

AHI AT THE US/MEXICO BORDER1

**APNEA-HYPOPNEA INDEX AT THE US/MEXICO BORDER:  
SOCIODEMOGRAPHIC AND HEALTH CORRELATES AND FREQUENT SNORING  
AS A MARKER OF RISK**

By

**NADIA IVANA SARDJEV**

---

A Thesis Submitted to The W.A. Franke Honors College

In Partial Fulfillment of the Bachelor's degree

With Honors in

Physiology and Medical Sciences

**THE UNIVERSITY OF ARIZONA**

**MAY 2026**

### **Abstract**

Despite being one of the most prevalent sleep disorders around the world, most knowledge on sleep apnea is based on studies conducted among clinical populations and people of predominantly non-Hispanic White ethnicity. The US-Mexico border communities, despite exposure to certain environmental, occupational, and social factors relevant to sleep and cardiometabolic outcomes, are understudied in this field. This thesis explores the association between the apnea-hypopnea index (AHI) and sociodemographic/health factors in a community sample of adults in Nogales, Arizona. In particular, this thesis addresses two research questions: which sociodemographic/health factors are associated with AHI in this population, and would frequent snoring affect this association or potentially serve as a low burden marker for predicting the risk of sleep apnea.

Data for this thesis come from the Nogales Cardiometabolic Health and Sleep (NoCHeS) study, which includes 992 adults aged 25-60 who underwent home testing for sleep apnea using the Embletta device and had their test results interpreted using the 3% hypopnea criteria by certified polysomnographic technicians. Information regarding socio-demographics and health behaviors were self-reported, while body mass index was measured objectively. Age, sex, body mass index, education, years in the US, financial status, hypertension, diabetes, and anglo and mexican acculturation according to the Acculturation Rating Scale for Mexican Americans II (ARSMA-II) were considered independent variables. Frequent snoring was defined as snoring on three or more days of the week. Log transformation was applied to AHI to deal with extreme right skewing. The multivariable linear regression analysis and least absolute shrinkage and selection operator (LASSO) regression analysis were employed to establish correlations between

### AHI AT THE US/MEXICO BORDER<sup>3</sup>

selected factors and the outcome of interest and to identify the most parsimonious set of predictors of AHI, respectively. Interactions explored if the effect of each sociodemographic factor on AHI was conditioned upon snoring status.

Overall, the mean AHI of the sample was 6.58 (SD = 9.67; median = 3.60; IQR = 5.90).

Frequent snoring was reported by 22.3% of participants. Being older, male, having higher BMI, being diagnosed with hypertension, and frequent or always snoring were positively associated with the levels of AHI. No association was found between education, financial status, and acculturation and AHI. Interaction terms indicated a dependency between snoring status and AHI among each of the following correlates: the length of stay in the US, education, hypertension, and Mexican acculturation. The LASSO regression included twelve predictors and accounted for 16.8% of the total variance of log AHI, with age, being male, higher BMI, hypertension, and frequent or always snoring as the strongest predictors.

Taken together, this thesis confirms the main risk factors for sleep apnea found in previous literature to hold in the border area. At the same time, their importance varies depending on snoring frequency. Frequent snoring correlated with high AHI overall and modified the relationships between various characteristics and the outcome variable, making it an important factor to consider beyond the symptom status. It also suggests that snoring can be used as a case finding method among border residents, whose access to healthcare facilities is limited.

*Keywords: sleep apnea, apnea-hypopnea index, snoring, Hispanic health, US-Mexico border, acculturation, LASSO regression*

## **Apnea-Hypopnea Index at the US/Mexico Border:**

### **Sociodemographic and Health Correlates and Frequent Snoring as a Marker of Risk**

Sleep is a basic biological necessity. Yet, how individuals sleep is impacted by their environments, occupations, and social structures (Grandner, 2017). This broader context becomes particularly relevant in sleep-disordered breathing due to its influence on the evaluation process, the diagnostic rate, and whether a disorder goes unnoticed throughout a person's lifetime. Current scientific knowledge about sleep apnea is based on samples of patients referred to clinics and populations consisting primarily of non-Hispanic Whites. Little is known about the communities near the US-Mexico border, which feature young working populations, high cardiometabolic disease burden, and restricted access to specialized sleep services. This dissertation analyzes AHI using a community sample from Nogales, Arizona, and explores whether frequent snoring acts as a reliable risk marker for sleep apnea.

### **AHI as the Outcome**

Apnea-Hypopnea Index is defined as the average number of apneas and hypopneas per hour of sleep and is considered as the gold standard quantitative measure of sleep disordered breathing (Mendelson, 2018). An apnea event is defined as a reduction in airflow by at least 90 percent for a minimum duration of 10 seconds. A hypopnea is scored if there is a reduction in airflow by at least 30 percent for a minimum duration of 10 seconds, along with an accompanying oxygen desaturation or a cortical arousal. The 3% threshold for desaturation applied in the current analysis is in line with scoring criteria recommended in the most recent scoring manual by AASM. AHI can be regarded as a count variable. In cases outside the clinical setting, most

individuals will score low values close to zero, whereas fewer individuals will report extremely high values due to a severe disease state.

### **The Border Community Context**

This research consisted of subjects who came from the border community of Santa Cruz County in Arizona. They were selected using the services provided by Mariposa Community Health Center located in Nogales. This community has a binational culture, meaning that its residents have deep-rooted cultural and language associations with Mexico and they are mostly Mexican-Americans. This background is important because it influences how one understands the concept of acculturation differently compared to what they would observe in a larger metropolitan sample. In a multiethnic urban sample, more years of residency in America and more Anglo orientation would mean greater exposure to the institutions of mainstream America. However, in the case of Nogales, Spanish speaking is widespread, bicultural social networks exist, and food culture as well as media consumption is still highly influenced by Mexican culture irrespective of the number of years spent in America.

### **Acculturation as a Bidimensional Construct**

Acculturation in this research has been evaluated using the Acculturation Rating Scale for Mexican Americans-II (ARSMA-II; Cuéllar et al., 1995). The ARSMA-II is a two-dimensional scale, generating a score for the Anglo Orientation Subscale (AOS) and a score for the Mexican Orientation Subscale (MOS). These two dimensions are not endpoints along a unidimensional spectrum, so an individual could be rated high on either subscale, or even both. The items on the test measure preference for English or Spanish language use, frequency of media usage, composition of one's social network, and self-categorization according to culture. This

two-dimensional measurement acknowledges the prevalence of bicultural identity formation in border areas, and provides analysis of Anglo and Mexican orientations separately.

### **Snoring as a Candidate Marker**

Snoring is indicative of the partial blockage of the upper airway passages while sleeping and is the most frequently endorsed complaint in adult patients with sleep apnea. However, the question remains whether snoring can provide more information than the simple bivariate association between snoring and AHI. In the context of a community with limited access to specialized sleep facilities, the presence of a low-cost and low-burden symptom which would help the clinician identify those adults who should undergo more detailed assessment can be considered useful. In this thesis, an approach was taken in which it was assumed that snoring does not only show a correlation with AHI on average but also modifies the relationships between AHI and other factors. When snoring changes the way demographic variables such as age, body mass index, education, migration experience, and acculturation influence AHI, it plays more than just a symptom role; it works as a marker variable.

