P382 | Inter- and intra-expert variability in sleep scoring: comparison between visual and automatic analysis

V. Muto[1]; C. Berthomier[2]; C. Schmidt[1,3]; G. Vandewalle[1]; J. Devillers[1]; S.L. Chellappa[1]; C. Meyer[1]; C. Phillips[1,4]; P. Berthomier[2]; J.P. Prado[2]; O. Benoit[2]; M. Brandewinder[2]; J. Mattout[5]; P. Maquet[1,6]

[1]*GIGA-Cyclotron Research Centre-In vivo Imaging, Sleep Research Group, University of Liège, Liège, Belgium,* [2]*PHYSIP SA, Paris, France,* [3]*Psychology and Neuroscience of Cognition (PsyNCog), University of Liège,* [4]*Department of Electrical Engineering and Computer Science, University of Liège, Liège, Belgium,* [5]*Lyon Neuroscience Research Center, INSERM U 1028, CNRS UMR 5292, University of Lyon 1, Lyon, France,* [6]*Department of Neurology, CHU Liège, Liège, Belgium*

Objectives/Introduction: Visual sleep scoring (VS) is affected by inter-expert (difference in scoring between experts working on the same polysomnographic recordings) and intra-expert variability (evolution in the way to score of a given sleep scorer when compared with a reference). Here our aim was to quantify inter and intra-expert sleep scoring variability in a group of 6 experts (working at the same sleep center and trained to homogenize their sleep scoring) by using the automatic scoring algorithm Aseega (AS) as a reference.

Methods: Polysomnographic data were collected in 24 healthy young male participants (mean age 21.6 ± 2.5 years) and scored according to the AASM criteria. The first data set (DS1) was composed of 4 recordings, each one scored by 6 experts and AS (28 scorings). Other 88 recordings (DS2) were scored few weeks later by the same experts and AS (176 scorings). The epoch-by-epoch agreement (concordance and Cohen kappa coefficient) was computed between VS and AS.

Results: Inter-expert agreement on DS1 decreased as the number of experts increased, from 86% for mean pairwise agreement down to 69% for all 6 experts. Adding AS to the pool of experts narrowly changed the kappa value, from 0.81 to 0.79. Intra-expert variability was highlighted by the systematic decrease of the agreement between AS and each single expert across datasets (-3.7% on average).

Conclusions: Whatever the approach used, visual scoring induces inter- and intra-expert variability. Even if autoscoring neither provides any ground truth, nor can improve the inter-scorer agreement (as several automated methods exist), it can efficiently cope with the intra-scorer variability, since the AS used is perfectly reproducible and largely insensitive to experimental conditions. These properties are mandatory when dealing with large dataset, making autoscoring methods a sensible option. A number of automatic methods are currently available; however precise assessment and comparison on common datasets using common metrics remain an open question.

Disclosure: C. Berthomier, P. Berthomier and M. Brandewinder have ownership and directorship in Physip, and are employees of Physip. This study was funded by Fonds National de la Recherche Scientifique (Belgium), University of Liège research funds, Swiss National Science Foundation (# 310030_130689) and the European Research Council (ERC).

P383 | Guidelines for the application of the objective sleepiness scale for drowsiness assessment

J. Taillard[1]; P. Berthomier[2]; M. Brandewinder[2]; C. Berthomier[2]

[1]*USR 3413 SANPSY, CNRS/Université de Bordeaux, Bordeaux,* [2]*PHYSIP SA, Paris, France*

Objectives/Introduction: Electroencephalography (EEG) remains the reference method to study drowsiness, but only few objective methods have been proposed based on the visual analysis of wake EEG. Muzet proposed the Objective Sleepiness Score (OSS) [1]. For each 20s-epoch this scale defines 5 states of drowsiness based on the visual analysis of EEGs (C3, O1, P3, Fz) and EOGs. An automatic algorithm (AA) following a single-EEG approach was developed and validated using a database visually scored by Muzet (VA1) using OSS [2]. The objective of this study was to improve and facilitate the use of the OSS by proposing guidelines that explicit the visual scoring criteria.

Methods: 19 volunteers (30 ± 6 years, 12 women) were recorded during 2 h of driving in a simulator, with a Cortical Brain State (CBS) test as a biocalibration at the beginning and end of the session. All recordings (8192 scoring decision epochs) were scored by VA1, which stands as the reference, and by AA. Two sample recordings of 422 and 438 epochs were randomly chosen and visually scored by another expert (VA2).

VA2 applied two scoring procedures: first, literal OSS scoring rules as stated in [1], then adapted the scoring criteria of EEG by taking into account the individual characteristics deriving from the biocalibration. The epoch-by-epoch agreement (concordance, C, and kappa coefficient, K) was calculated.

Results: The agreement between AV1 and AV2 increased from 15% (K = -.40) when applying literal rules to 83% (K = .60) when applying adapted rules. When including AA, the overall agreement variated from K = -.12 to K = .53.

Conclusions: Taking into account the CBS observed during the biocalibration phase is key to enable the scoring of drowsiness using OSS. Alpha and theta rhythms used to determine drowsiness states must be defined by their characteristics as they appear during the CBS test.

[1] Muzet et al. Preventing driver drowsiness at the wheel: can steering grip sensor measurement contribute to its prediction? Proc. of 4th Eur. Congress and Exhibition on Intelligent Transport Systems and Services, Budapest,24-26 May 04.

[2] Berthomier et al., Real-Time Automatic Measure of Drowsiness based on a Single EEG Channel. J. Sleep Res., 17:P434, 2008.

Disclosure: Nothing to disclose.