3-31-2008

AN OVERVIEW OF OBSERVATIONAL SLEEP RESEARCH WITH APPLICATION TO SLEEP STAGE TRANSITIONING

Brian S. Caffo
Department of Biostatistics, The Johns Hopkins Bloomberg School of Public Health, bcaffo@jhsph.edu

B. Swihart
Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health

A. Laffan
Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health

C. Crainiceanu
Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health

N. Punjabi
Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health and Johns Hopkins Bayview Medical Campus

An overview of observational sleep research with application to sleep stage transitioning

Caffo, B, Swihart, B, Laffan, A, Crainiceanu, C, Punjabi, N

March 31, 2008

1 Introduction

Sleep is a ubiquitous fact of our lives that we spend roughly one third of our lives doing. Fatigue, jet-lag, poor sleep and vivid dreams are frequent points of our morning discussions. We look and feel terrible after getting too little sleep; hence a twenty billion dollar industry of beds, pillows, pills and other tools has cropped up to help us sleep better (Myser, 2007). Correspondingly, there is plenty of things designed to keep us awake, helping desperate students prepare for all night cramming sessions.

Despite the importance of sleep in our lives, and the lives of so many other species, a definitive answer on the specific neuro-biological or physiologic purpose of sleep eludes researchers (see the discussion in Sejnowski and Destexhe, 2000; Rechtschaffen, 1998). However, substantial advancements in the field are uncovering the crucial role that sleep plays in our health, behavior and well being. One need only consider sleep deprivation, a well recognized form of torture, to highlight this fact. Consider the following quote (Conroy, 2001) from former Israeli Prime Minister Menachem Begin, who suffered forced sleep deprivation as a KGB prisoner:

> In the head of the interrogated prisoner, a haze begins to form. His spirit is wearied to death, his legs are unsteady, and he has one sole desire: to sleep, to sleep just a little, not to get up, to lie, to rest, to forget... Anyone who has experienced this desire knows that not even hunger and thirst are comparable with it.

Studies of sleep duration have found associations with variety of negative health outcomes. For example, short sleep duration correlates with impaired cognitive function, hypertension, glucose intolerance, altered immune function, obesity and mortality (Ohayon and Vecchierini, 2005; Nieto et al., 2000; Punjabi et al., 2004; Gottlieb et al., 2005; Van Cauter et al., 2000; Ayas et al., 2003; Patel et al., 2004; Cizza et al., 2005).

1

Quantity of sleep is only one measurable facet of sleep that is associated with health. Table 1 gives a few of the more common measurements of sleep and sleep disturbance. The next section describes sleep measurement in greater detail.

Sleep measure	Description
Arousal index	Number of arousals per hour slept
Epworth Sleepiness Scale	Aggregate measure of general sleepiness
Respiratory disturbance index	Number of apneas and hypopneas per hour
Sleep architecture	Proportion of time spent in various sleep states
Sleep efficiency	Time asleep as a proportion of time in bed
Sleep latency	Time until falling asleep
Total sleep time	Total time asleep in a night

Table 1: Measurements of sleep

A common sleep disorder that is of particular public health interest is sleep apnea. This is a chronic condition characterized by collapse of the upper airway during sleep. Complete collapses lead to so-called "apneas", where as partial collapses to so-called "hypopneas". Over the last decade, several groups have shown that these events lead to several physiologic derangements.

The respiratory disturbance index (RDI), sometimes also called the apnea/hypopnea index (AHI), is the principal measure of severity of sleep apnea. This rate index is the count of the number of apneas and hypopneas divided by the total time slept in hours. A severely affected patient may have an RDI of thirty events per hour or higher! Hence such a patient is, on average, having a disruption in their sleep and breathing every two minutes. As one can imagine, such frequent disruptions in sleep and oxygen intake can have negative health consequences. A high RDI has been shown to be associated with hypertension, cardiovascular disease, cerebrovascular disease, excessive daytime sleepiness, decreased cognitive function, decreased health-related quality of life, increased motor vehicle crashes and occupational accidents, and mortality (Young et al., 2002).

In this article, we relate measures of sleep apnea with transitions that occur between "sleep states". Sleep states are based on visual classification of brain electroencephalograms (EEGs) patterns (see below). Two major sleep states are rapid eye movement (REM) and non-REM. Sleep states can be seen as a categorical response time series. Crude summaries of these states, such as the percentage time spent in each one, are often used as predictors of health. Instead, we investigate the role that sleep apnea has on the rate of transitioning between the states. We emphasize that the rate of transitioning contains important additional information than the crude percentage of time spent in each state. Notably, we use matching to account for

2

other variables that might be related to both disease status and sleep behavior and hence compare a severely diseased group with a matched non-diseased group.

