

Designs adaptatifs avec sélection de traitement basée sur les données

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 sélection de traitement basée sur les données (fonction ADTreatSel).

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 sélection de traitement basée sur les données

2.1. Design de l'essai

Considérons un essai de phase III à plusieurs bras qui sera mené pour étudier l'efficacité de plusieurs doses ou schémas posologiques d'un traitement expérimental (les doses ou les schémas posologiques seront appelés traitements) par rapport à un groupe de contrôle, par exemple un placebo. Le principal critère d'efficacité de cet essai peut être un critère d'évaluation continu, binaire ou de type temps jusqu'à un évènement. Deux analyses intermédiaires seront menées pour effectuer une évaluation précoce de la futilité et identifier le traitement le plus prometteur, respectivement.

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 : Une évaluation de la futilité sera effectuée pour chaque traitement et un traitement sera abandonné à cette analyse intermédiaire s'il est peu probable qu'il soit efficace.
- Règle de sélection du traitement lors de la deuxième analyse intermédiaire : Un ou plusieurs traitements seront sélectionnés et les traitements restants seront abandonnés lors de cette analyse intermédiaire.

Les règles de décision (arrêt pour futilité et règles de sélection du traitement) sont non-contraignantes ("non-binding") et peuvent être ignorées par le sponsor de l'essai ou le comité de surveillance des données. Il est supposé dans ce module que l'essai ne sera pas arrêté à l'une ou l'autre analyse intermédiaire en raison de son efficacité supérieure ; cependant, le design adaptatif proposé est facilement étendu pour mettre en place une option permettant d'évaluer le profil d'efficacité de chaque traitement et de mettre fin à l'essai en raison de l'efficacité supérieure si un signal d'efficacité très fort est détecté dans au moins un groupe de traitement. Une autre extension potentielle repose sur une règle de réestimation de la taille de l'échantillon ou du nombre d'événements introduite 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). Une fois que le meilleur traitement a été choisi lors de la deuxième analyse intermédiaire, le nombre cible de patients ou d'événements pourrait être ainsi augmenté pour améliorer la probabilité de succès sur la base d'une règle de réestimation prédéfinie.

Les règles d'arrêt pour futilité et de sélection du traitement sont définies dans les Sections 2.2 et 2.3, respectivement, et la méthodologie du design adaptatif est présentée dans la Section 2.4. L'approche proposée pour définir des designs adaptatifs avec sélection de traitement basée sur les données 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 sera établie à l'aide de la puissance conditionnelle, voir, par exemple, Wassmer et Brannath (2016, chapitre 7). Pour chaque traitement individuel, il s'agit de la probabilité d'une amélioration significative par rapport au critère d'évaluation principal de l'efficacité lors de l'analyse finale, conditionnelle aux données intermédiaires dans le groupe de traitement et le groupe de contrôle sélectionnés.

Notons m le nombre total de traitements dans l'essai et considérons la comparaison du k ème traitement par rapport au contrôle, $k = 1, \dots, m$. La puissance conditionnelle pour cette comparaison est notée par CP_k . La dérivation de la puissance conditionnelle est fournie dans la documentation de la fonction ADSSMod et sera omise ici. Le k ème bras traitement sera abandonné lors de la première analyse intermédiaire pour futilité si la puissance conditionnelle spécifique à ce traitement ne dépasse pas un seuil prédéfini désigné par c , où $0 < c < 1$, c'est-à-dire

$$CP_k \leq c.$$

Le seuil de futilité est généralement fixé à une valeur assez faible, par exemple, il dépasse rarement 0.3. L'essai sera interrompu pour futilité à cette analyse intermédiaire si tous les traitements sont abandonnés.

2.3. Règle de sélection du traitement

Une règle de sélection du traitement sera appliquée lors de la deuxième analyse intermédiaire afin d'identifier le nombre prédéfini de traitements les plus performants. Les meilleurs traitements correspondent aux tailles d'effet les plus importantes en fonction du critère d'évaluation principal de l'efficacité (la taille de l'effet est définie comme la valeur négative du log-rapport de risque ("hazard ratio") si le critère d'évaluation principal de l'efficacité est de type temps jusqu'à un évènement).

