Friedemann Leibfritz

\nA LMI-Based Algorithm for Designing Suboptimal Static H\n

Bericht des Journal of Public Health

Kurzfassung

Background: Considerable time and energy are expended in the scientific community to discuss the validity, importance, and applicability of the results of clinical trials. Depending on the goals, perspectives, and other motivating factors, protagonists and skeptics come to different conclusions, even when using the same methods and tools for critical appraisal. The aim of this study was to complement existing methods and tools with minor modifications to provide a prototype instrument that generates commonly accepted versions of critical appraisals. Methods: As a pilot experiment, one university-based and one industry-based referee independently completed the twin assessment of five trials published in well-recognized journals. They identified the study questions, defined the simplest, i.e., ideal, study designs to answer these questions, and checked eight validity criteria. Identical positive or negative answers of both referees increased or decreased the validity score. A maximum of two disagreements (0 score) was allowed. This procedure, which had been tested by two referees in a pilot experiment, was repeated with 19 third-year medical students and their supervisor at the Universidade Federal Fluminense, Niterói/RJ, Brasil. Four students each played the roles of the industry-based and university-based referees and finally recorded their consensus. Results: The two referees of the pilot experiment agreed in all but one answer to the five investigated publications. The points of criticism differed in various papers. The consensus reached by the students considerably differed from the consensus reached by the referees. Conclusions: A consensus score generated by two referees or by two groups of students is feasible, but the achieved result is not necessarily reproducible. The critical appraisal of the study question in connection with the applied study design deserves special attention. It is time consuming but possible to identify and describe the possible flaws in the design conduct and report of clinical trials, but it is unlikely to reach a reproducible interpretation. These data indicate the problems with even evidence-based assessments and appraisals: the assessments may well be reproducible, but not the appraisals. Quality scores that include also the appraisal may therefore be interpreted with caution. Appraisals or quality scores may be used for interim decisions until data are available that confirm under real-world conditions what was predicted by the results generated under ideal but artificial conditions of a clinical trial.