Insufficient Evidence for the Use of Automated and Semi-Automated Scoring of Polysomnographic Recordings

Gary K. Zammit, PhD

Sleep Disorders Institute, Clinilabs, Inc., New York, NY

THE RECENT MANUSCRIPT ENTITLED EVALUATION OF AUTOMATED AND SEMI-AUTOMATED SCORING OF POLYSOMNOGRAPHIC RECORDINGS FROM A CLINICAL TRIAL USING ZOLPIDEM IN THE TREATMENT OF INSOMNIA (Sleep 2007;30:1562-1574) is the report of a study that compares automated and semi-automated scoring to manual sleep-stage scoring in healthy adult subjects. This report does not adequately address aspects of the study design that limit the generalizability of the data, nor does it accurately portray the limitations of automated or semi-automated polysomnographic scoring. The authors’ conclusion that automated or semi-automated polysomnographic scoring offers an alternative to “costly, time consuming, and intrasite and intersite variable manual scoring” is a claim that is not supported by the investigation, which did not include any empiric assessment of cost, time, or variability in visually scored polysomnographic data.

The data submitted to analysis in this study were obtained from a clinical trial of healthy subjects with no history of sleep, medical, or psychiatric disorders. Therefore, the study results cannot be generalized to multicenter trials that involve clinical populations or the screening of subjects with disturbed sleep. Further, these data were collected using highly standardized methods and processed under controlled conditions in a core polysomnography laboratory prior to their release to the authors. Only polysomnograms that met the core laboratory’s standards were selected for inclusion. Therefore, the automated analyses used in this study were performed following the careful sanctioning processes of the core laboratory, limiting the generalizability of these findings only to data obtained under those processes.

The AASM Manual for the Scoring of Sleep and Associated Events represents the current standard in the field. This manual provides specifications for the visual scoring of sleep that retain much of the framework of the original Rechtschaffen and Kales manual. A recent review of visual scoring by R & K criteria finds that the interrater reliability between human sleep-stage scorers is generally substantial (Cohen’s κ: 0.68-0.87) versus the modest reliability for fully automated scoring (κ: 54.8%-62.7%) reported by Svetnik et al. Performance of automated and semi-automated scoring is considerably poorer when considering their percentage agreement with visual scoring of Stage 1 sleep (16.6%-30.6%), a critical factor in determining outcome variables in clinical trials such as latency to persistent sleep or wakefulness after sleep onset. The current acceptability of any computer-assisted scoring method is further belied by the Standards for the Accreditation of Sleep Disorders Centers, which require that automatically scored tracings be reviewed on an epoch-by-epoch basis for accuracy. The fact that epoch-by-epoch review alone is sufficient to visually stage sleep renders automated scoring unnecessary. Finally, the potential for error when using automated processes is notable, given that up to 53.3% of the epochs scored using an automated system may need to be modified by full manual review, and the percentage of agreement reported for semi-automatic scoring may be lower than that obtained for visual scoring alone, even when the polysomnograms of normal healthy subjects (not clinical patients) are being reviewed.

The development of computerized scoring technologies is a worthwhile endeavor. However, before new technologies are applied in clinical practice and research, it is critically important that they be fully validated, reliable, and accepted as standards of practice, with clear guidelines for their use. Although the work of Svetnik et al provides one step toward that goal, it presents insufficient evidence to support the use of automated or semi-automated sleep stage scoring in polysomnographic processing. The currently accepted practice of visual sleep stage scoring in a core polysomnography laboratory remains the standard for use in clinical trials.

REFERENCES


2. Walsh JK, Deacon S, Dijk DJ, Lundahl J. The selective extrasynaptic GABAA agonist, gaboxadol, improves traditional hypnotic

