SLEEP ASSESSMENT

Sleep Continuity Measured By Survival Curve Analysis

Robert G. Norman, PhD1,2; Marc A. Scott, PhD3; Indu Ayappa, PhD3; Joyce A. Walsleben, PhD3; David M. Rapoport, MD1

1Division of Pulmonary & Critical Care Medicine, New York University School of Medicine, New York, NY; 2Department of Humanities and Social Sciences, New York University Steinhardt School of Education, New York, NY

Study Objectives: To develop and demonstrate the utility of measures of sleep continuity based on survival analysis techniques.

Design: Retrospective.

Setting: University sleep laboratory.

Patients: Anonymous nocturnal polysomnograms from 10 normal subjects, 10 subjects with mild sleep disordered breathing (SDB) (apnea-hypopnea index [AHI], 15-30/hr), and 10 subjects with moderate/severe SDB (AHI > 30/hr).

Interventions: N/A

Measurements and Results: Polysomnograms were analyzed to measure the lengths of episodes of contiguous sleep and processed using several common survival analysis techniques. Using separate survival curves for each group to describe the durations of continuous epochs of sleep (sleep run lengths), statistically significant differences were found between all three groups (p<.001) as well as between the normal and mild SDB groups (p<.001), suggesting differences in the stability of sleep. Using survival regression techniques applied separately to each subject, statistically significant differences were found among all three groups (p<.001) and, more importantly, between the normal and mild SDB groups (p<.005). Similarly, estimation of sleep continuity based on the pooled sleep run data for each group also showed statistically significant differences (normal vs mild, p<.001; Normal vs moderate/severe, p<.001). In addition, the latter technique showed that changes in the “stability” of sleep could be demonstrated as runs progressed.

Conclusion: Sleep continuity measured by survival curve analysis of the lengths of runs of contiguous sleep provides a potentially useful method of quantifying sleep continuity. The results suggest that sleep becomes more stable as sleep progresses in normal subjects and those with mild SDB and less stable in subjects with moderate/severe SDB.

Keywords: Sleep disordered breathing, sleep fragmentation, sleep continuity, survival analysis, sleep runs

Citation: Norman RG; Scott MA; Ayappa I et al. Sleep continuity measured by survival curve analysis. SLEEP 2006;29(12):1625-1631.

INTRODUCTION

IT IS WELL KNOWN THAT BOTH SLEEP FRAGMENTATION1-6 AND SLEEP DEPRIVATION7-9 RESULT IN A VARIETY OF DAYTIME CONSEQUENCES. BONNET demonstrated that the degradation of sleep latency measured with a Multiple Sleep Latency Test (MSLT) in sleep fragmentation studies was directly related to the induced fragmentation rate.6 Previously, Bonnet had developed the sleep continuity theory which posits that at least 10 minutes of uninterrupted sleep is needed to serve a recuperative function.4 Sleep restriction experiments, in which subjects were restricted for 2 weeks to either 4 or 6 hours of sleep per night, have shown a progressive decrement in performance (as measured by the Psychomotor Vigilance task lapses), and cognitive function (measured by digit symbol substitution and serial addition/subtraction tasks).9

The effects of sleep deprivation and fragmentation have been investigated with enforced protocols using specific restriction patterns and fixed rates of sleep fragmentation. However, diseases which fragment sleep, such as obstructive sleep apnea hypopnea syndrome, (OSAHS) are characterized by variable sleep fragmentation rates. This results in variable continuity of sleep throughout the sleep period (i.e., the lengths of uninterrupted periods of sleep are non-constant). Despite the demonstrated effects of a disruption of sleep continuity, it is incompletely captured by most conventional analyses of sleep. Some indices, such as total sleep time (TST), time (or % TST) in each sleep stage, time (or % total sleep period, TSP) of wake after sleep onset (WASO), and sleep efficiency (SE) only capture gross changes in sleep architecture. Other indices such as sleep stage shifts (SSh, the number of transitions between sleep stages), the number of awakenings from sleep, and the arousal index (Arl, the number of arousals per hour of sleep) capture the number of potentially fragmenting behaviors which might impair continuity and have been shown to correlate with daytime function.10,11 However, they do not directly measure the continuity of consecutive epochs of sleep.