### **The Present Study**

In the current study, we pursue two complementary objectives. The first is to examine the patterns of AHI in this border population and determine which sociodemographic and health-related factors exhibit the strongest relationship with AHI. In terms of potential predictors, these include age, sex, body mass index (BMI), level of education, duration of living in the U.S., financial situation, hypertension, diabetes, and the two ARSMA-II subscales measuring

## AHI AT THE US/MEXICO BORDER7

acculturation. The second objective is to determine whether frequency of snoring is a moderator of the relationships described in the first part of our research question and can thus serve as an indicator of the corresponding risks. Our research was guided by three specific hypotheses. According to the first hypothesis, classic biologic risk factors such as older age, being male, having high BMI, hypertension, and diabetes are positively associated with AHI. Hypothesis two assumes that social variables – including education, financial situation, history of immigration, and degree of acculturation – do not have universal impacts but may differ based on the specifics of the border population studied in the current paper. Finally, our third hypothesis states that frequency of snoring both strongly correlates with AHI and moderates the effect of several other variables. We employed conventional multivariable regression analysis in combination with a LASSO regression (Tibshirani, 1996).

## **Methods**

### **Study Design and Sample**

The analyses performed in support of this thesis are based on data collected through the NoCHeS study, a community-based research study examining the health status related to cardiometabolic disease and sleep disorders in working age adults living in the community of Nogales, AZ. This study was supported by grants from the National Institutes of Health (R01MD011600 and R01MH135978), and it was conducted in collaboration with Mariposa Community Health Center in Nogales, which served as the recruitment center and primary clinical contact site. Participants were recruited via clinic outreach, referral from community health workers, and advertising in the community. Inclusion criteria consisted of living in the area, being between the ages of 25 and 60 years old, and the ability to complete an in-home sleep monitoring study. These age parameters were selected because sleep apnea is highly prevalent among working age adults and tends to be under-diagnosed in primary care settings.

All procedures associated with this study were approved by the University of Arizona Institutional Review Board. Written informed consent was obtained from all participants, with materials and consent forms available in English and Spanish. Study staff conducted all participant contact, allowing participants to use either English or Spanish in the consent process. As is standard practice in community-based research studies, reimbursement was provided to participants.

The current analysis focused on an analytic sample of 992 individuals with complete data regarding the sleep testing and demographic/health factors relevant for modeling in the regression analyses. Some individuals in the parent study sample were omitted from the current

sample due to missing AHI scores and/or biologically implausible values detected as part of data cleaning, as discussed further in Data Cleaning below. The mean age of this sample was 43.1 years (SD=10.6), and 21.5% were male. The vast majority of the sample was adults of Mexican heritage, reflecting the demographics of the Nogales area. Additional demographic details are presented in Table 1.

### **Sleep Assessment**

The presence of sleep disordered breathing was assessed by means of an Embletta MPR home sleep test monitor (Natus Medical, Inc., Pleasanton, California). Emblettas are classified as Type III portable monitors. Embletta measures airflow via a pressure transducer cannula, respiratory effort via chest and abdominal belts, oxygen saturation and pulse rate via a fingertip oximeter, body position, and snoring. In contrast to polysomnographs, Type III monitors do not capture information about the electroencephalogram, which implies that sleep onset and arousals cannot be scored directly, although episodes of apneas and hypopneas can be scored easily based on the airflow, respiratory effort, and oxygen channels.

The devices were distributed to participants by Mariposa Community Health Center study staff. In-person training of participants included instructions on proper placement and care of the device during the night of recording, as well as take-home instructions. Recording took place on a regular night of sleep at the participant's home. In the morning, the participants returned the monitor to Mariposa Community Health Center, where study staff checked the recording quality and scheduled a follow-up test if deemed necessary.

Sleep monitoring data was scored for all participants by certified polysomnographic technologists at the University of Arizona Sleep Center according to the guidelines in the latest edition of the AASM manual (Berry et al., 2018), including the 3% hypopnea criterion. Apneas were scored when the airflow drops by 90% or more for more than 10 seconds. Hypopneas were scored if there is an airflow reduction of 30% or more and at least a 3% drop in oxygen saturation. Arousal-based hypopneas scoring was not possible due to the absence of electroencephalography. The AHI was computed as a sum of the number of apneas and hypopneas divided by the total recording duration. Polysomnographic technologists were blind to participant demographics and medical history.

### **Sociodemographic and Health Variables**

Sociodemographic and health factors were obtained from a standardized interview carried out by bilingual research assistants, as well as objective measures collected during the time of recruitment.

### ***Demographic and Socioeconomic Characteristics***

Age was reported in years. Sex was reported as either male or female according to self-report. Educational level was measured using a four-point scale consisting of less than high school, high school graduate, some college, and college graduate, with college graduate used as the reference category for analyses. Years lived in the U.S. was classified into 10 years or less, 11 to 15 years, 16 to 20 years, and over 20 years, with the latter two groups combined and 10 years or less used as the reference category. In contrast to simply measuring whether people were born in the U.S., this measure sought to capture information about migration background at the family level; it is

important to note that for most border dwellers, their families had been tied to both Mexico and the U.S. for generations. Financial condition was reported using a four-point perceived social class scale consisting of poor, lower middle, middle, and upper middle, with middle used as the reference group.

### ***Body Measurements and Cardiometabolic Conditions***

Height and body weight were recorded using a stadiometer and calibrated scales, respectively. Body Mass Index (BMI) was computed as body weight in kilograms divided by height in meters squared and modeled continuously in units of kg/m<sup>2</sup>. Hypertension was classified by either of the following: reporting a diagnosis of hypertension, use of hypertensive drugs, or blood pressure measurements consistent with hypertension thresholds on the day of the study visit. Diabetes mellitus was classified similarly by either of the following: reporting a diagnosis of diabetes, use of diabetic drugs, or hemoglobin A1C measurements that met criteria for diabetes on the day of the study visit. The combined definition ensured that people whose clinicians diagnosed hypertension or diabetes and who are not under treatment would not be missed.

### ***Acculturation***

The acculturation level was measured using the Acculturation Rating Scale for Mexican Americans-II (ARSMA-II; Cuellar et al., 1995). The ARSMA-II is an extensively used two-dimensional scale that generates two subscales: Anglo Orientation Subscale (AOS) and Mexican Orientation Subscale (MOS). These subscales include questions regarding language use, media exposure, social networks, and cultural identification. A score for each dimension can be independently calculated, permitting the identification of Anglo-oriented, Mexican-oriented,

bicultural (with high scores on both), and marginal (with low scores on both) individuals. In the current study, AOS and MOS were regarded as continuous variables in the regression analyses.