In the next section we give a brief overview of sleep measurement. Following that, we discuss an analysis of sleep transition rates comparing those with moderate to severe sleep apnea to those without. We end the article with a discussion.

2   Sleep measurement

The gold standard of sleep measurement is based on an overnight sleep study called a "polysomnogram." During polysomnogram, a patient has several physiologic recordings that are digitized and subsequently stored. One of these recordings includes the skull surface electroencephalogram, which measure the actual electrical activity from neurons firing. Because the EEG measures aggregate electrical activity in the cortex, they have poor spatial resolution; however, they have excellent temporal resolution, with hundreds of measurements per second.

Other physiologic recordings measure eye movement (an electro-oculogram), leg movement (electro-myogram), oxygen saturation (pulse-oximter), air flow, chin movement activity (electro-myogram), chest and abdominal movement (via belts around the torso), and heart rate and rhythm (electro-cardiogram). As such, a polysomogram produces an enormous amount of data, as each of these signals is recorded nearly continuously over a night of sleep. These signals are processed by trained technicians under the supervision of sleep physicians.

The technicians and sleep physicians distill this deluge of information to more manageable summaries. In clinical settings, these summaries are used to help patient care decisions. The summaries are also used in research, to investigate the causes and consequences of sleep related phenomena. Table 1 lists examples of summaries of the polysomnogram as well as the Epworth Sleepiness Scale, a subjective assessment of daytime sleep propensity.

Often a concern in sleep clinics is whether or not the subject has sleep apnea and, if so, to evaluate the severity of the disease. As previously mentioned, the primary measure of severity of sleep apnea is the number of apneas or hypopneas per hour.

Another important summary splits the sleep pattern into a few distinct sleep states. This is done visually, by trained and certified technicians and physicians, by grouping the data into thirty second "epochs". The states of interest are labeled: Wake, Stage I, Stage II, Stage III, Stage IV and REM. Stages I-IV are referred to as Non-REM sleep. Stages I and II represent light sleeping and encompass $3-8\%$ and $44-55\%$ of total sleep time, respectively. Stages III and IV represent deeper sleep and represent about $15-20\%$ of the total sleep

time. In REM sleep, which compromises approximately $20-25\%$ of total sleep time, the body is inactive while the brain manifests EEG patterns similar to wakefulness. Most dreaming occurs in REM sleep (for more information see Chokroverty, 1999).

A patient's "sleep architecture" is simply the person-specific percentage spent in each of the six states. Sleep architecture can vary between people and within a person as they age. For example, infants spend more than 80% of their sleeping time in REM (Ohayon et al., 2004; Chokroverty, 1999). It is generally accepted that sleep staging is relevant for understanding sleep's effect on health. We are particularly interested in the impact of the rate of transitioning between the various sleep states. Note that it is not the case that a person necessarily transitions from wakefulness through Stages I to IV in sequential order, and then to REM. Instead, people pass through the states in cycles, with transitioning from any state to another both possible and likely to occur. Wakefulness to REM is the transition that occurs the least frequently.

In addition to measuring night-time sleep signals, other behavioral measurements can be very valuable in sleep research. One of the more widely used measures is the Epworth Sleepiness Scale. This is an aggregated score of several self administered questions involving sleep behavior, and is used as a measure of daytime sleep propensity. For example, patients are scored on whether or not they fall asleep while sitting and reading, or watching television.

3 Sleep transition rates

In this section we present some analysis from sleep transition rates from a large multi-center cohort study, the Sleep Heart Health Study (SHHS). The SHHS added to and combined data from large well established longitudinal studies: the Atherosclerosis Risk in Communities Study, the Cardiovascular Health Study, the Framingham Heart Study, the Strong Heart Study and the Tucson Health and Environment study. At enormous expense, the SHHS recruited over six thousand subjects from these studies to undergo an abbreviated in-home polysomogram (see Quan et al., 1997, for more details on the study). Roughly four thousand of these subjects repeated this process around four years later. The sleep data were processed by trained SHHS technicians, including rigorous data quality checks. The SHHS offers a unique data set to understand sleep and health. However, being an observational study, analysis of the data is often challenging. Any effects or absence of effects seen might be due to subtle biases from the sampling or from (measured or unmeasured) variables either unaccounted for or improperly accounted for in the analysis.