2.4. Méthodologie du design adaptatif

Comme dans les autres modules de design adaptatif (fonctions ADSSMod et ADPopSel), une règle de décision basée sur les données sera appliquée lors de la deuxième analyse intermédiaire, ce qui a pour conséquence d'augmenter le risque d'erreur de première espèce. Pour résoudre ce problème, la comparaison du bras de traitement sélectionné par rapport au bras contrôle lors de l'analyse finale doit être ajustée comme décrit ci-dessous.

Pour définir cet ajustement, considérons d'abord le cas où aucune règle de sélection de traitement n'est appliquée lors de la deuxième analyse intermédiaire et tous les traitements qui n'ont pas été abandonnés lors de la première analyse intermédiaire seront conservés. Soit p_k la valeur-p unilatérale de l'effet du traitement correspondant à la comparaison du k ème traitement par rapport au contrôle à l'analyse finale, $k = 1, \dots, m$. Étant donné qu'il existe plusieurs possibilités de revendiquer un résultat positif de cet essai, un ajustement de la multiplicité (une procédure de tests multiples) doit être appliqué. Les options disponibles incluent les procédures de Bonferroni, Holm et Hochberg. Pour plus d'informations sur les ajustements de multiplicité couramment utilisés, voir par exemple Dmitrienko et D'Agostino (2018).

De retour au design adaptatif avec règle de sélection de traitement basée sur les données, notons $1 \leq s \leq m$ le nombre prédéfini de traitements choisis lors de la deuxième analyse intermédiaire. Il est important de noter que, même si un seul traitement est choisi pour être comparé au contrôle à l'analyse finale, il est toujours nécessaire d'ajuster la multiplicité. Cet ajustement représente une pénalité pour le biais de sélection en raison de la possibilité d'identifier le meilleur traitement lors de la deuxième analyse intermédiaire.

Les valeurs-p de l'effet du traitement pour les traitements sélectionnés seront calculées de manière standard lors de l'analyse finale et les valeurs-p pour les traitements restants seront fixées à 1. Pour protéger le risque global de première espèce dans l'essai, une procédure de tests multiples définie prospectivement sera appliquée aux valeurs-p résultantes. Par exemple, si un seul traitement est choisi pour l'évaluation finale, toutes les valeurs de p sauf une sont égales à 1 et l'ajustement de multiplicité qui en résulte prend généralement une forme très simple. Avec les procédures de Bonferroni, Holm ou Hochberg, un effet de traitement statistiquement significatif sera établi lors de l'analyse finale si la valeur-p pour le traitement sélectionné ne dépasse pas α/m , où α indique le niveau de significativité unilatéral dans l'essai ($\alpha = 0.025$).

3. Étude de cas

L'utilisation de règles de sélection de traitement basées sur des données dans les essais confirmatoires sera illustrée à l'aide d'un essai de phase III pour le traitement de la schizophrénie. Trois doses d'un traitement expérimental seront comparées à un placebo dans cet essai. Les doses sont marquées Dose L (faible dose), Dose M (dose moyenne) et Dose H (dose élevée). La durée de la période de traitement est de 6 semaines et l'analyse d'efficacité primaire est basée sur la variation de la Baseline à la semaine 6 pour le score total PANSS (Positive and Negative Syndrome Scale). Une réduction plus importante du score total PANSS indique un effet bénéfique.

Les données recueillies dans l'essai de phase II ne sont pas très fiables en raison d'une grande variabilité et le sponsor de l'essai aimerait tirer parti d'un design adaptatif pour effectuer une évaluation précoce de la futilité et réduire les doses non-performantes. Une autre analyse intermédiaire sera effectuée pour sélectionner la dose la plus prometteuse. Cette dose sera incluse dans l'analyse finale tandis que les autres doses seront abandonnées.

Le nombre de patients recrutés par bras a été fixé à 180. Un schéma de randomisation égal et un taux élevé d'abandon des patients de 25% ont été supposé (les patients sont susceptibles d'être perdus de vue dans les essais sur la schizophrénie).

