STAT 568 Project:
An Analysis of the Association Between Race and Other Covariates on Obstructive Sleep Apnea

Humza S. Haider

April 10, 2018

Introduction

Obstructive Sleep Apnea (OSA) is a disorder which causes a temporary inhibition of airflow during sleep, leading to poorer quality of sleep [4, 9]. In tune with the poor quality of sleep, OSA has been linked to detrimental effects on cognitive abilities including memory, learning ability, and attention [4, 18]. Not limited to cognitive impacts, OSA has been observed to be associated with Type II diabetes [7, 8], depression [6, 3], obesity [19, 2, 11, 17], and hypertension [10].

Prior studies have examined profiles of different races with regards to OSA [15, 5, 1, 13], yet lack to consider the impact that race may play on the risk of being diagnosed with OSA. Many times the evaluation that race may play on the risk of diagnosis is confounded by the inequality of socioeconomic status among different races.

Further, an income inequality among races would likely lead to imbalance of diagnoses of OSA as the diagnosing procedure is often expensive and requires an overnight stay at a sleep laboratory. The gold standard for diagnosing OSA is a polysomonogram (PSG) which records the number of times air flow is interrupted per hour of sleep. While serving as the gold standard, a PSG requires a laboratory setting, technicians, and is often expensive and unaccessible to many patients [12]. For these reason, we speculate that many individuals may unknowingly have OSA.

Using data from the National Health and Nutrition Examination Survey (NHANES) provided by the CDC, we examine the impact that race and other covariates many play on the risk of OSA. As few individuals in the NHANES dataset have a confirmed diagnosis of OSA we choose to introduce a new variable, probable-OSA (pOSA), derived from common symptoms of OSA, *e.g.* snoring and snorting during sleep. Prior work using this definition and a similar dataset was completed by Sands et al., however, they examined the role of pOSA and race on hypertension as opposed to pOSA being the primary dependent variable of interest [16].

Methods

Data was extracted from the publicly available NHANES 2007-2008[1] data which aims to provide a nationally representative sample of non-institutionalized residents of the United States. As NHANES follows a complex survey design, survey methodology (specifically weighting) was applied prior to analysis. A full list of methods and analytic guidelines for NHANES data is provided on their website [2].

For the data sample considered, all participants had non-missing age, race, and pOSA status. Given this constraints, the total sample size was N = 5995.

Our definition of pOSA is defined in the paper by Sands et al. [16] which examined how pOSA related to hypertension, however, we choose to reiterate the definition here for clarity. The NHANES survey data had three questions directly relating to symptoms commonly used to identify potential sleep apnea:

1. Of those diagnosed with a sleep disorder, was it sleep apnea?
2. How often do you snore while sleeping?
3. How often do you snort / stop breathing while sleeping?

[1] https://wwwn.cdc.gov/nchs/nhanes/ContinuousNhanes/Default.aspx?BeginYear=2007
[2] https://wwwn.cdc.gov/nchs/nhanes/analyticguidelines.aspx

The last two questions were scored on one of 4 levels: (1) "Never", (2) Rarely (1-2 nights/week), (3) Occasionally (3-4 nights/week), or (4) Frequently (5 or more nights/week). Any individual who reported snoring 3+ nights per week, snorted / stopped breathing 3+ nights per week or reported being diagnosed with sleep apnea were marked as positive for pOSA. We also have include an OSA variable which is the number of individuals who self reported OSA, *i.e* responded positively to question 1.

Other covariates included were race, age, household income to poverty ratio, martial status, gender, hypertension, Type II diabetes, depression, smoking status, BMI, and education. For interpretability, all continuous covariates were broken down into ordinal, factor variables. Martial status was coded as single or "married" where the married category included individuals whom lived with there partners. This covariate was included as there is potential that individuals who live alone may not be aware of there snoring, and/or snorting status. Hypertension, Type II diabetes, depression, and smoking are all known comorbidities of OSA and so were included as covariates. Similarly, BMI and age have been found to be strongly associated with OSA and were in turn included in the model. As we are interested in the effects of race on OSA, we wished to control for socioeconomic status as this commonly confounds any race effects present in the data. As a substitute for socioeconomic status we include income poverty ratio and education.