An example of the information lost in the traditional measures used to describe sleep continuity is schematized in Figure 1...
1, which shows two idealized hypnograms simplified to contain only the states of sleep and wake. Each hypnogram has the same numbers of epochs of sleep, epochs of wakefulness, and transitions between these states. By definition, the TST, SSh, and all other indices mentioned above would be equal in the two situations. In the first hypnogram the transitions are evenly spaced, while in the second the transitions are grouped at the end. The first occurrence of a contiguous sleep run (defined as a period of consecutive epochs of sleep bounded by wake) in the bottom hypnogram is nearly 4 times the length of any of the runs in the top hypnogram, a clear difference in the continuity of the sleep obtained, despite equal amounts of sleep in each situation.

Survival analysis, so named because it is often used to describe time to death, is useful for modeling the elapsed time until the occurrence of any definable event. Rather than providing a single measure of these elapsed durations, such as the median or mean time, survival analysis provides an estimate of the entire distribution of survival times within the sample. In the case of sleep continuity, the time of interest is the duration from the onset of sleep (or a specific stage thereof) to the transition to wake (or to some other stage). The period between the onset of sleep and a terminating transition can be used to define a “run” of sleep which can be characterized by its length. Using the standard techniques of survival analysis, the distribution of the durations of these runs can be characterized in several ways. We hypothesized that survival analysis techniques would prove more informative than the standard measures in describing sleep continuity. Ultimately, better measures of sleep continuity may correlate better with measures of daytime function.

The purpose of the present study was to develop and test measures of sleep continuity based on survival analysis techniques in order to capture information on the continuity of sleep that is lost when using the conventional measures of sleep. The rationale of this approach is to improve upon the measurement of sleep continuity and to provide quantifiable measures of abnormality that may be linked to clinical outcomes (e.g., sleepiness).

METHODS

Subjects

We used existing sleep stage records obtained during prior research studies. Full-night polysomnography consisting of EEG (C3/A2, O1/A2 and FZ/A2), EOG, submental EMG, anterior tibialis EMG, a lead II ECG, thoracic and abdominal motion, airflow using a nasal cannula, pulse oximetry and esophageal manometry had been performed. Sleep was scored by a single scorer according to standard criteria.12 We obtained sleep stage sequences from 10 normal subjects (with an AHI<5 and no sleep complaints and no identifiable abnormality, such as PLMs, on NPSG, age 30.4±4.7 years, BMI 23.4±3.4 kg/m²); 10 subjects with mild sleep disordered breathing (SDB) (with an AHI of 15-30/hr and no other identifiable causes of arousal, age 47.5±15.0 years, BMI 33.2±13.9 kg/m²); and 10 subjects with moderate/severe SDB (with an AHI >30/hr and no other identifiable causes of arousal age 45.0±11.1 years, BMI 42.1±8.4 kg/m²). Subjects with an AHI from 5-15/hr were excluded because it was felt that they would represent a heterogeneous group of normal subjects and subjects with mild disease. Use of the records was approved by the Institutional Board of Research Associates of the New York University School of Medicine.

Sleep Continuity Measured by Survival Analysis—Norman et al

The AHI was calculated from all apneas (airflow < 10% of baseline for >10 seconds) and all hypopneas (airflow < 50% of baseline for > 10 seconds or airflow < 80% of baseline for > 10 seconds with a 4% desaturation). For each subject, conventional indices of sleep (TST, SE, SSh, ArI, awakenings and WASO) were calculated from the sleep stage data. Stage shifts, Arousal Index and number of awakenings were expressed as a percentage of the TST and WASO as a percentage of total sleep period in order to control for differing record lengths.

