



institut**Curie**

# Évaluation et comparaison des règles de décision d'arrêt précoce pour survenue d'EIG dans les essais cliniques

Juliette Djadi-Prat, Yann De Rycke

Unité de biostatistique du Dr. B. Asselain  
Institut Curie  
26 rue d'Ulm  
75248 Paris Cedex 05

# Monitoring des EIG

---

- EIG : événements indésirables graves (critère binaire, survenue précoce)
- Décision par le Comité de surveillance de l'essai clinique :
  - Sans outil statistique spécifique
  - Outils statistiques utilisés :
    - Méthodes existantes adaptées à la toxicité et aux essais de phase III
    - Nouvelles méthodes dédiées
- Analyse séquentielle, indépendamment du critère de jugement principal
  - **Inflation du risque** de conclure à tort à une toxicité excessive ( $\alpha$ )
  - Difficulté logistique

# Caractéristiques des méthodes étudiées

---

- Visent à **contrôler le risque de première espèce** consenti pour le suivi des EIG (qui peut être différent de celui consenti pour le critère de jugement principal).
- Utilisent uniquement une frontière haute (traduisant des EIG trop fréquents), et pas de frontière basse (qui traduirait des EIG qui ne seraient pas trop fréquents).
- Le **nombre maximal de sujets inclus**  $N_{\max}$  dans l'essai est toujours déterminé à partir du critère de jugement principal  
→ la puissance peut être insuffisante en fin d'essai pour conclure.

# Comparaison de méthodes fréquentistes

---

## • Méthodes classiques :

- Développées pour l'évaluation de **l'efficacité**.
- Adaptées secondairement à ce type de surveillance.
  - Méthode de **Wald** adaptée aux EIG
  - Méthode de **Fleming** adaptée aux EIG (2, 3, 4 et 5 étapes)

## • Nouvelles approches :

- Pour les EIG, les cliniciens préfèrent un **suivi continu** des EIG, pour lesquels certaines méthodes plus flexibles peuvent être envisagées.
  - Méthode de **Goldman**
  - Méthode de **Kramar** (fonction de dépense du risque  $\alpha$ )

→ **Comparaison par simulations**

---

# Analyses intermédiaires

---

## Hypothèses de test (tests unilatéraux)

$$H_0: p \leq p_0$$

$$H_1: p > p_0$$

Avec :

$p$  : proportion théorique d'EIG

$p_0$  : proportion maximale d'EIG considérée acceptable

On choisit une valeur particulière  $p_1$  pour l'hypothèse alternative (*proportion d'EIG considérée comme inacceptable*)

- Nécessaire pour l'utilisation de certaines méthodes
- Nécessaire pour les calculs de puissance

# Principes

---

## • Méthodes classiques (Wald, Fleming)

- Les analyses sont réalisées en fonction du nombre de sujets inclus ( $n$ )
- La décision d'interruption est basée sur le **nombre d'EIG** ( $k$ ) : chaque méthode permet de calculer, en fonction du nombre de sujets inclus, un nombre  $K_n$  d'EIG seuil considéré comme excessif.

## • Méthodes dédiées (Goldman, Kramar)

- Les analyses sont réalisées en fonction du nombre d'EIG ( $k$ )
- La décision d'interruption est basée sur le **nombre de sujets inclus** ( $n$ ) : chaque méthode permet de calculer, en fonction du nombre d'EIG, un nombre de sujets inclus  $N_k$  en dessous duquel les EIG sont considérés comme trop fréquent.

# Configurations étudiées

---

- Pour chaque combinaison de :

- $N_{\max} = 25, 60, 140, 200$
- $p_0 = 0.03, 0.05, 0.10, 0.20$   
 $p_1 = p_0 + 0.15$

- Pour l'application des méthodes :

- 3 couples  $(\alpha, \beta) = (0.05, 0.20), (0.10, 0.20), (0.20, 0.10)$
- Fleming à 2, 3, 4, 5 étapes équidistantes, la dernière étant réalisée à la fin de l'essai
- Kramar:  $\gamma = -4, 1, 4$

➔ Au total : 16 types d'essai  $\times$  3 paramètres de test  $\times$  7 méthodes