Note that as this a retrospective study we are limited to the number of applicable designs and statistical models that we could apply, specifically, fractional factorial and any aliasing model is not applicable here. Clearly a full factorial design is not feasible as this would require hundred of thousands of observations given the number of covariates we included. Additionally, for a full factorial model, "treatments" or inclusion into the study would need to be designed in a non random way in order to estimate all effects. Given the non-optimal, random assignment of "treatments", we use a multiple factor ANOVA model for this data.

One challenge facing model building was the presence of missing data in many of the variables. Specifically, income to poverty ratio had almost 10% of data values missing. For this analysis we considered multiple imputation but ultimately decided it was out of the scope of this analysis due to already complex survey design implementation. For this reason, observations with missing data were ultimately removed prior to analysis.

Note here that our dependent variable, pOSA is binary, requiring a generalized linear model (GLM) approach as opposed to the classic linear model. For this reason we specify the `quasibinomial` link function for the GLM. Due to the survey methodology requirements, analysis was conducted in R version 3.4.4 using the `survey` package, specifically the `svyglm` function.

Due to the large number of covariates considered, manually checking all possible interactions is not feasible. Instead, we first checked common interactions found for medical work, *e.g.* interactions between age and BMI. Following this, the `step` function was used to identify potential interactions. Suggestions from the `step` function were added to the model up until a likelihood ratio test between the model with and without the interaction have a p-value above 0.001. While this threshold may seem severe, due to the large number of interactions, we wish not to have a false discovery. By including interactions in this manner we are able to utilize both AIC and likelihood ratio methods for variable selection.

Results

Demographics of the study population are available in Table 1. Note that this is the study population *after* accounting for survey weights, *i.e.* counts were generated using the `svytable` command. The breakdown of our dependent variable, pOSA, was almost exactly half positive (3055, 50.95%) and half negative (2940, 49.05%). We see that pOSA positive has a much higher frequency than confirmed OSA (286, 4.77%). Also of note was that a very small number of indiviuals who were underweight on the BMI scale (93, 1.57%). For this reason, prior to performing ANOVA, the underweight BMI group was collapsed into the normal BMI group as to not have severly unequal groups.

As mentioned in the previous section, the income poverty ratio is missing nearly 10% of the data. Following the building of the ANOVA model, the income poverty ratio variable was not significant ($p = 0.934$) and had no significant interactions (data not shown, note the significance test for this ANOVA model was the Rao-Scott Likelihood Ratio Test with a working likelihood due to the survey nature of the data). For this reason, we excluded the income poverty ratio and re-ran the ANOVA model and had no significant changes between the model with and without income poverty ratio. We report the full model output for the model without income poverty ratio in Table 2. In total, after removing data with missing covariates the model has a total of $N = 5340$ observations. As this is a GLM with a quasibinomial link our coefficients are in terms of odd ratios (OR) as opposed to traditional coefficients. In total, three interactions were included in the model, age by marital status, age by gender, and gender by marital status. Residual plots did not indicate any lack of fit nor influential outliers. Further, a Hosmer-Lemeshow test did not indicate any lack of fit.

The ANOVA table (data not shown) indicated race to not be a significant factor ($p = 0.548$). As white (Caucasian) is the dominant race in the United States we choose to use contrasts comparing every race against white. Doing so showed near significance for the "Other" category of race ($OR = 1.412$, $p = 0.064$). A full comparison among races is visible in Figure 1.

As suspected, marital status had a strong association with pOSA and additionally had an interaction with gender. By the odds ratios given in Table 2, one can observe that being married as a man is much more likely to have pOSA as opposed to a married woman. This interaction effect is highlighted in Figure 2.

