

Quand arrêter l'inclusion de patients ?

Une heuristique corrigée quand la valeur de convergence est inconnue

Hugues Genvrin

Complément aux *Éléments de statistiques* — octobre 2026

Le problème

L'heuristique initiale

Fixer la précision voulue

La règle de Chow et Robbins

Comparer deux traitements

Prendre en compte le coût

Synthèse

Le problème

La situation

- On inclut des patients un par un ; chaque inclusion a un **coût**.
- On ne connaît pas à l'avance l'écart-type σ de la population, c'est-à-dire la **valeur de convergence** de l'écart-type de la série.
- On voudrait une règle qui dise : « on en sait assez, inutile d'inclure de nouveaux cas ».

Ce que montre la diapositive 15

L'écart-type de la série converge vers σ , qui **n'est pas nul** en général : une loi normale ou de Poisson garde sa dispersion.

La question à se poser

Qu'est-ce qui doit avoir « convergé » pour que l'on puisse s'arrêter ?

Deux quantités à ne pas confondre

L'écart-type des individus

$$s_n \xrightarrow[n \rightarrow \infty]{} \sigma > 0$$

Il décrit la dispersion des patients. Il **ne tend pas vers zéro**.

L'erreur-type de la moyenne

$$\frac{s_n}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{} 0$$

Elle décrit l'incertitude sur la moyenne.
Elle **tend vers zéro**.

- C'est l'erreur-type qui détermine la précision d'une estimation et la puissance d'un test.
- Le bon critère d'arrêt porte donc sur la **précision**, pas sur la stabilité de la variance.
- La puissance **se calcule**, dès qu'on fixe la différence recherchée et qu'on estime σ .

L'heuristique initiale

Arrêter lorsque $\frac{V_n}{V_{2n}} \geq 0,95$.

Simulation sur des données normales (100 000 répétitions) :

n	Probabilité que $V_n/V_{2n} \geq 0,95$	Précision de la moyenne (IC 95 %)
10	49 %	$\pm 0,62 \sigma$
30	57 %	$\pm 0,36 \sigma$
100	67 %	$\pm 0,20 \sigma$
200	75 %	$\pm 0,14 \sigma$

Pourquoi elle ne convient pas

1. **Elle est remplie par hasard** une fois sur deux dès 10 patients, quand la moyenne n'est connue qu'à $\pm 0,6$ écart-type près.
2. **Elle demande d'avoir déjà inclus $2n$ patients** pour décider s'il fallait s'arrêter à n : elle ne fait rien économiser.
3. **Elle ne dépend pas de la question posée** : une variance stable ne dit pas si la moyenne est assez précise, ni si l'effet recherché est détectable.
4. **La variance se stabilise de toute façon** : sa fluctuation relative est d'environ $\sqrt{2/n}$, quelle que soit la précision voulue.

L'idée à garder

Utiliser l'estimation courante de la variance pour décider de l'arrêt : c'est exactement ce que font les règles séquentielles.

Fixer la précision voulue

Si l'on connaît σ

On fixe à l'avance une **précision** d : la demi-largeur souhaitée de l'intervalle de confiance à 95 % de la moyenne.

Effectif nécessaire

$$\bar{x} \pm z \frac{\sigma}{\sqrt{n}} \text{ avec } z \frac{\sigma}{\sqrt{n}} \leq d \iff n \geq n^* = \left(\frac{z \sigma}{d} \right)^2, \quad z = 1,96.$$

Précision d	$0,3 \sigma$	$0,2 \sigma$	$0,1 \sigma$
Effectif n^*	43	96	384

Le problème

σ est justement inconnu : c'est la valeur de convergence que l'on n'a pas.

L'idée : remplacer σ par son estimation courante

- À chaque nouveau patient, on recalcule l'écart-type s_n .
- On s'arrête dès que l'effectif atteint l'effectif nécessaire *estimé* :

$$n \geq \left(\frac{z s_n}{d} \right)^2 .$$

- Mais au début, s_n peut être très petit par hasard : on risquerait de s'arrêter trop tôt.

Deux garde-fous

1. un **effectif minimal** n_0 avant tout arrêt, par exemple 10 à 30 ;
2. un terme correctif $\frac{1}{n}$ ajouté à la variance, qui empêche un arrêt précoce quand s_n est accidentellement petit.

