Contemporary Clinical Trials 32 (2011) 16–24

Contents lists available at ScienceDirect

Contemporary Clinical Trials

journal homepage: www.elsevier.com/locate/conclintrial

Validation of a risk-assessment scale and a risk-adapted monitoring plan for academic clinical research studies — The Pre-Optimon study

Valérie Journot[a,b,*], Jean-Pierre Pignon[c], Claude Gaultier[d], Véronique Daurat[e], Annie Bouxin-Métro[f], Bruno Giraudeau[g], Pierre-Marie Preux[h], Jean-Marc Tréluyer[i], Sylvie Chevret[j], Valérie Plättner[k], Claire Thalamas[l], Stéphanie Clisant[m], Philippe Ravaud[n], Geneviève Chêne[a,b,o]
and on behalf of the Optimon Collaborative Group

[a] INSERM, U897, F-33076 Bordeaux, France
[b] INSERM, CIC-EC7, F-33076 Bordeaux, France
[c] Institut de Cancérologie Gustave-Roussy, Service de Biostatistiques et d'Épidémiologie, F-94805 Villejuif, France
[d] INSERM, Pôle Recherche Clinique, F-75013 Paris, France
[e] Assistance Publique-Hôpitaux de Paris, DRCD, F-75010 Paris, France
[f] ANRS, F-75013 Paris, France
[g] INSERM, CIC0202, F-37032 Tours, France
[h] CHU de Limoges, Unité Fonctionnelle de Recherche Clinique et Biostatistique, F-87025 Limoges, France
[i] Assistance Publique-Hôpitaux de Paris, Hôpital Necker-Enfants Malades, Unité de Recherche Clinique, F-75015 Paris, France
[j] Assistance Publique-Hôpitaux de Paris, Hôpital Saint-Louis, Département de Biostatistique et d'Informatique Médicale, F-75010 Paris, France
[k] Hospices Civils de Lyon, Direction à la Recherche Clinique, F-69229 Lyon, France
[l] INSERM, CIC9302, F31059 Toulouse, France
[m] Centre Oscar Lambret, Unité Intégrée de Recherche Clinique, F-59020 Lille, France
[n] Assistance Publique-Hôpitaux de Paris, Hôpital Hôtel-Dieu, Centre d'Épidémiologie Clinique, F-75181 Paris, France
[o] CHU de Bordeaux, Unité de Soutien Méthodologique à la Recherche Clinique et Épidémiologique, F33076 Bordeaux, France

ARTICLE INFO

Article history:
Received 2 July 2010
Accepted 6 October 2010

Keywords:
Clinical trial
Intraclass correlation coefficient
Kappa coefficient
Monitoring plan
Proportional odds model
Risk-assessment scale
Risk-based approach

ABSTRACT

Context: Good Clinical Practice regulates monitoring activities in clinical research. Due to question and design diversity, and limited resources, on-site monitoring is often less intensive in the academic context, and variable. Standardization is needed, and relies on definition and validation of tools accounting for risk.
Objective: To define, and validate tools, to implement a risk-based monitoring strategy for academic clinical research.
Methods: Working groups of experienced professionals searched the literature, and built a consensus risk-assessment scale (RAS), and a risk-adapted monitoring plan (RAMP). We allocated 200 protocols to 49 assessors. We assessed the RAS relevance *vs.* a visual analogue scale (VAS), and its reproducibility through Kraemer's kappa, and intraclass correlation coefficient (ICC) from a random proportional odds model. We identified sources of disagreement through a logistic regression. We described assessors' difficulties during assessment. We applied the RAMP to 10 protocols per risk level, and rated its feasibility (0 = easy to 4 = impossible).
Results: RAS and RAMP were defined in 4 levels. RAS relevance was good: RAS-risk levels were evenly distributed on VAS-risk (0.6, 2.6, 5.6, and 7.9). Reproducibility was moderate to good: kappa = 0.48, ICC = 0.70. Major disagreements (36%) arose from decision-makers, rather than hands-on managers. Most difficulties occurred in ill-written protocols (17%). RAMP was easily

Abbreviations: CE, "Conformité Européenne" — European Conformity; CRF, Case-Report Form; CRO, Clinical Research Organization; CTU, Clinical Trial Unit; ICC, intraclass correlation coefficient; RAMP, risk-adapted monitoring plan; RAS, risk-assessment scale; SAE, serious adverse event; VAS, visual analogue scale.

* Corresponding author. Centre de Recherche U897, case 11, Université Victor Segalen bordeaux 2, 146, rue Léo Saignat, 33076 Bordeaux cedex, France. Tel.: +33 5 57 57 92 45; fax: +33 5 57 57 11 29.
E-mail address: valerie.journot@inserm.fr (V. Journot).

1551-7144/$ – see front matter
doi:10.1016/j.cct.2010.10.001

feasible for most protocols (mean score: 0.2 to 0.9). We proposed a standard synopsis for evaluation purpose.
Conclusion: We defined, and validated risk-based tools. This risk-adapted strategy will be compared to an intensive one in a randomized trial, Optimon, to define a standard of practice for academic clinical research.

1. Introduction

In clinical research studies, sponsors are responsible for monitoring activities, though often delegating them to Clinical Research Organizations or Clinical Trials Units. These activities are regulated or framed by the Guideline for Good Clinical Practice (GCP) [1]. An interpretation of the GCP has led to the widespread notion that a monitoring strategy based on frequent on-site visits, and exhaustive checking of all collected data only can ensure that "the rights, safety and well-being of trial subjects are protected,…and that the clinical trial data are credible" [1]. Credible data has often been interpreted as complete and accurate data. This intensive monitoring has thus reached the status of a standard of practice. Yet, this status is increasingly contested, for it has never been formally evaluated, nor compared to other strategies [2–5].