In addition to the gender by marital status interactions, age had interactions with two variables, marital status, and gender, both of which are visualized in Figure 3. Most clear is the interaction between age and gender- as age increases, females risk of pOSA tends to increase but there is no clear relationship in males (Figure 3, bottom). In the age by marital status interaction we see that for married individuals there tends to be a slight linear increase in risk as age increases but for singles, there appears to be no change or even a decrease to risk as age increases.

Except for within interaction terms, we saw that none of the ordinal terms were significant past the linear term. In the interaction terms we see the quadratic and cubic terms of age being significant with the gender variable $p = 0.027$, and $p = 0.043$, respectively. However, since the cubic term is just past the threshold of significance, we speculate this may be the occurrence of overfitting as high polynomials often overfit the data's true nature.

Variable	N	Percent
OSA	Missing = 0	0.00%
Yes	286	4.77%
No	5709	95.22%
pOSA	Missing = 0	0.00%
Yes	3055	50.95%
No	2940	49.05%
Race	Missing = 0	0.00%
White	4172	69.60%
Black	676	11.28%
Mexican	503	8.39%
Other Hispanic	295	4.92%
Other	348	5.81%
Income Poverty Ratio	Missing = 594	9.91%
(0 - 1]	738	13.65%
(1 - 2]	1134	21.00%
(2 - 3]	861	15.95%
(3 - 4]	631	11.67%
(4 - 5]	2037	37.72%
Age	Missing = 137	2.29%
(18 - 25]	647	11.04%
(25 - 45]	2213	37.77%
(45 - 65]	2068	35.31%
(65 - 85]	930	15.87%
Martial Status	Missing = 291	4.85%
Single	2051	35.96%
Married	3653	64.04%
Gender	Missing = 0	0.00%
Male	2897	48.33%
Female	3097	51.67%
Hypertension	Missing = 294	4.90%
Yes	2042	35.81%
No	3659	64.18%
Diabetes (Type II)	Missing = 6	0.10%
Yes	608	10.15%
No	5381	89.85%
Depression	Missing = 0	0.00%
Yes	334	5.57%
No	5661	94.49%
Smokes	Missing = 294	4.90%
Yes	2699	47.33%
No	3002	52.66%
BMI	Missing = 111	1.85%
Underweight	93	1.57%
Normal	1779	30.24%
Overweight	2030	34.51%
Obese	1981	33.67%
Education	Missing = 6	0.10%
Less than High School	1219	20.34%
High School	1512	25.23%
Some College	1742	29.08%
College Graduate	1518	25.33%

Table 1: Demographics table for the NHANES 2007-2008 data with a total samplesize of N = 5995. The counts and percentages reported after adjusting for survey weights. Of note is that "Married" includes individuals who are married or living with their partner.

Coefficient	Odds Ratio	95% CI	P-value
Mexican	0.916	(0.748, 1.121)	0.394
Other Hispanic	1.166	(0.935, 1.455)	0.173
Black	0.876	(0.733, 1.046)	0.144
Other	1.412	(0.981, 2.033)	0.064 .
Age - Linear	0.647	(0.485, 0.864)	0.003**
Age - Quadratic	0.883	(0.656, 1.189)	0.413
Age - Cubic	0.778	(0.566, 1.068)	0.121
Married	2.491	(1.951, 3.180)	0.000***
Female	0.770	(0.604, 0.981)	0.034*
Less than high school	0.900	(0.735, 1.101)	0.306
Some college	0.886	(0.723, 1.087)	0.246
College graduate	0.799	(0.641, 0.997)	0.047*
BMI - Linear	2.617	(2.284, 3.000)	0.000***
BMI - Quadratic	1.030	(0.907, 1.168)	0.652
Hypertension	1.369	(1.145, 1.636)	0.001**
No Diabetes	0.953	(0.750, 1.211)	0.694
Depression	1.262	(0.940, 1.694)	0.121
Smoker	1.326	(1.134, 1.550)	0.000***
Age - Linear:Married	1.663	(1.242, 2.227)	0.001**
Age - Quadratic:Married	0.813	(0.573, 1.153)	0.246
Age - Cubic:Married	0.900	(0.602, 1.344)	0.606
Age - Linea:Female	1.427	(1.073, 1.896)	0.014*
Age - Quadratic:Female	1.465	(1.045, 2.053)	0.027*
Age - Cubic:Female	0.671	(0.457, 0.987)	0.043*
Married:Female	0.504	(0.369, 0.689)	0.000***

