

Designs adaptatifs avec réestimation de la taille de l'échantillon ou du nombre d'événements

MedianaDesigner package

1. Introduction

Ce document fournit une description de la méthodologie statistique utilisée dans le module de design adaptatif qui prend en charge la réestimation de la taille de l'échantillon ou du nombre d'événements (fonction ADSSMod).

Pour plus d'information sur le paquet MedianaDesigner, veuillez-vous référer aux pages internet suivantes :

<http://www.mediana.us/medianadesigner>

<http://medianasoft.github.io/MedianaDesignerFrench>

2. Designs adaptatifs avec réestimation de la taille de l'échantillon ou du nombre d'événements

2.1. Design de l'essai clinique

Considérons un essai de phase III avec deux bras (traitement expérimental versus contrôle). Le principal critère d'efficacité de cet essai peut être un critère d'évaluation continu, binaire ou de type temps jusqu'à un événement. Deux analyses intermédiaires seront mises en place pour effectuer des évaluations précoces de futilité et d'efficacité.

Les données d'analyse intermédiaire seront analysées en levant l'aveugle afin d'appuyer les règles de décision suivantes :

- Règle d'arrêt pour futilité lors de la première analyse intermédiaire : L'essai sera arrêté pour futilité si un effet significatif du traitement est peu susceptible d'être établi lors de l'analyse finale.
- Règle de réestimation de la taille de l'échantillon ou du nombre d'événements lors de la deuxième analyse intermédiaire : La taille totale de l'échantillon (pour des critères d'évaluation continus ou binaires) ou le nombre total d'événements (pour des critères d'évaluation de type temps jusqu'à un événement) sera augmenté si la probabilité de succès prédite est plus petite qu'attendue lors de cette analyse intermédiaire.

Notons que les règles d'arrêt pour futilité et de réestimation de la taille de l'échantillon ou du nombre d'événements sont non-contraignantes ("non-binding") et peuvent être ignorées par le sponsor de l'essai ou le comité de surveillance des données. De plus, l'essai ne sera pas terminé lors de l'une ou l'autre analyse intermédiaire en raison de son efficacité supérieure.

La règle d'arrêt pour futilité et la règle de réestimation de la taille de l'échantillon/du nombre d'événements sont définies dans les Sections 2.2 et 2.3, respectivement. La méthodologie du design adaptatif est présentée dans la Section 2.4. L'approche proposée pour construire des designs adaptatifs avec une réestimation de la taille de l'échantillon ou du nombre d'événements est illustrée dans la Section 3.

2.2. Règle d'arrêt pour futilité

La règle d'arrêt pour futilité à appliquer lors de la première analyse intermédiaire peut être mise en place à l'aide de toute définition pertinente de la probabilité de succès prédite, voir, par exemple, Wassmer et Brannath (2016, chapitre 7). Le plus souvent, la puissance conditionnelle est utilisée, celle-ci étant définie comme la probabilité d'établir un effet significatif du traitement pour le critère d'évaluation principal de l'efficacité lors de l'analyse finale, conditionnelle aux données de ce même critère disponibles lors de cette analyse intermédiaire.

Les notations suivantes seront utilisées pour calculer la puissance conditionnelle lors de la première analyse intermédiaire. En supposant que le critère d'évaluation principal de l'efficacité est continu ou binaire et en supposant un design d'essai équilibré, notons n_1 la taille totale de l'échantillon lors de la première analyse intermédiaire. De même, notons n_2 la taille totale de l'échantillon lors de l'analyse finale. La puissance conditionnelle est donnée par

$$CP = \Phi(aZ - bz_{1-\alpha}),$$

Où Z est la statistique de test intermédiaire,

$$a = \sqrt{\frac{n_2 - n_1}{n_1}} + \sqrt{\frac{n_1}{n_2 - n_1}}, \quad b = \sqrt{\frac{n_2}{n_2 - n_1}},$$

$z_{1-\alpha}$ est le $100(1 - \alpha)$ centile de la loi normale standard et $\Phi(x)$ est la fonction de distribution cumulative de la loi normale standard. Enfin, α indique le niveau de signification unilatéral dans l'essai, c'est-à-dire $\alpha = 0.025$. La formule de la puissance conditionnelle suppose que les données futures sont cohérentes avec les données intermédiaires. La puissance conditionnelle est dérivée de la même manière pour les essais avec critère d'évaluation de type temps jusqu'à un événement.