http://biostats.bepress.com/jhubiostat/paper168

3.1 Matching

We consider now the relationship between sleep transitions and sleep apnea. Hence we compare a group of severely diseased patients with sleep apnea, as defined by an RDI greater than 22.3 events per hour, with a group of healthy controls having RDIs of less than 1.33 events per hour. The groups were chosen so that each subject in the diseased group had a matching subject in the control group. This process helps control for confounding variables, not unlike methods such as regression analysis. However matching, unlike regression adjustment, forces a discussion of how alike or unlike the diseased and control groups are. In contrast, regression adjustment will happily plod along via linearity assumptions, even if their is no overlap in the confounding variables for the diseased and control groups. However, matching is not without its issues. Most notably, the SHHS population being studied is a subset of the population selected originally for study; hence a matched subset may lack the generalizability of results on the original data (see Rubin, 1973; Rosenbaum and Rubin, 1985, for more information about matching)

Not everyone in SHHS was eligible to be matched. For example, to adequately define disease status, only those of SHHS with outstanding sleep recording quality and without any history of coronary heart disease, cardiovascular disease, hypertension, chronic obstructive pulmonary disease, asthma, or stroke were eligible for matching. From there, current smokers were ineligible. These rigid qualification standards widdled down the original SHHS pool from over $6,000$ to 183 diseased and 458 non-diseased eligible for matching. These groups are not representative of the population, for it is common for conditions such as hypertension and cardiovascular disease to occur alongside of sleep apnea. In a way, our "diseased" group is healthy in many aspects - except for that of a high index of sleep apnea disease severity.

The matching variables included body mass index (BMI, a ratio a subjects weight to the square of their height), age, race, and sex. Exact matching was used for race and sex, while BMI and age were matched within a caliper (i.e. matched within an acceptable range). The matching procedure produced 60 pairs. Apart from body mass index, none of the variables were significantly different (using Gosset's T test and chi-square tests at the 5% level) when comparing across the two groups. While concern exists about the differential body mass indices, we note that obesity is the primary cause of sleep apnea and both the diseased and control groups , though having statistical different BMIs, were similar practically. Specifically, the average body mass indices for the diseased group was $30.7\ Kg/m^2$, while it was $29.2\ Kg/m^2$ for the non-diseased.

Of note is that traditional sleep architecture, while not considered for a matching parameter, showed no statistically significant difference between groups after the matching. This implies that the gold standard measurement summary of sleep states, sleep architecture, may not be affected by sleep apnea. Using this fact, we now consider how frequently the subjects transition between the various states.
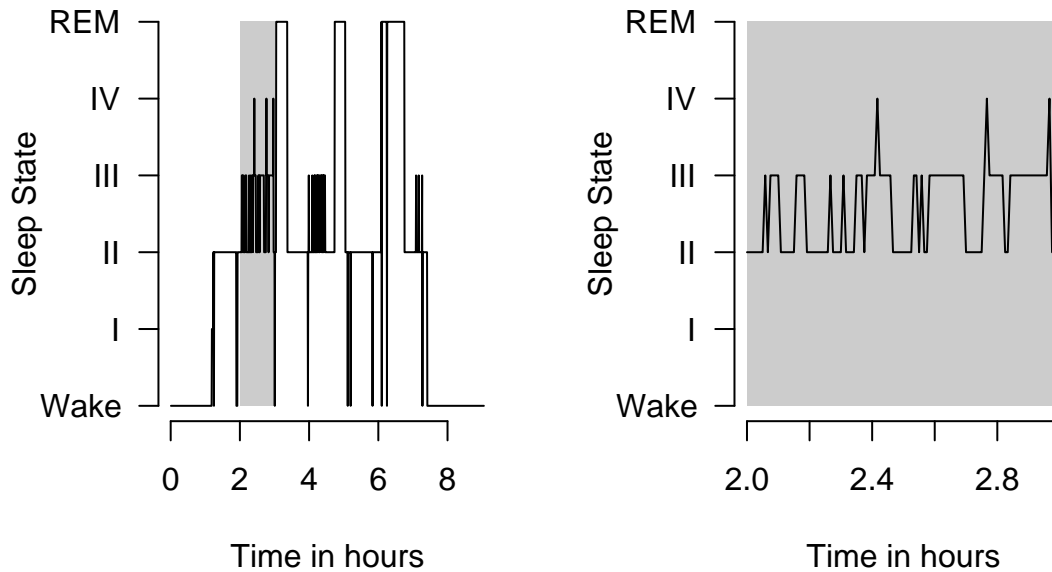
Figure 1: Hypnogram plot for a single subject. The left plot shows the full night where as the right plot highlights sleep between the second and third hours.

