8-7-2012

MODELING SLEEP FRAGMENTATION IN POPULATIONS OF SLEEP HYPNOGRAMS

Bruce J. Swihart
Johns Hopkins School of Public Health, bruce.swihart@gmail.com

Naresh M. Punjabi
Johns Hopkins Medical Institution

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

Modeling sleep fragmentation in populations of sleep hypnograms

Bruce J. Swihart, Naresh M. Punjabi, Ciprian M. Crainiceanu

August 6, 2012

Abstract

We introduce methods for the analysis of large populations of sleep architectures (hypnograms) that respect the 5-state 20-transition-type structure defined by the American Academy of Sleep Medicine. By applying these methods to the hypnograms of 5598 subjects from the Sleep Heart Health Study we: 1) provide the first analysis of sleep hypnogram data of such size and complexity in a community cohort with a 4-level comorbidity; 2) compare 5-state 20-transition-type sleep to 3-state 6-transition-type sleep for a check of feasibility and information-loss; 3) extend current approaches to multivariate survival data analysis to populations of time-to-transition processes; and 4) provide scalable solutions for data analyses required by the case study. This allows us to provide detailed new insights into the association between sleep apnea and sleep architecture. Supporting R as well as SAS code and data are included in the online supplementary materials.

Keywords: Competing risks, Multi-state, Recurrent event, Sleep-disordered breathing, Stratified.

1 Introduction

There is a wide and personal familiarity with sleep. Everyone has firsthand experience sleeping or facing the day with a lack thereof. Far less familiarity exists with sleep science, medicine, and epidemiology. Sleep science is the study and measurement of the sleep process itself: what biological and physiological happenings within an organism define sleep. Sleep medicine harnesses sleep science so that sleep-related conditions may be diagnosed and subsequently treated at an individual-level. At a population-level building upon the work of sleep medicine, sleep epidemiology studies associations of sleep-related conditions (i.e., insomnia, sleep-disordered breathing, sleep-deprivation, etc), not necessarily the sleep process itself, with non-sleep-related health outcomes, such as hypertension or all-cause mortality.

Sleep science conceptualizes an individual's sleep as a hypnogram, a discrete-state discrete-time stochastic process (Figure 1). Currently, the field of sleep science broadly generalizes a typical sleep progression after falling asleep from Wake (W, on the hypnogram axis) being Stage 1 (1) to Stage 2 (2) to Stage Slow-wave (S) back to Stage 2 and then Rapid Eye Movement (R). The progression makes up one sleep cycle, which lasts approximately 60-90 minutes and repeats through the night. Judging from the three examples of Figure 1, sleep does not evolve so straightforwardly in reality. The top panel fits the generalization well, but the other two do not, with many detours and interruptions to the charted course of a "typical" sleep: the middle panel has more alternations between Stage 2 and Stage Slow-wave; while the bottom panel has a duration in Wake before leaving Stage 1 and a much more fragmented Stage Slow-wave portion. Given that every 30 seconds the trajectory can change to any other of the four states or remain in the current state makes for a very diverse functional space for one hour snippets of three individuals, let alone for a typical overall sleep time of approximately 7 hours for thousands of individuals in sleep

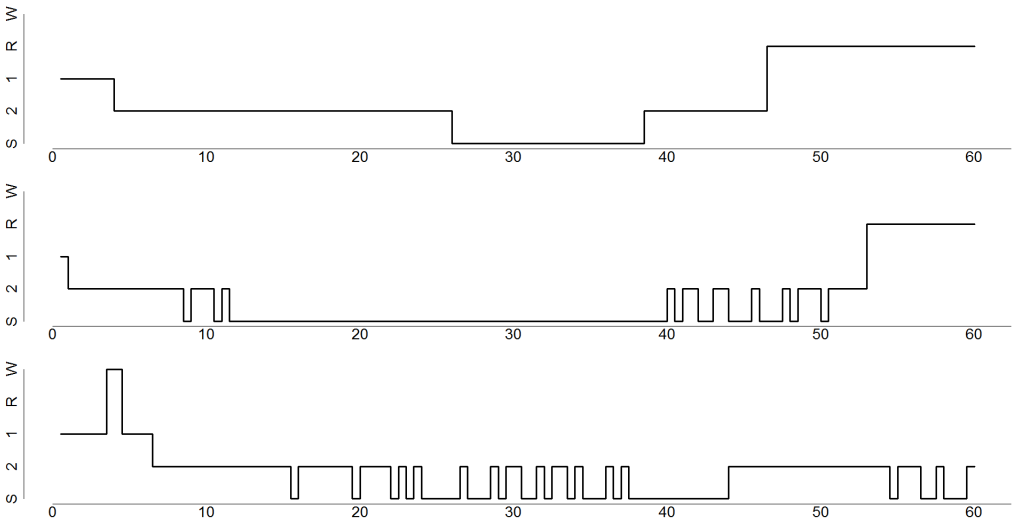


Figure 1: The first hour of sleep for three individuals, visualized by discrete-time discrete-state spaghetti plots known as sleep hypnograms. The vertical axis displays the five stages of sleep: Wake, Rapid-Eye Movement (REM), Stage 1, Stage 2, and Stage Slow-wave labeled "W", "R", "1", "2", "S", respectively. The horizontal axis denotes time from sleep onset in 30-second epochs for 60 minutes.

epidemiology investigations. In modeling and comparing group-defined populations of sleep hypnograms through relative metrics of transition-type specific rates, we can learn how the sleep process differs as a function of the group-defining metric. Making the group-defining condition a sleep-related one enhances sleep epidemiology by providing associations of the sleep process itself with a sleep-related condition, complementing the current widespread practice of associations of non-sleep related morbid conditions with sleep-related conditions. We consider the 4-levels of severity in sleep-disordered breathing as a group-defining condition.