---

# Résultats



institutCurie



# Méthode de Goldman

---

- Temps de calcul très longs
- En pratique, ces règles de décision n'ont pu être calculées et appliquées que pour les configurations où  $N_{\max}=25$ , et où  $p_0$  n'est pas trop élevé.
- Pour ces configurations particulières, on observe:
  - $\alpha_{\text{simul}}$  équivalent au risque  $\alpha_{\text{calc}}$  calculé lors du choix de la règle de décision
  - $\beta_{\text{simul}}$  équivalent au risque  $\beta_{\text{calc}}$  calculé lors du choix de la règle de décision

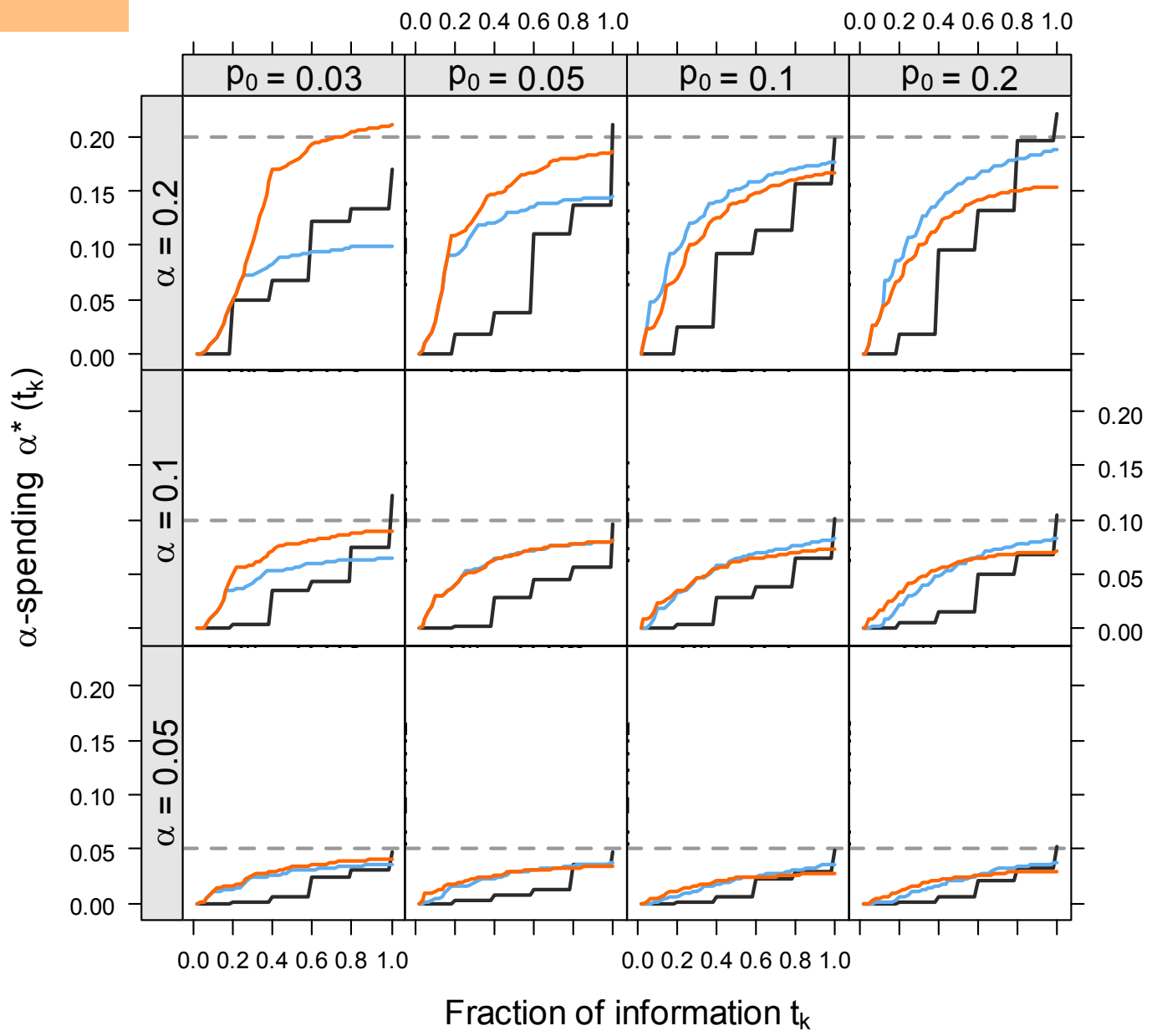
# Pour l'ensemble des méthodes

---

- Tendance à la sous-estimation du risque de 1<sup>ère</sup> espèce
- La puissance  $1 - \beta_{\text{simul}}$  est insuffisante quand
  - $N_{\text{max}}$  est faible
  - $p_0$  est faible
- À l'inverse, pour des grands effectifs, toutes les méthodes sont surpuissantes