### ***Snoring***

Snoring frequency was measured by self-report using a five-point question that inquired into the number of days per week the individual experienced snoring. The options were never, rarely (less than one day per week), sometimes (one to two days per week), often (three to four days per week), and always (five to seven days per week). For interaction analysis, frequent snoring was considered as three or more days per week, including both the often and always categories. The criterion of three days per week was selected since snoring at this frequency would draw attention from a bed partner and thus allow him/her to accurately report the event.

### **Data Cleaning and Relationship to Earlier Reported Values**

Two problems emerged when the distribution of AHI was first observed. First, a few AHI values appeared unrealistic and were marked as erroneous entries for further verification. Second, a few records included an AHI of zero that after verification turned out to represent missing data rather than actual AHI data without any events detected. Cases that satisfied any of the above criteria underwent additional scrutiny and revision by the senior investigator along with the scorer team, with either corrections made or missing value status assigned to cases where indicated.

The abstract used for presentation at the joint SLEEP and AADSM conferences was based on the first pass analysis of data. Statistics, tables and figures presented in the current thesis are based on the cleaned dataset, including the log-transformation process that will be described later.

Because of this, some of the descriptive statistics reported herein diverge from those presented in

the conference abstract. The most obvious discrepancy between statistics relates to the AHI mean: 3.6 for the former vs. 6.58 with a median of 3.60 for the latter. Both represent correct statistical description of their corresponding sets. However, the thesis figures should be considered the primary statistics, as per the common rule of thumb that conference abstracts are only preliminary results.

### **Statistical Analysis**

Descriptive statistics were computed for AHI and all candidate predictors. Continuous variables are presented as mean (standard deviation); categorical variables are presented as n (%). The distributional characteristics of AHI were assessed visually with histograms and numerically with skewness and kurtosis. The untransformed distribution of AHI demonstrated considerable right skewness (skewness = 3.48) and kurtosis (kurtosis = 18.38), with most cases having a very low number of apneas and hypopneas per hour and a long right tail consisting of individuals with extreme disease. The original conference abstract assumed linearity in AHI and thus modeled it directly as a continuous variable. Although common practice, this approach allows straightforward interpretation of event counts but fails to account for skew in AHI's distribution. Moreover, such a highly skewed distribution can lead to inflated regression coefficients in the presence of only a few outliers with very high AHI values. The follow-up analysis employed a log transformation of AHI, which is appropriate when modeling right-skewed data with a lower bound of zero. AHI was transformed by taking the natural logarithm of AHI plus one ( $\log[\text{AHI} + 1]$ ). The addition of one was necessary because AHI is equal to zero for some cases, and the standard log transformation is undefined at zero. Following log transformation, the distribution became approximately normal (skewness = 0.41; kurtosis = 2.85). The coefficients from models using this transformed dependent variable should be interpreted as approximate proportional

changes in AHI plus one rather than changes in events per hour. The transformation sacrifices some interpretability for clinicians but enhances estimation accuracy.

All statistical analyses were carried out using Stata version 18 (StataCorp, College Station, Texas). Linear multivariable regression models were fit to estimate the associations between each potential predictor and log AHI, both unadjusted and adjusted for age and sex. Regression coefficients are reported as B coefficients (unstandardized regression coefficients) with 95% confidence intervals and exact p-values (two-tailed). Where appropriate, parallel models on the untransformed AHI scale were fit to aid interpretation. Selected results are also presented as events per hour to facilitate the translation of findings to practice.

Interactions examined the relationships between frequent snoring and each demographic correlate, after adjustment for age, sex, and BMI. Each interaction was tested by adding a product term between snoring and the correlate of interest to a model containing age, sex, BMI, and the main effects of each predictor. Statistically significant interactions were followed up with stratified analyses and graphics depicting raw AHI plotted against the predictor of interest for each category of snoring frequency. The interaction plots used the five-level snoring variable rather than the binary frequent snoring indicator for greater precision. However, the interaction tests themselves employed the binary indicator, allowing for simpler hypothesis tests.

In addition to ordinary regression techniques, the competing predictors were evaluated using least absolute shrinkage and selection operator (LASSO) regression. LASSO regression is a type of penalized regression, which adds a penalty proportional to the sum of the absolute value of the coefficients to the loss function. As a result, redundant coefficients are shrunk towards zero, and predictors are automatically selected (Tibshirani, 1996). The tuning parameter lambda ( $\lambda$ ) was

estimated by tenfold cross-validation with respect to prediction error. Candidate predictors with nonzero coefficients at the optimal value of  $\lambda$  were included in a follow-up ordinary regression analysis that yielded standard inferential statistics. The resulting predictors and coefficients are listed in Table 4.

Inferential statistics are reported as B coefficients (unstandardized regression coefficients) with 95% confidence intervals and exact p-values. Variance explained is reported as  $R^2$  for the final unpenalized linear regression model. Formal power analyses were not performed, since the sample size was constrained by the parent study design. However, the sample size of 992 was sufficient to detect small to moderate associations between the candidate predictors and AHI in linear regression models.

## Results

### Sample Characteristics

Descriptive characteristics of the sample used in the analysis are shown in Table 1. The average age of the 992 participants in the sample was 43.1 years (SD = 10.6), and the average body mass index (BMI) was 31.1 kg/m<sup>2</sup> (SD = 6.0), making them fall into the obese range according to the criteria set by the WHO. Twenty-one and one-half percent of the sample considered themselves male. In terms of their educational level, the respondents spanned the entire spectrum, with 17.5% having graduated from college, 35.7% being college-educated without finishing, 31.2% completing high school, and 15.6% having less than a high-school education. Concerning the length of time spent in the US, the majority of them had lived in this country for at least twenty years; indeed, 64.2% of the respondents reported that their families had been living in the US for over twenty years, 10.4% for sixteen to twenty years, 7.3% for eleven to fifteen years, and 18.2% for ten years or less. Most of them classified themselves as belonging to the middle-class social stratum, with 54.4% identifying as such. At the same time, 32.4% identified themselves as coming from the lower middle class, 8.0% as poor, and 5.3% as belonging to the upper middle class.

The cardiometabolic conditions were widespread among the study population. The prevalence of hypertension and diabetes was 19.8% and 10.8%, respectively. The average score of the Anglo Orientation Subscale of the ARSMA-II was 1.98 (SD = 0.94) and of the Mexican Orientation Subscale – 3.20 (SD = 0.58), indicating the bicultural nature of the sample which was mostly of Mexican orientation. As for snoring, 36.1% never snored, 25.1% snored rarely (once a week or less), 16.5% did it somewhat regularly (one to two nights per week), 11.1% did it fairly often

(three to four nights per week), and 11.3% always snored (five to seven nights per week).

Overall, 22.3% snored frequently.