1.1 Background: sleep-disordered breathing

Sleep-disordered breathing (SDB) is characterized by recurrent collapse of the upper airway and is associated with recurrent episodes of intermittent hypoxemia (low levels of oxygen saturation in arterial blood) and arousals from sleep (abruptly moving from "deep" sleep to "lighter" sleep, yet not necessarily awakening). The characterizations are thought to be largely sequential and indicative of the work of the sympathetic nervous system: during sleep, the muscle tone in the upper airway relaxes, obstruction (full or partial) to the airflow ensues, the obstruction lowers oxygen levels in the blood prompting the sympathetic nervous system's "fight-or-flight" survival mechanism involving a central nervous system jolt, resulting in some stress on the heart and brain, and thus the arousal from deep to lighter sleep. The upper airway obstruction from sleep-disordered breathing has a range of consequences, from the nuisance and embarrassment of snoring to the ultimate adverse health outcome of death. Approximately 9% of women and 24% of men in the general population have sleep-disordered breathing and approximately 75% of those affected remain undiagnosed (Young et al., 2002). These prevalence estimates grow with the obesity epidemic, because the obese tend to have less muscle tone and more adipose tissue in the throat area, enabling

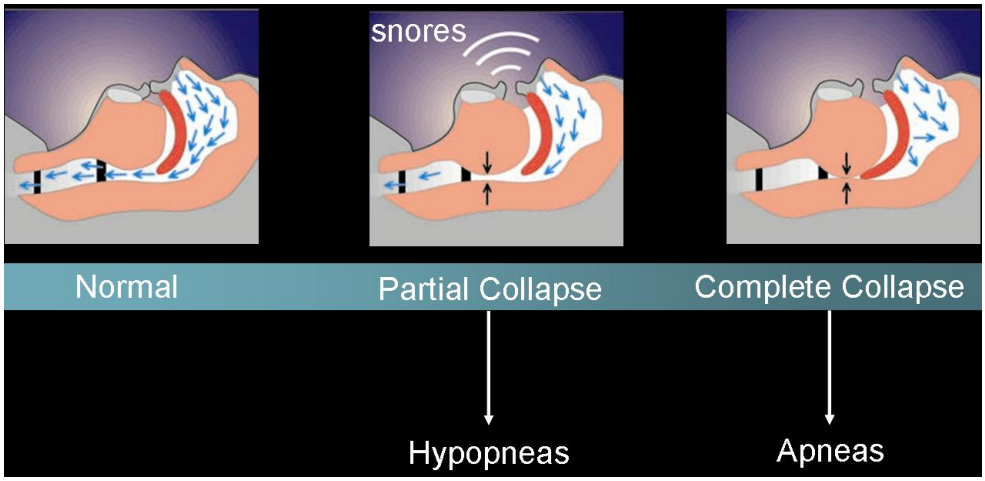


Figure 2: An instance of airflow obstruction in sleep-disordered breathing, illustrated. Without medical intervention, such recurrent obstructions are rectified through iterative sympathetic nervous system activation. Sympathetic nervous activation can produce arousal events in sleep. Such arousals may or may not be represented in the hypnogram. Courtesy of Johns Hopkins Sleep lab.

more often and severe upper airway obstructions.

Estimates of the increased likelihood of all-cause mortality due to SDB independent of other prevalent conditions and adjusted for demographic factors range from 1.3 - 3.8 times that of those without sleep-disordered breathing. For death by cardiac-event, the estimates jump to a factor of 5.2 (Young et al., 2002). The alarming circumstances of high and growing prevalence, high undiagnosed proportion of those affected, and the notable risk of death involved motivated the investigation of the SDB and mortality association.

SDB severity is measured by a composite measure known as the respiratory disturbance index at 4% oxygen desaturation (rdi4p). The composition is such that both the hypoxemia and arousals of SDB are taken into account, and the rdi4p has the units of events per hour of sleep. Didactically, suppose one slept for 10 hours and had 50 apneas (full obstructions), 50 hyponeas (partial obstructions), and 25

arousals. Say 33 of those apneas and 2 of the hypopneas caused a 4% or greater oxygen desaturation. Then the rdi4p is (33+2+25)/10 = 6 events / hr.

In turn, rdi4p is often categorized into four diagnostic groups of increasing SDB severity: absent, mild, moderate, and severe. Although the association between SDB and mortality has been established and the causal pathway suspected to be through the heart, the effect of SDB on sleep and whether the effects of sleep itself have any contribution to mortality has remained largely uninvestigated, probably due to a lack of accessible methods for modeling sleep at the population level.

1.2 Relations to other approaches

Our approach is motivated by epidemilogic studies, which constitute two important features: 1) thousands of subjects; and 2) group-defined populations of comparative interest. Hypnogram data have been the topic of previous analytic frameworks and data applications distinct from the one in current consideration, as many have been demonstrated on substantially smaller sample sizes of homogenous groups or 2-level groups. For example, an important (clinical) goal was to relate time-varying hormone levels to the sleep process modeled as a reduced set of transition-types for the number of states considered (for S states, a full and flat model paradigm would be all $S \times (S-1)$ pairwise transition-types). To remove potential confusion about the novelty of our approach, we provide an itemized list of publications related to our problem together with the problems that were addressed.

- Fahrmeir and Klinger (1998) relate time-varying binary cortisol levels to the propensity of making a transition in a 3-state 4-transition-type hierarchy for 1 homogenous group of 30 males using a nonparametric multiplicative hazard model for event history analysis.

- Yassouridis et al. (1999) relate time-varying binary cortisol levels to the propensity of making a transition in a 4-state 6-transition-type hierarchy for 1 homogenous group of 30 males using a nonparametric multiplicative hazard model for event history analysis.

- Aalen et al. (2004) illustratively uses 27 individuals for 1 homogenous group in a 2-state 1-transition-type conceptualization with counting processes using dynamic covariates to model time-varying cortisol levels.

- Norman et al. (2006) use 30 individuals, 10 each from 3 different group-defined populations of disease severity in a sleep-runs analysis (which is a 2-state 1-transition-formulation with no intra-subject repeated events clustering).

- Kneib and Hennerfeind (2008) extend the work of Fahrmeir and Klinger (1998) relating time-varying cortisol levels to the propensity of making a transition in a 3-state 4-transition-type hierarchy for 1 homogenous group of 70 individuals using a counting process representation of a semiparametric multi-state model implemented with full and empirical Bayesian methods.

- Swihart et al. (2008) use a matched group of 60 pairings of healthy to diseased subjects in a flat 3-state 6 transition-type paradigm with log-linear models for relative transition counts and multi-state survival models for relative transition rates.

- Kalus et al. (2009) use 47 healthy subjects of different sexes and ages to relate 11-transition-types (some hierarchical) to three distinct time-varying hormones using semiparametric multinomial logit models.

- Swihart et al. (2012) use a matched group of 51 pairings of healthy to diseased subjects in a flat 3-state 6 transition-type paradigm with a Bayesian implemen-

tation of a Poisson regression incorporating temporal and frequency transition information that was equivalent to a multi-state model and an instance of a strutured additive regression model. An analysis on 5614 subjects was also conducted as a time trial test for scalability.

Our approach uses 5598 subjects with 5-state 20-transition-type sleep and focuses on analyses that investigate many covariates of different types and their association with the transition process. We were unable to use or adapt the methods described above to reasonably analyze our data set. Thus, the goal of our paper is to provide statistical models that: 1) are not more complex than necessary for comparing (more than 2) populations; 2) characterize transition-type-specific features of the frequency and rate behaviors observed in Figure 1; and 3) do not over-simplify the sleep state transition process, as done by the currently accepted characterization of sleep cycles. Characterizing sleep itself in terms of well known statistical models fitted with widely accessible and established methods to largely generalizable populations is the case study presented herein. The chief contribution of this paper is the application of the statistical models to population sleep data, a result of appropriately acknowledging and handling the intertwining of hypnogram resolution (the number of states and transition-types considered), data size, computational feasibility, and modeling paradigm.

