

Validity and Reproducibility of a Risk Scale in Academic Clinical Research Studies.

V. Journot¹, J.P. Pignon², C. Gaultier³, A. Bouxin-Métro⁴, P. Ravaud⁵, G. Chêne⁶,
and the Optimon Study Group.

- 1 INSERM, U593 and CIC-EC, Bordeaux, France, valerie.journot@bordeaux.inserm.fr
- 2 IGR, Villejuif, France, jppignon@igr.fr
- 3 AP-HP, Hôpital R. Debré, Paris, France, claude.gaultier@rdb.ap-hop-paris.fr
- 4 ANRS, Paris, France, annie.metro@anrs.fr
- 5 AP-HP, Hôpital Bichat, Paris, France, philippe.ravaud@bch.ap-hop-paris.fr
- 6 CHU de Bordeaux, USMR, Bordeaux, France, genevieve.chene@isped.u-bordeaux2.fr

Background. Standards of quality for monitoring of academic clinical research studies should be risk-adapted. Several risk scales have been used but never formally assessed. **Objective.** To assess validity and reproducibility of a new risk scale for academic clinical research. **Methods.** Literature was searched for existing scales, which formats and items were discussed among clinical research experts. A new 4-level risk scale (4LRS) was designed. Assessors evaluated protocols' risk as an ordinal response through 4LRS (level A to D = lowest to highest risk), and as a continuum through an analogical visual scale (AVS). Sample size was 200 protocols, randomly allocated among at least 40 assessors (incomplete block design: 20 protocols per assessor). Protocols' and assessors' characteristics, and difficulties encountered by assessors were collected. Validity was assessed through the AVS distribution according to 4LRS. Reproducibility was estimated by standard multiple kappa, weighted kappa based on Kraemer's rank-ordering assessment method, intra-class correlation coefficient (ICC) estimated from a mixed proportional odds model on 4LRS, and ICC estimated from a mixed linear model on AVS. **Results.** Fifteen clinical research units/sponsors supplied 200 protocols, and 32 proposed 49 assessors. Protocols' risk assessment was: A: 27%; B: 24%; C: 26%; D: 23%. AVS median according to 4LRS was: A: 0.04; B: 0.24; C: 0.58; D: 0.80. Multiple kappa was 0.30. Weighted kappa was 0.48. ICC was 0.69 for 4LRS, and 0.62 for AVS. Assessors encountered difficulties for 42% assessments, mainly resulting from ill-written abstracts (14%) or protocols (17%), and scale complexity (15%). Yet, we failed to identify disagreement sources. **Conclusions.** Validity of 4LRS was good, thus confirming the assumption of 4 distinct ordered levels. Despite the assessment complexity, reproducibility estimated by the most appropriate methods was intermediate to good. Assessors' difficulties should be accounted for: a revised version of 4LRS should facilitate its use; we shall propose tools for protocols' harmonisation; selected and trained assessors should collectively assess risk. The revised 4LRS will be used in the Optimon trial, comparing two monitoring strategies for academic multicenter clinical studies: classical, as used in the industry, and optimised, i.e. risk-adapted and optimised on major scientific and regulatory principles.

Keywords. ACADEMIC CLINICAL RESEARCH, STUDY MONITORING, RISK SCALE, VALIDITY, REPRODUCIBILITY.