Table 2: Model Output from the summary() command on the ANOVA model. Significant labels are given as '.' for 0.05 - 0.10, '*' for 0.01 - 0.05, '**' for 0.001 - 0.01, and '***' for <0.001. Education was not coded as ordinal as we believe a categorical representation and allowing comparison to the base category "High school graduate" serves more useful than ordinal effects in education.

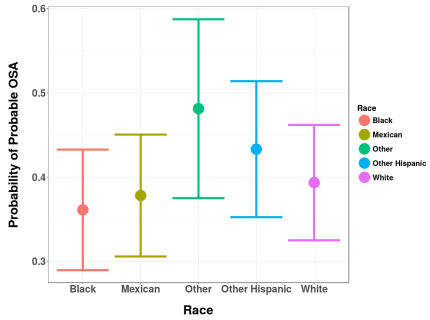

Figure 1: The predicted probabilities of pOSA for different races, with error bars representing the 95% CI. To predict these points we attempted to choose a representative individual, *i.e.* the probability was predicted for a married male, in the 25-45 age range, normal BMI, high school education, no hypertension, diabetes, depression and doesn't smoke.

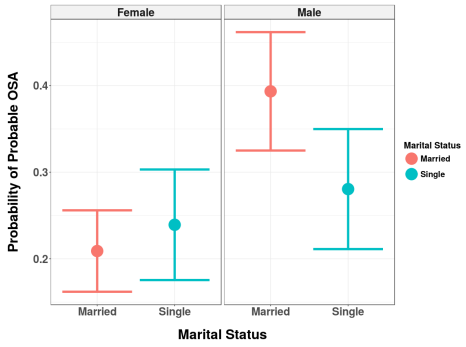

Figure 2: The predicted probabilities of pOSA for marital status across gender, with error bars representing the 95% CI. To predict these points we attempted to choose a representative individual, *i.e.* the probability was predicted for a white individual in the 25-45 age range, normal BMI, high school education, no hypertension, diabetes, depression and doesn't smoke.

