienced better health. Additionally, the data show starting with 100 Screen-positive Controls and 100 Diagnosed drivers at the PSG/matching date, after 18 months the study firm retained on average 13 more Diagnosed drivers than Screen-positive Controls (see Online Supplement Section XIII; the result is driven by the longer retention of Positive Adherence and OSA Negative drivers as compared to Controls). This shows that the OSA program also improved driver retention and lowered the study firm's turnover costs.

Trucking firm managers and their medical insurance providers should consider these findings on cost savings, along with earlier findings on the reduction in the risk of serious preventable truck crashes,[36] when considering OSA programs. Trucking safety regulators should consider them with respect to mandating OSA screening standards for commercial motor vehicle operators.

Acknowledgements

The authors wish to specifically acknowledge the contributions of the journal's associate editor and anonymous referees, and also the assistance of the executives and staff of the study firm, Schneider Enterprise Resources, LLC (dba Schneider National, Inc.) in acquiring operational and medical data utilized, as well as financial assistance to UMM's Truckers & Turnover Project from Schneider, the Roadway Safety Institute (the USDOT Region 5 University Transportation Center, which is funded by the USDOT Office of the Assistant Secretary for Research and Technology), the MacArthur Foundation's Research Network on the Origins of Norms and Preferences, the Sloan Foundation, and the University of Minnesota, Morris. The research received support from Harvard Catalyst, The Harvard Clinical and Translational Science Center (NIH Award #UL1 RR 025758, financial contributions from Harvard University and its affiliated academic health care centers), and support from National Surface Transportation Safety Center for Excellence (Project# 12-UI-017). The content is solely the responsibility of the authors and does not necessarily represent the views of the journal or of any of the research sponsors.

Disclosures: Mark Berger MD, FCCP reports that he is an owner and CEO of Precision Pulmonary Diagnostics, LLC, which has received fees for sleep-apnea-related services from the study firm. In addition, Dr. Berger has been issued US patents 7599892, US 7720696, US 8249896 and US 8200510. Charles A. Czeisler MD, PhD reports grants from Cephalon Inc., grants from Mary Ann & Stanley Snider via Combined Jewish Philanthropies, grants from National Football League Charities, grants from Optum, grants from Philips Respironics, Inc., grants from ResMed Foundation, grants from San Francisco Bar Pilots, grants from Schneider Inc., grants from Sysco, grants from Cephalon, Inc, grants from Jazz Pharmaceuticals, grants from Takeda Pharmaceuticals, grants from Teva Pharmaceuticals Industries, Ltd, grants from Sanofi-Aventis, Inc, grants from Sepracor, Inc, grants from Wake Up Narcolepsy, personal fees from Bose Corporation, personal fees from Boston Celtics, personal fees from Boston Red Sox, personal fees from Columbia River Bar Pilots, personal fees from Institute of Digital Media and Child Development, personal fees from Klarman Family Foundation, personal fees from Samsung Electronics, personal fees from Quest Diagnostics, Inc, personal fees from Vanda Pharmaceuticals, personal fees from American Academy of Sleep Medicine (AASM), personal fees from CurtCo Media Labs LLC, personal fees from Global Council on Brain Health/AARP, personal fees from Hawaii Sleep Health and Wellness Foundation, personal fees from Harvard School of Public Health (HSPH), personal fees from Maryland Sleep Society, personal fees from National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), personal fees from National Sleep Foundation (NSF), personal fees from New England College of Optometry, personal fees from University of Michigan, personal fees from University of Washington, personal fees from Zurich Insurance Company, Ltd, personal fees from Purdue Pharma, LP, personal fees from McGraw Hill, personal fees from Houghton Mifflin Harcourt/Penguin,