Indeed, the GCP allows for gradation of monitoring intensity according to "the objective, purpose, design, complexity, blinding, size, and endpoints of the trial" [1]. Yet its implementation usually depends on the type of sponsor. Pharmaceutical industries mainly aim at drug labeling through randomized trials, and apply an intensive monitoring strategy by fair of non compliance with Good Clinical Practice. Academic sponsors deal with a larger panel of scientific research requiring various types of study designs, but command lower funding. Therefore, they generally apply less intensive, or more properly, optimized strategies. As a consequence, monitoring practices are highly variable among academic sponsors, and there is a need for standardization of practices in monitoring activities.

The issue is not only, or not at all, to reduce costs and energy, but to apply them to the most crucial study components. Adaptations should tend toward a reduction in the most costly component that is on-site monitoring visits, by decreasing their frequency and length, provided however that patient's rights and safety are ensured by other means, such as investigator education and training, intensive database controls, and intensive remote monitoring. Therefore patient's risk must be assessed before the beginning of the study, and a corresponding minimum acceptable level of monitoring intensity must be defined.

The French Optimon (for OPTImization of MONitoring) trial will compare two monitoring strategies: an intensive, standard-like strategy *vs*. a strategy adapted on patient's risk, and consistent with the main scientific and regulatory principles, among academic clinical studies. In the Pre-Optimon project, we defined and evaluated a risk-assessment scale and a risk-adapted monitoring plan, preliminary to their use in Optimon. This paper presents the methods used for the definition of these tools. It reports the evaluation of the relevance and reproducibility of the risk-assessment scale, and the feasibility of a risk-adapted monitoring plan in the context of the academic clinical research.

2. Materials and methods

2.1. Risk-assessment scale

A first working group including clinical research experts from different French academic institutions, and French inspectors of the Agence Française de Sécurité Sanitaire des Produits de Santé was formed. They searched the literature for existing risk-assessment scales. They met several times, and discussed these scales items and formats. They proposed a new risk-assessment scale (RAS) in 4 levels, from level A: no risk or almost none, to level D: highest risk. This RAS was sent for comments to professionals of most French academic clinical research institutions, contacted through the e-mail lists of three national clinical research networks. The group modified the scale accordingly, and wrote a guideline for its use, with directions, examples, and a glossary.

Then relevance and reproducibility of the scale was formally assessed. Protocols were collected, and assessors volunteered. Protocols were randomly assigned to assessors through a partially incomplete blocks design: 200 protocols, randomly divided in 20 groups of 10; at least 40 assessors, randomly divided in 20 groups of 2 to 3; 20 protocols per assessor; 4 to 6 assessors per protocol. A protocol risk was assessed twice in the same questionnaire, through the RAS, and through a plain visual analogue scale (VAS) ranging from 0 (no risk) to 10 cm (maximal risk). The questionnaire also collected protocol characteristics (study clinical field and objectives, study design, and size in terms of number of pages of the protocol; existence and number of pages of the classical abstract and/or of the synopsis: a synopsis is a summary of both scientific and organizational information intended for protocol evaluation, while the classical abstract only contains scientific information), assessors' characteristics (sex, age, affiliation, position, occupation, experience, familiar clinical fields, objectives, and design), and difficulties encountered by assessors.

Relevance of the RAS was evaluated by describing the distribution of VAS-risk depending on RAS-risk. A multiple comparison test was performed by Scheffé's method at the overall 5% significance level.

Reproducibility of RAS-risk was assessed by several methods evaluating agreement between assessors. We used an extension of the kappa coefficient proposed by Kraemer, which allows for more than 2 assessors, more than 2 ordinal risk levels, and a varying number of assessments per protocol [6]. Each assessment is interpreted as a preference ordering made by the assessor among risk levels.

For instance, choosing the risk level A for a protocol is equivalent to choose the preference ranking order: "A is preferred to B, which is preferred to C, which is preferred to D". Each assessment may be consequently transformed into a vector of ranks: the rank 1 is allocated to the preferred level, the rank 2 to the preferred level but one, and so on up to rank 4. We used classical handling for ties: when 2 levels were in the same relative position to the preferred level, they both received the average of the 2 possible ranks. So we got the following transformations of risk level into vector of ranks: A = (1,2,3,4); B = (2.5,1,2.5,4); C = (4,2.5,1,2.5); D = (4,3,2,1). Then, a Spearman correlation coefficient was calculated for each paired assessment (same protocol assessed by 2 different assessors). Spearman coefficients were summarized into a kappa coefficient: $\kappa = (R_I - R_T)/(1 - R_I)$, with R_i the average of Spearman coefficients for all pairs of rank vectors for protocol i, R_I the average of the R_i, and R_T the average of Spearman coefficients for all pairs of rank vectors for all protocols. This method is based on the same principle as the weighted kappa, but it allows for more than 2 assessors. We used a 1000-sample Bootstrap on protocols to estimate 95% confidence intervals for Kraemer's kappa.

As an alternative method to evaluate agreement between assessors, we used an intraclass correlation coefficient estimated from a proportional odds model on RAS-risk levels, with a unique term in the model, a protocol random effect. The proportional odds model relies on the assumption of a latent continuous risk variable with a standard logistic distribution underlying the RAS-risk levels [7]. The assumption of an underlying continuous risk variable seems intuitively relevant. The logistic distribution has an analytical solution, and is therefore preferred to a Gaussian distribution. Besides, a standard logistic distribution has larger tails and lower probabilities of belonging to a risk level than a standard Gaussian distribution, which seems more relevant to our 4-level data. The standard distribution is used for identifiability reasons. In this model, the intraclass correlation coefficient for random protocol effect is thus: $ICC_{RAS} = \hat{\sigma}_P^2 / (\hat{\sigma}_P^2 + \hat{\sigma}_R^2)$, with $\hat{\sigma}_P^2$ the estimated variance of the normally distributed protocol effect, and $\hat{\sigma}_R^2 = \pi^2 / 3$ the residual variance of a standard logistic distribution [7].