**Dépense du risque  $\alpha_{\text{simul}}$**   
 $N_{\text{max}}=60, \beta=0.20$

Methods  
 Fleming — Wald — Kramar —



# Méthode de Wald

---

- $\alpha_{\text{simul}}$  est inférieur au risque  $\alpha$  initialement consenti

En particulier quand

–  $N_{\text{max}}$  est faible

–  $p_0$  est faible

–  $\alpha$  est grand

- La dépense du risque  $\alpha_{\text{simul}}$  est d'allure convexe:  
→ la dépense du risque se fait préférentiellement en début d'essai.
- $1 - \beta_{\text{simul}}$  est inférieure à la puissance des autres méthodes

# Méthode de Fleming

---

- Dans la plupart des configurations,  $\alpha_{\text{simul}}$  est proche de la valeur  $\alpha$  fixée initialement.  
Légèrement moins satisfaisant quand  $N_{\text{max}}=25$ .
- Dépense du risque  $\alpha_{\text{simul}}$  difficile à évaluer sur ces analyses groupées: plutôt concave? Avec une dépense favorisée en fin d'essai
- Méthode la plus puissante
- Nombre moyen d'EIG observés sous  $H_1$  en cas d'arrêt supérieur aux autres méthodes

# Méthode de Kramar

---

- $\alpha_{\text{simul}}$  est presque toujours sensiblement inférieur à  $\alpha$ , d'autant plus que  $p_0$  est grand.
- La dépense du risque  $\alpha_{\text{simul}}$  respecte l'allure de la fonction choisie (paramètre  $\gamma$ ).
- La puissance est au moins équivalente à celle des autres méthodes, à l'exception des configurations où  $p_0$  est grand et  $N_{\text{max}}$  petit.

# Discussion



# Discussion : à propos des méthodes

---

- **WALD** :

- Plan ouvert tronqué
- Borne inférieure supprimée
- Lourdeur de mise en œuvre.

- **MULTI-ÉTAPES DE FLEMING** :

- Borne inférieure supprimée
- Approximation normale
- Nécessité de planification du nombre et des temps de réalisation des analyses intermédiaires
- Procédure groupée, pas forcément adaptée aux EIG



# Discussion : à propos des méthodes

---

- **GOLDMAN** :

- Fiabilité en termes de risques
- Mise en œuvre limitée en pratique

- **KRAMAR** :

- Approximation normale dans le calcul des  $\alpha'$
- La dernière analyse n'a aucune raison d'avoir lieu après l'inclusion du dernier patient
  - une partie du risque  $\alpha$  ne peut pas être dépensé
- Il est possible qu'aucune analyse ne soit réalisée (pas d'EIG)

# Pour quels événements?

---

## Méthodes développées pour :

- EIG suivis définis lors de l'élaboration du protocole (événements rares ou inattendus) *événement responsable d'hospitalisation, de prolongation d'hospitalisation, d'incapacité, ou engageant le pronostic vital (EMA)*.
- Critère binaire** (présence/absence) pour ce type de suivi.
- De survenue **précoce** par rapport au rythme d'inclusion.
- Un seul temps d'évaluation de l'absence ou présence des EIG : on ne s'intéresse qu'à la survenue du premier événement (ou EIG unique).

## En pratique :

- Choix de certaines EIG d'intérêt selon la balance bénéfique/risque, la disponibilité des données (multi-centrique)...
- Extension à d'autres événements d'intérêt

# Avec quelles hypothèses $p_0$ $p_1$ ?

---

- Grande sensibilité des conclusions selon les hypothèses de test choisies (autres paramètres égaux par ailleurs).
- En particulier pour des fréquences si faibles.
- Difficulté de fixer des hypothèses précises.