personal fees from Koninklijke Philips Electronics, N.V., personal fees from Cephalon, Inc, personal fees from State of Washington Board of Pilotage Commissioners, personal fees from Ganesco Inc., holds an equity interest in Vanda Pharmaceuticals, outside the submitted work. In addition, Dr. Czeisler holds a number of process patents in the field of sleep/circadian rhythms (e.g., photic resetting of the human circadian pacemaker). Since 1985, Dr. Czeisler has also served as an expert on various legal and technical cases related to sleep and/or circadian rhythms including those involving the following commercial entities: Complete General Construction Company, FedEx, Greyhound, HG Energy LLC, South Carolina Central Railroad Co., Stric-Lan Companies LLC, Texas Premier Resource LLC and United Parcel Service (UPS). Dr. Czeisler's interests were reviewed and managed by Brigham and Women's Hospital and Partners HealthCare in accordance with their conflict of interest policies. Stefanos N. Kales MD, MPH has served as a consultant and expert witness on cases involving commercial drivers. Dr. Hickman has served as an expert witness on cases involving commercial drivers. Erin Mabry, PhD reports grants from National Surface Transportation Safety Center for Excellence. The other authors have indicated no financial conflicts of interest. Non-financial disclosure: none.

6. References

1. Lal C, Strange C, Bachman D. Neurocognitive impairment in obstructive sleep apnea. Chest. 2012; 141 (6): 1601-1610.
2. Naismith S, Winter V, Gotsopoulos H, Hickie I, Cistulli P. Neurobehavioral functioning in obstructive sleep apnea: differential effects of sleep quality, hypoxemia and subjective sleepiness. J Clin Exp Neuropsychol. 2004; 26 (1): 43-54.
3. Wong KK, Grunstein RR, Bartlett DJ, Gordon E. Brain function in obstructive sleep apnea: results from the Brain Resource International Database. J Integr Neurosci. 2006; 5 (1): 111-121.
4. Colten HR, Altevogt BM. *Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem.* Washington (DC): Institute of Medicine Committee on Sleep Medicine Research, National Academies Press (US); 2006.
5. Kendzerska T, Mollayeva T, Gershon AS, Leung RS, Hawker G, Tomlinson G. Untreated obstructive sleep apnea and the risk for serious long-term adverse outcomes: A systematic review. Sleep Med Rev. 2014; 18 (1): 49-59.
6. Neilan TG, Farhad H, Dodson JA, et al. Effect of sleep apnea and continuous positive airway pressure on cardiac structure and recurrence of atrial fibrillation. Journal of the American Heart Association. 2013; 2 (6): e000421.
7. Lüthje L, Andreas S. Obstructive sleep apnea and coronary artery disease. Sleep Med Rev. 2008; 12 (1): 19-31.
8. Ejaz SM, Khawaja IS, Bhatia S, Hurwitz TD. Obstructive sleep apnea and depression: A review. Innov Clin Neurosci. 2011; 8 (8): 17-25.
9. Harris M, Glozier N, Ratnavadivel R, Grunstein RR. Obstructive sleep apnea and depression. Sleep Med Rev. 2009; 13 (6): 437-444.
10. Lee W, Lee SA, Chung YS, Kim WS. The relation between apnea and depressive symptoms in men with severe obstructive sleep apnea: mediational effects of sleep quality. Lung. 2015; 193 (2): 261-267.
11. Björnsdóttir E, Janson C, Sigurdsson JF, et al. Symptoms of Insomnia among Patients with Obstructive Sleep Apnea Before and After Two Years of Positive Airway Pressure Treatment. Sleep. 2013; 36 (12): 1901-1909.
12. Ong JC, Crisostomo MI. The more the merrier? Working towards multidisciplinary management of obstructive sleep apnea and comorbid insomnia. J Clin Psychol. 2013; 69 (10): 1066-1077.
13. Smith S, Sullivan K, Hopkins W, Douglas J. Frequency of insomnia report in patients with obstructive sleep apnoea hypopnea syndrome (OSAHS). Sleep Med. 2004; 5 (5): 449-456.
14. Meslier N, Gagnadoux F, Giraud P, et al. Impaired glucose-insulin metabolism in males with obstructive sleep apnoea syndrome. Eur Respir J. 2003; 22 (1): 156-160.
15. Morgenstern M, Wang J, Beatty N, Batemarco T, Sica AL, Greenberg H. Obstructive Sleep Apnea: An Unexpected Cause of Insulin Resistance and Diabetes. Endocrinol Metab Clin North Am. 2014; 43 (1): 187-204.
16. Xie W, Chakrabarty S, Levine R, Johnson R, Talmage JB. Factors associated with obstructive sleep apnea among commercial motor vehicle drivers. J Occup Environ Med. 2011; 53 (2): 169-173.
17. Charles LE, Burchfiel CM, Fekedulegn D, Andrew ME, Violanti JM, Vila B. Obesity and sleep: the Buffalo Police health study. Policing. 2007; 30 (2): 203-214.