New Measures of Continuity

Three measures of sleep continuity were developed: one based on nonparametric survival curves calculated from the combined data within each group and two measures based on regression analyses—one performed separately in each subject and one performed on all subjects’ data simultaneously. The first step in each of these analyses was the identification of all periods consisting of contiguous epochs of sleep (sleep runs). In this paper a run of sleep was defined using only the sequence of epoch-based sleep stages represented in the hypnogram. A run began with a change from wake to any stage of sleep. A sleep run continued until one of the following conditions occurred: 1) there was a change from stages 2, 3, 4, or REM sleep to either stage 1 or wake; 2) if the entire run consisted of stage 1 then the run terminated with a transition to wake. If a run terminated with a change to stage 1 and the next epoch was also stage 1 a new run of sleep was begun at that point.

Nonparametric Survival Curves on Individual and Grouped Data

Following the identification of sleep runs in each subject, a survival curve representing sleep continuity in that subject can be obtained using the Kaplan-Meier estimates of survival.13 Figure 2 shows survival curves generated from one normal subject and one subject with moderate to severe SDB. Note the difference in the proportion of runs lasting more than 10 minutes between the two curves (45% for the normal subject, 4% for the SDB subject).

In order to compare sleep continuity between groups of subjects, all data from all subjects in a single SDB group were pooled and a group survival curve was generated using standard statistical techniques which take into account the multiple runs of sleep in each subject.14,15 This differs slightly from the usual Kaplan-Meier survival curve, which describes time to single event for each individual but produces similar results. Each group survival curve, characterizes the overall distribution of sleep run lengths for all subjects for a single SDB group.
Modeling of Survival on Individual and Grouped Data

A number of regression based measures describing the distribution of runs of sleep (survival) are possible. This paper uses two methods: 1) a subject-specific measure of continuity derived from sleep run data from each subject; and 2) a group-specific measure of continuity derived from pooled sleep run data from all subjects.

The first method (see Figure 3) fits the distribution of sleep run durations for each subject to an exponential survival curve that models survival as $e^{-\theta T}$. The single parameter $\theta$ characterizes the shape of the survival curve for each subject as a function of run length ($T$)—basically how quickly the curve drops. Other commonly used survival curve models (Weibull and gamma distributions) were also examined and compared to the exponential fit using goodness of fit statistics. Details of fitting the exponential curve and choice of models are described in the appendix.

The second method of modeling survival fits sleep run data for all subjects simultaneously, using survival regression with the group (normal, mild SDB, or moderate/severe SDB) as the independent variable. Similar to the individual subject analysis, regressions were first fit assuming an exponential distribution of pooled run lengths. The exponential distribution, by definition, has the same risk of ending a run of sleep at each point in the run (i.e., the stability of sleep does not change). In a separate analysis, a Weibull distribution was assumed which has an additional shape parameter allowing the risk of ending a run to be dependent on the

Figure 2—This figure shows hypnograms from one normal subject (upper left) and one subject with severe SDB (upper right). Following identification of sleep runs, survival curves are generated using the Kaplan-Meier method. The survival curves are visually distinct and show greater run lengths in the normal subject than in the SDB subject. Note the difference in the proportion of runs lasting more than 10 minutes (dotted lines) between the two curves (45% for the normal subject, 4% for the SDB subject).

Figure 3—This figure shows the fit of survival regression curve based on the exponential distribution. The solid line shows the Kaplan-Meier survival curve in one subject with severe SDB. The dashed line is the fitted survival regression curve $S(T) = e^{-\theta T}$. For this subject, the value of theta is 0.632, which provides a unidimensional measure of sleep continuity.
run length. Thus, in the Weibull model sleep can become more or less stable as a run progresses. An additional term in the model, the frailty, captures intersubject differences (see appendix for details).