La règle de Chow et Robbins

La règle (Chow et Robbins, 1965)

Règle d'arrêt

On choisit une précision d et un effectif minimal n_0 . On s'arrête au premier $n \geq n_0$ tel que

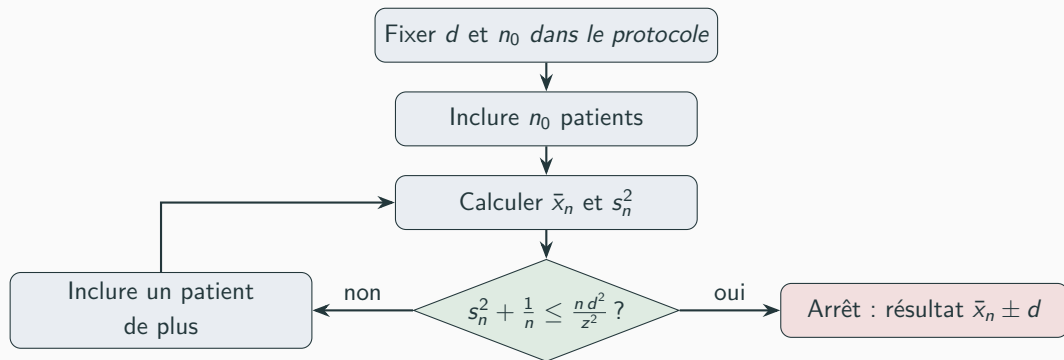
$$\boxed{s_n^2 + \frac{1}{n} \leq \frac{n d^2}{z^2}} \quad \text{où } s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2.$$

On donne alors l'intervalle $\bar{x}_n \pm d$.

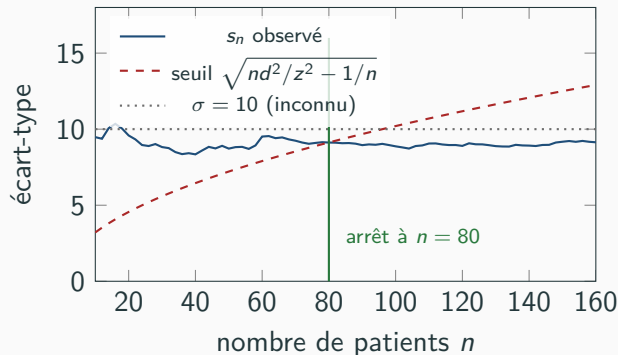
Garanties (quand $d \rightarrow 0$)

- l'effectif atteint vérifie $N/n^* \rightarrow 1$: on n'inclut **pas plus** de patients que si l'on avait connu σ ;
- l'intervalle $\bar{x}_N \pm d$ contient la vraie moyenne avec une probabilité qui tend vers 95 % ;
- **aucune hypothèse** sur la loi des données, hormis une variance finie.

L'algorithme pas à pas



Un exemple : un score de moyenne 50 et d'écart-type 10



- Précision voulue : $d = 2$ points.
- Si σ était connu : $n^* = 96$.
- La règle s'arrête à $n = 80$, quand la courbe bleue passe sous la courbe rouge.
- Résultat : $\bar{x} = 48,3 \pm 2$, qui contient bien la vraie moyenne 50.

Performance en simulation

Effectif moyen atteint et couverture réelle de l'intervalle $\bar{x} \pm d$ (cible : 95 %), avec $n_0 = 10$.

Loi des données	Précision d	n^* si σ connu	Effectif moyen	Couverture
Normale	$0,3 \sigma$	43	42	94 %
Normale	$0,2 \sigma$	96	96	95 %
Normale	$0,1 \sigma$	384	382	95 %
Poisson ($\lambda = 4$)	$0,2 \sigma$	96	95	94 %
Lognormale peu asymétrique	$0,2 \sigma$	96	93	93 %
Lognormale très asymétrique	$0,2 \sigma$	96	59	78 %

- La règle fonctionne très bien pour des données normales ou de Poisson.
- Pour des données **très asymétriques**, elle s'arrête trop tôt : les grandes valeurs, rares, n'ont pas encore été vues.

Données asymétriques : deux remèdes

1. Augmenter l'effectif minimal n_0

Lognormale très asymétrique, $d = 0,2 \sigma$:

n_0	10	30	50
Couverture	78 %	89 %	96 %

2. Travailler sur le logarithme des données

Pour des concentrations, des durées ou des coûts, on applique la règle à $\ln x$: les données deviennent à peu près normales, et l'on estime la moyenne géométrique avec une précision relative.

Une version exacte pour des données normales : Stein (1945)

Procédure en deux étapes

1. Inclure n_0 patients et calculer leur écart-type s_0 .
2. Calculer l'effectif total

$$N = \max \left(n_0, \left\lceil \frac{t_{n_0-1}^2 s_0^2}{d^2} \right\rceil \right),$$

où t_{n_0-1} est le quantile à 97,5 % de la loi de Student à $n_0 - 1$ degrés de liberté.