Le design adaptatif est mis en place comme suit :

- Une règle d'arrêt pour futilité basée sur la puissance conditionnelle sera appliquée lors de la première analyse intermédiaire, qui aura lieu à 30% de fraction d'information. Cela signifie que cette analyse intermédiaire sera effectuée après que 216 patients aient terminé la période de traitement ou abandonné l'essai avant la fin de la période de traitement. Le seuil de futilité (c) est fixé à 20%. Cette valeur est inférieure à un seuil optimal obtenu à l'aide de l'approche implémentée dans le module de futilité (fonction FutRule). Cette valeur plus faible a été choisie car elle entraîne un taux de sensibilité très élevé (probabilité de retenir correctement au moins un bras de traitement lors de cette analyse intermédiaire) en supposant une taille d'effet commune de 0.4 dans les trois bras de traitement. Dans cette hypothèse alternative d'effet bénéfique, le taux de sensibilité dépasse 90%.
- La deuxième analyse intermédiaire sera effectuée à 50% de la taille totale de l'échantillon, c'est-à-dire après que 360 patients aient terminé la période de traitement ou abandonné l'essai avant la fin de la période de traitement. La dose la plus prometteuse avec la plus grande taille d'effet sera choisie lors de cette analyse intermédiaire.
- L'analyse finale sera effectuée à l'aide d'un ajustement de multiplicité basé sur le test de Hochberg, ce qui signifie qu'un effet de traitement significatif sera établi si la valeur- p pour la dose choisie n'est pas supérieure à $\alpha/3$.

Il est utile de noter que la taille globale de l'échantillon dans le design adaptatif est bien inférieure au nombre de patients recrutés (180 patients) multiplié par le nombre de bras d'essai (4 bras). Une caractéristique importante du design adaptatif est que seuls deux bras sont retenus dans l'essai après la deuxième analyse intermédiaire. Si le recrutement des patients est suffisamment lent, la taille totale de l'échantillon pourrait être réduite de 25%, c'est-à-dire que seulement 540 patients seront recrutés dans l'essai.

Les scénarios d'effet de traitement suivants ont été pris en compte pour évaluer les caractéristiques fonctionnelles du design adaptatif :

- Scénario 1 : La taille de l'effet commune est de 0.4.
- Scénario 2 : La taille de l'effet pour les Doses L et M est de 0.3 et la taille de l'effet pour la Dose H est de 0.4.

Des résumés des caractéristiques fonctionnelles les plus importantes du design adaptatif sont présentés dans les Tables 1, 2 et 3. Les performances du design adaptatif sont comparées à celles de trois designs de référence. Ces designs sont des designs traditionnels à deux bras avec 180 patients recrutés par bras. Chaque design compare une dose unique du traitement expérimental (Dose L, Dose M ou Dose H) au placebo et applique la même règle d'arrêt pour futilité que dans le design adaptatif, mais n'utilise pas de sélection du traitement.

La Table 1 présente les probabilités d'abandon de chaque bras de traitement pour futilité lors de la première analyse intermédiaire dans le design adaptatif ou dans l'un des trois designs traditionnels à deux bras. La probabilité d'abandonner un bras de traitement est assez faible (environ 13%) dans le Scénario 1, ce qui est prévisible puisqu'un fort effet de traitement avec une taille réelle de l'effet de 0.4 est supposé. Dans le Scénario 2, l'effet du traitement est plus faible aux Doses L et M et, par conséquence directe, ces deux doses sont plus susceptibles d'être éliminées à l'analyse intermédiaire. La probabilité d'abandonner chacune de ces doses est d'environ 26%.

La Table 2 se concentre sur le design adaptatif et résume les caractéristiques de la règle de sélection du traitement lors de la deuxième analyse intermédiaire. Les probabilités de choisir chaque dose comme étant le traitement le plus performant sont présentées. Étant donné que les trois doses sont supposées être également efficaces dans le Scénario 1, chaque dose a autant de chances d'être sélectionnée et les trois probabilités indiquées dans la Table 2 sont essentiellement égales à 33.3%. Quand des tailles d'effet différentes sont considérées (Scénario 2), les Doses L et M sont nettement moins susceptibles d'être choisies lors de la deuxième analyse intermédiaire. La dose la plus prometteuse (Dose H) sera identifiée comme la meilleure dose à ce stade plus de 50% du temps.