# Recommandations ?

---

- **Pour les essais de petites tailles**

- Méthode de Goldman particulièrement pertinente si disponible

- **Pour les essais de grandes tailles**

- La méthode de Fleming semble être un bon compromis, en termes de résultats et d'organisation logistique
- Méthode de Kramar permet d'imposer la dépense du risque  $\alpha$ , mais méthode lourde à mettre en place
- La méthode de Wald doit être réservée au très gros essais, avec des fréquences élevées

# Conclusion

---

- **Chaque méthode présente ses avantages et ses limites propres.**

## **Le choix est fonction de :**

- L'effectif de l'essai
- Les fréquences  $p_0$  et  $p_1$  acceptées
- La logistique disponible (Kramar vs Wald ou Fleming)
- La priorité donnée
  - à l'exactitude des risques  $\alpha$  et  $\beta$
  - à la fonction de dépense du risque  $\alpha$
  - aux nombres moyens de sujets inclus et d'EIG attendus

## • **Aide à la décision**

## • **Perspectives :**

- Utilisation de ces méthodes pour des toxicités autres que les EIG
- Package récent développé par l'équipe d'A. Kramar



institut**Curie**

**Merci de votre attention**

JDP

# Références (1)

---

- Armitage P. Sequential tests in prophylactic and therapeutic trials. *Quaterly J. Med.* 1954; 23:255-74.
- Armitage P, McPherson CK, Rowe BC. Repeated significance tests on accumulating data. *J. R. Statist. Soc. A.* 1969;132:235-44.
- Fleming TR. One-sample multiple testing procedure for phase II clinical trials. *Biometrics.* 1982;38:143-51.
- Goldman AI, Hannan PJ. Optimal continuous sequential boundaries for monitoring toxicity in clinical trials : a restrictedsearch algorithm. *Statistics in Medicine* 2001; 20:1575-89.
- Goldman AI. Issues in designing sequential stopping rules for monitoring side effects in clinical trials. *Controlled Clinical Trials* 1987; 8:327-37.

## Références (2)

---

- Hwang IK, Shih WJ, De Cani JS. Group sequential designs using a family of type I error probability spending functions. *Stat. Med.* 1990;9:1439-45.
- Kramar A, Bascoul-Mollevi C. Early stopping rules in clinical trials based on sequential monitoring of serious adverse events. *Med Decis Making*, 2009 May-Jun;29(3):343-50.
- Schoenfeld DA. A simple algorithm for designing group sequential clinical trials. *Biometrics.* 2001 Sept;57(3):972-4.
- Wald. Sequential tests of statistical hypotheses. *Annals of Mathematical Statistics* 1945; 16(2):117-86.
- Wald A. Sequential Analysis. 1947 Wiley, New York.
- Wald A, Wolfowitz J. Optimum character of the sequential probability ratio test. *Ann. Math. Stat.* 1948; 19:326-39.