L'essai sera arrêté pour futilité lors de la première analyse intermédiaire si la puissance conditionnelle ne dépasse pas un seuil prédéfini désigné par c_1 , où $0 < c_1 < 1$. Ce seuil de futilité est généralement fixé à une valeur assez faible, par exemple, c_1 est généralement

inférieure à 0.3. En d'autres termes, un arrêt précoce est recommandé si la probabilité de succès prédite est trop faible.

Il est utile de noter que la règle d'arrêt pour futilité est facile à réécrire en termes de tailles d'effet intermédiaires si le critère d'évaluation principal de l'efficacité est continu ou binaire, ou de rapports de risque ("hazard ratio") intermédiaires si le critère d'évaluation principal de l'efficacité est un critère d'évaluation de type temps jusqu'à un événement. Par exemple, en supposant un critère d'évaluation continu ou binaire, la taille de l'effet à la première analyse intermédiaire est donnée par

$$Z/\sqrt{n_1/4}.$$

La taille de l'effet correspondant à la puissance conditionnelle de c_1 est égale à

$$\theta = \frac{\Phi^{-1}(c_1) + bz_{1-\alpha}}{a\sqrt{\frac{n_1}{4}}},$$

ce qui signifie que l'essai sera arrêté pour futilité si la taille de l'effet intermédiaire est inférieure à θ .

2.3. Règle de réestimation de la taille de l'échantillon/du nombre d'événements

Une règle de réestimation de la taille de l'échantillon ou du nombre d'événements sera appliquée lors de la deuxième analyse intermédiaire afin d'augmenter la probabilité de succès si la probabilité de succès prédite n'est pas suffisamment élevée. Cette règle sera également établie à l'aide de la puissance conditionnelle et celle-ci sera calculée lors de la seconde analyse intermédiaire de la même manière que dans la Section 2.1.

La règle de réestimation de la taille de l'échantillon/du nombre d'événements est construite à l'aide de deux seuils pour la puissance conditionnelle, notés par c_2 et c_3 . Ces seuils définissent « l'intervalle prometteur » ("promising interval") où il est le plus judicieux d'augmenter la taille de l'échantillon de l'essai ou le nombre cible d'événements. En supposant que l'essai n'a pas été interrompu pour futilité lors de la première analyse intermédiaire, le nombre total de patients ou d'événements peut être modifié lors de la deuxième analyse intermédiaire comme suit :

- Si $CP \leq c_2$, la taille de l'échantillon ou le nombre d'événements d'origine est conservé.
- Si $CP > c_2$ et $CP \leq c_3$, la taille de l'échantillon ou le nombre d'événements est augmenté pour atteindre le niveau souhaitable de puissance conditionnelle jusqu'à un plafond prédéfini.

- Si $CP > c_3$, la taille de l'échantillon ou le nombre d'événements d'origine est conservé.

Le niveau souhaitable de puissance conditionnelle est généralement fixé à la probabilité anticipée de succès dans l'essai, par exemple, si la puissance statistique de l'essai est de 90%, le niveau souhaitable de puissance conditionnelle sera fixé à 0.9. À l'aide des notations introduites dans la section 2.1, le nombre mis à jour de patients ou d'événements à l'analyse finale, noté par \tilde{n}_2 , est donné par

$$\tilde{n}_2 = n_1 + \frac{n_1}{Z^2} \left(z_{0.9} + z_{1-\alpha} \sqrt{\frac{n_2}{n_2 - n_1}} - Z \sqrt{\frac{n_1}{n_2 - n_1}} \right)^2$$

si le niveau souhaitable de puissance conditionnelle est de 0.9. Comme indiqué ci-dessus, la taille de l'échantillon ou le nombre d'événements mis à jour sera plafonné à une valeur prédéfinie. Ce plafond est souvent fixé à 20% ou 30%, c'est-à-dire que la taille de l'échantillon ou le nombre d'événements ne peut pas être augmenté de plus de 20% ou 30%. Cependant, le plafond peut parfois atteindre 50%.

2.4. Méthodologie du design adaptatif

Les inférences statistiques de l'analyse finale doivent être ajustées pour tenir compte de la réestimation de la taille totale de l'échantillon ou du nombre total d'événements basée sur les données lors de la deuxième analyse intermédiaire. Cet ajustement est nécessaire puisque les modifications de design basées sur les données sont connues pour augmenter le risque d'erreur de première espèce ; toutefois, si aucune modification n'est apportée lors de la deuxième analyse intermédiaire, l'analyse finale sera effectuée sans aucun ajustement.

