Investigating the proper adaptation of sleep scoring algorithms for pediatric trials

Georg Dorffner[1,2], Georg Gruber[2], Marco Hirner[2], Lukas Pinka[1], Patrick Erber[1] , Tawab Shoja[1], Ruth Luigart[3], Barbara Schneider[3] Peter Anderer[2]

[1:] Medical University of Vienna, Section for Artificial Intelligence, Vienna, Austria

[2:] The Siesta Group, Vienna, Austria, [3:] Children's Hospital St. Marien, Landshut, Germany

METHODOLOGICAL QUESTION

Can adapting algorithms for computer-supported sleep scoring lead to improved automated scoring in infant and child data?

INTRODUCTION

Computer-supported scoring is gaining importance in clinical trials when it comes to objectively measuring a drug's effect on nocturnal sleep, mainly due to the high reliability of such methods. While validated systems exist for approved use in adults, the validity of the underlying algorithms for infants, children or adolescents still needs to be proven. Sleep scoring, by standards of the American Academy for Sleep Medicine, is based on identifying features in electrophysiological recordings (EEG, EOG, EMG) and assigning sleep stages to epochs of 30s duration. With the exception of neonates, the qualitative criteria for such features and their classification into stages are largely the same for pediatric subjects as compared to adults. However, the quantitative basis of the EEG data changes considerably during neural development requiring adaptations in any algorithm that is to automatically detect sleep stages. Hence, our goal was to adapt scoring algorithms developed for adults to make them applicable to pediatric data.

DISCLOSURES

Georg Dorffner, Marco Hirner, Georg Gruber and Peter Anderer are employees and shareholders of The Siesta Group, a service provider for measuring electrophysiological signals including sleep in clinical trials.

MEDICAL UNIVERSITY OF VIENNA

METHODS

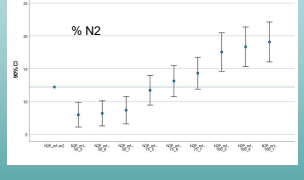
Based on a data set of around 150 anonymized recordings from a children's sleep laboratory, manually scored by two independent experts, covering several age groups from 1 to 12 years of age, we investigated the applicability of previously developed sleep scoring algorithms (1) to such data. Our hypothesis was that the main quantitative characteristics of infant electrophysiology can be reduced to two important values: (a) the average amplitude of the EEG, which can be expected to be higher in younger children, and (b) the cutoff value between alpha and theta bands in a spectral description of the data, which is expected to move to lower frequency values in younger children. While these effects can be expected to be of continuous nature, for better statistical comparison we considered several discrete values for (a) and (b) and divided the age range into three age groups known from literature to reflect major phases in brain and sleep development.

Based on that hypothesis, we applied our standard sleep scoring algorithm in 9 different settings, by varying (a) the average amplitude (original, scaled by a factor of 75% or 50%) and (b) the cutoff between alpha and theta (7Hz, as in the original algorithm, 6Hz, or 5Hz). We validated the results by calculating a set of standard sleep endpoints (sleep efficiency, percentage of time in each sleep stage) and comparing the values of these endpoints to those derived by the visual expert scorers, in terms of bias (average deviation between the two scoring methods) and consistency (correlation coefficient between the two methods). Subsequently, we selected the one setting, out of the nine different ones applied, that overall resulted in the smallest bias and the largest correlation, with an emphasis on the most important distinctions REM (rapid eye movement sleep) vs. non-REM and light vs. deep sleep. The band cutoff value was expected to largely influence the former distinction, whereas the amplitude scaling factor was expected to affect the latter.

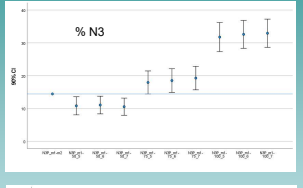
RESULTS

Results demonstrated that, indeed, varying the two main parameters led to much improvement in automated scoring, especially for younger age groups, as compared to the visual baseline. For all three age groups considered, which are known from literature to be relative stable periods (1-5 years = pre-school, 5-9 years = early school ages, 9-12 years = final years of neural plasticity) the optimal configuration turned out to be a 50% amplitude scaling and a 5Hz cutoff value for alpha/theta. With respect to the most important sleep stages, the results for bias and correlation reached a range (as quantified by 90% confidence intervals) comparable to that of the interrater agreement in the data set.

As an example, figures 1-3 show confidence intervals for three major sleep variables (percentage of stages N2, N3 and R) benchmarked against the average deviation of the two visual scorers, depending on the 9 different settings in the age group 5-9 years.



THE siesta GROUP



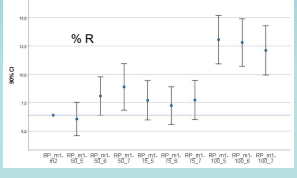


Fig. 1-3: 90% confidence intervals for three major sleep variables (percentage in stages N2, N3 and R) in the age group 5-9 years are depicted. The blue line represents the average deviation between the two visual scorers. If a CI lies below that line, the scoring can be considered statistically equivalent

CONCLUSION

These results constitute an important step toward full validation of computer-supported scoring of infant and child data in that the time needed for visual expert review can be reduced considerably. This points toward the availability of efficient and reliable sleep scoring also in pediatric CNS trials.

REFERENCE

(1) Anderer, P., et al. (2010). "Computer-assisted sleep classification according to the standard of the American Academy of Sleep Medicine: validation study of the AASM version of the Somnolyzer 24 x 7." Neuropsychobiology 62(4): 250-264.

Web: www.thesiestagroup.com

Contact: georg.dorffner@meduniwien.ac.at