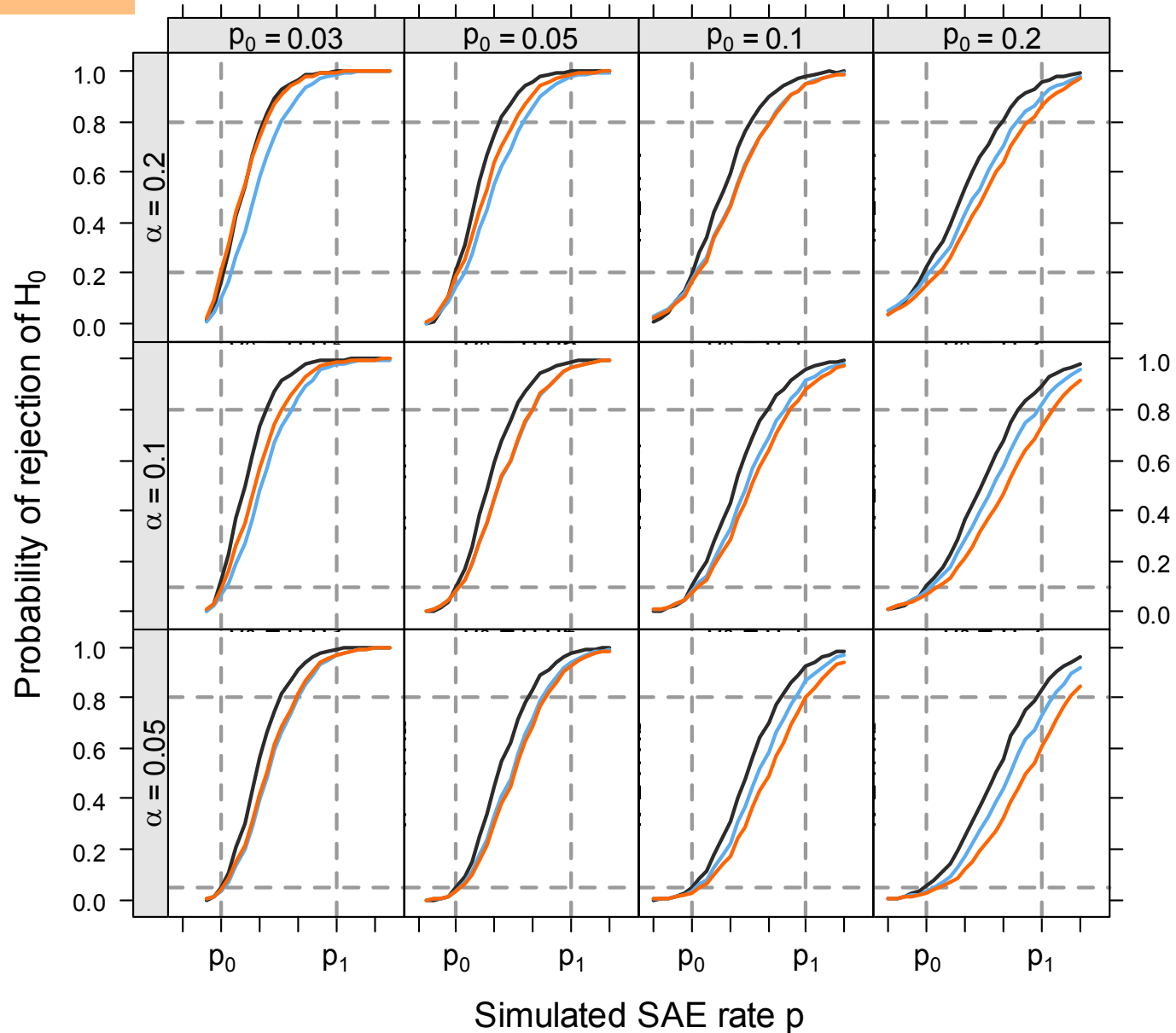
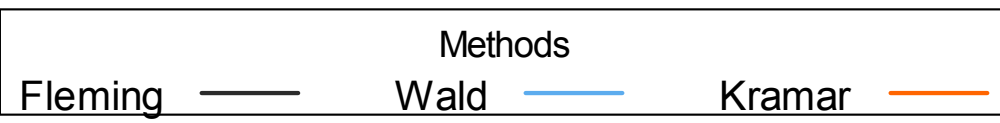


# Annexes



# Probabilité de rejet de $H_0$

$N_{\max}=60, \beta=0.20$



**Wald**



institut**Curie**

# La procédure séquentielle de Wald (SPRT)

---

- **Analyses séquentielles continues** : réalisation d'une analyse après l'inclusion de chaque nouveau patient
- Statistique de test : logarithme du rapport des vraisemblances RV.
- Règle de décision : les bornes d'interruption sont déterminées selon les erreurs de type I et II consenties :
  - Poursuite de l'essai tant que :

$$\left( \ln\left(\frac{\beta}{1-\alpha}\right) < \right) \ln RV < \ln\left(\frac{1-\beta}{\alpha}\right)$$

- Donc interruption dès que :

$$\ln\left(\frac{\text{Likelihood under } H_1}{\text{Likelihood under } H_0}\right) \geq \ln\left(\frac{1-\beta}{\alpha}\right) \qquad \ln\left(\frac{\mathcal{B}(k_n, n, p_1)}{\mathcal{B}(k_n, n, p_0)}\right) \geq \ln\left(\frac{1-\beta}{\alpha}\right)$$

# Calcul de la règle de décision à partir du SPRT (2)

---

- Dans la surveillance de la toxicité, seul un excès de toxicité peut amener à interrompre l'essai → la borne inférieure est ignorée.
- Il suffit de vérifier cette inégalité lors de la survenue de chaque événement indésirable.

