Pilot Study of Sleep/Wake Classification by Leg-Worn Actigraphy

Thomas Ferree, PhD, and Shai Gozani, MD, PhD; NeuroMetrix Inc., Waltham, MA, USA
John Winkelman, MD, PhD; Departments of Psychiatry and Neurology, Massachusetts General Hospital, Boston, MA, USA

INTRODUCTION

Motivation. Actigraphy is used widely to measure and study sleep characteristics. Common parameters that are clinically useful include time in bed, total sleep time, sleep onset, and wake after sleep onset. Most actigraphy devices are worn on the wrist. The goal of this study is to assess the accuracy of sleep/wake classification by leg-worn actigraphy in a clinical population.

Study Device. Quell® (NeuroMetrix, Inc., Waltham, MA) is a transcutaneous electrical nerve stimulation (TENS) device worn on the calf for the treatment of chronic pain. Because pain affects sleep quality, measuring sleep metrics accurately can provide valuable feedback to the user on the efficacy of pain therapy to improve sleep. In addition, real-time detection of sleep duration and quality can improve the user experience and efficacy, e.g., by modifying the stimulation amplitude and scheduling.

Sleep Classification Algorithm Development. The device contains a 3-axis accelerometer. Signal processing algorithms and sleep/wake state classifiers were developed on a data set comprised of 115 nights of sleep from 60 unique subjects, mostly adults without a known sleep disorder. This pilot validation tests classifier performance prospectively against polysomnography (PSG) in a clinical population with presumed sleep disorders.

METHODS

Subjects. Patients referred to the Massachusetts General Hospital Sleep Center were eligible for the study. Inclusion criteria were: age 18-75 years, expected wake time <30% of total recording time, and probable periodic leg movements of sleep. Exclusion criteria were: circadian sleep disorders, pregnancy, and implantable medical devices.

Study Design. Thirteen subjects were recruited; three were dropped due to technical errors in recording. Of the remaining 10, half wore one device, and the other half wore two devices, one on each leg. PSG data were collected following clinical standards and scored by the same technician.

Recording Session: Participating patients were prepared for PSG. The technician applied the Quell device(s) to the calf, and pressed the button to start recording.

PSG Data. PSG data were scored in 30-second epochs, with an arbitrary start time. Scored data were de-identified then provided to NeuroMetrix for further analysis.

METHODS, CONT'D

Quell Data. Accelerometer data were acquired at 50 Hz and processed to one-minute epochs aligned with clock time. Orientation and activity each epoch were stored in the device for further analysis. Epochs were scored awake if the mean activity exceeded a threshold fixed for all subjects. That threshold was set such that the average sleep efficiency in the training population was 95%.

Data Alignment. For minute-by-minute comparison, PSG data were down-sampled to even minutes by assigning each minute the score for the 30-second epoch covering the majority of that minute. In two subjects, detailed visual inspection revealed that epoch times were shifted. For these devices, their epoch times were corrected by 5 and 24 minutes. The figure below shows the aligned data for one subject which required the correction 24 minutes.

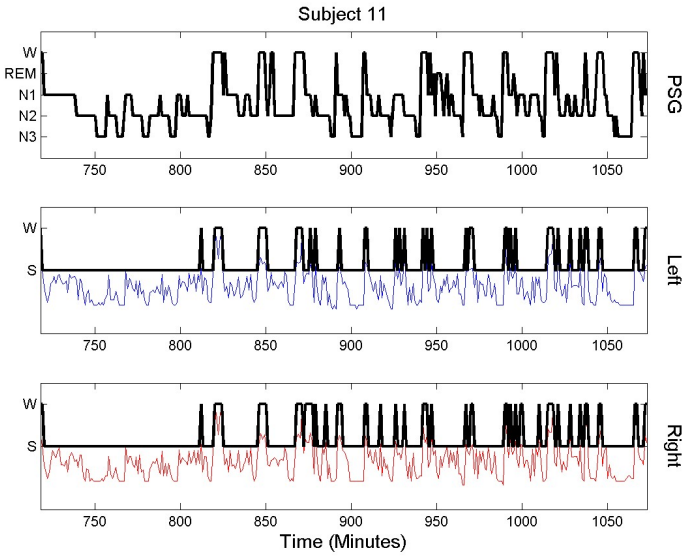

Figure. Data from one patient wearing Quell devices on each leg. Top panel: sleep stages from scored PSG. Bottom panels: $\log_{10}$ activity (blue, red) and actigraphy-based sleep/wake classification (black). The comparison window is from the time of Lights Out to Lights On as marked in PSG. See the minute-by-minute similarity between modalities and between limbs.

RESULTS

	1	2	3	4	6	7	8	11	12	13
ROC	0.82	0.72	0.83	0.82 / 0.80	0.66 / 0.74	0.76 / 0.76	0.67	0.81 / 0.79	0.56 / 0.62	0.69
ACC	0.88	0.85	0.83	0.86 / 0.87	0.89 / 0.89	0.88 / 0.88	0.91	0.89 / 0.86	0.82 / 0.82	0.85
SENS	0.95	1.00	0.96	0.89 / 0.89	0.99 / 0.99	0.92 / 0.93	0.98	0.94 / 0.91	0.97 / 0.97	1.00
SPEC	0.46	0.21	0.38	0.50 / 0.57	0.04 / 0.02	0.46 / 0.37	0.14	0.59 / 0.61	0.11 / 0.14	0.24

Prospective detector performance. For each device, minute-by-minute agreement in sleep/wake classification was quantified by the area under the receiver operating characteristic curve (ROC), accuracy, sensitivity and specificity (see **Table**). For the group, performance was summarized by the median and median absolute deviation: ROC 0.76 (0.06), accuracy 0.87 (0.02), sensitivity 0.96 (0.03), and specificity 0.37 (0.20).

Post-hoc Analysis. In post-hoc analysis with the same activity threshold, applying a peaked averaging window like that used by Cole et al. (1992) improved the group performance to ROC 0.89 (0.03), accuracy 0.87 (0.03), sensitivity 0.96 (0.03), and specificity 0.53 (0.25); note that specificity was much improved.

DISCUSSION

In this prospective study, the Quell device compared favorably with the gold standard PSG in a clinical population: a non-trivial finding, given that it algorithms were developed in a normal population without PSG for training. ROC quantifies the quality of the detector independent of threshold. For the selected threshold, accuracy was comparable to inter-rater variability in the PSG literature, and sensitivity was very good because activity is usually low during true sleep. Specificity is more variable because some subjects lie very still while awake. Improving and validating our ability to track sleep in users with normal and abnormal sleep can provide valuable feedback to users about the efficacy of pain management.

ACKNOWLEDGEMENTS

The authors thank Martin Moynihan for Quell programming, the PSG technician and study coordinators in the MGH Sleep Lab, and Xuan Kong for helpful suggestions.

Presented at American Academy of Neurology (April 21-27, 2018, Los Angeles, CA)
Contact: tom_ferree@neurometrix.com