Figure 1 displays the time series for the six sleep states over time for a single subject; such a plot is called a "hypnogram". Figure 2 displays a "Lasagna plot" comparing 60 diseased and 60 matched control subjects. In this plot, each color corresponds to a different sleep state, while the horizontal index is time and the vertical index is subject. It cohesively and saliently displays the information of many hypnograms. We have found such plots useful, especially when various sorting operations are performed (see Swihart et al., 2007).

3.2 Analysis of sleep transition rates

We restrict ourselves to studying the transitions starting at the first transition from wakefulness to sleep (usually to Stage 1) and ending at the last transition from a sleep state to wakefulness. That is, we discard time before sleep onset and after waking up. We note that the initial time in bed before sleep onset (sleep latency) between the matched pairs showed no difference (Student's t-test P-value of 0.70). We also discard the final wake states at the end of the monitoring time.

Figure 3 shows mean/difference plots for the log base two of the transition rates for the matched pairs, as well as the transition rate (see Tukey, 1988; Bland and Altman, 1986). By a transition, we mean a change
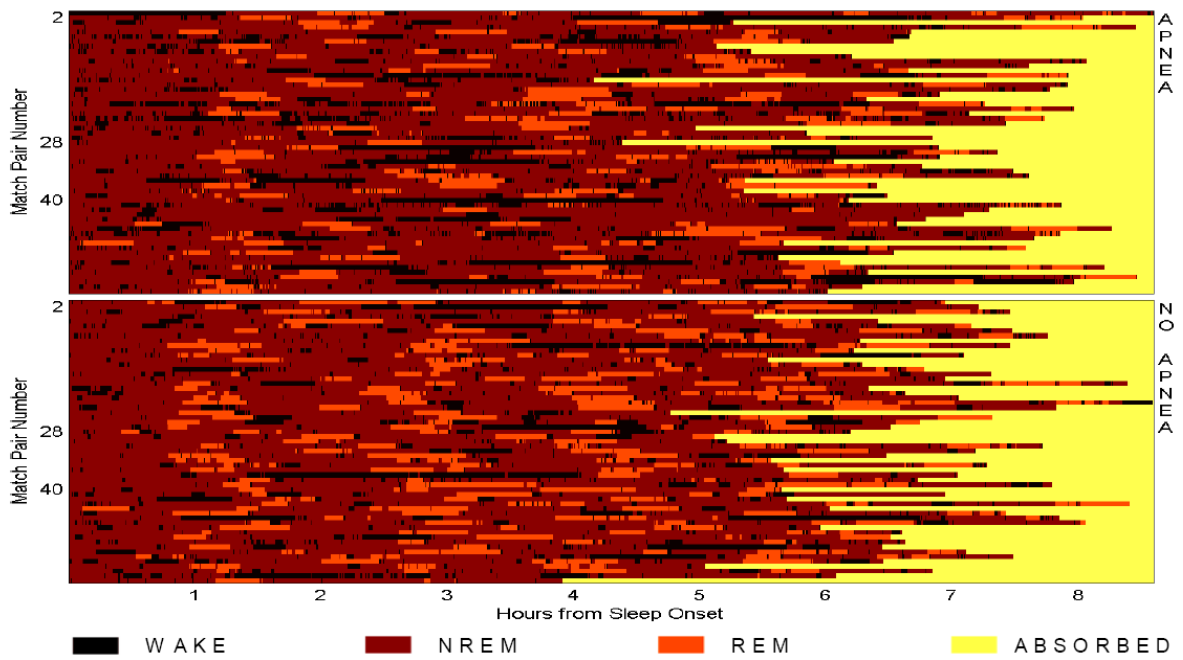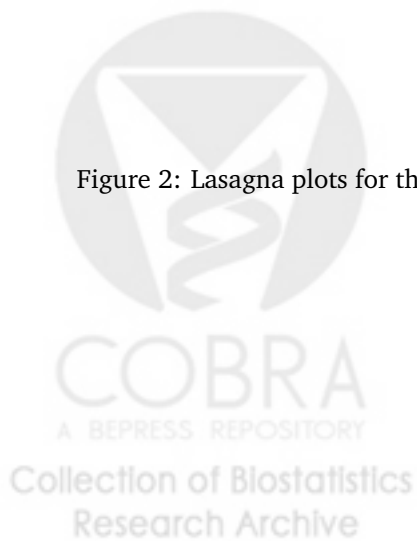
6