3. Inclure les $N - n_0$ patients restants, puis donner $\bar{x}_N \pm d$.

- Couverture **exactement garantie** d'au moins 95 % pour des données normales.
- Prix : un peu plus de patients (116 en moyenne contre 96 pour $d = 0,2 \sigma$, $n_0 = 15$).
- Avantage : l'effectif est fixé une fois pour toutes après la première étape.

Comparer deux traitements

Quand le but est un test

Pour comparer deux traitements, l'effectif nécessaire par groupe est

$$n = \frac{2 (z_{1-\alpha/2} + z_{1-\beta})^2 \sigma^2}{\Delta^2},$$

où Δ est la plus petite différence cliniquement importante. Il dépend lui aussi de σ inconnu.

Le recalcul d'effectif en cours d'étude (Wittes et Brittain, 1990)

1. Prévoir un effectif initial à partir d'une estimation prudente de σ .
2. Après une partie des patients, estimer σ **sans regarder la différence entre les groupes** (en aveugle).
3. Recalculer n avec cette estimation et poursuivre jusqu'à ce nouvel effectif.

On ne regarde pas l'effet du traitement : le risque de type I n'est pratiquement pas modifié.

Arrêter plus tôt : les plans séquentiels groupés

- On prévoit dans le protocole **quelques analyses intermédiaires**, par exemple à 25, 50 et 75 % des inclusions.
- À chacune, on peut arrêter :
 - pour **efficacité**, si la différence est déjà nette ;
 - pour **futilité**, si les chances de conclure sont devenues trop faibles.
- Les seuils sont plus exigeants aux premières analyses, pour que le risque global d'erreur de type I reste de 5 % : méthodes de Pocock, d'O'Brien et Fleming, de Lan et DeMets.

À ne pas faire

Regarder le résultat après chaque patient et s'arrêter dès que $p < 0,05$: le risque réel d'erreur de type I devient très supérieur à 5 %.

Prendre en compte le coût

Arrêter quand un patient de plus ne rapporte plus assez

Coût c par inclusion, coût $\lambda (\bar{x} - \mu)^2$ pour l'erreur d'estimation.

Coût total espéré après n patients

$$C(n) = c n + \lambda \frac{\sigma^2}{n}, \quad \text{minimal pour } n^* = \sigma \sqrt{\frac{\lambda}{c}}.$$

Règle séquentielle

Un patient de plus réduit l'erreur espérée de $\frac{\lambda \sigma^2}{n(n+1)}$. On s'arrête dès que ce gain, estimé avec s_n , devient inférieur au coût :

$$\frac{\lambda s_n^2}{n(n+1)} \leq c.$$

Exemple : $\sigma = 10$, $\lambda/c = 100$: $n^* = 100$. Un coût plus élevé fait s'arrêter plus tôt, en $\sqrt{1/c}$.

Synthèse

Quand stopper l'inclusion de cas ?

1. **Fixer dans le protocole** ce que l'on veut obtenir : une précision d sur la moyenne, ou la détection d'une différence Δ .
2. **Fixer un effectif minimal** n_0 : 10 à 20 pour des données normales, 30 à 50 pour des données asymétriques, ou travailler sur leur logarithme.
3. **Après chaque inclusion**, recalculer s_n et s'arrêter dès que $s_n^2 + \frac{1}{n} \leq \frac{n d^2}{z^2}$ (Chow et Robbins).
4. Pour une **comparaison** de traitements : recalcul d'effectif en aveugle, ou plan séquentiel groupé avec des seuils fixés d'avance.

Inutile de connaître la valeur de convergence σ : la règle l'estime au fur et à mesure.

- C. Stein, « A two-sample test for a linear hypothesis whose power is independent of the variance », *Annals of Mathematical Statistics* 16, 1945.
- A. Wald, « Sequential tests of statistical hypotheses », *Annals of Mathematical Statistics* 16, 1945.
- Y. S. Chow, H. Robbins, « On the asymptotic theory of fixed-width sequential confidence intervals for the mean », *Annals of Mathematical Statistics* 36, 1965.
- S. J. Pocock, « Group sequential methods in the design and analysis of clinical trials », *Biometrika* 64, 1977.
- P. C. O'Brien, T. R. Fleming, « A multiple testing procedure for clinical trials », *Biometrics* 35, 1979.
- K. K. G. Lan, D. L. DeMets, « Discrete sequential boundaries for clinical trials », *Biometrika* 70, 1983.
- J. Wittes, E. Brittain, « The role of internal pilot studies in increasing the efficiency of clinical trials », *Statistics in Medicine* 9, 1990.
- C. Jennison, B. W. Turnbull, *Group Sequential Methods with Applications to Clinical Trials*, Chapman & Hall, 2000.