### **Distribution of AHI**

The raw distribution and log-transformed distributions of AHI are shown in Table 2 and illustrated in Figure 1. The mean raw value of AHI was 6.58 events per hour (SD = 9.67), with a median value of 3.60 (IQR = 5.90), and a range extending from 0.0 to 89.2 events per hour. The distribution was extremely right-skewed (skewness = 3.48) and very leptokurtic (kurtosis = 18.38). Most participants had relatively low values, while fewer participants had much higher values indicative of more serious disease. Based on conventional clinical thresholds, a significant proportion of participants would have an AHI of 5 or greater (indicating mild OSA or worse, based on a raw threshold of 5 or higher events per hour), and some would be classified as having moderate to severe disease based on higher thresholds.

Following log transformation, the resulting  $\log(\text{AHI} + 1)$  variable had an approximately normal distribution, with mean = 1.55 (SD = 0.93), median = 1.53 (IQR = 1.24), skewness = 0.41, and kurtosis = 2.85. Log-transformation of the data allowed for proper use of linear regression for the subsequent analysis, and the results below are all based on  $\log(\text{AHI} + 1)$ .

### **Sociodemographic and Health Correlates of AHI**

Regression results for  $\log(\text{AHI} + 1)$  are presented in Table 3, both unadjusted and adjusted for age and sex. Older age was significantly associated with higher AHI (B = 0.0153 per year, 95% CI [0.0101, 0.0206],  $p < .001$  adjusted), corresponding to roughly a 1.5% proportional increase in AHI plus one for each additional year of life. Male sex was a strong correlate of

higher AHI ( $B = 0.4711$ , 95% CI [0.3359, 0.6062],  $p < .001$  adjusted), reflecting an approximately 60% higher AHI plus one for men compared to women. Higher BMI was significantly associated with higher AHI ( $B = 0.0338$  per  $\text{kg}/\text{m}^2$ , 95% CI [0.0246, 0.0431],  $p < .001$  adjusted).

Hypertension was associated with substantially higher AHI in unadjusted analyses ( $B = 0.5509$ , 95% CI [0.4065, 0.6953],  $p < .001$ ) and remained significant after adjustment for age and sex ( $B = 0.2998$ , 95% CI [0.1485, 0.4511],  $p < .001$ ). Diabetes was similarly associated with higher AHI ( $B = 0.4511$ , 95% CI [0.2624, 0.6397],  $p < .001$  unadjusted;  $B = 0.2160$ , 95% CI [0.0305, 0.4015],  $p = .023$  adjusted). The smaller adjusted coefficient for diabetes indicates that part of its bivariate association with AHI was carried by age and sex.

Education and financial status were not significantly associated with AHI in either unadjusted or age- and sex-adjusted models. None of the four education categories differed significantly from the college-graduate reference (all  $p > .50$ ), and none of the three non-middle-class financial categories differed significantly from middle class (all  $p > .12$ ). Length of time in the United States was likewise not significantly associated with AHI in either model, and the point estimates were small in magnitude. Anglo and Mexican acculturation also showed no significant main effects on AHI in this sample (all  $p > .13$ ).

Snoring showed a clear gradient with AHI in the unadjusted analysis. Compared with participants who reported never snoring, those who reported sometimes snoring had a higher AHI ( $B = 0.3281$ , 95% CI [0.1526, 0.5037],  $p < .001$  unadjusted), and those in the frequently and always snoring categories had substantially higher AHI (frequently:  $B = 0.6875$ , 95% CI [0.4860, 0.8890],  $p < .001$ ; always:  $B = 0.6864$ , 95% CI [0.4817, 0.8911],  $p < .001$ ). After adjustment for age and sex, the rarely and sometimes categories no longer differed significantly

from never, while the frequently and always categories remained robust (frequently:  $B = 0.4035$ , 95% CI [0.1987, 0.6083],  $p < .001$ ; always:  $B = 0.3893$ , 95% CI [0.1795, 0.5991],  $p < .001$ ). On the raw AHI scale, frequent snoring (combining frequently and always) was associated with an increase of approximately 4.3 events per hour relative to never snoring after adjustment, consistent with the magnitude reported in the original conference abstract.

### **Interactions with Snoring**

Interaction terms tested whether the relationship between AHI and each demographic correlate depended on snoring frequency, adjusting for all other covariates. Four specific dependencies were observed in the cleaned re-analysis underlying the present findings: the association between AHI and length of time in the US, education, hypertension, and Mexican acculturation each varied significantly according to snoring category. These interactions are depicted in Figure 2, with AHI plotted separately for each predictor according to snoring frequency.

For the interaction between length of time in the US, there was little variation in the association between immigration and AHI in non-rare or infrequent snorers, but a clear effect emerged in those with higher snoring frequencies, specifically those in the ‘frequent’ and ‘always’ categories. In frequent snorers, there was a positive association between years since arrival in the US and AHI – that is, individuals whose families had been in the US for the longest time periods exhibited higher AHI levels than individuals at the short end of the spectrum. Within the always snorers category, there was some evidence of an inverse relationship between length of time in the US and AHI. While the mechanisms involved cannot be inferred from the data at hand, the above findings indicate that snoring acts as a proxy for compromised upper airways that varies across environmental exposures to US conditions.

The interaction with education level indicated a lack of association in participants who never snored, an inverse relationship in rare and frequent snorers, and a positive relationship among always snorers. This pattern suggests that, within the studied population, the well-known educational gradient in risk for obstructive sleep apnea differs across snoring frequencies. One possible explanation is that the relationship between social gradients and AHI breaks down in highly snorers because their condition represents a near-certainty in upper airway disease.

Regarding the interaction with hypertension, the positive association between the latter and AHI was stronger in participants who snored sometimes, rarely, or frequently, and weaker among participants who snored never or always. The pattern indicates that, in the former two categories of snoring, hypertension may be a marker of cardiometabolic disease processes that are independent from upper airway disease – the pathway leading to the latter does not involve the former. Meanwhile, the inverse relationship between the two variables among always-snoring participants likely results from airways obstruction having advanced to such an extent that even hypertensive participants exhibit only minimally elevated AHI levels.

Finally, regarding the interaction with Mexican acculturation, we find that this factor exhibited very weak relationships with AHI across almost all snoring frequencies. There was, however, a strong negative correlation between Mexican acculturation and AHI within participants who snored frequently. We also noted a similar, but less pronounced correlation between anglo acculturation and AHI, although not within frequent snorers. These correlations are smaller than the BMI or sex effects, but consistent with previous work suggesting that the effects of acculturation may be more prominent among sub-groups than across entire populations, defined either by symptoms or other risk factors. The above findings suggest a protective role of

Mexican cultural ties among frequent snorers, although further research is needed to establish causality.

We also observed interaction terms involving financial status, diabetes, and Anglo acculturation, but none met the criteria for a dependency in the cleaned re-analysis used for the current study, in contrast to what was observed in the first-pass analysis presented in the conference abstract.

### **LASSO Selection of Parsimonious Predictors**

Results from LASSO variable selection are shown in Table 4 and Figure 3. The minimum  $\lambda$  via cross-validation was 0.022, which included 12 predictor variables in the model. At the cross-validation minimum, the mean squared error was 0.711, and the model explained 16.8% of the variance in the outcome  $\log(\text{AHI} + 1)$  ( $R^2 = .1676$ ). In terms of the raw scale, the multiple regression model that accounted for the LASSO-penalized predictors explained approximately 17.6% of the variance in the raw AHI.