**Statistical Comparisons**

Comparisons between the three grouped survival curves were performed using a permutation based log rank test. Pairwise post hoc analyses were performed by the same technique using a Bonferroni adjustment for multiple comparisons. Comparisons of the conventional sleep measures and the continuity (theta) values between the groups were performed by ANOVA using Bonferroni adjusted t tests to perform post hoc comparisons of each SDB group with the normal group. Transforms to stabilize variance or produce normality were evaluated where appropriate. All statistical analyses, as well as the calculations involved in the continuity measures, were performed in the R system for statistical computing and graphics.16

**RESULTS**

Results for the conventional indices of sleep within each group are shown in Table 1. Analysis of variance revealed statistically significant differences between the groups for all parameters. Post hoc analyses revealed statistically significant differences between the moderate/severe SDB group and the normal group for each parameter. As expected, the moderate/severe SDB group had a lower total sleep time and sleep efficiency, more time spent in WASO and greater numbers of arousals, sleep stage shifts, and significant differences between the normal group and the moderate/severe SDB group. The permutation based log rank test confirmed statistically significant differences between all three curves (P<.001) and between each of the possible pairs of SDB groups (P<.001 for each pair).

Figure 4 shows the nonparametric grouped survival curves for the normal, mild SDB and moderate/severe SDB groups using the pooled sleep runs in each group. These curves demonstrate, for example, that 20% of sleep runs last more than 10 minutes in the normal subjects. In contrast, only 10% of runs last more than 10 minutes in the mild SDB group and and 2% in the moderate/severe SDB group. The permutation based log rank test confirms statistically significant differences between all three curves (P<.001) and between each of the possible pairs of SDB groups (P<.001 for each pair).

Figure 5 shows the results for the modeling of individual sleep run survival curves (characterized by the exponential parameter labeled as theta). For each group, the box plot shows the median, 1st and 3rd quartiles, and minimum and maximum values of theta. ANOVA revealed statistically significant differences among the groups (P<.001). A post hoc analysis showed a statistically significant difference between the theta values of the normal and moderate/severe SDB groups (P<.001). Unlike the analyses based on the conventional sleep indices, there was a statistically significant difference between the mean theta values in the normal and mild SDB groups (P<.005).

For the regression based on fitting all subjects’ sleep run data simultaneously, the Weibull distribution based model fit significantly better than the exponential distribution based model (likelihood ratio test, \( \chi^2=46.0, df=1, P<.001 \)). In addition, the shape parameter of the Weibull model was statistically significant (shape=0.952, P<.01). For this model, there were statistically significant differences between the normal and both the mild (P<.02) and the moderate/severe SDB groups (P<.001). When the Weibull shape parameter was determined separately for each group, statistical significance for this parameter was found in each group (P<.001). The shape parameter was less than 1 and identical (0.856) in the normal and mild groups but was greater than 1 (1.12) in the moderate/severe group. Thus, the Weibull model captured differences between the groups in the stability of sleep: for the normal and mild SDB groups, the Weibull model predicted longer sleep runs (greater stability) than the exponential model and predicted shorter runs (less stability) for the SDB group.

**DISCUSSION**

Conventional summary measures of sleep did not differ between the normal and mild SDB groups in this study, though differences between the normal and moderate/severe groups were