Enfin, les résultats des calculs de puissance sont présentés dans la Table 3. Avec les designs à deux bras, les valeurs de puissance correspondent aux probabilités d'établir un effet significatif à

chaque dose par rapport au placebo sans ajustement de multiplicité. La Table 3 montre que, dans le Scénario 1, chaque design traditionnel a une puissance adéquate avec des valeurs supérieures à 80%. Le design adaptatif offre une amélioration considérable par rapport à chaque design à deux bras et la puissance pour la comparaison de la meilleure dose par rapport au placebo est de 90%. Dans le Scénario 2, le design adaptatif fonctionne beaucoup mieux que les designs à deux bras correspondant aux Doses L et M. Contrairement à ces deux designs, le design adaptatif a une option de basculer sur la Dose H qui a une plus grande taille d'effet, ce qui augmente les chances de succès. Lorsque le design à deux bras correspondant à la Dose H est comparé au design adaptatif, le premier offre un léger avantage de puissance par rapport au second (81.2% contre 79.1%). Cela est dû au fait que la règle de sélection du traitement utilisée dans le design adaptatif n'est pas parfaite et que les Doses L et M sont occasionnellement sélectionnées pour l'analyse finale, ce qui entraîne une perte de puissance par rapport à ce design particulier à deux bras.

Références

Dmitrienko, A., D'Agostino, R.B. (2018). Multiplicity considerations in clinical trials. *New England Journal of Medicine*. 378, 2115-2122.

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

Table 1. Arrêt pour futilité dans les designs traditionnels et adaptatifs

Scénario pour l'effet du traitement	Paramètre	Valeur
Scénario 1	Probabilité d'abandonner la Dose L pour futilité	13.4%
	Probabilité d'abandonner la Dose M pour futilité	13.6%
	Probabilité d'abandonner la Dose H pour futilité	13.1%
Scénario 2	Probabilité d'abandonner la Dose L pour futilité	26.0%
	Probabilité d'abandonner la Dose M pour futilité	25.6%
	Probabilité d'abandonner la Dose H pour futilité	13.6%

Scénario 1 : La taille de l'effet commune est de 0.4. Scénario 2 : La taille de l'effet pour les Doses L et M est de 0.3 et la taille de l'effet pour la Dose H est de 0.4.

Table 2. Sélection du traitement dans le design adaptatif

Scénario pour l'effet du traitement	Paramètre	Valeur
Scénario 1	Probabilité de choisir la Dose L comme traitement le plus performant	33.6%
	Probabilité de choisir la Dose M comme traitement le plus performant	32.9%
	Probabilité de choisir la Dose H comme traitement le plus performant	32.8%
Scénario 2	Probabilité de choisir la Dose L comme traitement le plus performant	19.6%
	Probabilité de choisir la Dose M comme traitement le plus performant	20.4%
	Probabilité de choisir la Dose H comme traitement le plus performant	54.7%

Scénario 1 : La taille de l'effet commune est de 0.4. Scénario 2 : La taille de l'effet pour les Doses L et M est de 0.3 et la taille de l'effet pour la Dose H est de 0.4.

Table 3. Calculs de puissance dans les designs traditionnels et adaptatif

Scénario pour l'effet du traitement	Paramètre	Valeur
Scénario 1	Puissance pour la Dose L (design traditionnel)	81.5%
	Puissance pour la Dose M (design traditionnel)	81.2%
	Puissance pour la Dose H (design traditionnel)	81.3%
	Puissance pour la meilleure dose (design adaptatif)	90.0%
Scénario 2	Puissance pour la Dose L (design traditionnel)	58.1%
	Puissance pour la Dose M (design traditionnel)	58.2%
	Puissance pour la Dose H (design traditionnel)	81.2%
	Puissance pour la meilleure dose (design adaptatif)	79.1%

Scénario 1 : La taille de l'effet commune est de 0.4. Scénario 2 : La taille de l'effet pour les Doses L et M est de 0.3 et la taille de l'effet pour la Dose H est de 0.4.