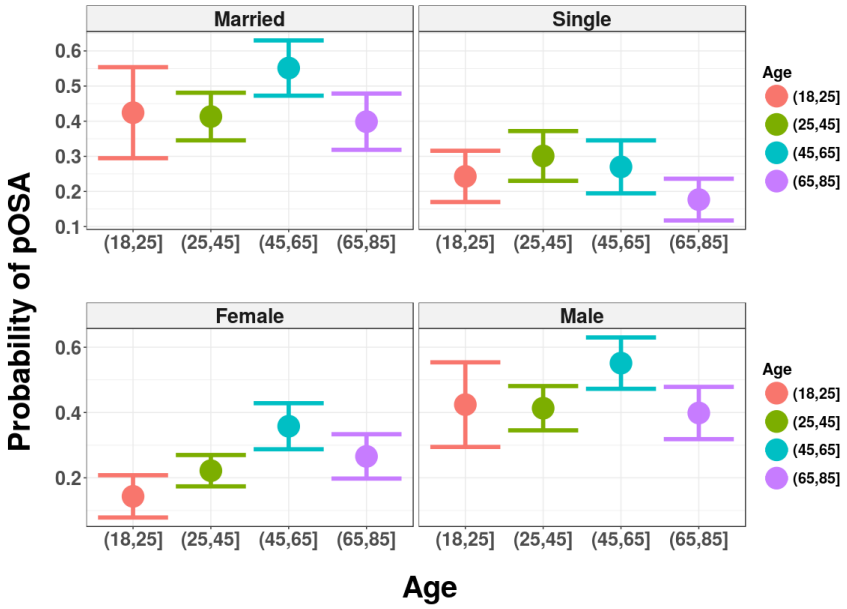

Figure 3: The predicted probabilities of pOSA for marital status across age, and gender across age with error bars representing the 95% CI. To predict these points we attempted to choose a representative individual, *i.e.* the probability was predicted for a white individual, normal BMI, high school education, no hypertension, diabetes, depression and doesn't smoke.

Discussion

While pOSA seemed to very rich in the associations between itself the chosen covariates, race was not included in these significant associations. Almost no differences were apparent after controlling for the other covariates in the model. The most significant race group was the "Other" group, however, this group contains many different races and so inference on how race effects risk of pOSA is not applicable.

While race showed no significant effects, the most powerful predictors of pOSA were BMI, hypertension, depression, and marital status. Of the first three, all are common comorbidities associated with OSA [6, 3, 19, 2, 11, 17, 10]. As we have coded "Living with Partner" to be qualified under the married group, we expect that married individuals have their spouse to tell them about their snoring/snorting and are more likely to report it, in turn being more likely to be at risk for pOSA. Those who are single may be unaware of their snoring/snorting during their sleep and would give an inaccurate answer to the survey questions. Interestingly, this effect only happened to married men - married women and single women showed no significant difference in their risk of pOSA. To further understand this phenomena requires more research to make a sound speculation.

Of note is that the demographics of pOSA is not representative of OSA in the population. A literature review in 2008 estimated the prevalence of OSA in adults to be between 3% and 7% [14]. While this does not deter the findings of the associated factors with pOSA, it does suggest that pOSA may be too liberal of a definition to be of much use. Future studies using a similar dataset should consider more robust definitions of pOSA and compare how results differ.

Another challenge of the data was missing data for many observations. Even after removing poverty to income ratio, our total sample size dropped from N = 5995 to 5340, about a 10% reduction. Unfortunately, GLM models are unable to handle missing data so there is no simple work-around for dealing with this issue. Future work with this dataset should attempt to perform multiple imputation prior to analysis, allowing for optimal usage of the full dataset.

Conclusion

Using the 2007-2008 NHANES dataset, a nationally representative sample, we were able to identify factors associated with pOSA using ANOVA on a GLM model. We found no association between race and pOSA besides that a generalized "Other" category may be at an increased risk of developing pOSA. In line with other research, we found significant associations between pOSA and hypertension, depression and BMI. When examining pOSA, marital status is a very important factor to include as it acts as an indicator of whether or not the individual is aware of their snoring/snorting during their sleep. Future work could extend this research by handling missing data and including a robustly defined pOSA and comparing results to the current definition of pOSA.

References

[1] S. Ancoli-Israel, M. R. Klauber, C. Stepnowsky, E. Estline, A. Chinn, and R. Fell. Sleep-disordered breathing in african-american elderly. *American journal of respiratory and critical care medicine*, 152(6):1946–1949, 1995.

[2] L. E. Charles, C. M. Burchfiel, D. Fekedulegn, M. E. Andrew, J. M. Violanti, and B. Vila. Obesity and sleep: the buffalo police health study. *Policing: An International Journal of Police Strategies & Management*, 30(2):203–214, 2007.

[3] M. Harris, N. Glozier, R. Ratnavadivel, and R. R. Grunstein. Obstructive sleep apnea and depression. *Sleep medicine reviews*, 13(6):437–444, 2009.