Pour appliquer l'ajustement statistique, l'effet du traitement doit être évalué séparément en deux étapes. Par exemple, en supposant un critère d'évaluation continu ou binaire, les deux étapes sont définies comme suit :

- L'étape 1 comprend tous les patients qui terminent la période de traitement ou abandonnent l'essai avant d'avoir terminé la période de traitement jusqu'à la deuxième analyse intermédiaire.
- L'étape 2 comprend tous les patients qui terminent la période de traitement ou abandonnent l'essai avant d'avoir terminé la période de traitement après la deuxième analyse intermédiaire.

Les preuves de l'efficacité du traitement des deux étapes de l'essai seront combinées en utilisant le principe de la fonction combinée (Wassmer et Brannath, 2016, chapitre 6), comme expliqué ci-dessous. Notons p_1 et p_2 les valeurs-p unilatérales de l'effet du traitement calculées à partir des données de l'étape 1 et de l'étape 2, respectivement. Le test de significativité lors de l'analyse finale sera effectué en utilisant la valeur-p combinée, qui est définie comme suit :

$$p = c(p_1, p_2),$$

et un effet significatif du traitement sera établi lors de l'analyse finale si $p \leq \alpha$. Les valeurs-p par étapes sont combinées à l'aide de la fonction de combinaison inverse-normale pondérée, c'est-à-dire,

$$c(x, y) = 1 - \Phi \left(\sqrt{w}\Phi^{-1}(1 - x) + \sqrt{1 - w}\Phi^{-1}(1 - y) \right),$$

$\Phi(x)$ désigne la fonction de distribution cumulative de la loi normale standard, et w et $1 - w$ sont les poids prédéfinis attribués aux étapes 1 et 2. Le plus souvent, le poids de l'étape 1 (w) est égal à la fraction d'information prévue lors de l'analyse intermédiaire. Par exemple, si la deuxième analyse intermédiaire devrait être effectuée après que 60% des patients aient terminé la période de traitement ou abandonné l'essai avant la fin de la période de traitement, w sera fixé à 0.6.

3. Étude de cas

Les designs adaptatifs avec réestimation de la taille de l'échantillon seront illustrés à l'aide d'un essai de phase III chez des patients atteints de polyarthrite rhumatoïde. Les profils d'efficacité et d'innocuité d'une dose unique d'un traitement expérimental seront comparés à ceux d'un placebo. Le critère d'évaluation principal de l'efficacité dans cet essai est binaire ; il est basé sur la définition de l'amélioration de l'American College of Rheumatology. Il est attendu que le traitement expérimental améliore le taux de réponse.

Les patients sont randomisés de manière égale dans les deux groupes de l'essai et le nombre total de patients recrutés dans l'essai est de 240 patients. Le calcul de la taille de l'échantillon a supposé que les taux de réponse réels dans les groupes placebo et traitement sont égaux à 35% et 60%, respectivement. En utilisant $\alpha = 0.025$ unilatéral et en supposant un taux d'abandon de 15%, la taille de l'échantillon de 240 patients garantit au moins 90% de puissance ; toutefois, le sponsor s'inquiète du fait que le taux de réponse au placebo pourrait être plus élevé que prévu et qu'un échantillon plus grand serait nécessaire pour maintenir une puissance de 90%.

Pour améliorer la robustesse de cet essai, un design adaptatif avec deux analyses intermédiaires sera utilisé. La première analyse intermédiaire, visant à évaluer la futilité, sera effectuée après que 40% des patients aient terminé la période de traitement ou abandonné l'essai avant la fin de la période de traitement. Une option permettant d'augmenter le nombre de patients recrutés jusqu'à 30%, c'est-à-dire jusqu'à 312 patients, sera permise lors de la deuxième analyse intermédiaire. Cette analyse intermédiaire sera effectuée après que 60% des patients aient terminé la période de traitement ou abandonné l'essai.

Les règles d'arrêt pour futilité et de réestimation de la taille de l'échantillon seront mises en place à l'aide de la puissance conditionnelle. Le seuil de futilité (c_1) lors de la première analyse

intermédiaire a été sélectionné à l'aide d'une approche optimale définie dans le module de futilité (fonction FutRule). En bref, un seuil optimal a été obtenu en maximisant les taux de sensibilité et de spécificité de la règle. Le taux de sensibilité a été calculé en supposant que les taux réels de réponse au placebo et au traitement sont égaux à 35% et 55%, respectivement. Le seuil de futilité optimal qui en résulte est d'environ 0.3 et garantit que les taux de sensibilité et de spécificité de la règle d'arrêt de futilité sont d'au moins 80%.

La règle de réestimation de la taille de l'échantillon lors de la deuxième analyse intermédiaire a été construite à l'aide des seuils suivants pour la puissance conditionnelle :

- $c_2 = 0.4$.
- $c_3 = 0.9$.

Les seuils définissent l'intervalle prometteur à cette analyse intermédiaire.