As a control, we also calculated an intraclass correlation coefficient estimated from a linear model on VAS-risk values with protocol as random effect: $ICC_{VAS} = \hat{\sigma}_P^2 / (\hat{\sigma}_P^2 + \hat{\sigma}_R^2)$, with $\hat{\sigma}_P^2$ the estimated variance of the protocol effect, and $\hat{\sigma}_R^2$ the estimated residual variance. The 95% confidence interval for both intraclass correlation coefficient estimates was approximated through the delta method [8].

We studied the sources of disagreement between assessors. Major disagreement was defined as an observed difference of 2 to 3 risk levels in the assessment of the same protocol by two assessors. We investigated the assessors' characteristics as fixed effects through a mixed logistic regression on major disagreement *vs.* no major disagreement with random protocol effect. We did not include protocols characteristics in the model, since we looked for agreement, whatever the protocol type or its risk level. Covariates selection consisted of univariate selection at a 25% significance level, with possibly grouping of categories, followed by multivariate backward selection at a 5% significance level. The large 25% level was adopted to avoid missing possible sources of disagreement, in an exploratory approach [9].

Finally, we described the difficulties encountered by assessors. Consequently, the RAS format was revised again, to facilitate and standardize its use.

2.2. Risk-adapted monitoring plan

A second working group including clinical research experts from different French academic institutions was formed. They worked at first with a monitoring plan used by the Paris Hospital Network (Assistance Publique-Hôpitaux de Paris). The group met several times, and discussed items and formats of the plan. They proposed a new risk-adapted monitoring plan, which was sent for comments to professionals of most French academic clinical research institutions. The group discussed the comments, and modified the plan accordingly. Finally, the group wrote a guideline and a glossary for its use.

We conducted a feasibility study of the risk-adapted monitoring plan. We randomly selected 40 protocols from the reproducibility study, 10 for each RAS-risk level, preferably among those uniformly assessed by the 4 to 6 different assessors concerned. We read each protocol, and evaluated the feasibility of the adaptation of the proposed risk-adapted monitoring plan to this protocol, considering the elements available in the documents. We graded feasibility from 0 (easy) to 4 (impossible), and we identified the impediments to adaptation.

2.3. Software

All analyses were carried out with the SAS® software, version 9.1 [10].

3. Results

3.1. Risk-assessment scale

The first working group found several examples of risk-assessment scales used for risk-adapted monitoring through the Internet, the most elaborated ones being developed and used by the Paris Hospital Network and the Medical Research Council [11–13]. A French workshop held in October, 2003 also stated the principle of a large approach of risk-assessment and risk-adapted monitoring [14]. Yet, no scale had already been formally evaluated. The group started with the Paris Hospital Network risk-assessment scale, and modified it during the discussions. Some items were revised, and, most importantly, we added the possibility to increase patient's risk depending on the existence of very risky interventions and/or investigations, or on the existence of an interaction between the target population-defining disease and the risk related to interventions and/or investigations. Consequently, intermediate and final risks were defined, both ranging from lowest level A to highest level D (Fig. 1).

In the reproducibility study, 15 clinical research institutions provided 223 protocols, among which 200 were selected: 10 did not have the adequate format (paper version, draft only or confidentiality condition), and 13 were randomly selected for exclusion among those originating from the two main providers.

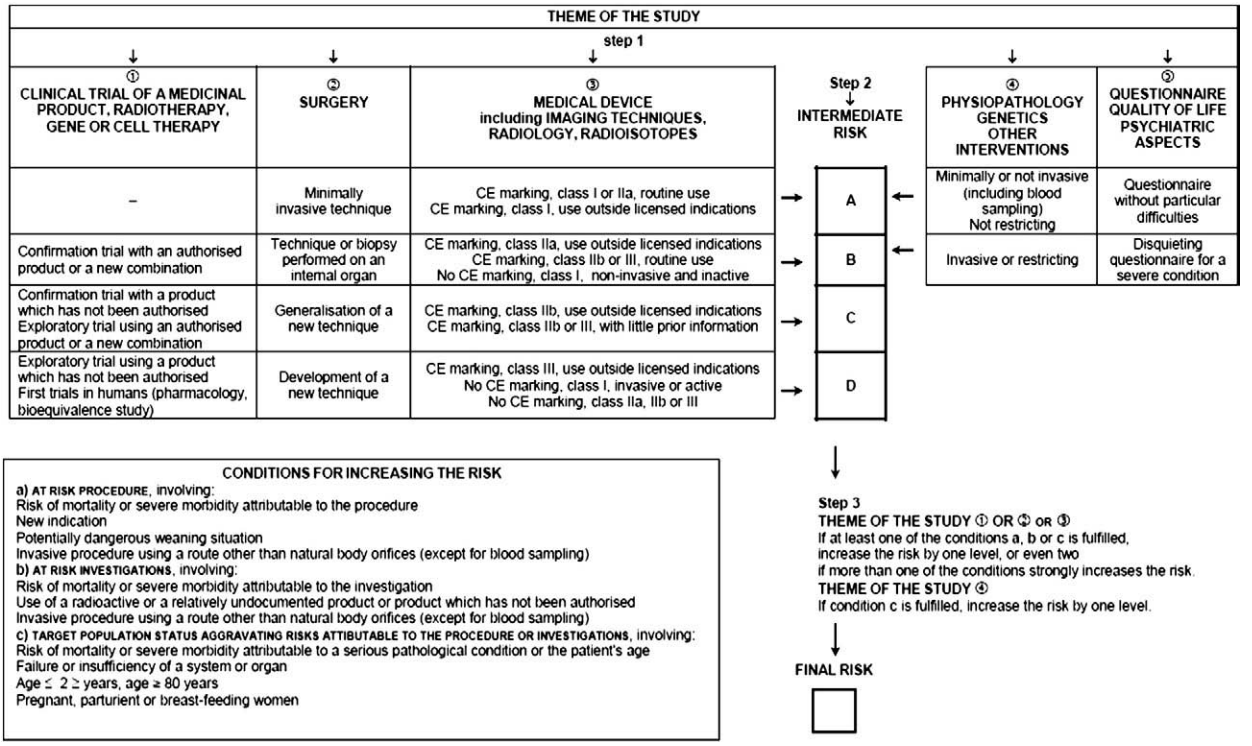

Fig. 1. Risk-assessment scale (main page) evaluated in the validity and reproducibility study — the Pre-Optimon study. The scale was improved following the results of Pre-Optimon. The instructions for use may be found on the Optimon website: https://ssl2.isped.u-bordeaux2.fr/optimon/Documents.aspx "Optimon Risk Assessment Scale". CE, Conformité Européenne — European Conformity.