[4] C. Lal, C. Strange, and D. Bachman. Neurocognitive impairment in obstructive sleep apnea. *Chest*, 141(6):1601–1610, 2012.

[5] B. Lam, M. Ip, E. Tench, and C. Ryan. Craniofacial profile in asian and white subjects with obstructive sleep apnoea. *Thorax*, 60(6):504–510, 2005.

[6] W. Lee, S.-A. Lee, Y.-S. Chung, and W. S. Kim. The relation between apnea and depressive symptoms in men with severe obstructive sleep apnea: mediational effects of sleep quality. *Lung*, 193(2):261–267, 2015.

[7] N. Meslier, F. Gagnadoux, P. Giraud, C. Person, H. Ouksel, T. Urban, and J. Racineux. Impaired glucose-insulin metabolism in males with obstructive sleep apnoea syndrome. *European Respiratory Journal*, 22(1):156–160, 2003.

[8] M. Morgenstern, J. Wang, N. Beatty, T. Batemarco, A. L. Sica, and H. Greenberg. Obstructive sleep apnea: an unexpected cause of insulin resistance and diabetes. *Endocrinology and Metabolism Clinics*, 43(1):187–204, 2014.

[9] S. Naismith, V. Winter, H. Gotsopoulos, I. Hickie, and P. Cistulli. Neurobehavioral functioning in obstructive sleep apnea: differential effects of sleep quality, hypoxemia and subjective sleepiness. *Journal of Clinical and Experimental Neuropsychology*, 26(1):43–54, 2004.

[10] T. G. Neilan, H. Farhad, J. A. Dodson, R. V. Shah, S. A. Abbasi, J. P. Bakker, G. F. Michaud, R. van der Geest, R. Blankstein, M. Steigner, 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*, 2(6):e000421, 2013.

[11] C. W. Ong, D. M. ODriscoll, H. Truby, M. T. Naughton, and G. S. Hamilton. The reciprocal interaction between obesity and obstructive sleep apnoea. *Sleep medicine reviews*, 17(2):123–131, 2013.

[12] A. I. Pack. Sleep-disordered breathing: access is the issue, 2004.

[13] S. Pranathiageswaran, M. S. Badr, R. Severson, and J. A. Rowley. The influence of race on the severity of sleep disordered breathing. *Journal of clinical sleep medicine: JCSM: official publication of the American Academy of Sleep Medicine*, 9(4):303, 2013.

[14] N. M. Punjabi. The epidemiology of adult obstructive sleep apnea. *Proceedings of the American Thoracic Society*, 5(2):136–143, 2008.

[15] S. Redline, P. V. Tishler, M. G. Hans, T. D. Tosteson, K. P. Strohl, and K. Spry. Racial differences in sleep-disordered breathing in african-americans and caucasians. *American journal of respiratory and critical care medicine*, 155(1):186–192, 1997.

[16] M. Sands-Lincoln, M. Grandner, J. Whinnery, B. T. Keenan, N. Jackson, and I. Gurubhagavatula. The association between obstructive sleep apnea and hypertension by race/ethnicity in a nationally representative sample. *The Journal of Clinical Hypertension*, 15(8):593–599, 2013.

[17] H. Tuomilehto, J. Seppä, and M. Uusitupa. Obesity and obstructive sleep apnea–clinical significance of weight loss. *Sleep medicine reviews*, 17(5):321–329, 2013.

[18] K. K. Wong, R. R. Grunstein, D. J. Bartlett, and E. Gordon. Brain function in obstructive sleep apnea: results from the brain resource international database. *Journal of integrative neuroscience*, 5(01):111–121, 2006.

[19] W. Xie, S. Chakrabarty, R. Levine, R. Johnson, and J. B. Talmage. Factors associated with obstructive sleep apnea among commercial motor vehicle drivers. *Journal of occupational and environmental medicine*, 53(2):169–173, 2011.