La Figure 1 présente un résumé graphique de la règle de réestimation de la taille de l'échantillon. La taille totale de l'échantillon lors de l'analyse finale est représentée dans cette figure en fonction de la taille de l'effet lors de l'analyse intermédiaire. On peut constater dans cette figure que la taille totale de l'échantillon à l'analyse finale est fixée à 240 patients si la taille de l'effet intermédiaire est supérieure à 0.358 ou inférieure à 0.232, que la taille totale de l'échantillon augmente linéairement pour atteindre 312 patients si la taille de l'effet intermédiaire est comprise entre 0.318 et 0.358 et, enfin, que la taille totale de l'échantillon est de 312 patients si la taille de l'effet intermédiaire est comprise entre 0.232 et 0.318.

Les principales caractéristiques fonctionnelles du design adaptatif proposé dans deux scénarios alternatifs d'effet du traitement sont résumées dans les Tables 1 et 2. Dans ces scénarios, le taux de réponse au traitement est conforme aux hypothèses initiales, c'est-à-dire qu'il est fixé à 60%, mais un taux de réponse placebo plus élevé est supposé, à savoir :

- Scénario 1 : Le taux de réponse au placebo est de 37.5%.
- Scénario 2 : Le taux de réponse au placebo est de 40%.

Les tables fournissent également des informations sur un design de référence (design traditionnel) avec la même règle d'arrêt pour futilité lors de la première analyse intermédiaire et une taille d'échantillon fixe de 240 patients.

En commençant par la Table 1, la règle d'arrêt pour futilité est plus susceptible d'être déclenchée dans le Scénario 2, à savoir que la probabilité d'arrêt pour futilité passe de 11.8% dans le Scénario 1 à 17.5% dans le Scénario 2. Cela n'est pas surprenant puisque l'effet du traitement est plus faible dans le Scénario 2 et qu'il y a donc plus de chances que la taille de l'effet soit plus petite lors de la première analyse intermédiaire. La probabilité d'augmenter la taille de l'échantillon est également plus élevée dans le Scénario 2, ce qui signifie que les résultats intermédiaires sont plus susceptibles de se situer dans l'intervalle prometteur. La puissance du

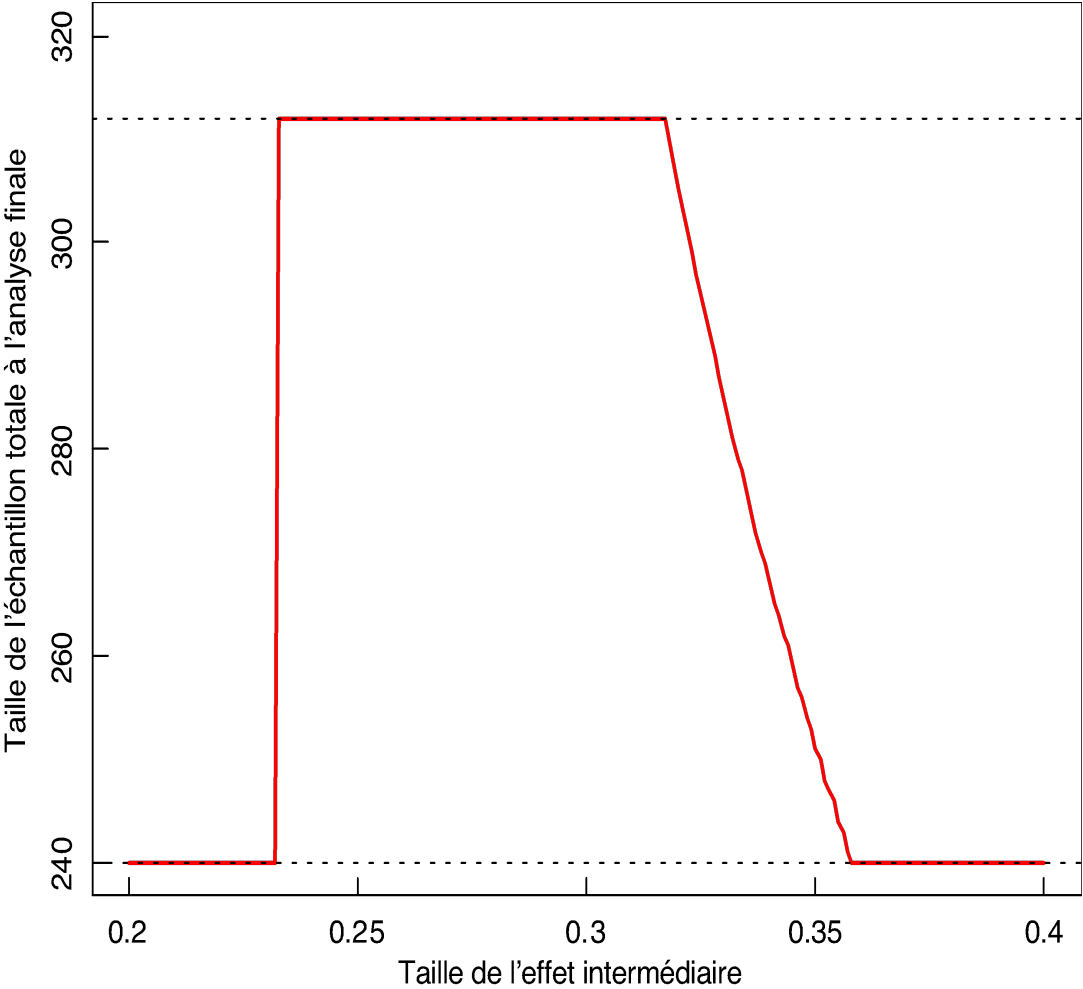
design des deux essais, traditionnel et adaptatif, est légèrement insuffisante dans le Scénario 1 (la puissance est égale à 83.6% et 84.9%, respectivement) et considérablement insuffisante dans le Scénario 2 (la puissance est égale à 74.0% et 76.0%). Ces résultats suggèrent qu'un échantillon de plus grande taille devrait être envisagé pour assurer une puissance adéquate dans chaque Scénario.