Protocol characteristics are shown in Table 1. They came from various types of academic sponsors, and covered a large variety of clinical fields, topics, and designs. Interestingly, designs other than trials were largely represented (60%). The classical abstract (93%), the contents table (81%), and the rationale (100%) were most generally present, whereas the synopsis (9%) was nearly always missing. Protocols were also highly heterogeneous in format and size (range of number of pages in protocols: 6–171; in abstracts: 1–9; in synopses: 1–15).

Fifty-one assessors from 32 clinical research institutions volunteered, among which 2 finally withdrew for lack of time, and 2 shared the work intended for one assessor. Assessors were mainly women in their middle age (Table 2). They often belonged to a multi-institution organization or to several institutions (80%). They generally had experience in a large variety of designs, topics and clinical fields.

Nine hundred and fifty-two protocol assessments were collected for final risk (93% of the expected for the 51 initially recruited assessors). The distribution of VAS-risk depending on RAS-risk is shown on Fig. 2. Median differences of VAS-risk between adjacent levels of RAS-risk ranged from 21 to 33 mm. Adjacent levels of RAS-risk were significantly different on VAS-risk by Scheffé's method for multiple comparisons at the 5% overall significance level: the VAS-risk distribution of each RAS-risk level is stochastically higher than the VAS-risk distribution of the lower adjacent RAS-risk level. This indicates good relevance for the assumption of a gradient of risk with 4 distinct levels [15].

Considering these nearly evenly spaced-out distributions of VAS-risk depending on RAS-risk, we chose ranks – that is 1, 2, 3, 4, with classical ties handling, – as scores for the estimation of the Kraemer's kappa coefficient for RAS-risk. Kraemer's kappa was 0.48 (Bootstrap 95% confidence interval: 0.41–0.53), indicating moderate agreement [16]. The intraclass correlation coefficient estimated from a random proportional odds model on RAS-risk was 0.70 (Delta method 95% confidence interval: 0.63–0.76), corresponding to substantial agreement. As a control, the intraclass correlation coefficient estimated from a random linear model on VAS-risk was 0.62 (Delta method 95% confidence interval: 0.56–0.68).

Major disagreements between assessments on the same protocol were frequent (36% protocols). In the univariate analysis, sources of major disagreement between assessments on the same protocol were selected for the multivariate analysis (Table 3): a male sex ($p=0.08$), an age of 40 or above ($p=0.001$), a generalist ($p=0.06$) and/or multiple ($p=0.002$) affiliation, a decision-making occupation ($p=0.004$), a clinical ($p=0.11$) and/or no methodological education ($p=0.02$), no familiarity with experimental designs ($p=0.20$), but familiarity with observational designs ($p=0.17$), and an overall experience above 12 years ($p=0.02$), increased the risk of major disagreement. Yet, in the multivariate analysis, only a male sex (odds ratio [95% confidence limits]: 2.05 [1.40–3.01]; $p=0.0002$), a multiple affiliation (1.75 [1.23–2.48]; $p=0.002$), a decision-making occupation (1.71 [1.22–2.40]; $p=0.002$), no methodological

Table 1
Protocols characteristics in the reproducibility study (N = 200) — the Pre-Optimon study.

Characteristic	#	%
Sponsor		
Hospitals from the Paris Hospital Network	91	45
Hospitals outside Paris area	67	33
Institute of oncology Gustave Roussy	23	12
French National Agency for AIDS Research and Viral Hepatitis	16	8
French National Institute of Health and Medical Research	3	2
Study design		
Phase I–II trial	69	34
Phase III–IV trial	43	21
Cohort	48	24
Cross-sectional study	29	15
Case–control study	11	6
Study main objective		
Pathophysiology	31	16
Diagnostic	28	14
Prognostic	11	6
Efficacy	106	53
Others *epidemiology*, *registry*, *safety*, *compliance*, *medico-economics*	24	12
Study topics [a]		
Drug	76	38
Medical device	41	21
Imaging — radiology	20	10
Genetics	18	9
Other pathophysiology	56	28
Others *radiotherapy*, *genes and cells therapy*, *surgery*, *quality of life*, *psychiatry*, *other interventions*	17	9
Clinical fields [a]		
Cancer	55	28
Infections	30	15
Paediatrics	22	11
Cardio-vascular disease	16	8
Intensive care — anaesthesia	15	8
Nephrology	12	6
Hepatology–gastroenterology	12	6
Pneumology	12	6
Neurology–neurosurgery	11	6
Others *biology of reproduction*, *diabetes*, *endocrinology*, *geriatrics*, *gynaecology*, *haematology*, *immunology*, *obstetrics*, *odontology*, *ophthalmology*, *orthopaedics*, *otorhinolaryngology*, *rheumatology, nutrition*, *psychiatry*, *sexology*, *stomatology*, *urology*	43	22
Content [a]		
Abstract	185	93
Synopsis [b]	17	9
Table of contents	162	81
Rationale	199	100

Size of protocol	Median	Interquartile range	Range
Number of pages in			
Protocol	38	26–54	6–171
Abstract	1	1–2	1–9
Synopsis [b]	2	1–4	1–15
Number of paragraphs in			
Abstract	7	5–11	1–23
Synopsis [b]	12	9–16	3–50

[a] Some protocols may fall into several categories for some characteristics, so that the overall proportion may exceed 100% for these characteristics.
[b] The synopsis is a summary of both scientific and organizational information intended for protocol evaluation, while the abstract only contains scientific information.

