

PRE-OPTIMON

Validation de la grille de l'AP-HP évaluant les risque encouru par les patients dans les études de recherche clinique institutionnelles

Version du 22 février 2006

Composition du groupe de travail

A. BOUXIN-METRO	Recherches Cliniques et Thérapeutiques, ANRS
P.H. BERTOYE	Direction de l'Inspection et des Etablissements, AFSSAPS
G. CHENE	Unité de Soutien Méthodologique à la Recherche Clinique et Epidémiologique, CHU de Bordeaux CIC-EC et Unité 593, INSERM
V. DAURAT	DRRC Ile-de-France, AP-HP
C. GAULTIER	DRRC Ile-de-France, GHU Nord, AP-HP Hôpital Robert Debré CIC, INSERM
V. JOURNOT	Unité 593, INSERM
E. MOTTEZ	Département de la Recherche Clinique et Thérapeutique, INSERM
J.P. PIGNON	Service de Biostatistiques, Institut Gustave Roussy
P. RAVAUD	URC Paris Nord, AP-HP Hôpital Bichat Unité 738, INSERM

1. Contexte

Depuis 2001, le Département de la Recherche Clinique et du Développement (DRRC Ile-de-France) de l'Assistance-Publique-Hôpitaux de Paris (AP-HP) utilise une grille standardisée afin d'évaluer le niveau de risque encouru par les patients dans les recherches cliniques dont elle assure la promotion, et d'adapter le niveau de monitoring de ces recherches selon ce niveau de risque.

La mise en place de cette grille a fait l'objet d'une procédure datant de juin 2001, et son utilisation pratique a débuté en septembre 2001. En mai 2005, 386 recherches étaient en préparation ou en cours à la DRRC Ile-de-France, dont 325 avaient été classées selon cette grille (données communiquées par V Daurat).

Cependant cette grille, élaborée par des experts, n'a fait l'objet d'aucune validation formelle. Or de nombreux promoteurs ou acteurs de la recherche clinique, en France et à l'étranger, souhaiteraient utiliser une grille de ce type dans leur pratique de mise en place de leurs recherches. En effet, les modalités de monitoring dépendent du niveau de risque de l'étude. De plus, dans l'essai OPTIMON, cette grille sert de critère d'éligibilité (les recherches à risque de niveau D sont exclues), et de critère de stratification dans l'analyse statistique.

Il est donc important de valider cette grille formellement.

2. Objectifs

Evaluer la pertinence des items composant la grille.

Evaluer la reproductibilité de la grille en tant qu'outil d'évaluation des recherches.

Evaluer la validité d'une grille en 4 niveaux tels que définis.

3. Méthodes

Ces trois objectifs nécessitent des méthodes différentes. La validation de la grille se fera donc en trois étapes.

3.1. *Pertinence des items de la grille*

La première étape consistera à évaluer la pertinence des items constituant la grille et à les modifier si nécessaire.

- *Objectif*

Mettre au point une version améliorée de la grille.

- *Participants*

Un groupe de travail, composé d'un petit nombre de personnes utilisatrices actuelles de la grille ou compétentes dans l'évaluation de protocoles, sera mis en place. Il s'agit de : Pierre-Henri Bertoye, Geneviève Chêne, Véronique Daurat, Claude Gaultier, Valérie Journot, Annie Métro, Estelle Mottez, Jean-Pierre Pignon, Philippe Ravaud.

- *Méthode*

Ce groupe de travail passera en revue toutes les grilles existantes, puis discutera chacun des items composant la grille : types de recherches, critères de classification, niveaux de risque. Eventuellement, s'il le juge nécessaire, il modifiera certains items, en justifiant cette

modification. Il rédigera en outre un guide d'utilisation de la grille en précisant clairement l'interprétation à donner à chaque item.

La grille éventuellement modifiée et son guide d'utilisation seront relus par un groupe plus large d'experts, qui commentera et proposera éventuellement des modifications. Le groupe de travail reverra les commentaires et modifications demandées et les intégrera éventuellement.

La grille finale sera celle utilisée dans l'étude de reproductibilité.

3.2. *Reproductibilité de la grille*

Une étude de reproductibilité sera mise en place. Des protocoles de recherche seront collectés et évalués selon la grille par différents évaluateurs. La reproductibilité de la mesure sera évaluée par le coefficient d'agrément kappa pondéré global [1]. Par ailleurs, un modèle à cotes proportionnelles permettra d'identifier les sources de variabilité de l'évaluation et d'évaluer leur part respective dans la variabilité totale, et un coefficient de corrélation intraclasse sera calculé.