18. Ong CW, O'Driscoll DM, Truby H, Naughton MT, Hamilton GS. The reciprocal interaction between obesity and obstructive sleep apnoea. Sleep Med Rev. 2013; 17 (2): 123-131.
19. Tuomilehto H, Seppä J, Uusitupa M. Obesity and obstructive sleep apnea—Clinical significance of weight loss. Sleep Med Rev. 2013; 17 (5): 321-329.
20. Young T, Finn L, Peppard PE, et al. Sleep disordered breathing and mortality: eighteen-year follow-up of the Wisconsin sleep cohort. Sleep. 2008; 31 (8): 1071-1078.
21. AlGhanim N, Comondore VR, Fleetham J, Marra CA, Ayas N. The Economic Impact of Obstructive Sleep Apnea. Lung. 2008; 186 (1): 7-12.
22. Leger D, Bayon V, Laaban JP, Philip P. Impact of sleep apnea on economics. Sleep Med Rev. 2012; 16 (5): 455-462.
23. Skaer TL, Sclar DA. Economic implications of sleep disorders. Pharmacoeconomics. 2010; 28 (11): 1015-1023.
24. Kapur V, Blough DK, Sandblom RE, et al. The medical cost of undiagnosed sleep apnea. Sleep. 1999; 22 (6): 749-755.
25. Bureau of Labor Statistics. Table 1.2 Employment by detailed occupation, 2016 and projected 2026. https://www.bls.gov/emp/tables/emp-by-detailed-occupation.htm. Accessed July 21, 2018.
26. Berger M, Varvarigou V, Rielly A, Czeisler CA, Malhotra A, Kales SN. Employer-Mandated Sleep Apnea Screening and Diagnosis in Commercial Drivers. J Occup Environ Med. 2012; 54 (8): 1017-1025.
27. Gurubhagavatula I, Maislin G, Nkwuo JE, Pack AI. Occupational screening for obstructive sleep apnea in commercial drivers. Am J Respir Crit Care Med. 2004; 170 (4): 371-376.
28. Howard ME, Desai AV, Grunstein RR, et al. Sleepiness, sleep-disordered breathing, and accident risk factors in commercial vehicle drivers. Am J Respir Crit Care Med. 2004; 170 (9): 1014-1021.
29. Moreno CR, Carvalho FA, Lorenzi C, et al. High risk for obstructive sleep apnea in truck drivers estimated by the Berlin questionnaire: prevalence and associated factors. Chronobiol Int. 2004; 21 (6): 871-879.
30. Talmage JB, Hudson TB, Hegmann KT, Thiese MS. Consensus criteria for screening commercial drivers for obstructive sleep apnea: evidence of efficacy. J Occup Environ Med. 2008; 50 (3): 324-329.
31. Federal Motor Carrier Safety Administration. *2017 Pocket Guide to Large Truck and Bus Statistics.* Washington, DC2017.
32. Parks PD, Durand G, Tsismenakis AJ, Vela-Bueno A, Kales SN. Screening for obstructive sleep apnea during commercial driver medical examinations. J Occup Environ Med. 2009; 51 (3): 275-282.
33. Hoffman B, Wingenbach DD, Kagey AN, Schaneman JL, Kasper D. The Long-Term Health Plan and Disability Cost Benefit of Obstructive Sleep Apnea Treatment in a Commercial Motor Vehicle Driver Population. J Occup Environ Med. 2010; 52 (5): 473-477.
34. Potts KJ, Butterfield DT, Sims P, Henderson M, Shames CB. Cost savings associated with an education campaign on the diagnosis and management of sleep-disordered breathing: a retrospective, claims-based US study. Population Health Management. 2013; 16 (1): 7-13.
35. Dimick JB, Ryan AM. Methods for evaluating changes in health care policy: the difference-in-differences approach. J Am Med Assoc. 2014; 312 (22): 2401-2402.
36. Burks SV, Anderson JE, Bombyk M, et al. Nonadherence with Employer-Mandated Sleep Apnea Treatment and Increased Risk of Serious Truck Crashes. Sleep. 2016a; 39 (5): 967-975.