Figure 2: Lasagna plots for the sixty diseased (apneic) and matched control subjects.
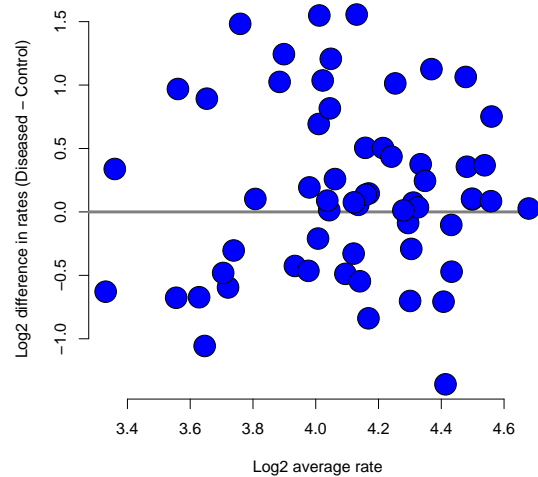
Figure 3: Mean/difference plot of log base two of transition rates with diseased minus matched controls on the vertical axis and pair-wise average log base two transition rates on the horizontal axis.

from any sleep state to another; hence the transition rate is the number of changes divided by the total time asleep in hours. Such plots highlight whether there is a difference between the two matched groups, while plotting against the average highlights whether any such difference is dependent on the magnitude of the transition rates. Log base two is used simply to work with ratios and because powers of two are easier to work with than powers of Euler's number, e, and represent smaller increments than the other option of using base 10. We highly recommend this convention.

Recall that the transition rate is defined as the number of transitions per hour of sleep. We note that a reasonable discussion could be had on which of the two measures, the transition rate or the raw number of transitions, is more important. It may be that a certain raw number of transitions is important for health, regardless of the rate. However, clearly a person who sleeps longer has more opportunities to transition between states, suggesting the use of rates. Regardless, we focus only on the analysis of the rates (see Swihart et al., 2008, for further analysis of the rates and transition numbers). In Figure 3, 39 of the 60 observations lie above the horizontal line, potentially indicating that diseased subjects transition more frequently than non-diseased. For example, under a null hypothesis of no difference in transition rates between the diseased and non-diseased groups, the binomial probability of 39 or more pairs out of 60 lying above the horizontal line is only $.014$ (the two sided P-value would double this number).

A useful summary of each subject's data would be a three-by-three table that displays counts of their previous sleep state by their current sleep state. Table 2 displays the combination of such summary tables

| | Previous state | | | | | |
| | Disease | | | Controls | | |
Current state	N	R	W	N	R	W
Non-REM (N)	31,880	160	1,733	32,592	134	1,376
REM (R)	346	7,609	175	351	8,784	114
Wake (W)	1,588	358	6,079	1,210	324	5,775
Total epochs	33,814	8,127	7,987	34,153	9,242	7,265
Total in hours	281.78	67.73	66.56	284.61	77.02	60.54

Table 2: Disease and control counts.

across subjects, with the non-REM sleeping states (Stages I-IV) aggregated. Shown are counts of the previous state by the current state, cross-classified by disease status. Transition counts occur in the off-diagonal cells, with the diagonal cells representing instances where the subjects stayed in the same state from one epoch to the next. For example, in the diseased subjects, there were 346 transitions from non-REM (N) to REM (R), while there were 175 transitions from wake (W) to REM. Column totals in this data are special, representing the time at-risk for various kinds of transitions. The column totals, which are counts of epochs, are converted into hours for convenience. Therefore, there were 346 transitions of type $N \rightarrow R$, and there were 281.78 total hours spent in non-REM where this type of transition could be made.

The most frequent transition in both groups is $W \rightarrow N$, with rates of $1,733/66.56 = 26.0$ and $1,376/60.54 = 22.7$, respectively. The next most frequent transitions are $N \rightarrow W$ and $R \rightarrow W$. The data paint a picture that a person, diseased or not, spends the majority of their time in the non-REM state. From there, they often transition to REM, but then spend a little time in REM before transitioning to wakefulness (more likely) or back to non-REM (less likely). In addition, from non-REM, they often briefly wake up, then transition (more likely) back to non-REM.