- *Objectif*

Évaluer la reproductibilité de la grille.

Évaluer l'influence des caractéristiques des protocoles et des évaluateurs sur l'évaluation.

Si possible, évaluer la variabilité de l'évaluation liée à une évaluation individuelle ou collective par structure.

- *Protocoles à évaluer*

Les protocoles seront des protocoles de recherche clinique de tous types, dans tous domaines, dont le promoteur est institutionnel. Un protocole détaillé devra avoir été rédigé, selon un plan présentant au moins certains nombres d'items (liste à définir à l'avance), et comporter un résumé.

La collecte des protocoles se fera auprès de la Direction à la Recherche Clinique de Paris et de Bordeaux, de l'ANRS, de l'INSERM, de la FNCLCC, et auprès de représentants d'organismes ou de pays ayant montré un intérêt pour la grille AP-HP. Les protocoles seront collectés des plus récents aux plus anciens, jusqu'à avoir collecté un nombre suffisant de protocoles.

- *Nombre de protocoles nécessaire*

Le coefficient d'agrément kappa a été développé initialement pour estimer l'accord entre 2 évaluateurs lorsqu'ils évaluent plusieurs sujets sur une réponse en 2 catégories. Au sein de l'accord observé, ce coefficient permet de distinguer l'accord dû au hasard de l'accord réel entre évaluateurs.

Il est possible de calculer un coefficient kappa dans le cas d'un nombre d'évaluateurs supérieur à 2 mais constant par protocole, et d'un nombre supérieur à 2 catégories de réponses^[1]. Par contre, il est impossible de calculer un nombre de protocoles nécessaire dans ce cas. On se ramènera donc au cas simple de 2 évaluateurs avec une réponse en 2 catégories.

Le bilan de la répartition des protocoles promus par la DRRC Ile-de-France entre 2001 et 2005 montre que la proportion des différents niveaux de risque de la grille AP-HP était de 36% pour le niveau A, 25% pour le niveau B, 12% pour le niveau C et 27% pour le niveau D. On fera donc les calculs selon deux cas de figure : la proportion de réponse dans une catégorie est de 0,30, ou bien de 0,15.

Pour un kappa calculé pour 2 notateurs et une réponse à 2 niveaux, le logiciel nQuery Advisor 6.0 donne les résultats suivants : sous l'hypothèse que la proportion de protocoles classés dans

l'un des niveaux de risque est de 0,30 (respectivement 0,15), un test pour grand échantillon bilatéral au risque $\alpha=0,05$ de l'hypothèse nulle " $\kappa=0,50$ " aurait une puissance de 80% de détecter un κ alternatif de 0,70 avec un nombre de protocoles de 173 (respectivement 273).

Compte tenu de ce résultat et de la lourdeur de la tâche d'évaluation pour un tel nombre de protocoles, un nombre de protocoles de 200 semble un bon compromis.

- *Evaluateurs*

Les évaluateurs seront des personnes ayant une activité en recherche clinique (méthodologistes, chefs de projets, statisticiens, ARC seniors) et issus des différents organismes et pays participants. Les investigateurs coordonnateurs des études pour lesquelles il est envisagé l'inclusion dans Optimon pourront également participer aux évaluations, mais leurs évaluations n'entreront pas dans le plan d'analyse au même niveau. D'autre part, les évaluateurs d'une même structure pourront participer en tant que structure, à condition qu'il y ait au moins trois évaluateurs par structure. Dans ce cas, une évaluation individuelle sera faite par chaque évaluateur, puis une évaluation collective consensuelle sera faite par la structure sur les mêmes protocoles.

Chaque évaluateur remplira un questionnaire permettant de décrire ses caractéristiques (sexe, âge, nationalité, formation initiale, métier, fonction, expérience pratique de la recherche clinique, domaine habituel de recherche, implication de terrain ou administrative,...), ce qui permettra de mieux étudier les sources de variabilité dans le modèle à cotes proportionnelles.

Avant les premières évaluations, les évaluateurs recevront une formation standard à l'utilisation de la grille, et disposeront ensuite d'un guide d'utilisation.

Chaque évaluateur produira une évaluation pour chaque protocole à évaluer.

Par ailleurs, une évaluation consensuelle sera établie par le groupe de travail initial (évaluation séparée, puis réunion téléphonique pour arriver à un consensus en cas de discordances).

- *Plan d'expérience*

Il est impossible de demander aux évaluateurs d'évaluer chacun la totalité des protocoles. Le nombre de protocoles qu'un évaluateur peut évaluer est fixé à 20. Par ailleurs, il semble possible de recruter 40 évaluateurs pour cette étude.