- Donc: rejet de  $H_0$  si :  $\ln RV_n \geq \ln\left(\frac{1-\beta}{\alpha}\right)$

$$e \geq \frac{\ln\left(\frac{1-\beta}{\alpha}\right) - n \times [\ln(1-p_A) - \ln(1-p_0)]}{(\ln p_A - \ln p_0) - [\ln(1-p_A) - \ln(1-p_0)]}$$

$$n(e) \leq \frac{\ln\left(\frac{1-\beta}{\alpha}\right) + e \times [\ln(1-p_A) - \ln(1-p_0) - \ln p_A + \ln p_0]}{\ln(1-p_A) - \ln(1-p_0)}$$

# Calcul des nombres d'EIG seuil $K_n$

---

- En exprimant le rapport de vraisemblance, à partir des distributions binomiales du nombre d'EIG sous l'hypothèse nulle et sous l'hypothèse alternative :

$$K_n = \frac{\ln\left(\frac{1-\beta}{\alpha}\right) - n \times \ln\left(\frac{1-p_1}{1-p_0}\right)}{\ln\left(\frac{p_1}{p_0}\right) - \ln\left(\frac{1-p_1}{1-p_0}\right)}$$

- Permet d'obtenir la droite frontière du nombre d'EIG  $k$ , en fonction du nombre de sujets inclus  $n$ .

**Fleming**



institut**Curie**

# Fleming

---

- On fournit initialement  $p_0$ ,  $(p_1)$ ,  $N_{\max}$ ,  $\alpha$  (et  $\beta$ ),  $E$ ,  $\{n_1, n_2, \dots, n_E\}$ .
- Pour chaque analyse  $e$  : calcul des bornes inférieure ( $a_e$ ) et supérieure ( $b_e$ ) du nombre d'événements observés entre lesquelles l'essai est poursuivi :

$$b_e = n_e \times p_0 + z_{1-\alpha} \sqrt{N_{\max} \times p_0 (1 - p_0)}$$

$$a_e = n_e \times p_a - z_{1-\alpha} \sqrt{N_{\max} \times p_a (1 - p_a)}$$

$$\text{avec } p_a = \frac{\left[ \sqrt{N_{\max} \times p_0} + z_{1-\alpha} \sqrt{1 - p_0} \right]^2}{N_{\max} + z_{1-\alpha}^2}$$

- Règles de décision : à l'étape  $e$  :
  - si le nombre d'événements  $k_e \geq b_e$ , stop et rejet de  $H_0$
  - (si le nombre d'événements  $k_e \leq a_e$ , stop et non rejet de  $H_0$ )
  - Si ( $a_e <$ ) nombre d'événements  $k_e < b_e$ , on poursuit jusqu'à l'étape  $e+1$



**Goldman**



institut**Curie**

# Méthode « Goldman » (1987, 2001) (1)

---

- Repose sur la borne sup du SPRT
- Changement de point de vue : on exprime le nombre de sujets inclus en fonction du nombre d'EIG observé  
→ règle d'interruption pour les événements successifs

$$n_k \leq N_k = \frac{\ln\left(\frac{1-\beta}{\alpha}\right) + k \times [\ln(1-p_1) - \ln(1-p_0) - \ln p_1 + \ln p_0]}{\ln(1-p_1) - \ln(1-p_0)}$$

$$n_{obs} \leq R_{SPRT} = \{n_1, n_2, n_3 \dots\}$$

## Méthode « Goldman » (2)

---

- A partir d'une règle donnée (finie), de  $p_0$  et  $p_1$ , on calcule les valeurs exactes de  $\alpha$  ( $\alpha_{\text{calc}}$ ) et  $\beta$  ( $\beta_{\text{calc}}$ ), l'espérance du nombre de sujets inclus (ASN) sous  $H_1$  et une fonction de coût ERL dépendant de  $\alpha_{\text{calc}}$  et  $\beta_{\text{calc}}$ .
- Ces éléments permettent de choisir une règle parmi d'autres.
- Les auteurs ont mis au point un algorithme permettant de tester toutes les règles d'interruption possible à partir de ces critères.