La Table 2 montre qu'au lieu d'adopter dès le départ un échantillon de plus grande taille, le sponsor de l'essai peut s'appuyer sur le design adaptatif pour augmenter le nombre total de patients dans l'essai lorsque les résultats semblent prometteurs lors de la deuxième analyse intermédiaire. Dans le Scénario 1, si les résultats intermédiaires se retrouvent dans l'intervalle prometteur, le design traditionnel avec 240 patients ne peut pas garantir une puissance de 90%. Avec ce design, la puissance lors de l'analyse finale sera réduite d'environ 2%. En utilisant une règle de réestimation de la taille de l'échantillon basée sur les données, le design adaptatif peut facilement améliorer la puissance, et dans les faits, la puissance à l'analyse finale devrait dépasser 95%. La même tendance est observée dans le Scénario 2 lorsque l'effet du traitement est encore plus faible. Lorsque le design traditionnel est appliqué, la puissance dans l'intervalle prometteur est considérablement réduite, jusqu'à 82.4%, tandis que le design adaptatif peut ici aussi augmenter la puissance et garantir une probabilité de succès supérieure à 90%.

Références

Wassmer, G., Brannath, W. (2016). Group Sequential and Confirmatory Adaptive Designs in Clinical Trials. New York: Springer.

Figure 1. Règle de réestimation de la taille de l'échantillon lors de la deuxième analyse intermédiaire



Les lignes pointillées sont tracées au niveau de la taille de l'échantillon d'origine de 240 patients et au niveau de la taille maximale de l'échantillon de 312 patients.

Table 1. Caractéristiques générales des designs traditionnels et adaptatifs

Scénario pour l'effet du traitement	Paramètre	Valeur
Scénario 1	Probabilité d'arrêt pour futilité lors de la première analyse intermédiaire	11.8%
	Probabilité d'augmenter la taille de l'échantillon lors de la deuxième analyse intermédiaire	17.6%
	Design traditionnel : Puissance	83.6%
	Design adaptatif : Puissance	84.9%
Scénario 2	Probabilité d'arrêt pour futilité lors de la première analyse intermédiaire	17.5%
	Probabilité d'augmenter la taille de l'échantillon lors de la deuxième analyse intermédiaire	20.8%
	Design traditionnel : Puissance	74.0%
	Design adaptatif : Puissance	76.0%

Scénario 1 : Le taux de réponse au placebo est de 37.5% et le taux de réponse au traitement est de 60%. Scénario 2 : Le taux de réponse au placebo est de 40% et le taux de réponse au traitement est de 60%.

Table 2. Puissance des designs traditionnels et adaptatifs dans l'intervalle prometteur

Scénario pour l'effet du traitement	Design	Puissance
Scénario 1	Design traditionnel	88.2%
	Design adaptatif	95.5%
Scénario 2	Design traditionnel	82.4%
	Design adaptatif	91.8%

Scénario 1 : Le taux de réponse au placebo est de 37.5% et le taux de réponse au traitement est de 60%. Scénario 2 : Le taux de réponse au placebo est de 40% et le taux de réponse au traitement est de 60%.