37. Ellen RL, Marshall SC, Palayew M, Molnar FJ, Wilson KG, Man-Son-Hing M. Systematic review of motor vehicle crash risk in persons with sleep apnea. Journal of Clinical Sleep Medicine : JCSM : official publication of the American Academy of Sleep Medicine. 2006; 2 (2): 193-200.
38. Tregear S, Reston J, Schoelles K, Phillips B. Obstructive sleep apnea and risk of motor vehicle crash: systematic review and meta-analysis. Journal of Clinical Sleep Medicine : JCSM : Official Publication of the American Academy of Sleep Medicine. 2009; 5 (6): 573-581.
39. Tregear S, Reston J, Schoelles K, Phillips B. Continuous positive airway pressure reduces risk of motor vehicle crash among drivers with obstructive sleep apnea: systematic review and meta-analysis. Sleep. 2010; 33 (10): 1373-1380.
40. FMCSA. Summary-Medical Review Board Meeting of January 2008. *Office of Medical Programs.* 2008; 2008(February 24). http://www.mrb.fmcsa.dot.gov/documents/Final_Meeting_Revised_Updated_93008.pdf.
41. Hersman DAP, Hart CA, Sumwalt RL. Safety Recommendations H-09-015 &-016 to the Federal Motor Carrier Safety Administration Regarding Obstructive Sleep Apnea in Commercial Drivers. In. Washington, DC: National Transportation Safety Board; 2009.
42. Pervall GC, Hernandez S. *Medical Review Board and the Motor Carrier Safety Advisory Committee joint recommendations regarding Federal Motor Carrier Safety Administration's and Federal Railroad Administration's Advanced Notice of Proposed Rulemaking (ANPRM) on Obstructive Sleep Apnea.* Washington, DC: Federal Motor Carrier Safety Administration;2016.
43. Jaillet J. President signs bill forbidding FMCSA to use sleep apnea 'guidance'. *Commercial Carrier Journal - Fleet Management Magazine.* 2013; (October 17). http://www.ccjdigital.com/president-signs-bill-forbidding-fmcsa-to-use-sleep-apnea-guidance/.
44. Federal Motor Carrier Safety Administration and Federal Railroad Administration. Evaluation of Safety Sensitive Personnel for Moderate-to-Severe Obstructive Sleep Apnea. In: USDOT, ed. Vol 81 FR 12642 Washington, DC: Federal Register; 2016:pp. 12642-12647.
45. Schremmer M. FMCSA's withdrawal of proposal halts pursuit of new apnea regulation. http://www.landlinemag.com/Story.aspx?StoryID=33784#.WXuzTlGQxGo. Accessed July 28, 2017.
46. Sleep Review Staff. After Steps Forward, FMCSA Withdraws Rulemaking on Sleep Apnea. http://www.sleepreviewmag.com/2017/07/steps-forward-fmcsa-withdraws-rulemaking-sleep-apnea/. Accessed July 28, 2017, 2017.
47. Mabry JE, Baker S, Hickman J, Hanowski R. *Case Study on the Impact of Treating Sleep Apnea in Commercial Motor Vehicle Drivers--Sleep Apnea Programs from Two Leading U.S. Carriers and Focus Group Findings.* Blacksburg, VA: National Surface Transportation Safety Center for Excellence;2012. 12-UI-017.
48. Burks SV, Belzer M, Kwan Q, Pratt S, Shackelford S. Trucking 101: An Industry Primer. *Transportation Research Circular Number E-C146.* 2010.
49. Bearth DP, Green B, Clevenger S. *Top 100 For-Hire Carriers 2017, Gearing Up for Growth.* Alexandria, VA: American Trucking Associations;2017.
50. Epstein LJ, Kristo D, Strollo PJ, Jr., et al. Clinical guideline for the evaluation, management and long-term care of obstructive sleep apnea in adults. Journal of Clinical Sleep Medicine : JCSM : Official Publication of the American Academy of Sleep Medicine. 2009; 5 (3): 263-276.
51. Burks SV, Monaco K. Is the U.S. labor market for truck drivers broken? Mon Labor Rev. 2019; (March).