## Méthode « Goldman » (3)

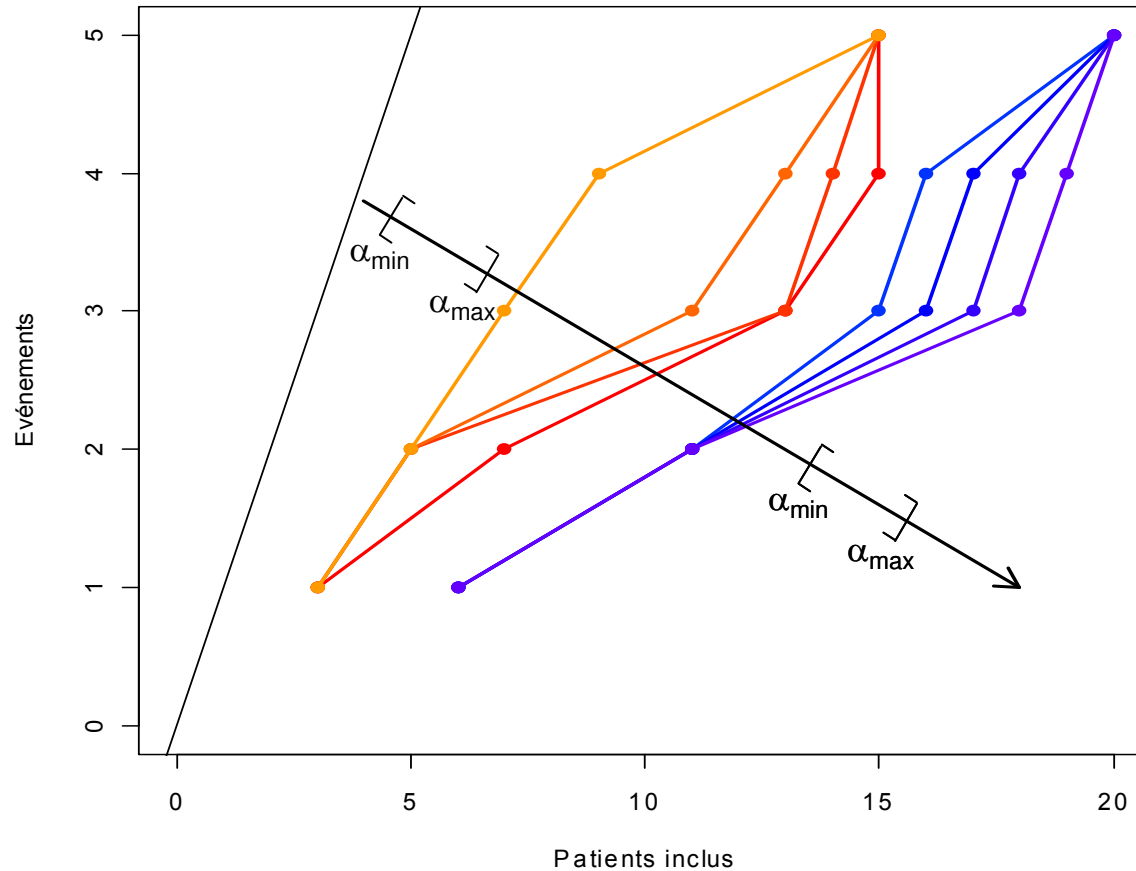
---

- Il reste une seule règle choisie, qui optimise  $\alpha_{\text{calc}}$ ,  $\beta_{\text{calc}}$ , ASN et ERL :
  - toutes les règles
  - parmi lesquelles certaines respectent  $\alpha_{\text{calc}} \in [\alpha - \delta; \alpha + \delta]$  (par ex: [0.049; 0.051])
  - parmi lesquelles certaines respectent  $1 - \beta_{\text{calc}} \geq \text{Puissance}_{\text{min}}$
  - parmi lesquelles on choisit la règle la plus pertinente (fonction de  $\alpha_{\text{calc}}$ ,  $\beta_{\text{calc}}$ , ASN et ERL)
- On choisit dans ce travail la règle qui minimise:

