ISCTM Working Group „Sleep Methodology in Clinical Trials"

Meeting, Oct 22, 2025, Amsterdam, NL

Chairs: Margaret Moline, Georg Dorffner
Major contributor: David McLaughlin

Participants: Julie Adams, Sam Berry, Martina De Lillo, Chris Edgar, Michaela Gold, Nanco Hefting, Michael Lagler, Dave Matthews, Niall McGowan, Sofie Mesens, Matthias Mohse, Thomas Pollmächer, Corey Puryear, Gary Zammit, Peter Zhang

The chairs reiterated the original goals of the WG, presented the status quo of the literature search exercise and explained the next steps toward the first two deliverables of the working group:

Deliverable 1: Review paper on the current evidence for objective sleep endpoints as biomarkers for neurological or psychiatric disorders in CNS trials

- Summaries for 6 disorders (AD, MDD, MS, PD, PTSD) and a set of original papers (in pdf: AD, GAD, MS, PD, PTSD, SZ) had been received
- David McLaughlin presented an AI-based analysis of those summaries (file: "ISCTM Sleep Structure Comparison 12NOV2025.xlsx" in folder "Draft Mansucript") highlighting the remaining gaps. In particular, he suggested to also include the focus on sleep microstructure variables.
- Georg Dorffner presented an AI-based summary analysis in tabular form based on all original papers (file: LiteratureResearchResults-unvalidatedDraft-Oct25.xlsx in folder "Draft Manuscript")
- New volunteers to complete the exercise were found:
    - Niall McGowan to perform exercise for ADHD
    - Sofie Mesens to focus on pediatric data
    - Julie Adams for epilepsy and sleep
    - Martina de Lillo to help in Parkinson's disease

- Homework for all volunteers (**until Dec 15, 2025**):
    - Based on the summary by David McLaughlin:
      Please see whether you can fill the gaps identified and potentially re-focus your research
    - Based on the tabular summary of identified papers by Georg Dorffner:
      Please check veracity of each entry
      Potentially extend list of papers
    - Complete analysis and/or paper upload for MDD, BD

Deliverable 2: A consensus paper on criteria and validation strategies for measurement instruments, as compared to gold standard (PSG)

- Again, several measures of performance (intraclass correlation, Bland-Altman statistics including bias) and the framework of equivalence and/or non-inferiority testing was proposed. These require proper tolerance intervals for defining how large a deviation between a novel instrument and the gold standard PSG is acceptable.

- To arrive at proper tolerance interval, the suggestion is that we orient ourselves along the well-known inter-rater variability when PSG recordings are scored by several raters. Indications for such variability can again be found in literature (see below).
- The goal of the working group is thus to find consensus on such tolerance intervals for the main sleep endpoints identified in deliverable 1.
- Some open questions in this context are:
    - What defines tolerance: is it the worst possible result in a multi-rater study of an objective measure such as PSG, or their average deviation, or something else)?
    - Should we differentiate between different patient and/or drug groups when defining tolerance intervals?
    - Does treatment effect play a role here?
    - What about occasional outliers – should they be part of the criteria
- Homework for all (until **Feb 1, 2026**)_:
    - Please extend the list of references on inter-rater variability (see below), especially toward addressing different patient groups, effect sizes and the like.
    - Please add potentially other performance measures and/or statistical validation frameworks that might be important for our endeavor.

References (inter-rater variability):

Cesari, Matteo, Ambra Stefani, Thomas Penzel, Abubaker Ibrahim, Heinz Hackner, Anna Heidbreder, András Szentkirályi, u. a. 2021. „Interrater Sleep Stage Scoring Reliability between Manual Scoring from Two European Sleep Centers and Automatic Scoring Performed by the Artificial Intelligence–Based Stanford-STAGES Algorithm“. *Journal of Clinical Sleep Medicine* 17 (6): 1237–47. https://doi.org/10.5664/jcsm.9174.

Danker-Hopfe, Heidi, Peter Anderer, Josef Zeitlhofer, Marion Boeck, Hans Dorn, Georg Gruber, Esther Heller, u. a. 2009. „Interrater Reliability for Sleep Scoring According to the Rechtschaffen & Kales and the New AASM Standard“. *Journal of Sleep Research* 18 (1): 74–84. https://doi.org/10.1111/j.1365-2869.2008.00700.x.

Magalang, Ulysses J., Ning-Hung Chen, Peter A. Cistulli, Annette C. Fedson, Thorarinn Gíslason, David Hillman, Thomas Penzel, u. a. 2013. „Agreement in the Scoring of Respiratory Events and Sleep Among International Sleep Centers“. *Sleep* 36 (4): 591–96. https://doi.org/10.5665/sleep.2552.

Malhotra, Atul, Magdy Younes, Samuel T. Kuna, Ruth Benca, Clete A. Kushida, James Walsh, Alexandra Hanlon, Bethany Staley, Allan I. Pack, und Grace W. Pien. 2013. „Performance of an Automated Polysomnography Scoring System Versus Computer-Assisted Manual Scoring“. *Sleep* 36 (4): 573–82. https://doi.org/10.5665/sleep.2548.

Nikkonen, Sami, Pranavan Somaskandhan, Henri Korkalainen, Samu Kainulainen, Philip I. Terrill, Heidur Gretarsdottir, Sigridur Sigurdardottir, u. a. 2024. „Multicentre Sleep-stage Scoring Agreement in the Sleep Revolution Project“. *Journal of Sleep Research* 33 (1): e13956. https://doi.org/10.1111/jsr.13956.

Pitkänen, Henna, Sami Nikkonen, Marika Rissanen, Anna Sigridur Islind, Heidur Gretarsdottir, Erna Sif Arnardottir, Timo Leppänen, und Henri Korkalainen. 2024. „Multi-centre Arousal Scoring Agreement in the Sleep Revolution“. *Journal of Sleep Research* 33 (4): e14127. https://doi.org/10.1111/jsr.14127.

Punjabi, Naresh M., Naima Shifa, Georg Dorffner, Susheel Patil, Grace Pien, und Rashmi N. Aurora. 2015. „Computer-Assisted Automated Scoring of Polysomnograms Using the Somnolyzer System“. *Sleep* 38 (10): 1555–66. https://doi.org/10.5665/sleep.5046.

Younes, Magdy, Samuel T. Kuna, Allan I. Pack, James K. Walsh, Clete A. Kushida, Bethany Staley, und Grace W. Pien. 2018. „Reliability of the American Academy of Sleep Medicine Rules for Assessing Sleep Depth in Clinical Practice“. *Journal of Clinical Sleep Medicine* 14 (02): 205–13. https://doi.org/10.5664/jcsm.6934.