The remainder of the article is organized as follows. In Section 2 we explicitly define the outcome process, group definitions, and the models to estimate the association between the two. Application to the Sleep Heart Health Study data occurs in Section 4, with some pre-application considerations highlighted in Section 3. In Section 5 we provide further insight into our modeling strategy and Section 6 ends the article with a discussion. The scientific findings indicate SDB severity affects distinct transition types differently and that collapsing from 5-state to 3-state can have

undesirable consequences when trends in the analogue set are in opposite directions. Thus, 5-state analyses might be preferred, underscoring the importance of having the accessible methods herein to handle the increase in data complexity.

2 Data Description and Methods

2.1 Data Origination and Manipulation

A brief overview of the science and measuring of sleep starts with polysomnography. Polysomnography is a multi-faceted monitoring process that produces polysomnograms. A polysomnogram is a collection of simultaneous time series which are the measured biosignals intrinsic to defining sleep. The biosignals that comprise a polysomnogram include the electroencephalogram (EEG), electro-oculogram (EOG), electromyogram (EMG), electrocardiogram (ECG), airflow, chest and abdominal effort, oxyhemoglobin saturation, and body position. The juxtaposition of these time series is the lens through which we see one's sleep. The data of a polysomnogram is voluminous and complex. The field of sleep traditionally summarizes the simultaneous time series of a polysomnogram into five stages of sleep known as the R and K system, as put forth by Rechtschaffen and Kales in 1968 and updated by the American Academy of Sleep Medicine (AASM) in 2007 (Rechtschaffen and Kales, 1968). Extensively, computer algorithms have supplanted sleep physicians for the task of translating the simultaneous curves of the polysomnogram into the five stages of sleep (Penzel and Conradt, 2000). The summarization occurs dually over time and across signals: 1) continuous time is discretized into sequential bins called epochs and 2) within each epoch the information across all time series is combined to declare one of the five stages of sleep: Wake, Stage 1, Stage 2, Stage Slow-wave, and Rapid Eye Movement (REM). A sleeper passes through these states in a recurrent fashion many times throughout the night. The R and K system facilitates a tremendous data

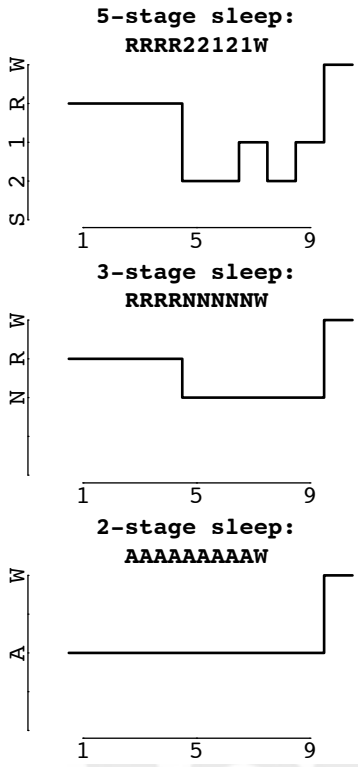
reduction, producing one discrete-time discrete-state process, the hypnogram, from many continuous-time and continuous-state time series.

The polysomnogram data is supplied by the Sleep Heart Health Study (Quan et al., 1997). The result of processing the polysomnogram into a discrete-time discrete-state hypnogram was an ASCII file of one line, displaying a symbol to represent the occupied state for sequential and mutually exclusive 30-second epochs. For example, RRRR22121W could be the 10 epoch tail-end of a string, where one R is 30-seconds of REM sleep, 2 is Stage 2, 1 is Stage 1, and W is wake (Figure 3). A transition occurs whenever there is a change of symbol adjacent to one another. There are 5 transitions in this string, chronologically: a REM-Stage 2 (labeled R2) transition, Stage 2-Stage 1 (labeled 21) transition, Stage 1-Stage 2 (labeled 12) transition, Stage 2-Stage 1 (labeled 21) transition, and a Stage 1-Wake (labeled 1W) transition. The time at risk (interchangeably, duration in state, transition time, survival time, failure time, or time-to-event) for R2 was 2 minutes, transition-type 21 had a time at risk of 1 minute, transition-type 12 had a time at risk of 0.5 minutes, the second occurrence of transition-type 21 had a time at risk of 0.5 minutes, 1W had 0.5 minutes, and no transition out was recorded of the final Wake epoch.

The stages of sleep are collapsable, yielding hypnograms of fewer states. The collapsability is biologically motivated. For instance, to go from a 5-state to 3-state hypnogram, Stage 1, Stage 2 and Stage Slow-wave are combined into Non-REM (NREM) sleep stage. In our example, RRRR22121W becomes RRRRNNNNNW. Continuing in this vein, both NREM and REM sleep stages of 3-state sleep can be collapsed into an "Asleep" stage, making sleep a 2-state process, where RRRRNNNNNW becomes AAAAAAAAAW.

For each hypnogram resolution (5-,3-, or 2-stage), two well-known modeling conceptualizations are available, each requiring a different data format: survival analysis

and Poisson regression. A time-to-event survival analysis requires a format detailing chronologically each possible transition per row, whether the transition was observed, and the time to transition (Figure 3). Poisson regression requires transition-type specific total counts of occurrence and total time at risk (tar) for those counts, which is straightforwardly borne of summing the observed (obs) and time-to-event (tte) variables of the properly constructed survival analysis format, respectively, by transition-type (shift).



Competing Risks format					*Poisson format*				
resolution	order	shift	obs	tte	resolution	type	shift	count	tar
5-stage	1	R2	1	2.0	5-stage	1	1R	0	1.0
	1	R1	0	2.0		2	2R	0	1.5
	1	RS	0	2.0		3	SR	0	0.0
	1	RW	0	2.0		4	1W	1	1.0
	2	21	1	1.0		5	2W	0	1.5
	2	2R	0	1.0		6	SW	0	0.0
	2	2S	0	1.0		7	R1	0	2.0
	2	2W	0	1.0		8	R2	1	2.0
	3	12	1	0.5		9	RS	0	2.0
	3	1R	0	0.5		10	RW	0	2.0
	3	1S	0	0.5		11	W1	0	0.5
	3	1W	0	0.5		12	W2	0	0.5
	4	21	1	0.5		13	WS	0	0.5
	4	2R	0	0.5		14	WR	0	0.5
	4	2S	0	0.5		15	S1	0	0.0
	4	2W	0	0.5		16	S2	0	0.0
	5	1W	1	0.5		17	12	1	1.0
	5	12	0	0.5		18	1S	0	1.0
	5	1R	0	0.5		19	21	2	1.5
	5	1S	0	0.5		20	2S	0	1.5
	6	W1	0	0.5					
	6	W2	0	0.5					
	6	WR	0	0.5					
	6	WS	0	0.5					
3-stage	1	RN	1	2.0	3-stage	1	NR	0	2.5
	1	RW	0	2.0		2	NW	1	2.5
	2	NW	1	2.5		3	RN	1	2.0
	2	NR	0	2.5		4	RW	0	2.0
	3	WN	0	0.5		5	WN	0	0.5
	3	WR	0	0.5		6	WR	0	0.5
2-stage	1	AW	1	4.5	2-stage	1	AW	1	4.5
	2	WA	0	0.5		2	WA	0	0.5

Figure 3: Sleep hypnograms of the same sleep trajectory as represented in 5-stage, 3-stage, and 2-stage sleep resolution with accompanying Poisson and survival analysis data formats.