It is of interest to compare whether these rates differ across disease groups. This is a somewhat challenging task, given that one must account for the correlation induced by matching as well as the correlation of the various rates within a particular subject. For example, in a subject with a high rate of non-REM to wake transitions, it is reasonable to believe that there would be a correspondingly high rate of wake to non-REM transitions; hence these two rates would be correlated. We fit a model that assumes a constant rate of transitioning over the time at-risk for transitioning, a so-called exponential hazard model, accounting for these kinds of correlations. While we relegate a more careful exploration of this data and model to later work, we present some of the results below.

Table 3 shows the estimated relative rates, i.e. the estimated rate of transitioning for the diseased divided

9

Transition type	Relative rate
$N \to R$	$_{0.84}1.04_{1.29}$
$N \to W$	$_{1.17}1.39_{1.65}$
$R \to N$	$_{1.11}1.46_{1.93}$
$R \to W$	$_{1.08}1.34_{1.67}$
$W \to N$	$_{0.87}1.04_{1.22}$
$W \to R$	$_{0.95}1.27_{1.68}$

Table 3: Relative rates comparing diseased (numerator) to non-diseased (denominator) subjects. The smaller numbers represent 95% credible intervals (Louis and Zeger, 2007).

by that of the controls, and 95% credible interval. A credible interval is a Bayesian analog to the confidence interval; readers unfamiliar with Bayesian analysis can simply think of them as confidence intervals. The data suggests that the rate of transitioning from non-REM to wakefulness, REM to non-REM, and REM to wakefulness all differ between the two groups. Notably, all of the estimates represent increases in the transition rates for the diseased subjects. This suggests that the disruption of sleep continuity in responses to the airway collapses during sleep may cause increased transitions between wakefulness and the other states.

4 Discussion

In this manuscript, we briefly reviewed an area of observational sleep research with a particular emphasis to analyzing sleep transitions and their relationship with sleep apnea. The analysis showed some potential differences between the diseased and non-diseased group with respect to the amount of transitioning; though significant work remains to be done to fully understand this problem. Especially, it would be important to consider how the transition rates differ over the night, relaxing the assumptions of the exponential model we presumed (see Fahrmeir and Klinger, 2003; Fahrmeir, 1999), improving the matching algorithm and applying the methods to the large, unmatched data using other adjustments. Hence, this study presented herein represents a small snippet of understanding of how sleep is influenced by a specific disease.

We emphasize, however, that the overarching focus of our research, is to better exploit the full information contained in polysomnograms from large-scale observational studies. This includes functional data analysis of the sleep state and EEG signals. We believe that there is important information omitted by only considering the standard epidemiological summaries, which in most cases were designed as simple clinical indices and hence may be improved upon for research purposes.

5 Further reading

The book by Chokroverty contains an excellent summary of sleep medicine. The manuscripts by Swihart et al. provide introductions to the display and analysis of sleep transitions while the article by Young et al. gives an overview of sleep disordered breathing.

Chokroverty, S. (1999). *Sleep disorders medicine: basic science, technical considerations, and clinical aspects*. Butterworth-Heinemann Boston, MA.

Swihart, B., Caffo, B., Bandeen-Roche, K., and Punjabi, N. (2008). Quantitative characterization of sleep architecture using multi-state and log-linear models. to appear in the *Journal of Clinical Sleep Medicine*.

Swihart, B., Caffo, B., Strand, M., and Punjabi, N. (2007). Novel Methods in the Visualization of Transitional Phenomena. *Johns Hopkins University, Dept. of Biostatistics Working Papers*.

Young, T., Peppard, P., and Gottlieb, D. (2002). Epidemiology of Obstructive Sleep Apnea. A Population Health Perspective. *American Journal of Respiratory and Critical Care Medicine*, 165(9):1217-1239.

References

Ayas, N., White, D., Manson, J., Stampfer, M., Speizer, F., Malhotra, A., and Hu, F. (2003). A Prospective Study of Sleep Duration and Coronary Heart Disease in Women. *Archives of Internal Medicine*, 163(2):205–209.

Bland, J. and Altman, D. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, 1(8476):307–310.