Table 2
Assessors' characteristics in the reproducibility study (N = 49) – the Pre-Optimon study.

Characteristic	#	%
Sex		
Female	35	71
Age (years)		
18–30	8	16
31–40	18	37
41–50	18	37
51 and more	5	10
Affiliations [a]		
Hospital	28	57
French National Institute of Health and Medical Research	20	41
French National Agency on AIDS Research and Viral Hepatitis	9	18
Oncology Institutes	8	16
Other institutions	5	10
Organizations [a]		
Clinical research center	43	84
Sponsor's clinical research department	9	18
Occupations [a]		
Methodologist — statistician	25	51
Project leader — clinical research assistant	19	39
Head	11	22
Sponsor — quality insurance	6	12
Investigator	4	8
Education [a]		
Statistics — epidemiology	28	57
Medical	25	51
Pharmacy — pharmacology	12	24
Life sciences	7	14
Clinical research	7	14
Familiar with clinical fields [a]		
Oncology	17	35
Infectious diseases	18	37
Cardio-vascular disease	16	33
Pediatrics	14	29
Neurology — neurosurgery	10	20
Familiar with study designs [a]		
Phases I–II trial	39	78
Phase III trial	41	84
Phase IV study	21	43
Diagnostic	7	4
Others	117	59
Familiar with study topics [a]		
Drug	39	80
Questionnaire — quality of life	28	57
Pathophysiology	19	39
Surgery	18	37
Medical device	16	33
Imaging — radiology	14	29
Radiotherapy	14	29
Gene — cell therapy	11	22
Genetics	11	22

Experience in clinical research (years)	Median	Interquartile range	Range
Overall	10	5–12	1–32
In the present occupation	3	2–6	0–16

[a] Some assessors may belong to several categories for some characteristics, so that the overall proportion may exceed 100% for these characteristics.

education (1.63 [1.05–2.53]; $p = 0.03$), and an overall experience of more than 12 years (1.53 [1.04–2.26]; $p = 0.03$) increased the risk of major disagreement.

Assessors indicated difficulties for 41% of assessments, less frequently for protocols of risk level A (31%), and more frequently for protocols of risk levels B and C (46% each). Difficulties were mainly due to missing or incomplete abstracts (14%), to incomplete or poorly-written protocols (17%, increasing from 11% to 23% with risk level A to D), and

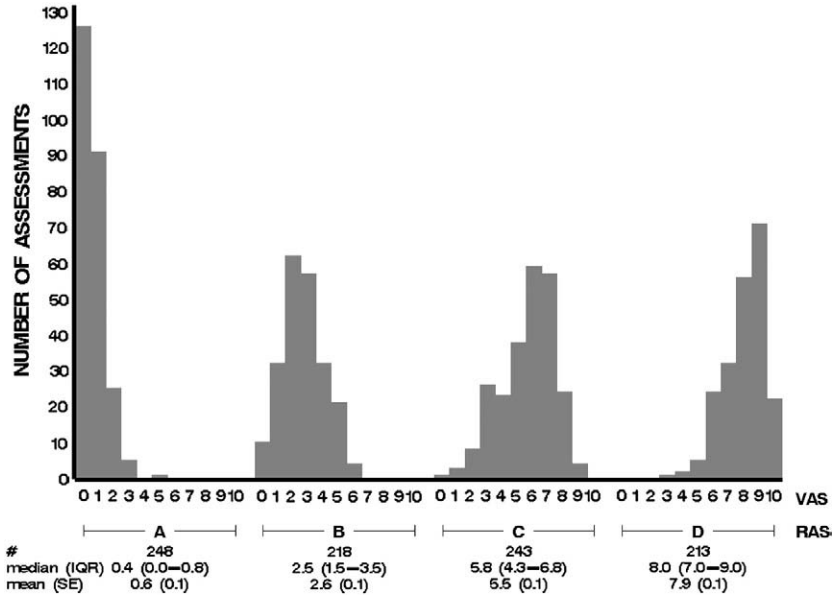

Fig. 2. Distribution of risk assessed by visual analogue scale (VAS), depending on risk assessed by risk-assessment scale (RAS): median and interquartile range (IQR), mean and standard error (SE) — the Pre-Optimon study.

to inadequacy or complexity of the RAS (15%). The ill-defined study type (9%), and the unknown clinical field (5%, increasing from 3% to 8% with risk level) were less frequently encountered difficulties. Assessors most generally encountered no difficulties with the instructions for use of the scale (98%). Several assessors noticed that detailed toxicity of drugs or invasive interventions, or European Conformity classes for medical devices, were missing information in abstracts and protocols. Secondly, the RAS complexity was mainly related to the risk increase conditions: there was often confusion between the baseline risk due to the defining condition of the target population, and the risk added by the interaction between this condition and interventions or investigations. Moreover, there was some variability in the choice of a range for the risk increase when conditions implied it. We also clarified these two issues in the final RAS version that may be found on the Optimon website: https://ssl2.isped.u-bordeaux2.fr/optimon/Documents.aspx.

3.2. Risk-adapted monitoring plan

The new working group was made of 20 experienced professionals in clinical research coming from different institutions. The group started with a first version proposed

Table 3
Risk of major disagreement between assessors (by 2 to 3 risk levels) in risk assessment for the same protocol, depending on assessors' characteristics: univariate ($p<0.25$) and backward ($p\leq0.05$) multivariate logistic regression, with random protocol effect — the Pre-Optimon study.