Figure 3 depicts three hypnogram resolutions for one subject's 10-epoch portion

of sleep, visualized with spaghetti plots. For populations of sleep hypnograms, visualizing several hypnogram trajectories in a spaghetti plot is prone to over-plotting. A lasagna plot, by contrast, is a heat map of a matrix where element S_{ij} is the state occupied by the i^{th} subject at the j^{th} epoch (Swihart et al., 2010). Therefore, a lasagna plot is a heatmap that displays clustered longitudinal data, with clusters in the rows and time in the columns and eliminates the overlapping of trajectories that plagues spaghetti plots. In addition, lasagna plots are capable of dynamic sorting. Figure 4 displays three lasagna plots for each of three different state resolutions for 5598 subjects over 1218 epochs (10 hours, 9 minutes). The top panel for a given resolution is unsorted with respect to subjects. The middle panel shows the same lasagna plot where subjects are organized into the four SDB groups (in descending order of severity) and within SDB group by total sleep time. The bottom panel is a within-column within-SDB-group sorting of the lasagna plot in the middle panel, which shows group-level temporal behavior. Note, as the legends of Figure 4 collapse (from left to right) how much information is lost: Stage Slow-wave has well defined peaks that alternate with REM across disease severity, and the prevalence of each group being in Stage 1 in the first epoch of sleep onset is decidedly over 50%, decreasing drastically and then stabilizing over the night.

2.2 Collapsing States and Information Mapping

The hypnogram is a 5-state stochastic process, comprising of Wake and 4 distinct stages of sleep. However, a 2-state stochastic process can be rendered by collapsing all the non-Wake stages into an "Asleep" state. With the 2-state rendering, there are only two transition types to consider: Wake to Asleep (WA) and Asleep to Wake (AW). The transitions can recur through the night and there are no competing risks – i.e., when in the Asleep stage, the only transition that can occur is the one to Wake.

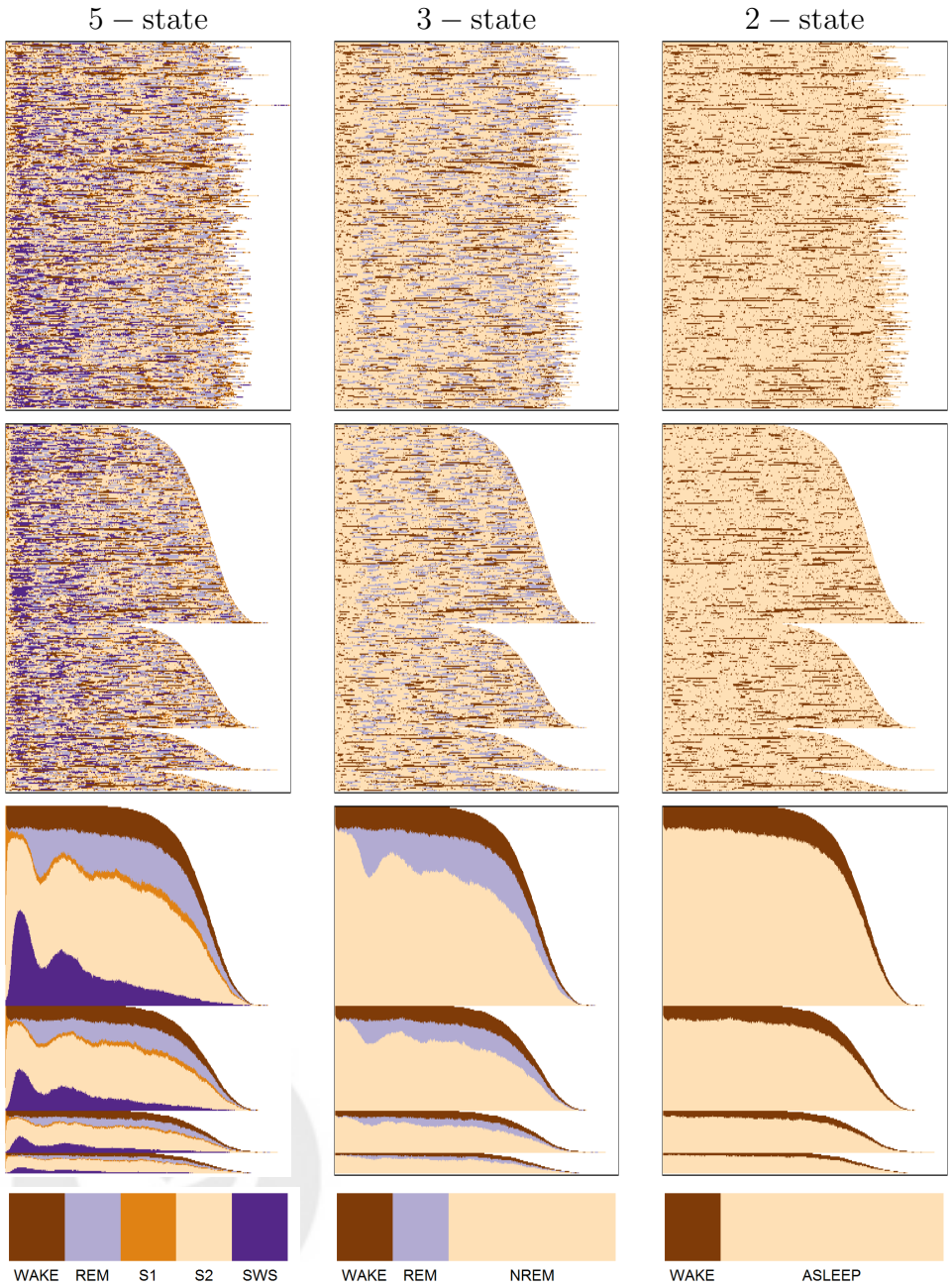


Figure 4: Lasagna plots for 5-state, 3-state, and 2-state sleep. Each lasagna plot has 5598 rows (subjects) and 1218 columns (epochs). The top row of lasagna plots display subjects in no particular order. The second row shows subjects grouped into SDB severity group (absent to severe; top to bottom within the plot) and ranked by total sleep time within severity group. The bottom row of lasagna plots are those of the middle row sorted within columns within severity group, highlighting the group-level temporal dynamics of Stage Slow-wave in dark purple.