$$(\alpha - \alpha_{\text{calc}})^2 + (\beta - \beta_{\text{calc}})^2$$

# Règles ordonnées (Goldman)

Règles d'interruption ordonnées



→ Sens d'augmentation d' $\alpha$  et de la puissance

R = { 3,5,7,9,15 }

R = { 6,11,18,19,20 }

**Kramar**



institut**Curie**

# Méthode « Kramar » (2005) (1)

---

- Lorsque  $k$  EIG sont observés parmi  $n_k$  sujets inclus, on peut établir l'intervalle de confiance unilatéral (exact) au risque  $\alpha$  de la proportion observée, à partir de la distribution binomiale :

$$\sum_{x=k}^n C_n^x p_{\text{inf}}^x (1 - p_{\text{inf}})^{n-x} = \alpha$$

$$p_{\text{obs}} \rightarrow \text{IC}_{1-\alpha} = [p_{\text{inf}} ; 1]$$

- On compare  $p_0$  à la borne inférieure de l'intervalle de confiance unilatéral (au seuil choisi) du taux de toxicité observée :
  - si  $p_{\text{inf}} > p_0 \rightarrow p_{\text{obs}}$  est incompatible avec  $p_0 \rightarrow$  arrêt de l'essai
  - si  $p_{\text{inf}} \leq p_0 \rightarrow p_{\text{obs}}$  est compatible avec  $p_0 \rightarrow$  poursuite de l'essai

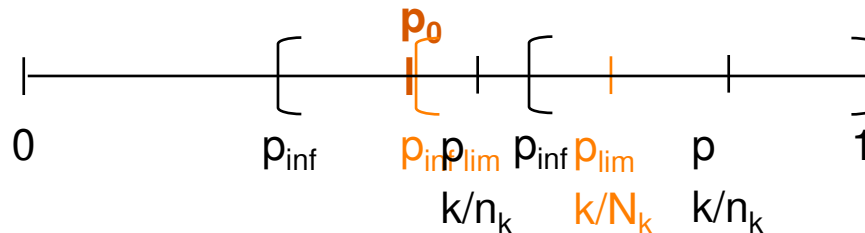
## Méthode « Kramar » (2)

- Par itération sur  $n$ , à  $k$  fixé, on détermine l'effectif MAXIMAL  $N_k$  de sujets inclus lors de la survenue de  $k$  EIG, qui vérifie :

$$p_0 < p_{\text{inf}}$$

- La règle de décision devient :
  - si  $n_k \leq N_k \rightarrow p_{\text{obs}}$  est incompatible avec  $p_0 \rightarrow$  arrêt de l'essai
  - si  $n_k > N_k \rightarrow p_{\text{obs}}$  est compatible avec  $p_0 \rightarrow$  poursuite de l'essai :

A  $k$  fixé :



- Les résultats sont fournis sous forme de table, pour différentes valeurs de  $p_0$ .



# Méthode « Kramar » (4) : Contrôle du risque $\alpha$

---

- **Choix d'une méthode de consommation du risque  $\alpha$  :**

– Fonction de la famille  $\gamma$  (Hwang, Shih, DeCani, 1990), dépendant de la fraction d'information  $t_k$ :

$$\alpha^*(\gamma, t_k) = \alpha_0 \frac{1 - e^{-\gamma \times t_k}}{1 - e^{-\gamma}}$$

– préférence des auteurs pour une consommation prédominante en début d'essai : choix de  $\gamma=4$

- **A partir des  $\alpha^*(t_k)$ , on calcule les  $\alpha'_k$  nominaux qui servent à établir les effectifs maximaux de sujets inclus, en utilisant l'algorithme d'Armitage (1969) :**

– A partir des  $\alpha^*(t_k)$  calculés, on détermine les  $c_k$  associés :

$$P_0(W(t_1) \geq c_1) = \alpha^*(t_1)$$

puis  $\alpha^*(t_1) + P_0(W(t_1) < c_1 \text{ \& } W(t_2) \geq c_2) = \alpha^*(t_2) \dots$

– A partir des  $c_k$ , on obtient les  $\alpha'_k$  :

tel que  $\alpha'_k = 1 - \Phi(c_k)$  ( $\Phi$  : fonction de répartition de  $N(0, 1)$ )

Fonctions de dépense du risque  $\alpha$   
Famille des fonctions  $\gamma$  (Hwang IK, Shih WJ, De Cani JS)  
(exemple :  $\alpha=5\%$ )

