

Designs adaptatifs avec sélection de population basée sur les données

MedianaDesigner package

1. Introduction

Ce document fournit une description de la méthodologie statistique utilisée dans le module de conception adaptative qui prend en charge la sélection de population basée sur les données (fonction ADPopSel).

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 population basée sur les données

2.1. Design de l'essai clinique

Les designs adaptatifs avec sélection de population basée sur les données sont utilisés dans les essais avec deux populations de patients ou plus (ces essais sont connus sous le nom d'essais multi-populations). Considérons un essai de phase III à deux bras pour évaluer le profil d'efficacité d'un traitement expérimental 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. L'évaluation de l'efficacité sera réalisée dans deux populations, à savoir la population globale de patients présentant la maladie d'intérêt ainsi qu'un sous-ensemble de patients spécifié prospectivement. Le sous-ensemble est défini à l'aide d'un classificateur binaire basé sur une caractéristique de Baseline du patient. Étant donné que les biomarqueurs sont le plus souvent utilisés à cette fin, les patients du sous-ensemble sélectionné seront appelés patients biomarqueur-positifs et les patients du sous-ensemble complémentaire seront appelés patients biomarqueur-négatifs. Deux analyses intermédiaires seront utilisées dans l'essai. La première analyse appuiera une évaluation de la futilité et la deuxième analyse se concentrera sur l'identification de la ou des meilleures populations pour l'analyse finale.

Les analyses intermédiaires effectuées en levant l'aveugle mettront en place les règles de décision suivantes :

- Règle d'arrêt pour futilité lors de la première analyse intermédiaire : Une évaluation de la futilité sera effectuée pour évaluer le profil d'efficacité dans la population globale. L'essai sera terminé pour futilité si l'effet du traitement dans la population globale est peu susceptible d'être significatif lors de l'analyse finale.
- Règle de sélection de la population lors de la deuxième analyse intermédiaire : La ou les populations les plus prometteuses seront sélectionnées pour évaluer l'effet du traitement lors de l'analyse finale.

Il est important de noter que les règles de décision (arrêt pour futilité et sélection de la population) sont non-contraignantes ("non-binding") et peuvent être ignorées par le sponsor de l'essai ou le comité de surveillance des données. Comme dans le module de design adaptatif avec sélection de traitement (fonction ADTreatSel), le design adaptatif considéré dans ce module peut être étendue en incorporant une règle d'arrêt pour efficacité à la première ou à la deuxième analyse intermédiaire (pour arrêter l'essai plus tôt en raison de preuves solides d'efficacité dans les populations globales ou biomarqueur-positifs) ou une règle de réestimation de la taille de l'échantillon/du nombre d'événements (pour augmenter le nombre cible de patients ou d'événements si la probabilité prédite de succès dans une ou l'autre des populations de patients est plus faible que prévu).

Les règles d'arrêt pour futilité et de sélection de la population sont définies dans les Sections 2.2 et 2.3, respectivement, et la méthodologie du design adaptatif est présenté dans la Section 2.4. Les designs adaptatifs avec sélection de la population basée sur les données sont illustrés dans la Section 3.

2.2. Règle d'arrêt pour futilité

Comme dans les autres modules de design adaptatif (fonctions ADSSMod et ADTreatSel), la règle d'arrêt pour futilité lors de la première analyse intermédiaire sera mise en place à l'aide de la puissance conditionnelle, voir, par exemple, Wassmer et Brannath (2016, chapitre 7). La puissance conditionnelle sera calculée dans la population globale de patients comme la probabilité d'un effet global significatif par rapport au critère d'évaluation principal de l'efficacité lors de l'analyse finale conditionnelle aux données intermédiaires dans la population globale.

Notons CP la puissance conditionnelle dans la population globale évaluée lors de la première analyse intermédiaire (la dérivation de la puissance conditionnelle est fournie dans la documentation de la fonction ADSSMod et sera omise ici). L'essai sera terminé pour futilité lors de cette analyse intermédiaire si la puissance conditionnelle ne dépasse pas un seuil prédéfini désigné par c , où $0 < c < 1$, c'est-à-dire

$$CP \leq c.$$

Comme indiqué dans les autres modules de design adaptatif, le seuil de futilité est généralement fixé à une valeur assez faible, par exemple, ce seuil dépasse rarement 0.3.

2.3. Règle de sélection de la population

La règle de sélection de la population à appliquer lors de la deuxième analyse intermédiaire visera à choisir les populations les plus pertinentes pour l'analyse finale. Cette règle de décision sera définie à l'aide des conditions d'influence et d'interaction (voir, par exemple, Millen et al., 2012). Ces conditions ont été introduites pour faciliter le processus de formulation d'allégations réglementaires significatives dans les essais multi-populations.

