Computational Statistics and Data Analysis 89 (2015) 1–11

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Modeling sleep fragmentation in sleep hypnograms: An instance of fast, scalable discrete-state, discrete-time analyses

Bruce J. Swihart[a,*], Naresh M. Punjabi[b], Ciprian M. Crainiceanu[a]

[a] *Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, 21205, United States*
[b] *Johns Hopkins University, Baltimore, MD, United States*

HIGHLIGHTS

- We explore associations of sleep-disordered breathing and sleep structure.
- 5-state hypnograms are systematically compared to 3-state.
- We analyze a community cohort of 5598 subjects (2.7 million rows total).
- We reduce analysis time from 8 hours to 30 s.

ARTICLE INFO

Article history:
Received 5 July 2013
Received in revised form 4 March 2015
Accepted 4 March 2015
Available online 14 March 2015

Keywords:
Competing risks
Multi-state
Poisson regression
Recurrent event
Sleep-disordered breathing
Stratified

ABSTRACT

Methods are introduced for the analysis of large sets of sleep study data (hypnograms) using a 5-state 20-transition-type structure defined by the American Academy of Sleep Medicine. Application of these methods to the hypnograms of 5598 subjects from the Sleep Heart Health Study provide: the first analysis of sleep hypnogram data of such size and complexity in a community cohort with a range of sleep-disordered breathing severity; introduce a novel approach to compare 5-state (20-transition-type) to 3-state (6-transition-type) sleep structures to assess information loss from combining sleep state categories; extend current approaches of multivariate survival data analysis to clustered, recurrent event discrete-state discrete-time processes; and provide scalable solutions for data analyses required by the case study. The analysis provides detailed new insights into the association between sleep-disordered breathing and sleep architecture. The example data and both R and SAS code are included in online supplementary materials.

1. Introduction

An individual's sleep is conceptualized as a hypnogram, a discrete-state discrete-time stochastic process (Fig. 1). Currently, the field of sleep science broadly generalizes a typical sleep progression as Wake (W, on the hypnogram axis) to 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 min and repeats through the night.

Judging from the three examples of Fig. 1, sleep is often more complex than the purportedly 'typical' pattern shown of the top panel. The top panel fits the generalization well, but the other two do not, with many detours and interruptions to the

* Corresponding author. Tel.: +1 443 287 8782; fax: +1 410 955 0958.
E-mail address: bruce.swihart@gmail.com (B.J. Swihart).
URL: http://www.biostat.jhsph.edu/~bswihart/ (B.J. Swihart).

http://dx.doi.org/10.1016/j.csda.2015.03.001
0167-9473/

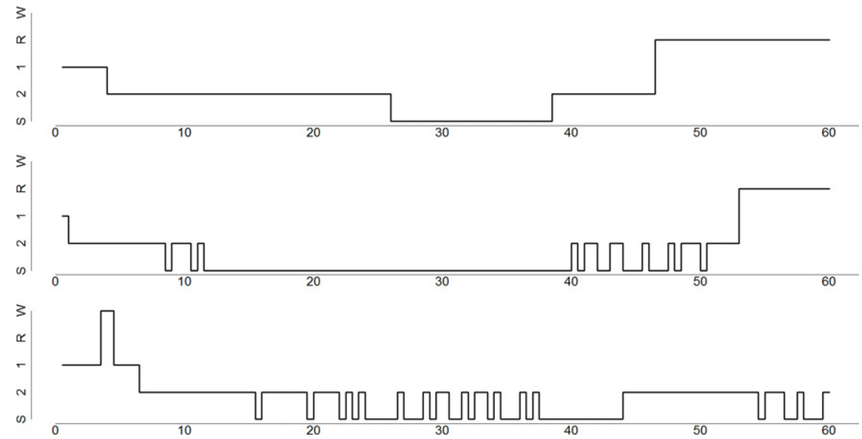

Fig. 1. The first hour of sleep for three individuals, visualized by discrete-time discrete-state spaghetti plots known as sleep hypnograms. The vertical axis displays the five stages of sleep: Wake, Rapid-Eye Movement (REM), Stage 1, Stage 2, and Stage Slow-wave labeled "W", "R", "1", "2", "S", respectively.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

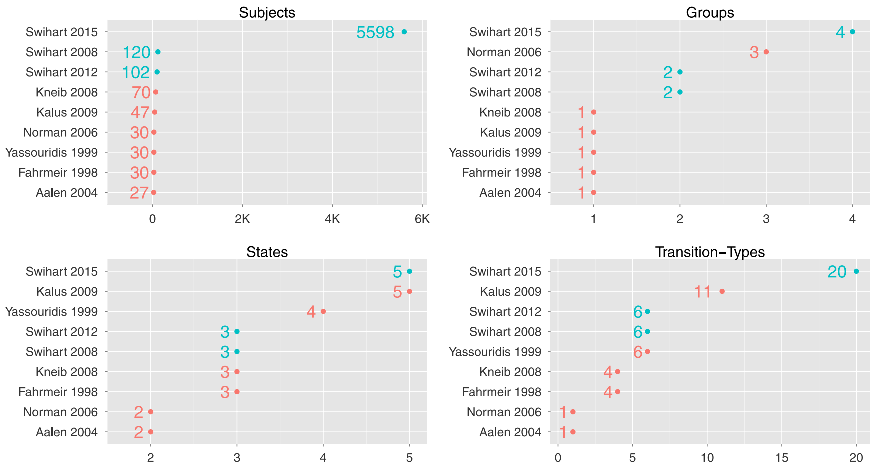

Fig. 2. A summarization of statistical methodology literature with sleep applications. Four facets are displayed: Number of Subjects, Number of Groups, Number of States, and Number of Transition-Types. On the vertical axis are the labels for the cited paper, and in each facet they are ordered based on the attribute of the facet. The proposed approach, Swihart 2015, is handling more subjects, more groups, more states and more transition-types. Green points represent a full and flat paradigm, where all $S(S-1)$ transition-types are modeled for the S states considered. This figure appears in color in the electronic version of this article. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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 s the trajectory can change to any other of the four states or remain in the current state makes for a diverse functional space for one hour snippets of three individuals, let alone for an overall sleep time of typically 7 h for the thousands of individuals in sleep epidemiology investigations.

1.1. Previous methods used in the analysis of hypnogram data

To remove potential confusion about novelty, we provide a brief review of related publications with the problems they addressed, along with a summarizing graphic (Fig. 2). Hypnogram data have been the topic of previous analytic frameworks and data applications demonstrated on modest sample sizes involving one homogeneous group or for comparing no more than two groups. For example, previous statistical methodology for sleep focused on an important clinical goal of relating time-varying hormone levels to the sleep process modeled as a reduced set of transition-types for the number of states considered (see Fahrmeir and Klinger, 1998; Yassouridis et al., 1999; Aalen et al., 2004; Kneib and Hennerfeind, 2008; Kalus et al., 2009). Norman et al. (2006) established differences in the stability of sleep and severity of sleep-disordered breathing in a sleep-runs analysis (which is a 2-state 1-transition-formulation with no intra-subject repeated events clustering). For S states, a full and flat model paradigm would include all $S \times (S-1)$ pairwise transition-types. Swihart et al. (2008) analyzed the full 3-state 6 transition-type paradigm with Poisson regression for relative transition counts and multi-state survival