A less aggressive summary of the 5-state is the 3-state hypnogram, where Stage 1, Stage 2, and Stage Slow-wave are condensed into Non-REM sleep (NREM) whereas REM along with Wake is left as-is. The 3-state hypnogram has a maximum of 6 pairwise transition-types: NREM to REM (NR), NREM to Wake (NW), REM to NREM (RN), REM to Wake (RW), Wake to NREM (WN), Wake to REM (WR). Unlike the 2-state process, the 3-state process has competing risks. For example, whilst in Wake, a sleeper is simultaneously at risk for WN and WR. With 5 states, there are 20 different transition-types. The collapsing scheme of the states dictates that NR of 3-state has its information mapped to the analogue set of {1R, 2R, SR} in 5-state, and likewise NW has {1W, 2W, SW}, RN has {R1, R2, RS}, RW has {RW}, WN has {W1, W2, WS}, and WR has {WR}. The remaining transitions are intra-NREM and have no analogue in 3-state: {S1, S2, 12, 1S, 21, 2S}.

Regardless of resolution, the key features of hypnogram fragmentation are represented in terms of all pairwise state transitions and their frequency, as well as the time at risk for the observed and potentially observed transition-types for each observed transition. These features of the hypnogram represent the stability of a sleep state: fewer counts along with more time at risk for transition-types originating in the same state indicate fewer exits of and longer durations in that state. Contrariwise, more counts in less time at risk indicates less state stability.

2.3 Group definitions

The task is to model the outcome of a relative measure of sleep stage transitioning as a function of group membership. In order to get transition-type specific group comparisons, the interaction between design variables representing group membership and transition-type are necessitated. For G groups and H transition-types, a model will produce $(G-1) \times H$ estimates of interest. For each transition-type h, the $G-1$

group comparisons can be seen as a "dose-response" or "trend" pattern relating the association of the underlying group defining characteristic and the stability of the starting state with respect to transition-type h.

2.4 Statistical Analysis

The estimates of focus are relative transition-type specific transition rates among the group-defined populations, which necessitate the inclusion of interactions of binary indicator variables for the transition-types and non-reference groups. In each of the following two models discussed, main effects play a different role. In a multi-state survival model stratified on transition-type, the interaction between group and transition-type essentially becomes the group indicator in that stratum, rendering the inclusion of the group main effect unnecessary, as well of course the main effect for transition-type. The baseline hazard acts as the referent group. However, in the log-linear analysis, the model is not stratified and thus the main effects of group and transition-type are included to provide backing for the interaction terms. For instance, if the main-effects were omitted in the log-linear analysis, the effect for the design variable for $g : h$ would be the transition rate for group g of transition-type h compared to all other transition-types $\{1, \ldots, h-1, h+1, \ldots, H\}$ for the reference group. That is, without stratification to restrict the comparison, the effect is not fully transition-type specific.

2.4.1 Multi-state survival models

For a survival analysis yielding a transition-type specific log-hazard $\alpha_h(t)$ and effect $\beta_{g:h}$, the Cox regression model is stratified on transition type h and regressed upon the interaction term involving h:

$$\alpha_h(t) = \alpha_{h0}(t) + g : h,$$

where $g : h$ represents the interaction terms sans the main effects, as discussed in Therneau and Grambsch (2000). Sleep viewed as a transitory multi-state system requires a stratified, recurrent event, competing risks proportional hazards model. Ultimately, the proportional hazards assumption is precariously situated by giving each transition-type a hazard and there being potentially many transition-types. A violation of this assumption can be tested. Such a test involves testing for the inclusion of variables that are the interaction of $\log(t) \times (g : h)$. If the test suggests the global inclusion is significant, then inclusion is the remedy for the violation.

2.4.2 Log-linear GEE model

The mean of a Poisson process can be modeled log-linearly with a log-offset of the total time at risk (tar),

$$\log \lambda^{(gh)} = g + h + g : h + \text{offset}\{\log(tar_h)\}.$$

The quantity $\lambda^{(gh)}$ is the rate for group g and type h and will yield relative *rates* (rr) of the overall counts between groups for transition-type h.

Noting that each individual has the same number of rows and that each (ordered) row within individual corresponds to the same attribute (transition-type), these log-linear models are well-suited for fitting with generalized estimating equations (GEEs). The method of GEE modeling is widely used, computationally fast, and can potentially model correlation structure. The $H \times H$ correlation matrix conveys the correlation of the time-adjusted frequencies of a transition-type occurring. Intuitively, negative correlations could be expected due to the competing risks nature of transition-types sharing the same starting state; whereas positive correlation could be anticipated for transition-types that share the same state as their ending and starting state. Common structures ("exchangeable" or "AR-1") do not admit both negative and positive correlations, and the unstructured specification is computation-

ally difficult. The realization that correlation is a nuisance in the GEE framework and that the analytic goals are point estimates and confidence intervals gives motivation to initially take an independent structure. Consistent estimates are produced regardless of correlation structure, and the possibility of bootstrapping subjects to correct confidence intervals is explored. Further discussion on this approach is found in section 5.

3 Pre-application considerations

3.1 Statistical challenges and solutions

The modeling task is to quantify transition-type-specific transition rates as a function of the group defining condition. To anchor the task in the display of Figure 4, consider the middle panel lasagna plot, which is organized into four SDB severity groups. We wish to quantify all rates of one color changing to another for each group and have inference for whether the rates differ between the groups. Prima facie, the task of modeling population-level transition-rates as a function of group status seems straightforward and well established, however there are nuances and subtleties that warrant special consideration in this case study. While one may be tempted to proceed by choosing a resolution, data format, and format-corresponding model – the science of sleep, data size, and computational feasibility add some twists and turns to the analysis path.