Pour définir les règles de décision en fonction des conditions d'influence et d'interaction, notons $\hat{\theta}_+$ et $\hat{\theta}_-$ les estimations intermédiaires de la taille de l'effet dans les populations de patients biomarqueur-positifs et biomarqueur-négatifs, respectivement. Comme dans les autres modules de design adaptatif, la taille de l'effet pour les critères d'évaluation de type temps jusqu'à un événement est définie comme la valeur négative du log-rapport de risque ("hazard ratio").

Les conditions d'influence et d'interaction sont appliquées séquentiellement. La condition d'influence stipule que l'analyse finale sera effectuée dans la population globale (et potentiellement dans la population de biomarqueur-positifs) s'il existe des preuves d'un effet significatif du traitement au sein de la population de patients biomarqueur-négatifs, ce qui signifie qu'un effet de traitement bénéfique n'est pas limité au sous-ensemble de patients biomarqueur-positifs. La condition d'influence est satisfaite si l'ampleur de l'effet intermédiaire dans la population biomarqueur-négatif est supérieure à un seuil cliniquement pertinent, c.-à-d.

$$\hat{\theta}_- \geq c_{inf},$$

où $c_{inf} \geq 0$ est connu sous le nom de seuil d'influence. Si la condition d'influence n'est pas remplie, l'effet global est susceptible d'être entraîné par un fort effet de traitement dans le sous-ensemble biomarqueur-positif et donc l'analyse finale sera limitée à la population biomarqueur-positif. Si le recrutement des patients n'est pas terminé au moment où la deuxième analyse intermédiaire est effectuée, seuls les patients biomarqueur-positifs seront recrutés après cette analyse intermédiaire.

Deuxièmement, si la condition d'influence est remplie, la condition d'interaction est appliquée pour déterminer si l'analyse finale sera effectuée dans les deux populations ou seulement dans la population globale. Cette condition stipule que l'évaluation finale inclura les deux populations de patients si un effet de traitement différentiel est présent, c'est-à-dire que le bénéfice du traitement dans la population de patients biomarqueur-positifs est beaucoup plus fort que dans la population de patients biomarqueur-négatifs. Mathématiquement, cela signifie que

$$\hat{\theta}_+/\hat{\theta}_- \geq c_{int},$$

où $c_{int} > 1$ est connu sous le nom de seuil d'interaction. Si cette condition n'est pas remplie, l'effet du traitement est susceptible d'être homogène entre les populations de patients biomarqueur-positifs et biomarqueur-négatifs et seul l'effet global sera examiné lors de l'analyse finale. Pour plus d'informations sur la sélection des seuils d'influence et d'interaction, veuillez-vous référer à Dmitrienko et Paux (2017).

2.4. Méthodologie du design adaptatif

Comme dans les autres modules de design adaptatif, un ajustement approprié doit être appliqué à l'analyse finale pour tenir compte de la règle de décision basée sur les données (règle de sélection de population) lors de la deuxième analyse intermédiaire. Cet ajustement garantit que le risque d'erreur de première espèce dans l'essai est conservé au niveau nominal, c'est-à-dire $\alpha = 0.025$ unilatéralement.

Pour définir cet ajustement, supposons d'abord que les conditions d'influence et d'interaction introduites dans la Section 2.2 sont toutes deux remplies lors de la deuxième analyse intermédiaire, ce qui signifie que l'effet du traitement sera évalué lors de l'analyse finale dans la population globale et dans le sous-ensemble des patients biomarqueur-positifs. Notons p_0 and p_+

les valeurs-p unilatérales de l'effet traitement dans la population globale et celle des patients biomarqueur-positifs, respectivement. Étant donné qu'il y a deux occasions de conclure que le résultat de l'essai est positif, un ajustement de la multiplicité sera utilisé pour contrôler le risque d'erreur de première espèce. Par exemple, si un ajustement de la multiplicité basé sur le test de Hochberg est choisi, un effet significatif sera établi dans la population globale si

$$\min(2\min(p_0, p_+), \max(p_0, p_+)) \leq \alpha \text{ et } p_0 \leq \alpha.$$

De même, un effet significatif sera établi dans la population de patients biomarqueur-positifs si

$$\min(2\min(p_0, p_+), \max(p_0, p_+)) \leq \alpha \text{ et } p_+ \leq \alpha.$$