52. Burks SV, Carpenter J, Götte L, Rustichini A. Cognitive Skills Affect Economic Preferences, Social Awareness, and Job Attachment. Proc Natl Acad Sci U S A. 2009; 106 (19): 7745-7750.
53. Arrighi HM, Hertz-Picciotto I. The evolving concept of the healthy worker survivor effect. Epidemiology. 1994; 5 (2): 189-196.
54. Li CY, Sung FC. A review of the healthy worker effect in occupational epidemiology. Occup Med. 1999; 49 (4): 225-229.
55. Burks SV, Anderson JE, Bombyk M, et al. Supplementary Material to Nonadherence with Employer-Mandated Sleep Apnea Treatment and Increased Risk of Serious Truck Crashes. *Sleep.* 2016b; 39(5); 967-975. Accessed Mar 21, 2016.
56. Kribbs NB, Pack AI, Kline LR, et al. Objective measurement of patterns of nasal CPAP use by patients with obstructive sleep apnea. The American Review of Respiratory Disease. 1993; 147 (4): 887-895.
57. Bureau of Economic Analysis. Table 2.3.4 Price Indexes for Personal Consumption Expenditures by Major Type of Product. Interactive table selected to show annual PCE values for all major expenditure types, including health care, 2005-2015. http://www.bea.gov/iTable/iTable.cfm?reqid=9&step=1&acrdn=2#reqid=9&step=3&isuri=1&904=2000&903=64&906=a&905=2015&910=x&911=1. Accessed March 18, 2016.
58. Efron B, Tibshirani R. Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. Statistical Science. 1986; 1 (1): 54-75.
59. Belloti F, Deb P, Manning WG, Norton EC. twopm: Two-part models. The Stata Journal. 2015; 15 (1): 3-20.
60. Deb P, Norton EC. Modeling Health Care Expenditures and Use. Annu Rev Public Health. 2018; 39: 489-505.
61. Deb P, Norton EC, Manning WG. *Health Econometrics Using Stata.* Vol 3. 2018/01/13 ed. College Station, TX: Stata Press; 2017.
62. Babyak MA. What You See May Not Be What You Get: A Brief, Nontechnical Introduction to Overfitting in Regression-Type Models. Psychosom Med. 2004; 66 (3): 411-421.
63. Manning WG, Basu A, Mullahy J. Generalized modeling approaches to risk adjustment of skewed outcomes data. J Health Econ. 2005; 24 (3): 465-488.
64. Sieber WK, Robinson CF, Birdsey J, et al. Obesity and other risk factors: The National Survey of U.S. Long-Haul Truck Driver Health and Injury. Am J Ind Med. 2014; 57 (6): 615-626.
65. Thiese MS, Moffitt G, Hanowski RJ, Kales SN, Porter RJ, Hegmann KT. Commercial driver medical examinations: Prevalence of obesity, comorbidities, and certification outcomes. J Occup Environ Med. 2015; 57 (6): 659-665.
66. (HCUP) HCCaUP. Clinical Classifications Software (CCS) for ICD-9-CM. www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp. Accessed March 24, 2018.
67. Efron B, Tibshirani R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Statistical Science. 1986; 1 (1): 54-77.
68. Frost & Sullivan. *Hidden Health Crisis Costing America Billions; Underdiagnosing and Undertreating Obstructive Sleep Apnea Draining Healthcare System.* Darien, IL: American Academy of Sleep Medicine; 2016 2016.
69. Rajaratnam SM, Barger LK, Lockley SW, et al. Sleep disorders, health, and safety in police officers. JAMA. 2011; 306 (23): 2567-2578.
70. Barger LK, Rajaratnam SM, Wang W, et al. Common sleep disorders increase risk of motor vehicle crashes and adverse health outcomes in firefighters. J Clin Sleep Med. 2015; 11 (3):