The predictors retained by LASSO were age, male, BMI, hypertension, diabetes, length of stay in the U.S. (with more than one level retained), financial status (with more than one level retained), Anglo acculturation, Mexican acculturation, and snoring (at rarely, sometimes, frequently, and always levels). Once an unpenalized linear regression was run on the LASSO-retained set, age ( $B = 0.0092$ , 95% CI [0.0028, 0.0155],  $p = .005$ ), male ( $B = 0.3576$ , 95% CI [0.2144, 0.5009],  $p < .001$ ), BMI ( $B = 0.0223$ , 95% CI [0.0122, 0.0325],  $p < .001$ ), hypertension ( $B = 0.2473$ , 95% CI [0.0847, 0.4100],  $p = .003$ ), frequent snoring ( $B = 0.3654$ , 95% CI [0.1604, 0.5704],  $p < .001$ ), and always snoring ( $B = 0.3462$ , 95% CI [0.1348, 0.5576],  $p = .001$ ) became significant predictors of AHI. While diabetes had approached significance here, it had not attained significance in this model ( $B = 0.1875$ , 95% CI [-0.0106, 0.3856],  $p = .064$ ).

Length of stay in the U.S., financial status, and acculturation variables were retained by LASSO, but they failed to attain statistical significance in the unpenalized regression analysis. The fact that they were retained means that they were explaining some nontrivial variance even though their coefficients in the regression analysis remained relatively small.

### **Summary of Findings**

In summary, there were three major trends found across all analyses. First, known risk factors for obstructive sleep apnea, such as advanced age, being male, increased body mass index, hypertension, and snoring were all positively associated with increased AHI in the study population, after adjustment for covariates. These associations conformed to expectations regarding their magnitude and consistency. Second, the commonly analyzed factors in studies on health in Hispanics and Latinos, such as education level, socioeconomic status, duration of stay in the US, and acculturation, did not appear to have an unambiguous main effect on AHI in the current study. Finally, many of the variables mentioned above were also observed to interact significantly with snoring frequency, which suggests that their association with AHI was conditional upon the frequency of snoring.

## **Discussion**

This thesis investigated apnea-hypopnea index (AHI) and related sociodemographic and health factors among 992 adults of working age living near the US-Mexico border. Three recurring themes emerged in the analysis results. The first is that each of the usual suspects for apnea risk – advanced age, male sex, obesity, hypertension, and frequent snoring – had a unique positive association with AHI in this border sample. The second is that each of the contextual or social factors commonly investigated in Hispanic and Latino health research – years of education, financial status, length of time in the United States, and acculturation – did not have a statistically meaningful main-effect relationship with AHI. The third is that specific dependencies on frequent snoring appeared in the interaction analyses: the relationship between AHI and years in the United States, years of education, hypertension, and Mexican orientation each depended on the respondent being a snorer.

### **Conventional Risk Factors at the Border**

All of the usual apnea risk factors studied here had a relationship with AHI as expected from past studies. The age coefficient increased AHI by a small amount for each additional year. Male sex was strongly associated with increased AHI, even though men represented only 21.5% of the sample. Body mass index had a positive effect on AHI per unit of weight-for-height. The mean value of 31.05 made the average participant obese, although most were only overweight. Hypertension was positively associated with AHI in unadjusted analysis, and it remained a robust predictor in the final LASSO model. All of these cardiometabolic markers of health worked as expected in this border sample, which is interesting in its own right because it

suggests that a standard risk profile applies to a population whose health needs have rarely been investigated.

### **Why Social Variables Did Not Show Main Effects**

There are several reasons why education, financial status, years since immigration, and acculturation were not independently associated with AHI in the sample. First, the great majority of participants had resided in the United States for over 20 years, so the variable is essentially capped at 20. Second, education had a full range, but the cardiometabolic risk load is already high in the sample, leaving little variance to be explained by socioeconomic differences. Third, the sample is intentionally acculturated biculturally, with high scores for Mexican orientation ( $M=3.20$ ,  $SD=.58$ ) and moderate scores for Anglo orientation ( $M=1.98$ ,  $SD=.94$ ), representing the realities of a border community where the cultural and linguistic environment is rich in either direction regardless of generational depth. In such a community, acculturation differences that have affected health outcomes elsewhere cannot be expected to matter much.

It would be inappropriate to assume that acculturation, education, or immigration history are therefore irrelevant to sleep apnea risk in this population. There are specific dependencies on snoring, as discussed in subsequent sections. Also, despite their nonsignificant main-effect associations with AHI, several of these factors were retained in the LASSO analysis.

### **Snoring as a Marker**

One of the most striking aspects of the analysis is how the predictor that snored frequently was found not only to be a strong correlate of AHI in its own right but also to modulate the relationships between four other predictors and AHI. Frequent snoring was present in 22.3% of

the sample, and it had a significantly larger mean AHI value compared to the sample in both unadjusted and adjusted models. The interaction analyses asked whether the relationship between AHI and each demographic correlate was dependent on whether participants snored and yielded four dependencies where the answer was yes: the relationship with AHI differed according to snoring status in cases of length of time in the United States, education, hypertension, and Mexican acculturation.

As far as the implications of the dependencies go, these can be considered fairly straightforward as well. In the case of the United States' residency, the relationship between AHI and years spent there was negligible or non-existent among nonsnorers and rare snorers, but became much more variable among those with more frequent snoring, meaning that snoring marks the subgroup among which environmental impact on AHI can be seen the most vividly. In terms of education, the relationship went through the usual positive gradient, but was reversed among the always snorers, showing that the relationship between higher education and lower likelihood of OSA does not apply to this population in the same manner as others. Among people with hypertension, the relationship with AHI was most evident among intermediate snoring groups and least evident among the extremes. Finally, in case of Mexican acculturation, the orientation showed little relationship to AHI in all snoring classes save the one where participants snored often, in which case higher orientation was associated with lower AHI values. This finding can be seen as part of the larger trend in the data set showing that acculturation plays some role in the AHI relationship in ways not accounted for by main effects models.

Overall, the results show that, rather than just being a symptom of the disease, snoring in this particular population acts as a marker, helping identify the subgroup among which further analysis can prove fruitful.

### **Methodological Considerations**

A couple points should be made about the methodology of the study. First, log transformation of AHI plus one was employed because AHI in raw form was found to be significantly skewed (3.48) and kurtotic (18.38), violating the conditions of linear regression. This transformation resulted in an approximately normally distributed measure, reducing the influence of outliers, making it more accurate for estimation of effect sizes. The downside of it is that regression coefficients are not easily interpretable as additive changes, but log transformation more accurately reflects the relationship between predictor and response variable.