3.1.1 Hypnogram resolution and the science of sleep

Hypnogram resolution of the analytic dataset is not a convenience of choice. A higher resolution reveals a finer structure of sleep. The associations of a finer level may not persist to the coarser level, due to the combining of transition-types when states collapse as well as the complete loss of information for transitions within the collapsed

state. Figure 3 highlights some key observations: higher resolution hypnograms potentially contain more information. Comparing the higher resolution 5-stage to the lower resolution 3-stage hypnogram in Figure 3 demonstrates the correspondence of transition-types in different resolutions. The states {1,2,S} are collapsed into {N}, and thus transition-types {1R, 2R, SR} are combined into {NR} (compare the first three rows of the Poisson format at 5-stage to 3-stage resolution). Likewise, {1W, 2W, SW} are combined into {NW}, {R1, R2, RS} are combined into {RN}, and {W1, W2, WS} are combined into {WN}. Exhaustively, {RW} and {WR} remain unaffected, diametrically opposite of the intra-NREM 5-state transition-types {S1, S2, 12, 1S, 21, 2S} which have no analogue in 3-state sleep. Rooting the concept in the hypnograms of Figure 3, the intra-NREM information lost to the lower resolution includes the time at risk and the occurence of the two {21} transitions and the {12} transition.

Figure 4 demonstrates a similar loss of information at the population and group-level when states are aggregated. The bottom panel lasagna plot displays group-level temporal information, highlighting slow-wave sleep temporal dynamics for 5-state resolution. The first two peaks (dark purple) are very distinct across SDB groups and are completely lost at lower resolutions. The resolution-level stands to be impactful on the associations being modeled by virtue of the combining of states and subsequent loss of transition information. Thus, choosing an analytic resolution should not be done *a priori* but rather as a result from comparing the associations of different resolutions and determining how compatibly the underlying process is represented.

3.1.2 Data format and data size

Hypnogram resolution is a determining factor in data size and affects the two data formats differentially. As put forth in the previous section, lowering the resolution

is not an option for data reduction prior to an analysis as a full analysis of these data encompasses the analyses of varying levels of resolution. The size of population hypnogram data does not depend on the number of subjects alone, but grows as more states and thus more transition-types are considered. The higher the resolution, the larger the survival analysis dataset becomes per individual, involving greater numbers of competing risks for each observed transition. Contrastingly, the Poisson format's size is robust to the number of observed transitions and stays fixed at the number of transition-types.

Treating the data as repeated measures count data in the Poisson format, the growth is straightforward because the number of rows per individual is the number of transition-types considered. A data set of N subjects having a S-state hypnogram will have $N \times H$ rows, where $H = S(S-1)$ transition-types. Considering the size of the data set when the hypnogram data is modeled with a multistate survival model is a little more complex, as it is a function for each individual of how many of each transition-type was observed, $C_i^{(h)}$, and how many transitions-types were at-risk, $(S-1)$. Thus, the total number of rows in a multistate survival analysis would be $(S-1)\sum_{i=1}^{N}\sum_{h=1}^{H} C_i^{(h)}$.

For the SHHS, 6369 polysomnograms were processed into 5-state hypnograms. The Poisson format dataset has $6369 \times 20 = 127,380$ rows and the survival format 3,075,248 rows, which brings the average observed number of transitions to $3,075,248/(6369 * (5-1)) \approx 121$. Lowering the resolution by collapsing to a 3-state hypnogram yields smaller datasets, $6369 \times 6 = 38,214$ and $839,154$ rows for the Poisson and survival formats, respectively. Consequently, the average number of transitions observed in 3-state sleep is $839,154/(6369 * (3-1)) \approx 66$.

3.1.3 Computational feasibility and model choice

A method of analysis is proposed for each data format: Generalized Estimating Equations (GEE) for the Poisson format and a proportional hazards survival model for the time-to-event format. Each are widely available, accessible and shovel-ready with respect to the corresponding dataset. Fitting GEEs is much more computationally feasible than multi-state survival models and give strikingly similar results. This advantage of GEEs is largely due to the more compact data size of the Poisson format, robustness properties, and fewer modeling assumptions. Prudent software and hardware choices are explored (in SAS and R) and can facilitate the fitting of both models to the full data, rendering population-level associations.

4 Application to the SHHS data

A two-model analysis is conducted each on 3-state and 5-state resolution data. The two models are competitors, in a sense: the stratified, recurrent event, competing risks multi-state survival analysis stands to honor the sleep process better, but may be more computationally intensive and/or fickle in terms of algorithm convergence. The application will show GEE gives computationally faster yet similar results. The two resolutions are competitors as well. If the 3-state analogues of the 5-state estimates reflect direction, magnitude, and significance of results then consideration could be given to using the 3-state resolution of sleep, as collapsing states did not obscure finer-level effects. In addition to comparing the analogues, the 5-state resolution transition-types with no 3-state analogue must also be analyzed for effects when deciding to use exclusively a lower resolution.

The subjects are standardized for each model-resolution combination. In the Sleep Heart Health Study, 6369 subjects had polysomnograms, and 5639 had polysomno-

grams of high enough quality to be reliably processed into 5-state hypnograms. Of the 5639, 5598 had complete demographic and covariate information (age, race, rdi4p, sex, and smoking status). For the 5-state resolution, all 20 possible transition-types were formulated into a Poisson format dataset (111,960 rows) and a survival format (2,716,188 rows). For the 3-state resolution, all 6 possible transition-types were formulated into a Poisson format (33,588 rows) and a survival format (728,966 rows).

Each model will be adjusted for the covariate information of age, sex, race and smoking status as well as model the association between sleep structure and increasing SDB severity. SDB typically is categorized into 4 bins of rdi4p events/hr: [0,5), [5,15), [15, 30), [30, ∞), with [0,5) (SDB-absent) serving as the reference group.

For each transition-type, the three groups' 95% confidence intervals and point estimates of the relative rate ratios (RR) for the GEE models and Hazard Ratios (HR) for the survival models can be clustered as three vertical lines, left-to-right increasing in terms of SDB severity. Those clusters by transition-type can then be organized in plots with other clusters to give view of the modeled relationship of SDB on sleep itself. As given by the two models, these dose response clusters are visualized in Figure 5, using an entering-exiting state arrangement for the 5-state resolution. An entering-exiting state arrangement gathers all the transition-types involving one state. For instance, the top left panel of Figure 5 displays the three group comparisons for each of the 4 transition-types entering Wake and the 4 transition-types exiting Wake in the same plot. Looking at the information from a state-centric point of view helps give a sense for what SDB severity is doing to the sleep process. For instance, in the middle plot displaying the entering-exiting transition-types for Stage 2, SDB severity increases the rate of exiting Stage 2 to lighter stages of sleep (Wake and Stage 1, increasing clusters in green) and concurrently decreases the exiting from Stage 2 to REM and Stage Slow-wave (decreasing in green). A 3-state resolution

would completely omit the relationship seen for intra-NREM transitions of type 21 and 2S. Given they visually show such a strong trend gives weight to analyzing the 5-state as opposed to the 3-state. The two columns of plots look very similar, indicating that using a GEE approach with independence working correlation will give similar results to the multi-state survival approach.