Assessors' characteristics		Selection step					
		Univariate			Multivariate		
		Odds ratio	95% Confidence limits	*p*	Odds ratio	95% Confidence limits	*p*
Sex	Male *vs.* female	1.37	0.97–1.95	0.08	2.05	1.40–3.01	2.10^{-4}
Age	≥40 years *vs.* <40 years	1.18	1.23–2.38	1.10^{-3}			
Affiliation	Generalist *vs.* specialized	1.41	0.99–2.01	0.06			
	multiple *vs.* single	1.73	1.23–2.44	2.10^{-3}	1.75	1.23–2.48	2.10^{-3}
Occupation [a]	Decision-making *vs.* data and quality	1.62	1.16–2.25	4.10^{-3}	1.71	1.22–2.40	2.10^{-3}
Education [b]	Clinical *vs.* no clinical	1.34	0.94–1.92	0.11			
	No methodological *vs.* methodological	1.62	1.08–2.45	0.02	1.63	1.05–2.53	0.03
Familiarity with designs [c]	Experimental *vs.* none	1.43	0.83–2.48	0.20			
	Observational *vs.* none	1.41	0.87–2.76	0.17			
Overall experience	>12 years *vs.* ≤12 years	1.60	1.09–2.35	0.02	1.53	1.04–2.26	0.03

[a] Occupation: decision-making: head, project leader, sponsor; data and quality: methodologist, statistician, clinical research assistant, data manager, quality insurance, investigator.
[b] Education: clinical: medicine, pharmaceutics; methodological: statistics, life science, clinical research.
[c] Designs: experimental: phase I to III trials, studies with interventions (drugs, radiotherapy, genes or cells therapy, surgery, medical device, or any other); observational: phase IV, epidemiologic, pathophysiology, diagnosis, imaging, genetics, quality of life or psychiatry studies, studies with questionnaires.

in the initial Optimon protocol, and modified it during the discussions. Firstly, the set-up and closure parts of the plan were simplified and standardized, as compared to usual practice. Secondly, on-site monitoring was made less intensive, and restrained to audits on selected data. Thirdly, a special device allowing remote and real time control of consent signature during the screening and inclusion process was set up: after the participant wrote his/her name and signed, a special duplicate is faxed to the study coordinating center; the name and signature are partially masked, so that they may be clearly perceived but not read. The final version of the risk-adapted monitoring plan in French and English may be found on the Optimon website (see summary in Fig. 3).

We evaluated the feasibility of the proposed risk-adapted monitoring plan on 40 protocols for the main items of the plan. We could find 10 protocols unanimously assessed in one risk level for level A only. For the other levels, we selected 10 protocols among those mainly assessed in the target level, and marginally assessed in an adjacent level. The plan seemed quite feasible for most protocols, with a mean overall score ranging from 0.2 to 0.9 depending on risk level. Difficulties mainly arose from incomplete protocols, which were more frequent in lower risk levels. The drugs and samples handling (mean score 2.0 for risk level A and 1.8 for risk level D), serious adverse events reporting (2.4 for risk level A), and the planned intensity of on-site visits (1.0 for risk level A), were rarely detailed in the protocols, so that it was difficult to define the monitoring plan properly.

3.3. Protocol synopsis

We identified the need for a standard protocol synopsis in the relevance and reproducibility study of the RAS, and in the feasibility study of the risk-adapted monitoring plan. This synopsis should be a technical and regulatory-oriented synopsis collecting all information necessary for protocol evaluation. It should contain information both on scientific and organizational aspects. We suggest that a standard synopsis should contain i) protocol identification: title, sponsor and coordinating centre coordinates; ii) scientific aspects: research question, objectives, design, population (main inclusion and exclusion criteria, and informed consent specificities), trial and control interventions(s), main outcome and main secondary outcomes, investigations (socio-demographic characteristics, clinical and other investigations, and risk related to investigations), sample size, and expected impact of study results; iii) organization aspects: governance, staff's experience at the sponsor's, at the coordinating centre's, and at the investigator sites', study management (feasibility evaluation, accrual and follow-up schemes, intervention allocation and management, data management, quality assurance and quality control systems), resources (type and timing of funding, timing of recruitment of staff at

	ACTION	RISK-ADAPTED MONITORING STRATEGY		
		risk level A	risk level B	risk level C
SET UP	**Initial contact** - contact between CRO/CTU and investigator site	- systematically - at meeting, by mail / e-mail / phone - documented		
	Verification of resources adequacy at investigator site - availability of material resources and skills at site	- systematically - if site known and experienced: standard questionnaire by mail / e-mail / phone - if site not known or experienced: standard questionnaire on-site, possibly combined with initiation visit		
	Study initiation - presentation of scientific and regulations aspects - presentation and discussion of practical aspects	- systematically - before inclusion of first patient - if site known and experienced: by phone - if site not known or experienced: on-site visit		
MONITORING	**Verification of CRF and database correctness compared with source file, and verification of compliance with procedures**	- no on-site visit	- 10% patients - sampling plan defined at start of study - on 100% key points - additional visit if major problem	- at least 1 visit per site - # patients for 1 day of monitoring - on 100% key points - additional visit if major problem
	Verification of understanding of internal and external circuits by investigator site	- systematically - after inclusion and reception of CRF forms by CRO/CTU for first patient - by phone		
	Verification of consent form - patient's existence - completeness and correctness of form filling	- remotely by masked copy of consent form		
			- on-site during the following visit or upon site closure	
	Searching out unreported SAEs - detection of SAEs not reported spontaneously	- systematically on-site or remotely		
	Corrections of CRF forms on-site	- remotely managed	- during each visit for key points - remotely managed for other data	
	Verification of CRF forms at CRO/CTU - full reception and completeness of expected forms	- systematically before data entry for forms not checked on-site		
	Remote or centralized monitoring - computerized checks - queries - remote contact with key people on-site to check for compliance with procedures	- systematically - 100% patients - 100% data + compliance with procedures - 100% sites - by e-mail / mail / phone		
SITE CLOSURE	**Administrative closure of site** - collection of consent forms (at sponsor's request) - validation/closure of CRF - accounting/destruction of drugs units - closure of site	- systematically - 100% patients - 100% sites		
		- by mail		- on-site

Fig. 3. Summary of the risk-adapted monitoring plan — the Pre-Optimon study. The whole risk-adapted monitoring plan may be found on the Optimon website: https://ssl2.isped.u-bordeaux2.fr/optimon/Documents.aspx "Optimon Risk-Adapted Monitoring Plan". CRF, Case-Report Form; CRO, Clinical Research Organization; CTU, Clinical Trial Unit; SAE, serious adverse event.

the sponsor's, at the coordinating centre's, and at the investigator sites').