233-240.
71. Sullivan JP, O'Brien CS, Barger LK, Rajaratnam SM, Czeisler CA, Lockley SW. Randomized, Prospective Study of the Impact of a Sleep Health Program on Firefighter Injury and Disability. Sleep. 2017; 40 (1).

Figure captions list

Figure 1. Retention on the Job After the PSG/matching Date by Study Sub-Group (Kaplan-Meier Survival Curves)

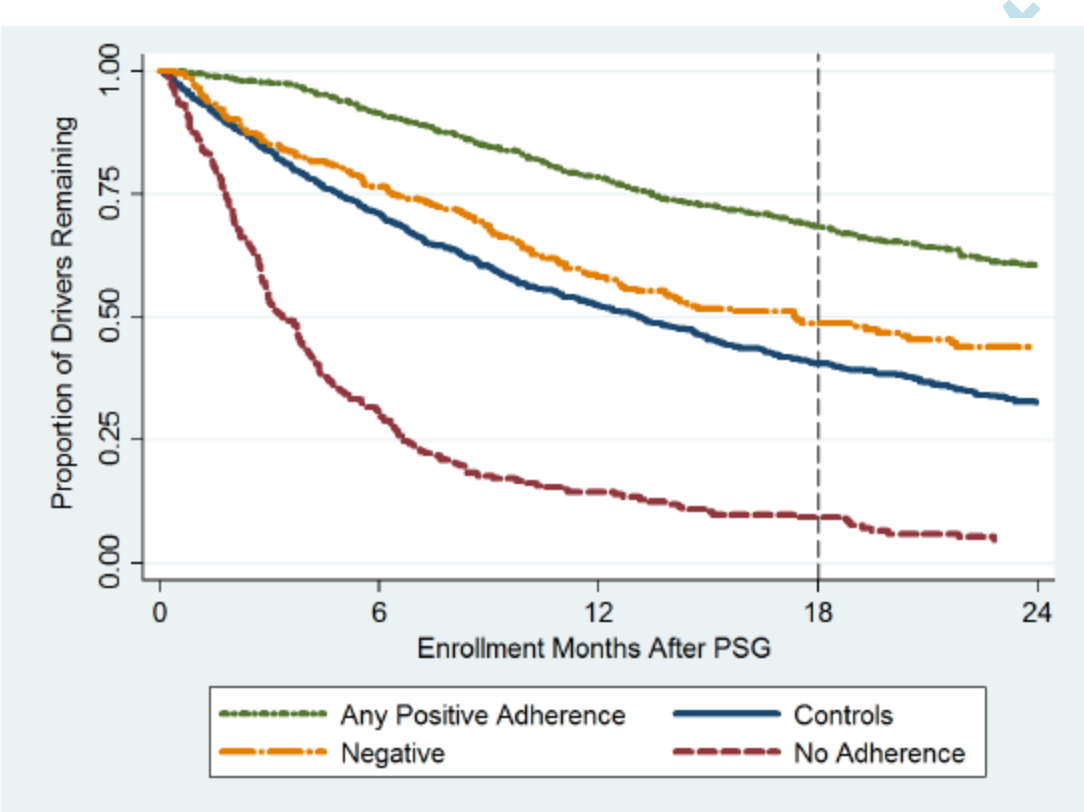

Table 1. Characteristics of Participants:

Screen-Positive Controls and Diagnosed Drivers

Category	Screen-Positive Controls N	%	Diagnosed Drivers N	%
Total	1,516		1,516	
Gender				
Female	40	2.6%	83	5.5%
Male	1,476	97.4%	1,433	94.5%
Age				
Under 40	583	38.5%	533	35.2%
40 - 49	409	27.0%	484	31.9%
50 or more	524	34.6%	499	32.9%
Race				
White	1,126	74.3%	1,158	76.4%
African-American	230	15.2%	202	13.3%
Other	160	10.6%	156	10.3%
Marital Status*				
Married	729	48.1%	667	44.0%
Single	787	51.9%	849	56.0%
Geography				
Midwest	529	34.9%	567	37.4%
Northeast	170	11.2%	100	6.6%
South	662	43.7%	727	48.0%
West	155	10.2%	122	8.0%
BMI				
Under 25	39	3%	61	4%
25 - 34.9	682	45.0%	702	46.3%
35 or more	795	52.4%	753	49.7%
Mean AHI	NA		27.3	
Mean Enrolled Months				
Before PSG	18.3		17.7	
After PSG	11.4[a]		13.9[a]	
Mean Tenure Months				
Before PSG	43.9		44.0	
After PSG	12.4[b]		14.4[b]	

Abbreviations “AHI” for “Apnea Hypopnea Index”, ”BMI” for “Body Mass Index,” “PSG” for “polysomnogram.” All characteristics not labeled as "before" or "after" reflect data after the PSG/matching date except Age and AHI, which are recorded on the PSG/matching date. *Marital Status is the most frequent status within the after period. Within a row,

columns containing the same superscript(s) are significantly different (t-tests, $p \leq .05$, not adjusted for multiple comparisons).

Table 2. Characteristics of Participants:

Screen-Positive Controls and Sub-groups of Diagnosed Drivers

Category	Controls		OSA Negative		Any Adherence		No Adherence	
	N	%	N	%	N	%	N	%
Total	1,516	100.0 %	292	19.3 %	932	61.5 %	292	19.3 %
Gender								
Female	40	2.6%	26	8.9%	43	4.6%	14	4.8%
Male	1,476	97.4%	266	91.1 %	889	95.4 %	278	95.2 %
Age								
Under 40	583	38.5%	122	41.8 %	286	30.7 %	125	42.8 %
40 - 49	409	27.0%	77	26.4 %	304	32.6 %	103	35.3 %
50 or more	524	34.6%	93	31.9 %	342	36.7 %	64	21.9 %
Race								
White	1,126	74.3%	222	76.0 %	717	76.9 %	219	75.0 %
African-American	230	15.2%	36	12.3 %	125	13.4 %	41	14.0 %
Other	160	10.6%	34	11.6 %	90	9.7%	32	11.0 %
Marital Status*								
Married	729	48.1%	137	46.9 %	419	45.0 %	111	38.0 %
Single	787	51.9%	155	53.1 %	513	55.0 %	181	62.0 %
Geography								
Midwest	529	34.9%	119	40.8 %	333	35.7 %	115	39.4 %
Northeast	170	11.2%	23	7.9%	67	7.2%	10	3.4%
South	662	43.7%	132	45.2 %	451	48.4 %	144	49.3 %
West	155	10.2%	18	6.2%	81	8.7%	23	7.9%
BMI								

Under 25	39	3%	38	13%	17	2%	6	2%
25 - 34.9	682	45.0%	169	57.9%	403	43.2%	130	44.5%
35 or more	795	52.4%	85	29.1%	512	54.9%	156	53.4%
Mean AHI	NA		1.8[a]		34.3[a]		30.2[a]	
Mean Enrolled Months								
Before PSG	18.3[b]		17.9[c]		19.5[b]		11.6[bc]	
After PSG	11.4[a]		12.9[a]		17.0[a]		5.2[a]	
Mean Tenure Months								
Before PSG	43.9[a]		41.7[d]		50.3[ad]		26.5[ad]	
After PSG	12.4[b]		13.7[c]		17.2[bc]		6.0[bc]	

Abbreviations “AHI” for “Apnea Hypopnea Index”, ”BMI” for “Body Mass Index,” “PSG” for “polysomnogram.” All characteristics not labeled as "before" or "after" reflect data after the PSG/matching date except Age and AHI, which are recorded on the PSG/matching date. *Marital Status is the most frequent status within the period. Within a row, columns containing the same superscript(s) are significantly different (t-tests, $p \leq .05$, not adjusted for multiple comparisons).