To explore the connection between resolutions, analogue plots can be made. Figure 6 shows the 1 member set to 3 member set mappings for the 3-state and 5-state models. In the top left panel, we see that for most severe SDB groups, there is a discordance of significant findings in 5-state sleep (1R and SR are significantly higher rates in severe SDB relative to SDB-free; 2R is significantly lower) and they "cancel" out in 3-state sleep as seen by the NR insignificance across SDB severities. As for the other plots, we see similar shapes between the resolutions, possibly indicating that 1W and 2W are drivers of NW; R1 of RN; and W1 and W2 of NW.

5 Modeling insights

To fit the survival analysis model, prudent software and hardware choices are suggested. Regardless of OS platform, 64-bit SAS and R are recommended. To remedy a violation of proportional hazards with $\log(t)$ interactions requires R version 2.13 or later of `coxph()` for its implementation of `tt()`. Running 64-bit SAS in Windows utilizing a 3.40 GHz quad-core processor with 16 GB of RAM, the 5-state resolution GEE with independent working correlation structure of the previous section took 13 seconds compared to 8.5 hours for the multistate model (13 hours for $\log(time)$ interactions proportional hazards correction). Given the GEE gave similar results, the GEE can be used quickly and repeatedly as an exploratory tool for an investigation, and when a final model is suspected then fit the corresponding survival model or bootstrap subjects for corrected GEE intervals (bootstrapping 1000 times was on

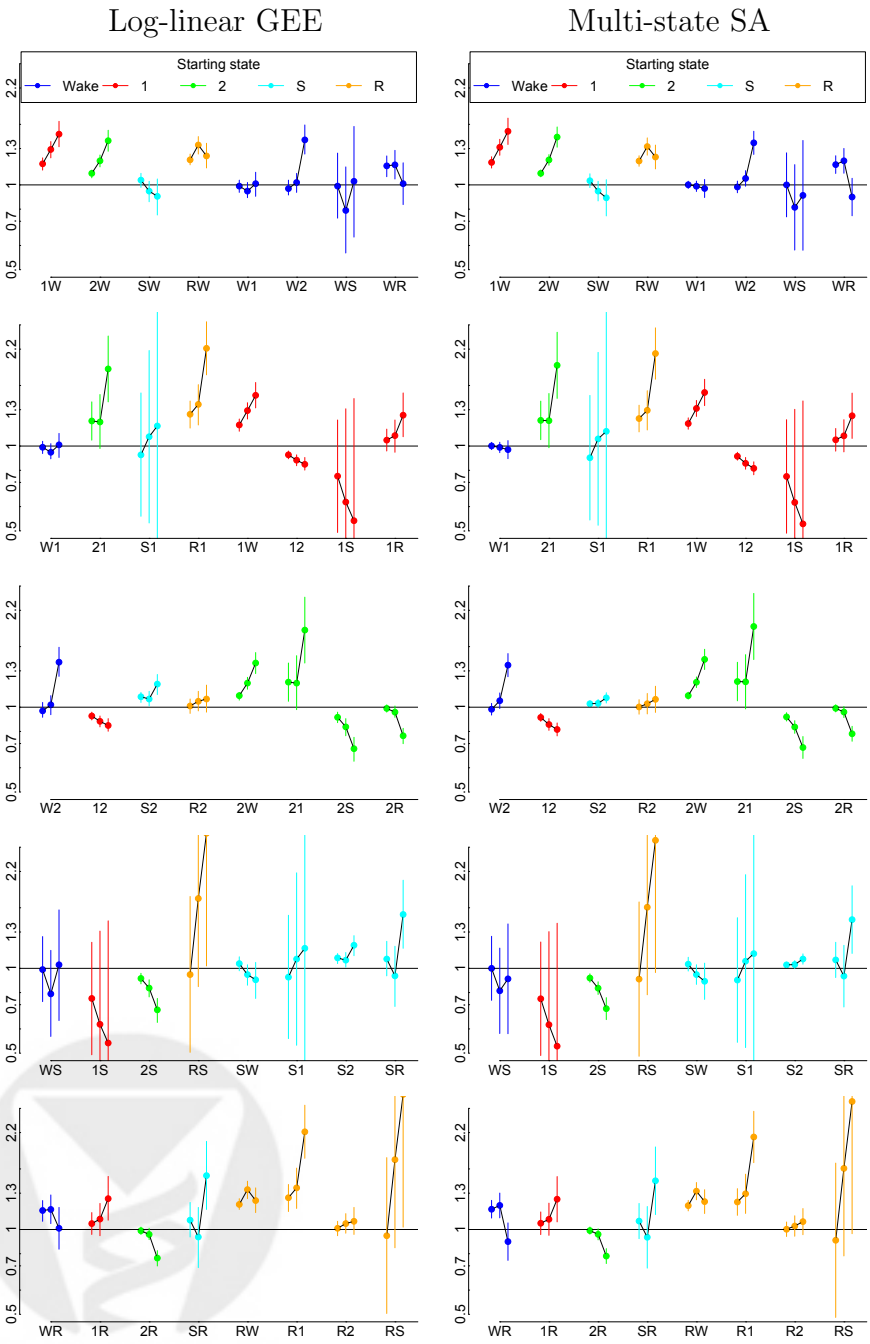


Figure 5: On the left, Relative Rates as a function of sleep disordered breathing. On the right, Hazard Ratios. Each of the 5 plots in a column is made displaying the 8 transition-types involving the entering and exiting of a state (top to bottom: Wake, Stage 1, Stage 2, Slow-wave, REM). Comparing plots within rows demonstrates how similar the estimates are between the modeling approaches.

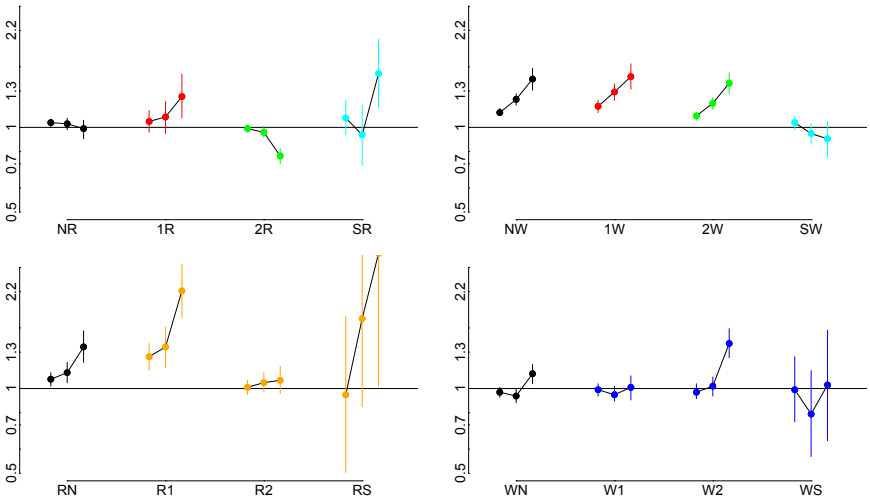


Figure 6: Compare the estimated trends of 3-state resolution (in black) to the corresponding 5-state analogues.

par computationally with fitting the survival model) (Sherman and le Cessie, 1997). Modeling of the correlation can be attempted, but proves challenging. The unstructured working correlation structure has many parameters and the specifiable common structures struggle to reflect the competing risks nature of the process. If estimating the unstructured correlation specification is prohibitive, a sample correlation matrix (or, an appropriately found nearest positive definite matrix to that data-based calculation) can be "user-specified".

