28th Annual Conference of the International Society for Clinical Biostatistics

C6

Validity and Reproducibility of a Risk Scale in Academic Clinical Research Studies
PRE-OPTIMON

Valérie Journot
INSERM U593 – case 11
Université Victor Segalen bordeaux 2
146 rue Léo Saignat
33076 Bordeaux cedex
Valerie.Journot@bordeaux.inserm.fr

V. JOURNOT[1], J.P. PIGNON[2], C. GAULTIER[3], A. BOUXIN-METRO[4], P. RAVAUD[5], G. CHENE[6], AND THE OPTIMON STUDY GROUP

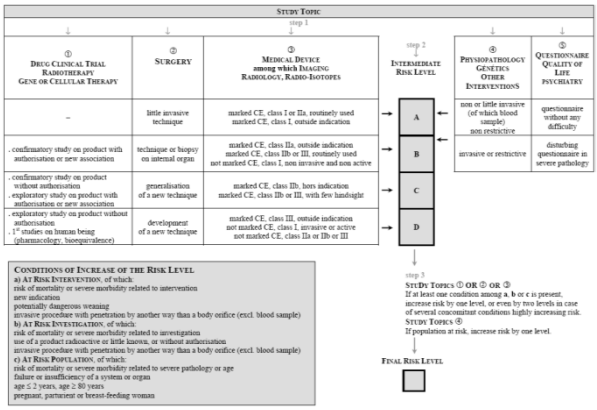
1. INSERM, U593, Bordeaux, France
2. Institut Gustave Roussy, Villejuif, France
3. AP-HP, Hôpital Robert Debré, Paris, France
4. ANRS, Paris, France
5. AP-HP, Hôpital Bichat, Paris, France
6. CHU Bordeaux, USMR, Bordeaux, France

BACKGROUND AND OBJECTIVES

It is more and more accepted that monitoring intensity in clinical research studies may be adapted depending on patients' risk. The French law already authorises sponsors to lighten on-site monitoring procedures, provided that appropriate remote procedures are applied. Therefore, standardised procedures are needed for monitoring adaptation and risk assessment. To prepare a trial comparing classical versus risk-adapted monitoring strategies, a panel of experts searched the literature for existing scales, discussed and established a new risk-assessment scale by consensus. We wish to evaluate validity and reproducibility of the proposed scale.

METHODS

proposed 4-level scale (A-lowest risk to D-highest risk)



study design

- 200 study protocols assessed by 49 assessors
- protocol allocation to assessors by an incomplete block design:
  . 20 groups of 10 randomly selected protocols
  . 20 groups of 2-3 randomly selected assessors
- risk assessment for each protocol by 4-6 assessors, using:
  . the proposed 4-level scale (4LS)
  . a visual analogic scale (VAS)

validity evaluation for the 4LS

distribution of VAS-assessed risk by 4LS-assessed risk level

reproducibility estimations

- multiple kappa agreement coefficient [1]

$$k_m = \frac{\sum_{l=1}^{4} p_l(1-p_l)k_l}{\sum_{l=1}^{4} p_l(1-p_l)}$$

with p_l the proportion of assessments in level l
k_c the classical kappa coefficient for level l (vs others)
but k_m does not allow for ordered levels

- weighted kappa agreement coefficient [2]

We suppose that when an assessor chooses one level, he/she expresses a preference ranking towards all (4) possible levels, so that each assessment may be transformed into a rank vector, with possible ties, as follows:

level A → (1 ; 2 ; 3 ; 4) level C → (4 ; 2.5 ; 1 ; 2.5)
level B → (2.5 ; 1 ; 2.5 ; 4) level D → (4 ; 3 ; 2 ; 1)

Spearman rank correlation coefficients between pairs of rank vectors is used as measure of agreement, and coefficients for several pairs are averaged.

$$k_w = \frac{r_I - r_T}{1 - r_I}$$

with r_i the mean of coefficients for pairs of rank vectors for protocol i
r_I the mean of r_i
r_T the mean of coefficients for all pairs of rank vectors

- intraclass correlation coefficient estimated by a mixed proportional odds model for 4LS-assessed risk

$$CCI = \frac{\sigma_P^2}{\sigma_P^2 + \sigma_R^2}$$

with σ_P^2 the variance of the random protocol effect
σ_R^2 the residual variance (for logistic distribution : $\pi^2/3$)

disagreement sources between assessments

sensibility analysis to characteristics of protocols and assessors

encountered difficulties during assessment

description of number and nature difficulties

RESULTS

PROTOCOLS

The 200 protocols came from 15 different organisations, 36% were phase II or III trials.

study topic

drug	38%
physiopathology	28%
medical device	21%
imaging - radiology	10%
genetics	9%

clinical field

cancer	28%
infection	15%
pediatrics	11%
cardio-vascular disease	8%
reanimation – anaesthesia	8%
nephrology	6%
neurology – neurosurgery	6%
hepato-gastro-enterology	6%

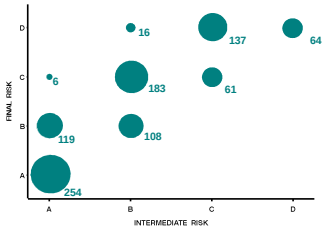
ASSESSORS

The 49 assessors came from 32 different organisations, 71% were women, 43% were less than 40, and 22% were head of their organisation.

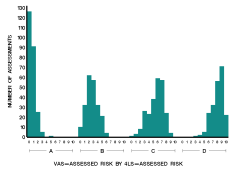
position

methodologist / statistician	51%
project leader / CRA	39%
sponsor / quality assurance	12%
investigator	8%

RISK ASSESSMENTS



VALIDITY



REPRODUCIBILITY

multiple kappa coefficient [1]	0.30
weighted kappa coefficient [2]	0.48
intra-protocol correlation coefficient by a mixed proportional odds model	0.69

DISAGREEMENT SOURCES AND DIFFICULTIES

No source of disagreement between assessors could be identified among protocols' or assessors' characteristics.

Assessors encountered difficulties for 42% of assessments. These difficulties aroused from badly written abstracts (14%), badly written protocols (17%), or from the 4-level scale complexity (15%).

DISCUSSION AND FUTURE PROSPECTS

The validity of the proposed scale is good, and the assumption of 4 ordered levels seems to be relevant.

To estimate its reproducibility, appropriate methods must be used to take the ordering of risk levels into account. Kraemer's approach fits well the objectives and the implicit assumptions of our design. The mixed proportional odds model is a more powerful option. Reproducibility of the proposed scale was found medium to good. Yet, with such a complex tool applied to such complex objects, reproducibility could not be expected to be perfect.

The scale was revised to standardise its use and improve risk assessment. The revised scale will be used in the French OPTIMON trial comparing two monitoring strategies: classical vs risk-adapted consistently with the main scientific and regulatory principles.

Besides, a consensus definition of a risk-assessment scale is undergoing among the European Clinical Research Infrastructures Network (ECRIN).

REFERENCES 1. Fleiss J.L. Statistical methods for rates and proportions. New York:John Wiley & Sons, 1981;321p.
2. Kraemer H.C. Extensions of the kappa coefficient. Biometrics 1980;36:207-16.

OPTIMON WEB SITE
https://ssl2.isped.u-bordeaux2.fr/optimon/