The LASSO regression was used to select the optimal number of predictors from the initial candidate pool (Tibshirani, 1996). LASSO regression does not test hypotheses; it seeks out predictors that add predictive weight to the model. The twelve variables selected in the process included age, sex, BMI, hypertension, diabetes, United States' residency (with several levels preserved), financial status (several levels preserved), Anglo and Mexican acculturation, and four snoring classes. Unpenalized analysis revealed six of the predictors that were independently significant for AHI: age, sex, BMI, hypertension, frequent snoring, and constant snoring.

Variables that remained part of the set yet failed to reach statistical significance include length of residency, financial status, and acculturation. These should not be excluded from the analysis as they contributed predictive power, perhaps through interaction or in conjunction with other predictors, which was shown in the interaction analyses.

### **Position Within the Broader NoCHeS Work**

The analyses presented above occurred concurrently with two separate projects analyzing the same data set. The first analyzed which characteristics could predict underdiagnosis of OSA among patients meeting AHI criteria for clinical diagnosis and what proportion of cases is left undiagnosed. Preliminary results shared by the lab indicated that the overwhelming majority of AHI-positive cases remain undiagnosed and that Mexican acculturation increases odds of underdiagnosis. The second effort attempted to analyze whether AHI can be predicted by different sets of variables depending on AHI category (minimal, mild, moderate, and severe AHI) and found that the optimal set varies depending on severity, with Mexican acculturation showing a stronger effect among low but still clinically relevant levels of AHI.

The current analyses stand as the basis for the two analyses discussed above. They establish the nature of correlations between sociodemographic and health-related characteristics and AHI and the nature of modification of said relationships by snoring. Diagnostic gap analysis represents another approach to the same data and attempts to discover what factors lead to underdiagnosis of OSA in this community. Severity stratification analysis attempts to find the predictors of AHI severity level.

### **Limitations**

There are several limitations associated with this study. First, since it has been designed as cross-sectional, the temporal relationship between the factors examined cannot be established. There are likely to be bi-directional interactions between AHI, hypertension, diabetes, and BMI; therefore, these analyses do not allow determining which variable preceded the other one.

Second, sleep-disordered breathing was assessed with a Type III home sleep apnea test rather than with full polysomnography. This type of device calculates total recording time but not the total sleep time; therefore, it cannot record arousal-based hypopneas without electroencephalography. In addition, it should be noted that each participant only completed a single night of recording, and this can have led to some amount of individual misclassification due to night-to-night variability.

Third, several factors included in the study such as snoring were self-reported. Self-reporting can lead to errors in the information provided; for example, it depends on the ability of participants to recall the necessary information. The use of the five-level snoring question did not allow to assess the intensity and positioning dependence of snoring as well as partner reports about the severity of snoring. The acculturation was assessed using the validated ARSMA-II scale, but it is a complex construct, and the biculturalism of the studied sample might limit variability on both scales.

Finally, the sample represented a single geographic location within the United States (single county and community health center catchment area). While the number of participants involved in this research can be considered relatively high for a community-based sample from the border,

the results cannot be generalized to any other community or Mexican American population. High representation of women limits conclusions about sex-related differences in the data.

The data presented in this abstract was analyzed on the basis of the preliminary data set. Cleaned data and updated analyses used in this thesis lead to different descriptive statistics, including mean AHI of 6.58 rather than the 3.6 reported in the abstract.

### **Future Directions**

There are several avenues for further research based on the results presented in this paper. One direction involves the ongoing work on the diagnostic gap between OSA cases identified by objective AHI criteria and cases diagnosed in the clinical realm. In preliminary findings from the same dataset, it appears that there were numerous underdiagnosed cases of OSA among the study population, and a higher degree of Mexican acculturation was associated with underdiagnosis. A deeper exploration of the pathway by which an adult is referred, evaluated, and diagnosed for OSA would enable the pinpointing of potential diagnostic gaps in this region.

The second line of inquiry involves stratified modeling of AHI across severity levels. Although different sets of predictors seem to be useful for different severities of the condition, the optimal variables for predicting a minimal but clinically significant amount of apneas and hypopneas may not be the same as those predicting the severe forms of the disorder. Stratification of the predictive model would provide additional insights into the usefulness of snoring as a predictor of specific severities of the condition.

The cross-sectional design of the parent study makes causal inferences about the direction of the association impossible. To address this issue, longitudinal analyses should be conducted using the NoCHeS dataset. For example, it would be interesting to examine the correlation of AHI over time with changes in weight, newly emerging cardiometabolic disorders, and clinical care. It would also be worthwhile to conduct a follow-up after a certain period and explore the predictive ability of snoring for the emergence of clinically recognized sleep apnea in currently mild cases.

Finally, it is worth investigating the relative importance of different components of acculturation as defined by ARSMA-II. As demonstrated above, the interactions between snoring and certain elements of acculturation seem to exist. In order to gain clarity about the aspects of cultural orientation that are especially relevant for sleep apnea risk, it would be helpful to include in the predictive models some subdomains of ARSMA-II, as well as dietary and occupational factors, along with language of clinical care.

## **Conclusion**

This thesis presented the apnea-hypopnea index in 992 working-age adults from a border region in the United States and Mexico along with the main predictors and moderators associated with the condition. Traditional risk factors for sleep apnea, including older age, being male, body mass index, hypertension, and frequent snoring, have effects on this population along the border that align with the sleep apnea literature. Neither social nor contextual factors, including education, socioeconomic position, immigration, and acculturation, were significant main-effect predictors of AHI, but many interacted with frequent snoring in terms of predicting AHI, suggesting a conditional rather than generalizable effect. Frequent snoring not only predicted AHI independently but also moderated the association between AHI and other factors. This result highlights the value of frequent snoring as a simple and easy-to-obtain measure of risk. The results of the current analysis provide the basis for further work in this dataset regarding diagnostic issues and severity-based prediction, and combined these two studies suggest that sleep apnea in this border population is driven by biological risk factors, enhanced by snoring, and confounded by social factors.

### References

- Berry, R. B., Brooks, R., Gamaldo, C., Harding, S. M., Lloyd, R. M., Quan, S. F., Troester, M. T., & Vaughn, B. V. (2018). The AASM manual for the scoring of sleep and associated events: Rules, terminology and technical specifications (Version 2.5). American Academy of Sleep Medicine.
- Cuéllar, I., Arnold, B., & Maldonado, R. (1995). Acculturation Rating Scale for Mexican Americans–II: A revision of the original ARSMA scale. *Hispanic Journal of Behavioral Sciences*, 17(3), 275–304. <https://doi.org/10.1177/07399863950173001>
- Grandner, M. A. (2017). Sleep, health, and society. *Sleep Medicine Clinics*, 12(1), 1–22. <https://doi.org/10.1016/j.jsmc.2016.10.012>
- Mendelson, W. B. (2018). *The science of sleep: What it is, how it works, and why it matters*. University of Chicago Press.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>