---

Table 1—Conventional Measures of Sleep

<table>
<thead>
<tr>
<th></th>
<th>Normal Mean (SD)</th>
<th>Mild OSAHS Mean (SD)</th>
<th>P*</th>
<th>Mod/Severe OSAHS Mean (SD)</th>
<th>P*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Sleep Time</td>
<td>385 (12.0)</td>
<td>441 (92.0)</td>
<td>.006</td>
<td>312 (56.0)</td>
<td>.041</td>
</tr>
<tr>
<td>Sleep Efficiency</td>
<td>89.2 (4.4)</td>
<td>85.4 (14.2)</td>
<td>.227</td>
<td>73.4 (10.1)</td>
<td>.006</td>
</tr>
<tr>
<td>Stage Shifts</td>
<td>21.6 (6.0)</td>
<td>27.6 (6.1)</td>
<td>.126</td>
<td>34.2 (9.0)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Arousal Index</td>
<td>14.8 (5.1)</td>
<td>23.9 (10.7)</td>
<td>.357</td>
<td>56.6 (26.3)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Awakenings/HR/min</td>
<td>3.0 (0.9)</td>
<td>4.5 (2.7)</td>
<td>.113</td>
<td>6.7 (2.7)</td>
<td>.004</td>
</tr>
<tr>
<td>TST (%)</td>
<td>7.23 (2.5)</td>
<td>15.11 (11.7)</td>
<td>.068</td>
<td>21.3 (11.7)</td>
<td>.013</td>
</tr>
<tr>
<td>WASO/TSP    %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* P-values represent comparison to the normal group, adjusted for multiple comparison.
The delineation of contiguous runs used in the analyses in this paper was based only on sleep stage scores, where an epoch of wake or stage 1 ended a run for any condition tested. Brief arousals, which have been shown to fragment sleep and produce physiologic consequences, were not included in this delineation. It might be supposed that the inclusion of arousals might alter the results. However, when we performed a repeat analysis including arousals as end-of-run criteria the nature of the results was unchanged. This may have been due to the fact that many of the post-arousal epochs were scored as sleep stage 1, thus providing the same end of run indication. In addition, the inter-rater agreement on arousals is worse than that of sleep staging and would introduce a source of unwanted variability in the run definitions.

Certainly other definitions of a run of sleep are possible. The current definition was chosen based on the authors’ experience and the transitional nature of stage 1 sleep. Other possible definitions of sleep runs might take into account transitions among the deeper stages of sleep, treat REM and nonREM sleep differently, or might require a change to stage wake to terminate the run. The determination of the “best” definition for a run of sleep will require comparison to an appropriate criterion or outcome variable and is the subject of further research. Nonetheless, with the definition used in the current analysis, differences in sleep continuity between normal subjects and those with mild SDB could be demonstrated.

An exponential distribution of sleep run lengths between arousals and/or awakenings was shown in our results for the survival curves fit to each individual. This distribution is often termed the “memoryless” probability distribution and has interesting consequences. Under an exponential distribution, the risk of a run of sleep ending is the same at all points within the run. This means that sleep stability does not change with time spent asleep. Thus, under the exponential model the groups differed only in the value of the (constant) risk of sleep run termination. However, there was evidence from the Weibull regression applied to the grouped data that the run length distribution was not that of a pure “memoryless” exponential. The better fitting Weibull survival regression produced a statistically significant shape parameter that represents changing stability of sleep during a run of sleep. When an analysis was performed separately for each group, the normal and mild group demonstrated a decreased risk of run termination as the run progressed while the moderate/severe group demonstrated an increasing risk of ending a sleep run as the run progressed. This was likely due to the existence of repetitive apneas (in the latter group) that produce sleep fragmentation and are more likely to recur the longer they have been absent. To put the effect of the shape parameter into perspective, the exponential model indicates that only 21% of sleep runs last 10 or more minutes in the normal subjects while the Weibull model predicts that 30% of the runs will last 10 or more minutes. Moreover, for the moderate/severe SDB group, the exponential model predicts that 19% of runs will last 5 or more minutes while the Weibull model predicts only 11% of runs will last this long. There is face validity, based on clinical experience, to these differences in sleep continuity described by the Weibull model.