Chokroverty, S. (1999). *Sleep disorders medicine: basic science, technical considerations, and clinical aspects*. Butterworth-Heinemann Boston, MA.

Cizza, G., Skarulis, M., and Mignot, E. (2005). A link between short sleep and obesity: building the evidence for causation. *Sleep*, 28(10):1217–20.

Conroy, J. (2001). *Unspeakable Acts, Ordinary People: The Dynamics of Torture*. University of California Press.

Fahrmeir, L. (1999). Modelling and exploring human sleep with event history analysis. *Journal of Sleep Research*, 8:25–36.

Fahrmeir, L. and Klinger, A. (2003). A nonparametric multiplicative hazard model for event history analysis. *Biometrika*, 85(3):581–592.

Gottlieb, D., Punjabi, N., Newman, A., Resnick, H., Redline, S., Baldwin, C., and Nieto, F. (2005). Association of Sleep Time With Diabetes Mellitus and Impaired Glucose Tolerance. *Archives of Internal Medicine*, 165(8):863.

Louis, T. and Zeger, S. (2007). Effective communication of standard errors and confidence intervals. *Johns Hopkins University, Dept. of Biostatistics Working Papers*, page 151.

Myser, M. (2007). $20 billion for a good night's rest we're sleeping less each year - and that's creating a big opportunity to sell everything from naps to high-tech beds. CNN Business 2.0 Magazine, March 15.

Nieto, F., Young, T., Lind, B., Shahar, E., Samet, J., Redline, S., D'Agostino, R., Newman, A., Lebowitz, M., and Pickering, T. (2000). Association of Sleep-Disordered Breathing, Sleep Apnea, and Hypertension in a Large Community-Based Study. *JAMA*, 283(14):1829–1836.

Ohayon, M., Carskadon, M., Guilleminault, C., and Vitiello, M. (2004). Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: developing normative sleep values across the human lifespan. *Sleep*, 27(7):1255–73.

Ohayon, M. and Vecchierini, M. (2005). Normative sleep data, cognitive function and daily living activities in older adults in the community. *Sleep*, 28(8):981–9.

Patel, S., Ayas, N., Malhotra, M., White, D., Schernhammer, E., Speizer, F., Stampfer, M., and Hu, F. (2004). A prospective study of sleep duration and mortality risk in women. *Sleep*, 27(3):440–4.

Punjabi, N., Shahar, E., Redline, S., Gottlieb, D., Givelber, R., and Resnick, H. (2004). Sleep-Disordered Breathing, Glucose Intolerance, and Insulin Resistance The Sleep Heart Health Study. *American Journal of Epidemiology*, 160(6):521–530.

Quan, S., Howard, B., Iber, C., Kiley, J., Nieto, F., O'Connor, G., Rapoport, D., Redline, S., Robbins, J., Samet, J., et al. (1997). The Sleep Heart Health Study: design, rationale, and methods. *Sleep*, 20(12):1077–85.

Rechtschaffen, A. (1998). Current perspectives on the function of sleep. *Perspect Biol Med*, 41(3):359–90.

Rosenbaum, P. and Rubin, D. (1985). Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score. *The American Statistician*, 39(1):33–38.

Rubin, D. (1973). Matching to Remove Bias in Observational Studies. *Biometrics*, 29(1):159–183.

Sejnowski, T. and Destexhe, A. (2000). Why do we sleep? *Brain Research*, 886(1-2):208–223.

Swihart, B., Caffo, B., Bandeen-Roche, K., and Punjabi, N. (2008). Quantitative characterization of sleep architecture using multi-state and log-linear models. *to appear in the Journal of Clinical Sleep Medicine*.

Swihart, B., Caffo, B., Strand, M., and Punjabi, N. (2007). Novel Methods in the Visualization of Transitional Phenomena. *Johns Hopkins University, Dept. of Biostatistics Working Papers*.

Tukey, J. (1988). Exploratory Data Analysis. *The Collected Works of John W. Tukey: Graphics 1965-1985, Volume V*.

Van Cauter, E., Leproult, R., and Plat, L. (2000). Age-Related Changes in Slow Wave Sleep and REM Sleep and Relationship With Growth Hormone and Cortisol Levels in Healthy Men.

Young, T., Peppard, P., and Gottlieb, D. (2002). Epidemiology of Obstructive Sleep Apnea. A Population Health Perspective. *American Journal of Respiratory and Critical Care Medicine,* 165(9):1217–1239.