**Table 1**

*Sociodemographic and Health Characteristics of the Analytic Sample (N = 992)*

<b>Variable</b>	<b>Category / Units</b>	<b>M (SD) or %</b>
Age	Years	43.09 (10.63)
Sex	Male	21.49%
Body mass index	kg/m <sup>2</sup>	31.05 (5.98)
Education	College graduate	17.46%
	Some college	35.72%
	High school	31.18%
	Less than high school	15.64%
Length of time in U.S.	10 years or fewer	18.16%
	11–15 years	7.27%
	16–20 years	10.39%
	More than 20 years	64.18%
Financial status	Middle class	54.39%
	Poor	7.97%
	Lower middle class	32.39%
	Upper middle class	5.25%
Hypertension	Yes	19.78%
Diabetes	Yes	10.80%
Acculturation	Anglo Orientation Scale (AOS)	1.98 (0.94)
	Mexican Orientation Scale (MOS)	3.20 (0.58)
Snoring frequency	Never	36.13%
	Rarely (< 1 night/week)	25.08%
	Sometimes (1–2 nights/week)	16.46%
	Frequently (3–4 nights/week)	11.05%
	Always (5–7 nights/week)	11.27%

*Note.* Continuous variables are presented as M (SD); categorical variables as percentages. Acculturation was assessed with the Acculturation Rating Scale for Mexican Americans–II (Cuellar et al., 1995).

**Table 2***Distribution of Apnea-Hypopnea Index (AHI) and Log-Transformed AHI*

<b>Statistic</b>	<b>AHI (events/hour)</b>	<b>Log(AHI + 1)</b>
Mean	6.58	1.55
Standard deviation	9.67	0.93
Median	3.60	1.53
Interquartile range	5.90	1.24
Skewness	3.48	0.41
Kurtosis	18.38	2.85

*Note.* AHI = apnea-hypopnea index, expressed as events per hour. The log transformation was applied as  $\ln(\text{AHI} + 1)$  to reduce right skew while preserving zero values.

**Table 3**

*Linear Regression of Log(AHI + 1) on Sociodemographic and Health Characteristics*

<b>Variable</b>	<b>Category</b>	<b>B</b>	<b>95% CI</b>	<b>p</b>
	<i>Age- and sex-adjusted models</i>			
Age	per year	0.0153	[0.0101, 0.0206]	< .001
Sex	Male	0.4711	[0.3359, 0.6062]	< .001
Body mass index	per kg/m <sup>2</sup>	0.0338	[0.0246, 0.0431]	< .001
Education	Some college	0.0217	[-0.1414, 0.1848]	.794
	High school	-0.0126	[-0.1820, 0.1568]	.884
	< High school	-0.0084	[-0.2074, 0.1906]	.934
Time in U.S.	11–15 years	-0.1977	[-0.4458, 0.0505]	.118
	16–20 years	0.0087	[-0.2079, 0.2253]	.937
	> 20 years	-0.0796	[-0.2318, 0.0725]	.305
Financial status	Poor	0.1704	[-0.0478, 0.3887]	.126
	Lower middle	0.0127	[-0.1106, 0.1360]	.840
	Upper middle	0.0172	[-0.2373, 0.2718]	.894
Hypertension	Yes	0.2998	[0.1485, 0.4511]	< .001
Diabetes	Yes	0.2160	[0.0305, 0.4015]	.023
Acculturation	Anglo (AOS)	-0.0161	[-0.0765, 0.0442]	.600
	Mexican (MOS)	-0.0435	[-0.1429, 0.0558]	.390
Snoring	Rarely	0.0565	[-0.0942, 0.2071]	.462
	Sometimes	0.0879	[-0.0898, 0.2656]	.332
	Frequently	0.4035	[0.1987, 0.6083]	< .001
	Always	0.3893	[0.1795, 0.5991]	< .001

*Note.* All regression models adjusted for age and sex. B = unstandardized regression coefficient on the log(AHI + 1) scale. Reference categories are college graduate (education), 10 years or fewer (time in U.S.), middle class (financial status), and never (snoring).

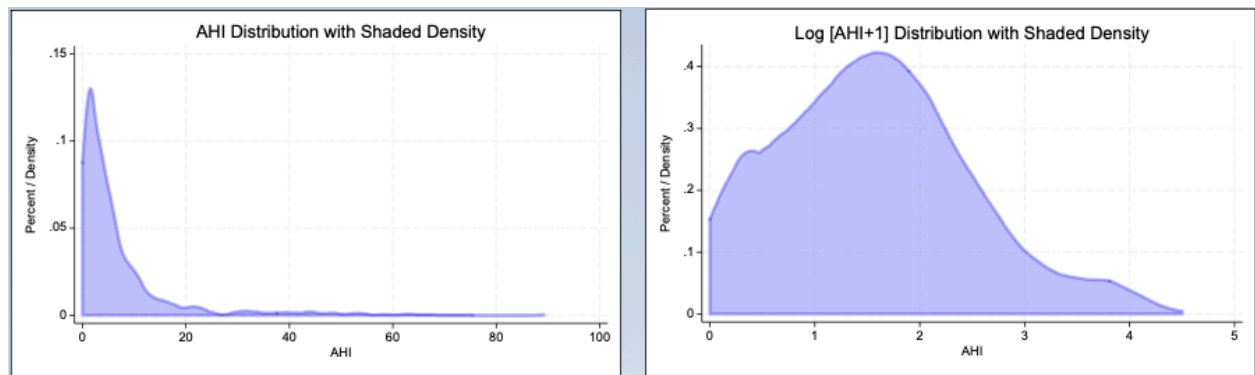
**Table 4***Final Linear Regression Model Following LASSO Variable Selection*

<b>Variable</b>	<b>Category</b>	<b>B</b>	<b>95% CI</b>	<b>p</b>
Age	per year	0.0092	[0.0028, 0.0155]	.005
Sex	Male	0.3576	[0.2144, 0.5009]	< .001
Body mass index	per kg/m <sup>2</sup>	0.0223	[0.0122, 0.0325]	< .001
Time in U.S.	11–15 years	-0.1856	[-0.4409, 0.0698]	.154
	16–20 years	0.0129	[-0.2123, 0.2382]	.910
	> 20 years	-0.0463	[-0.2205, 0.1280]	.602
Financial status	Poor	0.1656	[-0.0562, 0.3874]	.143
	Lower middle	0.0101	[-0.1193, 0.1395]	.878
	Upper middle	0.0247	[-0.2410, 0.2905]	.855
Hypertension	Yes	0.2473	[0.0847, 0.4100]	.003
Diabetes	Yes	0.1875	[-0.0106, 0.3856]	.064
Acculturation	Anglo (AOS)	-0.0197	[-0.0903, 0.0509]	.585
	Mexican (MOS)	-0.0767	[-0.1809, 0.0276]	.149
Snoring	Rarely	0.0446	[-0.1063, 0.1955]	.562
	Sometimes	0.0758	[-0.1008, 0.2524]	.400
	Frequently	0.3654	[0.1604, 0.5704]	< .001
	Always	0.3462	[0.1348, 0.5576]	.001

*Note.* LASSO regression was conducted with 10-fold cross-validation. The cross-validation minimum penalty was  $\lambda = 0.022$ , yielding 12 retained predictor terms. The follow-up unpenalized model explained  $R^2 = 0.168$  of the variance in  $\log(\text{AHI} + 1)$ .