In this paper, three different methods of analyzing sleep continuity are presented. The choice between these methods depends on the goals of the analysis and the relative strengths and weaknesses of each. The grouped curve analysis makes no assumption about the form of the survival curve, is relatively resistant to the effect of outliers and provides for tests of group differences. However, this technique does not provide a simple summary measure to each subject, or to each group, describing their sleep continuity and uses statistical methods that are not present in all statistical packages. The first regression-based method, the fit of an exponential survival curve to each subject, is mathematically simple and produces a measure of continuity for each subject that can be used for hypothesis testing, measures of association and models of relationships. For example, a model of age, gender and sleep continuity as predictors of MSLT measured sleepiness could be...
easily performed using the continuity (theta) values. However, the incorporation of the uncertainty associated with the measure is more difficult (see appendix). The final regression-based method, the use of a grouped survival regression, provides a statistically valid model for group comparisons (or other factors of interest) that incorporates the uncertainty of estimation into the model. In addition, it is likely that the use of more data in this model (than in the individual subject regression fits) allowed for the detection of the statistically better fit of the Weibull model, with its non-constant risk of ending a run. Even though this method uses statistical techniques that are easily implemented in only a few statistical packages, we believe it offers an important lesson regarding the scoring/tabulation of sleep and the lost information which may have an effect on outcomes.

In conclusion, survival analysis techniques can be used to produce potentially useful measures of sleep continuity that capture differences among subjects that are not identified by the conventional measures used to summarize sleep. The demonstration of the usefulness of these techniques will require further investigation, but we speculate that they may play a role in: 1) the clinical understanding of causative factors of excessive daytime somnolence and the effect of severity of illness; 2) providing an index of improvement in sleep continuity, especially in pharmaceutical trials where subtle changes (which are not tracked by conventional measures) may be valuable; 3) other situations in which a sensitive measure of sleep continuity is desired.

REFERENCES

13. Kaplan EL, Meier P. Nonparametric estimation from incomplete ob-
APPENDIX

1) Fitting an exponential curve to each subject

In fitting the exponential model to each subject’s sleep run distribution, we used a least squares’ approach. An alternative would be to use a maximum likelihood estimation for theta, which is provided by the inverse of the average run length (the variance of theta is the square of this value). In the present data, the maximum likelihood estimation produced a value of theta that was identical to the regression approach. However least squares estimates may be preferable as they produce values for theta that are more resistant to outliers than the maximum likelihood estimates of theta. For these data, the exponential model resulted in substantially smaller mean standard errors (0.027) than the gamma (0.094) or Weibull (0.093) based models and was thus preferable.

2) Effect of multiple runs of sleep within each subject on regressions performed on grouped data.

Although fitting a regression model to the pooled data from all subjects is superficially similar to fitting a regression model to a single subject, there are additional considerations when dealing with repeated measures. When repeated measures from each subject are combined together, models must deal with the induced intra-subject correlation among measures created by the repeated measurements. In the survival regression models presented here, an additional term, the frailty (or unobserved heterogeneity), was used to capture these intra-subject correlations. This frailty term, though statistically significant, is not of physiologic interest in this analysis and is treated as a nuisance term and left uninterpreted.

3) Evaluation of the impact of the uncertainty in $\theta$ on tests of group differences

In the group analyses performed on the values of the exponential parameter theta determined for each subject it was implicitly assumed that the $\theta$ values are measured values with unknown error. However, the uncertainty (error) in the value of $\theta$ is known for each subject (as a result of its estimation procedure). To evaluate the impact of this uncertainty on the group differences a Monte Carlo experiment was performed. For each subject, a Gaussian sampling distribution for theta was assumed, with mean and variance obtained from the estimation procedure for theta. This sampling distribution represents the possible values of theta after incorporating the uncertainty of its value. The Monte Carlo experiment consisted of 500 repeated evaluations. For each iteration of the experiment, a randomly sampled value for $\theta$ was chosen for each subject from that subject’s sampling distribution. From these sampled values for all subjects, a T test comparing the normal and mild SDB group was performed including the same Bonferonni adjustment used in the initial group comparisons. Fifty three percent of the samples resulted in a $P$ value of <.01 and 97% had a $P$ value less than .05. Thus, the statistical significance of the difference between the normal and mild SDB groups was resistant to any affect of the uncertainty in the measurement of $\theta$. 

SLEEP, Vol. 29, No. 12, 2006