On constituera donc 20 groupes de 2 évaluateurs ou structures, et 20 groupes de 10 protocoles. On construira un plan en blocs incomplets équilibrés sur les groupes d'évaluateurs et les groupes de protocoles, avec 2 groupes de protocoles évalués par groupe d'évaluateurs ou structures, selon le schéma suivant.

		groupes de 10 protocoles																				
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
groupes de 2 évaluateurs et/ou structures	1	x	x																			
	2		x	x																		
	3			x	x																	
	4				x	x																
	5					x	x															
	6						x	x														
	7							x	x													
	8								x	x												
	9									x	x											
	10										x	x										
	11											x	x									
	12												x	x								
	13													x	x							
	14														x	x						
	15															x	x					
	16																x	x				
	17																	x	x			
	18																		x	x		
	19																			x	x	
	20	x																				x

Figure : Plan en blocs incomplets équilibrés sur les groupes d'évaluateurs et les groupes de protocoles. Pré-Optimon.

De plus, la constitution des groupes d'évaluateurs ou structures et de protocoles sera aléatoire. L'attribution des numéros de groupe à ces groupes sera aléatoire.

Ce type de plan nécessite absolument d'identifier à l'avance les évaluateurs et les protocoles. Il y a donc une phase importante de préparation (contacts, demandes d'accord de participation, demande de protocoles, collecte des protocoles, envoi et récupération des questionnaires) avant la phase d'évaluation proprement dite. Si le nombre d'évaluateurs et/ou structures identifiés atteint 40, on pourra accepter les évaluateurs d'une même structure en tant que structure. Sinon, ces évaluateurs seront considérés à titre individuel et répartis dans les groupes au hasard.

Il est possible que ce projet rencontre un intérêt très fort de la communauté de la recherche clinique, et que le nombre d'évaluateurs et/ou structures souhaitant participer dépasse 40. Dans ce cas, on augmentera la taille des groupes d'évaluateurs et/ou structures. Par ailleurs, les investigateurs coordonnateurs qui participeront ne seront pas comptés parmi les 40 évaluateurs et/ou structures, mais augmenteront la taille des groupes d'évaluateurs et/ou structures. Les groupes d'évaluateurs et/ou structures pourront être de taille inégale, mais leur constitution sera effectuée au hasard.

- *Support de l'évaluation*

En première approche, l'évaluation se fera à partir du résumé du protocole. Si cela avère impossible parce que le protocole n'est pas assez détaillé, l'évaluation se fera sur le protocole complet. L'information du support ayant servi à l'évaluation et les difficultés rencontrées seront collectées et analysées.

- *Analyse statistique*

Dans une première étape, on évaluera la reproductibilité de la grille avec un mode d'utilisation individuel. Si une structure participe, on tirera au sort un de ses évaluateurs afin d'avoir une évaluation par un même nombre d'évaluateurs pour tous les protocoles. Dans une deuxième étape, on évaluera la concordance entre évaluation individuelle et collective et on comparera les variabilités intra- et interstructures.

Un coefficient d'agrément pondéré kappa sera calculé globalement ^[1]. On calculera également un coefficient kappa pour chaque paire de niveau de risque.

On construira un modèle à cotes proportionnelles avec différentes sources de variabilité : les protocoles, les évaluateurs et leurs caractéristiques. On calculera un coefficient de corrélation intraclasse de façon à estimer l'accord global entre les évaluateurs. On affinera ensuite l'analyse en introduisant dans le modèle différents facteurs fixes ou aléatoires correspondant aux caractéristiques des évaluateurs, de façon à identifier les sources de variabilité les plus fortes, et éventuellement proposer des conditions optimales d'utilisation de la grille.