**Figure 1**

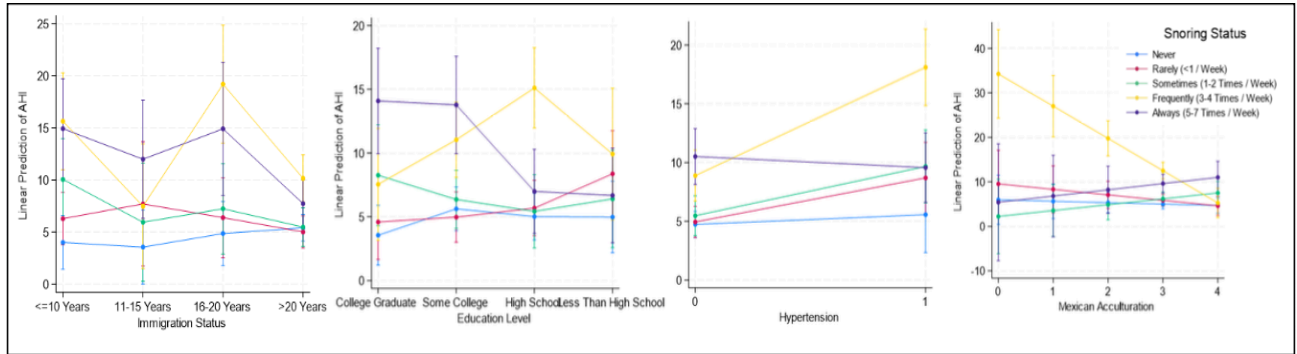
*Distribution of Apnea-Hypopnea Index in the Sample (Raw and Log-Transformed)*



*Note. The left panel depicts the raw distribution of the AHI score, which had a pronounced right skew (skewness = 3.48; kurtosis = 18.38). In contrast, the right panel shows log-transformed AHI ( $\ln(\text{AHI} + 1)$ ) scores, creating a roughly normally distributed variable (skewness = 0.41; kurtosis = 2.85). The transformed AHI variable was used as the outcome in all models.*

**Figure 2**

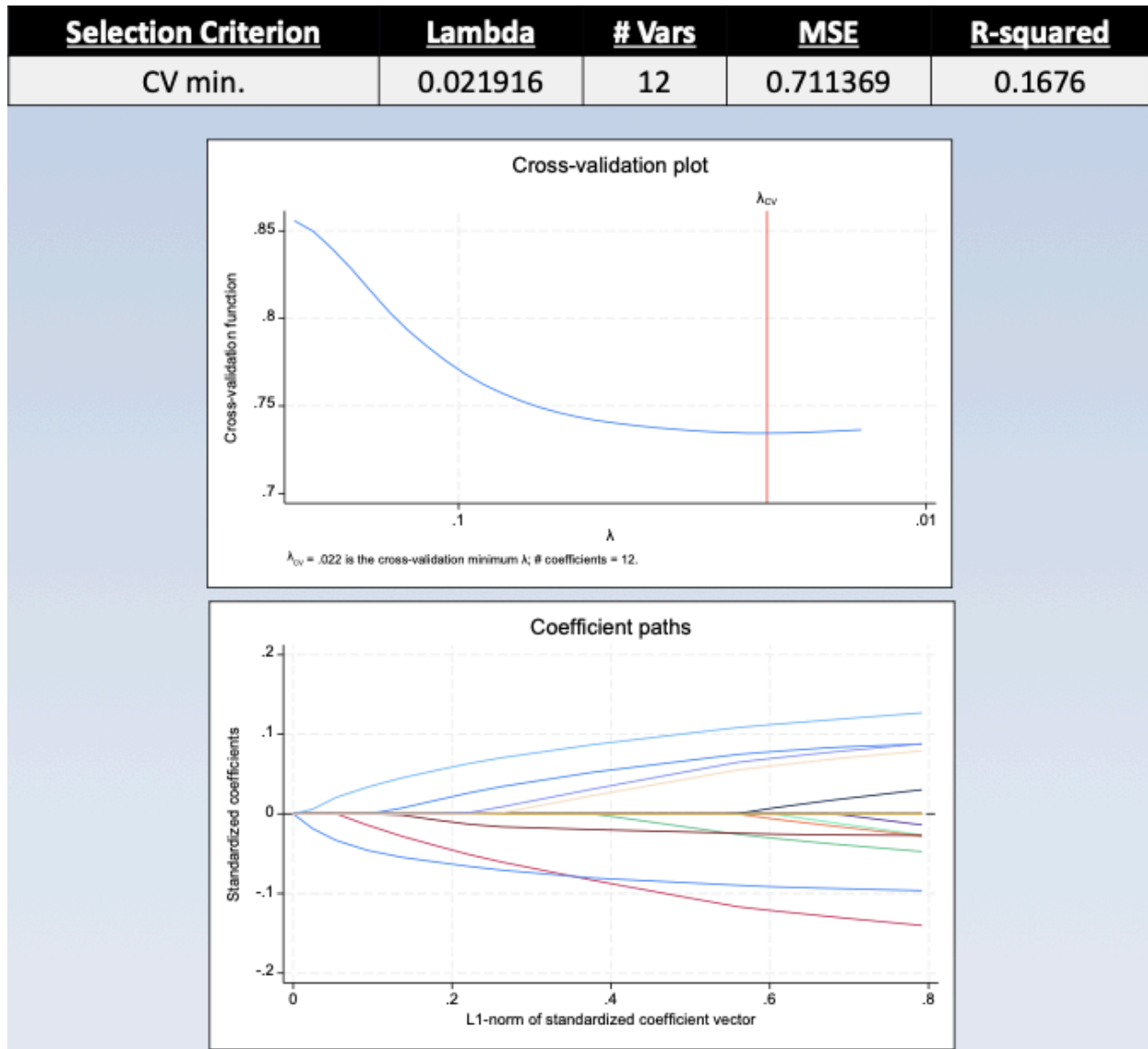
*Snoring × Predictor Interactions for AHI*



*Note. Raw AHI scores were predicted according to time living in the US, education, having high blood pressure, and Mexican acculturation, stratified by snoring frequency. Error bars show 95% confidence intervals around mean predicted AHI scores. Snoring frequency was categorized into five levels of frequency ranging from never (0 days/week) to always (5-7 days/week). However, raw AHI scores were used for visualization only, since statistical tests were performed using  $\log(AHI + 1)$ .*

**Figure 3**

*LASSO Cross-Validation and Coefficient Path Plots*



*Note. Top panel: The cross-validation curve is shown as a function of the penalty factor  $\lambda$ . The vertical dashed red line shows the point of lowest error, with  $\lambda = 0.022$  yielding 12 predictor variables with nonzero coefficients. Bottom panel: The plot shows the coefficient paths of potential predictors according to L1-norm penalization level. Predictors that retained nonzero coefficients at the optimal value  $\lambda$  were used in the final reduced model in Table 4.*