Si les données lors de la seconde analyse intermédiaire appuient la décision de sélectionner une seule population pour l'analyse finale, une seule des deux valeurs-p de l'effet du traitement sera définie. En particulier, si la condition d'influence n'est pas remplie, l'analyse finale se concentrera sur la population de patients biomarqueur-positifs, ce qui signifie que p_0 ne sera pas définie. Cette valeur-p sera fixée à 1 et donc l'effet du traitement dans la population de patients biomarqueur-positifs sera significatif si

$$p_+ \leq \alpha/2.$$

Aussi, si la condition d'influence est satisfaite, mais que la condition d'interaction n'est pas satisfaite lors de la deuxième analyse intermédiaire, l'évaluation finale sera effectuée dans la

population globale. Cela implique que p_+ sera fixée à 1 et qu'un effet significatif sera conclu dans la population globale si

$$p_0 \leq \alpha/2.$$

3. Étude de cas

Pour illustrer le processus de conception d'essais cliniques de phase III avec des règles de sélection de population, un essai de phase III multi-population en oncologie est utilisé. Cet essai est mené pour étudier l'efficacité d'un traitement expérimental par rapport au contrôle, par exemple en évaluant les meilleurs soins d'accompagnement. Le profil d'efficacité de ce traitement sera examiné dans la population globale de patients ainsi que dans un sous-ensemble prédéfini de patients ayant un statut de biomarqueur positif. Un effet thérapeutique accru est prévu au sein de la population biomarqueur-positif par rapport au sous-ensemble complémentaire de patients biomarqueur-négatifs. L'analyse principale d'efficacité est formulée en termes de survie globale.

Le sponsor de l'essai souhaite évaluer les avantages potentiels d'un design adaptatif à l'aide de deux analyses intermédiaires qui appuient respectivement l'évaluation de la futilité et la sélection de population, et le comparer à un design traditionnel axé sur l'évaluation de l'efficacité dans l'ensemble de la population.

Les designs adaptatifs et traditionnels seront guidés par les événements, c'est-à-dire, les analyses intermédiaires et finales seront effectuées après un nombre prédéfini d'événements. Le design adaptatif utilise les règles de décision suivantes :

- Une règle d'arrêt pour futilité basée sur la puissance conditionnelle sera appliquée lors de la première analyse intermédiaire. Cette analyse sera effectuée après avoir accumulé 40% du nombre total d'événements dans la population globale. Le seuil de futilité (c) est fixé à 20%. Cette valeur a été dérivée en utilisant l'approche implémentée dans le module de futilité (fonction FutRule). Une valeur optimale pour ce seuil dans la population globale a été choisie en supposant des temps de survie médians de 7.5 et 11 mois dans les bras contrôle et de traitement. Avec ce seuil de futilité, le taux de spécificité est proche de 80% et le taux de sensibilité approche les 90%.
- Une règle de sélection de population est appliquée lors de la deuxième analyse intermédiaire, qui est effectuée à 60% du nombre total d'événements dans la population globale. Cette règle est basée sur les conditions d'influence et d'interaction avec les seuils suivants :
 - Seuil d'influence $c_{inf} = 0$.
 - Seuil d'interaction $c_{int} = 1.3$.

La condition d'influence qui en résulte est assez libérale étant donné que la décision de procéder à l'analyse de la population globale sera prise à moins qu'un effet négatif du traitement ne soit observé dans la population des patients biomarqueur-négatifs lors de la deuxième analyse intermédiaire. La condition d'interaction basée sur le seuil sélectionné conclura qu'un effet de traitement différentiel est présent et recommandera donc que l'effet du traitement soit évalué dans les deux populations lors de l'analyse finale si l'ampleur de l'effet dans le sous-ensemble biomarqueur-positif est supérieure de plus de 30% à celle du sous-ensemble complémentaire.

- L'ajustement de la multiplicité à l'analyse finale sera basé sur le test de Hochberg.

Le design traditionnel utilise une seule analyse intermédiaire, qui repose sur la même règle d'arrêt de futilité que dans le design adaptatif. Comme indiqué ci-dessus, dans le design traditionnel, seul l'effet global sera évalué lors de l'analyse finale.

Un calcul initial de la puissance a été effectué dans cet essai en supposant que les temps de survie médians dans les bras contrôle et traitement sont égaux à 7.5 et 11 mois, respectivement. Ces temps de survie médians correspondent au rapport de risque ("hazard ratio") de 0.68. En supposant une puissance de 90% et aucune évaluation intermédiaire de la futilité, l'analyse finale devrait être effectuée après 290 événements. La taille de l'échantillon dans l'essai a été choisie en utilisant un ratio de randomisation de 1:2, à savoir, il est supposé que 140 patients seront affectés au groupe contrôle et 280 patients seront affectés au bras de traitement.