3.3. Validation des niveaux de risque

- *Objectif*

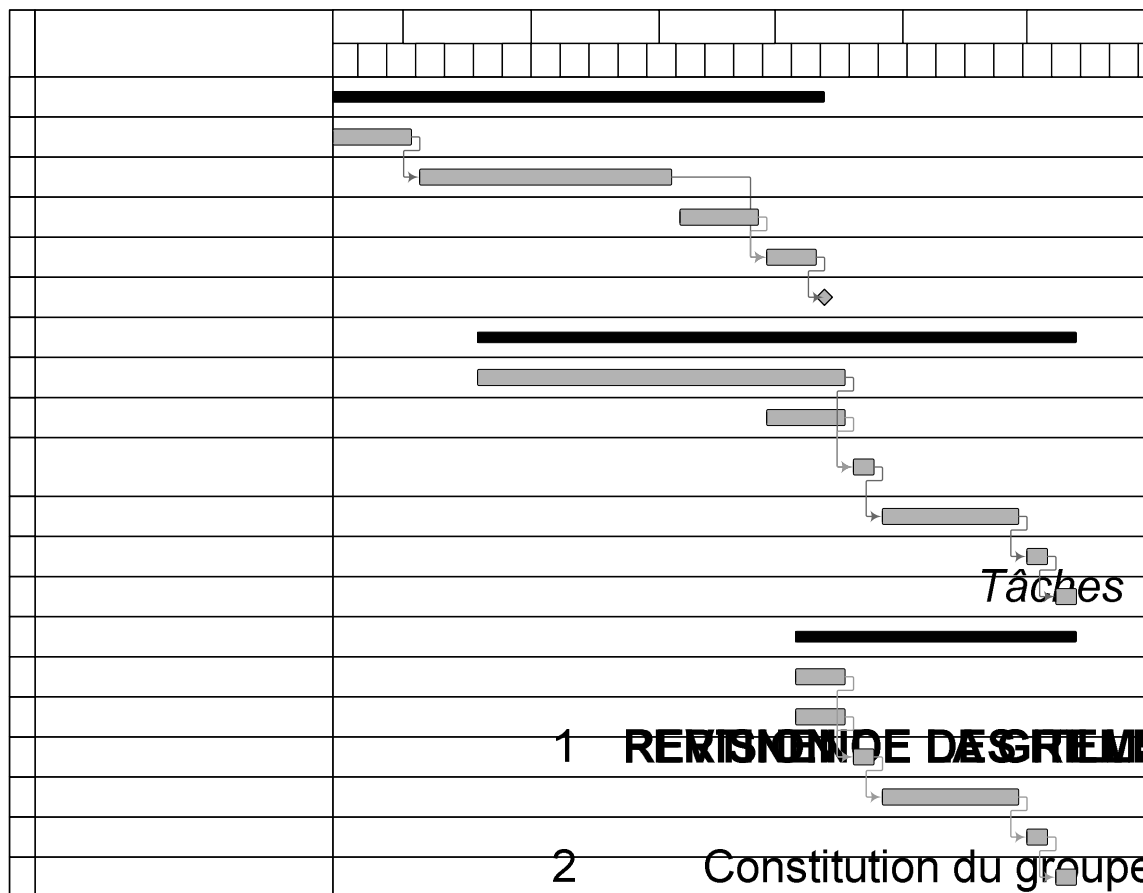
Evaluer la validité d'un découpage en 4 niveaux de risque tels que définis.

- *Méthode*

On comparera une évaluation de mêmes protocoles par la grille et par une échelle visuelle analogique, de façon à valider l'existence d'une gradation en 4 niveaux. On sélectionnera par tirage au sort 50 protocoles parmi les 200 évalués grâce à la grille, de façon à obtenir une répartition homogène des 4 niveaux de risque. Les évaluateurs appartiendront à 3 groupes de métier : investigateurs, méthodologistes, promoteurs.

On calculera pour chaque niveau de risque déterminé par la grille la valeur moyenne (ou médiane selon la distribution observée) sur l'échelle visuelle analogique. On effectuera une analyse de tendance globale, en essayant différentes tendances et en particulier une tendance linéaire. Par ailleurs, on effectuera une analyse de variance. Si les niveaux s'ordonnent bien en A, B, C, D, on comparera les niveaux successifs deux à deux (A vs B, B vs C, C vs D), avec ajustement du risque α par la méthode de Bonferroni, afin de vérifier que tous les niveaux sont bien distincts. Si l'ordre attendu n'est pas retrouvé, on comparera les niveaux par la méthode de comparaisons multiples de Scheffé, afin de définir les niveaux distincts. Pour chaque analyse, on vérifiera les hypothèses de normalité.

4. Calendrier



O : équipe Optimon ; GT : groupe de travail ; R : relecteurs

5. Références

1 Fleiss J L. Statistical methods for rates and proportions. New York: John Wiley & Sons, 1981; 321p.

- 3 Révision de la grille
- 4 Rédaction d'un guide d'utilisation
- 4 Diffusion aux relecteurs (R)
- 5 Relecture de la grille et du guide
- 6 Révision de la grille et du guide
- 7 **REPRODUCTIBILITE
ETUDE DE LA REPRODUCTIBILITE
VALIDATION DES NIVEAUX DE RISQUE**
- 8 Identification des évaluateurs (E)
- 9 Collecte des protocoles
- 10 Création du plan d'expérience
- 10 Diffusion de la grille et des protocoles