4. Discussion

4.1. Risk-assessment scale

First, we were able to define a risk-assessment scale applicable to any type of academic clinical research. Such scales have already been used by academic sponsors for several years, but to our knowledge, they were never formally validated or published.

It is very likely that there is a latent continuous risk variable, but we showed that the 4-level scale is relevant. Indeed, VAS-risk and RAS-risk were assessed in the same questionnaire, so that the agreement found between the two measures was probably overestimated. Yet, doing 2 separate assessments would have greatly complicated our study design, so as to render it unfeasible. Moreover, it is likely that good results would have been obtained with 3 or 5 levels too. Yet, we focused on 4 levels, because it is convenient for clinical studies management purpose: it is not too highly partitioned, but it allows a sensitive gradation of monitoring intensity.

We estimated the reproducibility of the scale with Kraemer's kappa coefficient. This method accounts for all particularities in our data: multiplicity of protocols and assessors, uneven number of assessments per assessor, and ordinal response. Following the Landis and Koch interpretation, the kappa found would indicate a moderate reproducibility of the RAS [16]. Yet, this largely spread interpretation rule has been criticized as too dogmatic. With our linear scoring system, the kappa coefficient decreases when the number of response classes increases [17]. Besides, a clinical research protocol is a complex object, leading to a complex assessment tool, so that we did not really expect our kappa to be that close to 1. Therefore, we consider this estimation as a good one, considering the context.

We also estimated an intraclass correlation coefficient through a random proportional odds model. This method is consistent with our data, but relies on several assumptions: we verified the odds proportionality; the existence of a latent continuous risk variable seems very likely, but we obviously could not verify the logistic distribution assumption of this variable; yet, if this assumption is correct, then it is possible to transform the latent variable as to obtain a standard logistic distribution.

Several assessors' characteristics were identified as sources of major disagreement: a male sex, multiple affiliations, no methodological education, a decision-making position, and a longer experience in clinical research. This may designate assessors having to take quick and various decisions, and unused to regulatory evaluation, rather those managing studies hands-on. Based on these results, we recommend to select and/or to train people for risk assessment.

Our study relies on 200 protocols and 49 assessors, both very heterogeneous, so that our results may be generalized. Yet, we focused on participant's risk only. One could also consider the risk for the validity of study results, the risk for the clinical research organization, and the risk for public health. The validation of a more general risk-assessment scale is ongoing among the European Clinical Research Infrastructures Network , where 2 primary risks (for the participant, and for the validity of study results) are considered, and 19 items belonging to 5 themes (participants, validity of study results, organization, governance, and impact on public health) are assessed [18–20].

4.2. Risk-adapted monitoring plan

Secondly, we proposed a risk-adapted monitoring plan in 4 levels of intensity. The most intensive one corresponds to the supposed gold standard in current practice: many on-site visits, and 100% data checked. It is relevant for clinical studies with the highest participant's risk. On the contrary, the less intensive one, without any on-site visit, is relevant for studies with no or few participant's risk, and is possible but associated with intensive remote monitoring: frequent data entry and computerized controls, and frequent queries.

The intensive standard is more and more frequently questioned [3–5], though some are reluctant to change this practice. Yet, even if this strategy has never been validated, it is already successfully used by the Paris Hospital Network and the Medical Research Council [11–13], and most probably by other academic institutions. Besides, as a first approach of monitoring intensity optimization, Liénard et al conducted a randomized study – with *vs*. without on-site initiation visit – nested among a clinical randomized trial [5]. They found no difference in terms of accrual, quantity or quality of data, and duration of follow-up. This is a first indication that some practices may be changed without any consequences on the quality of data.

The proposed risk-adapted monitoring plan is relevant for paper CRF. It was the frame designed for the Optimon trial, for e-CRFs were not frequently used among academic institutions by that time. The use of an e-CRF greatly simplifies acquisition of data into a database, queries handling, and data checks and correction. This will in fact facilitate, and thus promote the implementation of less intensive on-site monitoring strategies. However, the core of the risk-based approach for monitoring activities is still to define the intensity of on-site monitoring depending on patient's risk, not on technical facilities.

4.3. Standard synopsis

Finally, we proposed a standard technical and regulatory-oriented synopsis collecting all information necessary for protocol evaluation. We found no standard synopsis in the literature, and very few among the 200 French protocols from 15 academic sponsors used in our study. Protocol writers often implicitly assumed a wide understanding of clinical field, interventions, and investigations from the readers. This is probably proper for the protocol management staff, but certainly not for the academic protocol evaluators, in particular for sponsors or staff dealing with multiple clinical fields. Besides, the classical abstract is obviously not sufficient, neither for risk evaluation, nor for monitoring plan definition. So that academic sponsors would certainly benefit from promoting the use of a standard synopsis.

5. Conclusions

It is possible to use a validated risk-assessment scale and a risk-adapted monitoring plan for academic clinical research studies. They are now in use in the Optimon randomized non inferiority trial comparing two monitoring strategies: intensive *vs*. risk-adapted consistently with the main scientific and regulatory principles.

Acknowledgements

We thank all persons and institutions that actively participated in the various steps of this work: members of both working groups; institutions having provided protocols, and assessors of the relevance and reproducibility study; the resident in Public Health who handled the feasibility study of the risk-adapted monitoring plan. We also thank all professionals of clinical research that made valuable comments on different versions of the risk-assessment scale, and on the risk-adapted monitoring plan.

Pre-Optimon Steering Committee

Bouxin-Métro A (ANRS), Chêne G (INSERM, CHU de Bordeaux), Gauthier C (INSERM) , Journot V (INSERM), Pignon JP (IGR), Ravaud P (AP-HP)

Risk assessment scale - relevance and reproducibility study

First working group: Bertoye PH (AFSSAPS), Bouxin-Métro A (ANRS), Chêne G (INSERM, CHU Bordeaux), Daurat V (AP-HP), Gauthier C (INSERM), Journot V (INSERM), Pignon JP (IGR), Ravaud P (AP-HP).