Table 3. Testing Hypothesis 1: Aggregate Costs, Cost Differences, and Difference-in-Differences Savings Estimate for a Pool of 100 Drivers over 18 Months

Study Subgroup	Estimated Cost for 100 Drivers BEFORE	Estimated Cost for 100 Drivers AFTER	Cost Difference (After - Before)
Controls	$344,231	$530,925	$186,694
95% CI	(291,465, 404,707)	(411,120, 676,368)	(56,153, 339,195)
Diagnosed	$324,409	$358,061	$33,652
95% CI	(265,927, 391,084)	(292,625, 430,829)	(-54,686, 122,784)
Cost Difference (Controls - Diagnosed)	$19,822	$172,864	$153,042
95% CI	(-62,337, 100,716)	(31,722, 331,151)	(-5,352, 330,525)

Abbreviation: "PSG" for "polysomnogram." Table reports results from 10,000 empirical bootstrap iterations using costs in the 18 months before, and the 18 months after, the PSG/matching date. Costs are adjusted from dollars at the time of the study to 2018 dollars using the Personal Consumption Expenditure (PCE) index for health care expenses. The lower right table cell provides the difference-in-differences estimate of the aggregate cost savings. This estimate is statistically significant at the p=.06 level.

Table 4: Testing Hypothesis 2: Model-adjusted Estimates

of the Effect of OSA Treatment on a Typical Driver

Treatment Effect	Estimated PMPM Cost Difference BEFORE	Estimated PMPM Cost Difference AFTER	PMPM Cost Difference-in-Differences
Effect of Treatment (on Treated Drivers)	-$134	-$575	-$441
95% CI	(-$290, $22)	(-$972, -$179)	(-$861, -$21)
Effect of Treatment (on All Drivers Offered Treatment)	-$123	-$546	-$423
95% CI	(-$273, $27)	(-$915, -$176)	(-$816, -$30)

The two-part multivariate model jointly estimates the probability of positive costs and the level of costs if positive, and both parts of the model adjust for the following individual driver characteristics: length of per-period exposure (observed insurance enrollment), sex, age, race/ethnicity, marital status, geographic location, BMI, job type, study subgroup membership, and time-period, along with multiple interactions (e.g. time-period with other covariates, interactions of study subgroup with age and time period). The treatment effect for treated drivers (Rows 1 & 2) gives the estimated total PMPM cost savings and 95% CI associated with treatment, for those who accepted treatment. It is generated by using the estimated model coefficients and the model's two equations to calculate predicted cost values, with individual covariate values set at the mean for members of the Positive Adherence subgroup, and all exposures set at 18 months. The first calculation sets subgroup indicators for the "counterfactual" condition that the "average Positive Adherence driver" did not adhere to treatment, and the second calculation sets subgroup indicators for the condition that the "average Positive Adherence driver" did adhere to treatment, and the estimated treatment effect is the difference of the two predicted costs. The treatment effect for all drivers offered treatment (an "intention to treat" analysis"; Rows 3 & 4) repeats the calculations for Rows 1 & 2, except that the mean values for all covariates are set at the levels associated with all drivers offered treatment (i.e. including both the Positive and No Adherence subgroups). The difference between the after period and before period differences is the estimated treatment effect (difference-in-differences; last column). Stata's "margins" command was used to generate the estimated cost differences and confidence intervals. Costs are in 2018 dollars, adjusted using the Personal Consumption Expenditure (PCE) index for health care expenses. AHI $\geq$ 5 is the positive OSA diagnosis threshold level. Abbreviations: "Per-member per-month" is "PMPM," "polysomnogram" is "PSG," "CI" is "confidence interval."