Matching has the benefit of reducing the data size thereby easing computational burden, however matching changes the generalizability of the results. Data reduction via fewer subjects is not a guaranteed gain in computational feasibility. The plentitude of subjects in the full sample eases the problem of rare transition-types in this transition-type specific analysis. Not all transition-types occur equally, and the discrepancy is exacerbated at higher resolutions, which is where data reduction via fewer subjects would be most helpful in terms of computational feasibility. For instance,

transitioning from REM-Slow-wave (RS) is so rare that only 80 subjects experienced at least one and no one experienced more than 6 in their sleep (distribution: 0-6269, 1-66, 2-10, 3-2, 4-1, 5-0, 6-1). From 5-state to lower resolution 3-state sleep, {RS} is combined with more common transitions {R1, R2} to render {RN}, which has only 671 individuals experiencing 0. In addition, matching for groups of more than two levels is involved, and often diagnostic groupings are of more than two-levels. If matching is desired in a two-group situation, one can use propensity score methods (Ho et al., 2011).

The survival model parameterization requires no linear combinations, however, PROC PHREG requires that interaction variables be manually coded. The log-linear GEE predictor requires linear combinations of the group main effect and interaction terms (see web appendix). Handling multiple groups in the group-defining condition and several transition-types for when comparing resolutions takes organizational care. We advocate keeping rows of different resolutions analogous to one another for ease of comparison among the resolutions as well as making entering-exiting state plots for learning the "story" of the data analysis.

For situations that could be of interest to model counts without the temporal information as a function of group status, a third model could be fitted. The mean of a Poisson process can be modeled without the log-offset term:

$$\log \lambda_{\mathrm{rf}}^{(gh)} = g + h + g : h,$$

can easily be considered and gives only relative count information exclusive of any temporal information.

6 Discussion

The equivalence between a log-linear GLMM with log(time at risk) and multistate survival modeling assuming exponential survival times and piecewise constant hazards is well known. Thus, middle ground exists between the GEE and survival models put forth, but implementation is not as straightforward (Swihart et al., 2012). The implementation is computationally more feasible than the proportional hazards model, but requires extensive data manipulation and must be manually coded in WinBUGS.

The methods put forth stand to aid the investigation of sleep itself with sleep-related and non-sleep-related health outcomes. In the application we analyzed SDB predicting changes in sleep stage structure. Future work would be to continue down hypothesized causal pathways and connect the transition-type-specific count, time at risk, and rate features of sleep and predict a non-sleep related outcome, say, heart rate variability. Another direction of research would be to account for the longitudinal aspects of SHHS, as the sleep-EEG feature extraction work has (Crainiceanu et al., 2009). Doing so may ultimately provide better diagnostic tools and further our understanding of how sleep interacts with our health.

References

Aalen, O., J. Fosen, H. Weedon-Fekjær, Ø. Borgan, and E. Husebye (2004). Dynamic analysis of multivariate failure time data. *Biometrics 60*(3), 764–773.

Crainiceanu, C., B. Caffo, C. Di, and N. Punjabi (2009). Nonparametric signal extraction and measurement error in the analysis of electroencephalographic activity during sleep. *Journal of the American Statistical Association 104*(486), 541–555.

Fahrmeir, L. and A. Klinger (1998). A nonparametric multiplicative hazard model for event history analysis. *Biometrika 85*(3), 581.

Ho, D., K. Imai, G. King, and E. Stuart (2011). Matchit: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software 42*(8), 1–28.

Kalus, S., T. Kneib, A. Steiger, F. Holsboer, and A. Yassouridis (2009). A new strategy to analyze possible association structures between dynamic nocturnal hormone activities and sleep alterations in humans. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology 296*(4), R1216–R1227.

Kneib, T. and A. Hennerfeind (2008). Bayesian semi parametric multi-state models. *Statistical Modelling 8*(2), 169.

Norman, R., M. Scott, I. Ayappa, J. Walsleben, and D. Rapoport (2006). Sleep continuity measured by survival curve analysis. *Sleep 29*(12), 1625–31.

Penzel, T. and R. Conradt (2000). Computer based sleep recording and analysis. *Sleep Medicine Reviews 4*(2), 131–148.

Quan, S., T. Howard, C. Iber, J. Kiley, F. Nieto, G. O'Connor, D. Rapoport, S. Redline, J. Robbins, J. Samet, et al. (1997). The sleep heart health study: Design, rationale, and methods. *Sleep(New York, NY) 20*(12), 1077–1085.

Rechtschaffen, A. and A. Kales (1968). A manual of standardized terminology. *Techniques and Scoring System for Sleep Stages of Human Subjects. Washington, US Government Printing Office*.

Sherman, M. and S. le Cessie (1997). A comparison between bootstrap methods and generalized estimating equations for correlated outcomes in generalized linear models. *Communications in Statistics-Simulation and Computation 26*(3), 901–925.

Swihart, B., B. Caffo, K. Bandeen-Roche, and N. Punjabi (2008). Characterizing sleep structure using the hypnogram. *Journal of clinical sleep medicine: JCSM: official publication of the American Academy of Sleep Medicine 4*(4), 349–355.

Swihart, B., B. Caffo, C. Crainiceanu, and N. Punjabi (2012). Mixed effect poisson log-linear models for clinical and epidemiological sleep hypnogram data. *Statistics in medicine 31*(9), To Appear.

Swihart, B., B. Caffo, B. James, M. Strand, B. Schwartz, and N. Punjabi (2010). Lasagna plots: A saucy alternative to spaghetti plots. *Epidemiology (Cambridge, Mass.) 21*(5), 621–625.

Therneau, T. and P. Grambsch (2000). *Modeling Survival Data: Extending the Cox Model*. Springer.

Yassouridis, A., A. Steiger, A. Klinger, and L. Fahrmeir (1999). Modelling and exploring human sleep with event history analysis. *Journal of sleep research 8*(1), 25–36.

Young, T., P. Peppard, and D. Gottlieb (2002). Epidemiology of obstructive sleep apnea: a population health perspective. *American Journal of Respiratory and Critical Care Medicine 165*(9), 1217–1239.