Assessors: Alberti C, Allain L, Aouba A, Benhamou E, Blanc E, Boussuge J, Buchmuller A, Bulach C, Calvas F, Carrat F, Chabaud S, Clisant S, Cornu C, Costa Y, Doussau de Bazignan A, Driss H, Dussart S, Fagard C, Félin A, Fougerou C, Fournier I, Gauteul P, Germain C, Gibel L, Giraudeau B, Guiguet M, Jean-Marie C, Jolly N, Juteau N, Lafontaine C, Laporte S, Lauwers-Cances V, Legrand M, Leizorovicz A, Loos-Ayav C, Maison P, Marin B, Mathoulin-Pélissier S, Nguyen JM, Parienté A, Pauporté I, Perez P, Poizat G, Preux PM, Quet F, Rialland A, Roederer T, Serreau R, Thalamas C.

Protocol providers: Agence Nationale de Recherches sur le Sida et les hépatites virales, Assistance Publique – Hôpitaux de Marseille, Assistance Publique – Hôpitaux de Paris (Hôpital Européen Georges Pompidou, Hôpital Lariboisière, Hôpital Saint-Louis, Hôpital Cochin, Hôpital Bichat), Centres Hospitaliers Universitaires (Bordeaux, Limoges, Lyon, Nancy, Rennes, Toulouse), Institut de Cancérologie Gustave Roussy, Institut National de la Santé et de la Recherche Médicale (INSERM).

Risk-adapted monitoring plan – definition and evaluation

Second working group: Alberti C, Aouba A, Baloul S, Blazejewski S, Chambarétaud V, Clisant S, Cornu C, Daurat V, Deygas B, Durette C, Félin A, Fougerou C, Laviolle B, Le Corvoisier P, Nedelec AS, Olazabal L, Rialland A.

Feasibility study: Fleur Grégoire.

Financial support: The Optimon trial and the Pre-Optimon study are supported by the French National Hospital Clinical Research Program (PHRC).

References

[1] International conference on harmonization of technical requirements for registration of pharmaceuticals for human use. ICH Harmonized Tripartite guideline – Guideline for Good Clinical Practice – E6(R1). http://www.ich.org/2007 Accessed June 15, 2010.
[2] Grimes DA, Hubacher D, Nanda K, Schulz KF, Moher D, Altman DG. The Good Clinical Practice guideline: a bronze standard for clinical research. Lancet 2005;366:172–4.
[3] Brosteanu O, Houben P, Ihrig K, Ohmann C, Paulus U, Pfistner B, et al. Risk analysis and risk adapted on-site monitoring in noncommercial clinical trials. Clin Trials 2009;6:585–96.
[4] European Science Foundation, European Medical Research Council. Forward look — investigator-driven clinical trials. http://www.esf.org/activities/forward-looks/medical-sciences-emrc/fl-07-001-investigator-driven-clinical-trials.html2009 Accessed June 15, 2010.
[5] Liénard JL, Quinaux E, Fabre-Guillevin E, Piedbois P, Jouhaud A, Decoster G, et al. Impact of on-site initiation visits on patient recruitment and data quality in a randomized trial of adjuvant chemotherapy for breast cancer. Clin Trials 2006;3:1–7.
[6] Kraemer HC. Extensions of the kappa coefficient. Biometrics 1980;36:207–16.
[7] Hedeker D, Gibbons RD. Longitudinal data analysis. New York: John Wiley & Sons, Inc; 2006.
[8] Billingsley P. Probability and measure. In: Wiley, editor. Probability and mathematical statistics. 3rd ed. New York: John Wiley & Sons; 1995.
[9] Hosmer DW, Lemeshow S. Applied logistic regression. Wiley series in probability and statistics. Second ed. New York: John Wiley & Sons; 2000.
[10] SAS Institute Inc. SAS/STAT® 9.1 user's guide. Cary, NC: SAS Institute Inc.; 2004.
[11] Département de la Recherche Clinique et du Développement de l'Assistance Publique — Hôpitaux de Paris. Tableau récapitulatif de niveau de risque. http://www.drrc.aphp.fr/recherche_clinic/classification/recap_risque.php 2007 Accessed June 15, 2010.
[12] Ménard J, Daurat V. Différences et similarités des promotions industrielles et institutionnelles. Med Sci 2006;22:881–6.
[13] Department of Health of the Medical Research Council. Notes on good practice for research organisations in the management of a portfolio of trials 2: assessment of risk. http://www.ct-toolkit.ac.uk/_db/_documents/MPTrials2.pdf2004 Accessed June 15, 2010.
[14] Bertoye P-H, Courcier-Duplantier S, Best N. Adaptation of the application of good clinical practice depending on the features of specific research projects. Therapie 2006;61:279–85.
[15] Agresti A. Analysis of ordinal categorical data. New York: John Wiley & Sons, Inc; 1984.
[16] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.
[17] Sim J, Wright CC. The Kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther 2005;85:257–68.
[18] Demotes-Mainard J, Ohmann C. European Clinical Research Infrastructures Network: promoting harmonisation and quality in European clinical research. Lancet 2005;365:107–8.
[19] Déti EK, Kubiak C, Chêne G, Journot V, ECRIN Transnational Working Group on Study Monitoring. Elaboration d'un outil d'évaluation du risque des études de recherche clinique académiques. 2ème Conférence Francophone d'Epidémiologie Clinique Nancy 2008. abstract no 90. Rev Épidémiol Santé Publique 2008;56:S77.
[20] Journot V, Gaynor S, Alberti C, Friis-Bach K, Brosteanu O, Joppi R, et al. Risk-based monitoring approach in academic clinical research. 30th meeting of the Society for Clinical Trials Atlanta, GA 2009. abstract no A20. Clin Trials 2009;6:493–524.