Pour évaluer la performance du design adaptatif proposé, le nombre total d'événements dans la population globale a été fixé à 290, ce qui signifie que les première et deuxième analyses intermédiaires seront effectués après que 114 et 174 événements auront été accumulés. Le nombre cible d'événements au sein de la population de patients biomarqueur-positifs a été fixé à 190. Ce nombre d'événements est requis dans le design adaptatif si la population de biomarqueur-positifs est la seule population choisie pour l'analyse finale. Ce nombre cible d'événements se traduit par une puissance de 90% selon une hypothèse optimiste selon laquelle les temps de survie médians sont égaux à 7.5 et 12 mois dans les groupes contrôle et traitement, c'est-à-dire que le rapport de risque est de 0.63. La prévalence de la population des patients positifs au biomarqueur a été supposée être de 50% et le taux d'abandon annuel a été fixé à 5%.

Les caractéristiques fonctionnelles des designs traditionnels et adaptatifs ont été calculées selon trois scénarios d'effet de traitement. Dans tous les scénarios, un temps de survie médian commun de 7.5 mois a été envisagé dans le groupe contrôle, quel que soit le statut du biomarqueur. Les temps de survie médians supposés dans le groupe de traitement sont résumés dans la Table 1. Comme le montre cette table, un effet de traitement constant et fort, qui correspond au rapport de risque de 0.63, est supposé au sein de la population biomarqueur-positif. En revanche, l'effet du traitement s'affaiblit au sein de la population complémentaire, c'est-à-dire que le rapport de risque varie entre 0.75 dans le Scénario 1 et 0.94 dans le Scénario 3.

Table 1. Hypothèses relatives à l'effet du traitement

Scénario pour l'effet du traitement	Paramètre	Valeur
Scénario 1	Temps de survie médian dans le groupe de traitement (sous-ensemble biomarqueur négatif)	10 mois
	Temps de survie médian dans le groupe de traitement (sous-ensemble biomarqueur positif)	12 mois
Scénario 2	Temps de survie médian dans le groupe de traitement (sous-ensemble biomarqueur négatif)	9 mois
	Temps de survie médian dans le groupe de traitement (sous-ensemble biomarqueur positif)	12 mois
Scénario 3	Temps de survie médian dans le groupe de traitement (sous-ensemble biomarqueur négatif)	8 mois
	Temps de survie médian dans le groupe de traitement (sous-ensemble biomarqueur positif)	12 mois

Les caractéristiques fonctionnelles principales des deux designs sont présentées dans les Tables 2 et 3.

La Table 2 fournit une comparaison globale des designs traditionnels et adaptatifs et présente la probabilité commune d'arrêt pour futilité lors de la première analyse intermédiaire ainsi que la probabilité de succès pour chaque design. Comme prévu, la probabilité d'arrêt pour futilité augmente à mesure que le rapport de risque présumé dans la population de patients biomarqueur-négatifs approche 1. Un signal d'efficacité plus faible dans cette population influence directement l'effet global du traitement et la règle d'arrêt pour futilité est déclenchée plus fréquemment dans le Scénario 3 par rapport au Scénario 1. Ensuite, en comparant les valeurs de puissance pour les deux designs dans la Table 2, il est utile de noter que, si le design adaptatif sélectionne deux populations pour l'analyse finale, la probabilité de succès est définie comme la probabilité d'établir un effet de traitement significatif dans l'une ou l'autre population. Il ressort de la Table 2 que les deux designs garantissent pratiquement le même niveau de puissance (environ 80%) si un fort effet de traitement est attendu chez les patients biomarqueur-négatifs (Scénario 1). Dans les autres scénarios, l'effet du traitement au sein de la population négative au biomarqueur est supposé être plus faible et l'approche adaptative démontre un avantage de

puissance par rapport au design traditionnel. Ce gain de puissance est dû à la règle de sélection flexible de la population utilisée par le design adaptatif. Si le traitement expérimental semble ne bénéficier qu'aux patients biomarqueur-positifs, le design adaptatif sélectionnera très probablement cette population pour l'analyse finale. La capacité d'identifier la meilleure population de patients d'une manière basée sur les données se traduit par une probabilité de succès plus élevée par rapport à l'approche traditionnelle qui évalue toujours l'effet global dans l'essai. Bien sûr, la puissance des deux designs est insuffisante dans les Scénarios 2 et 3 et le nombre cible d'événements doit être augmenté si un signal d'efficacité plus faible est prévu dans la population des patients biomarqueur-négatifs.

La règle de sélection flexible de la population utilisée dans le design adaptatif est illustrée dans la Table 3. Cette table présente les probabilités de résultats individuels lors de la deuxième analyse intermédiaire dans les trois scénarios d'effet du traitement. En commençant par le Scénario 1, l'effet du traitement est supposé fort dans la population de biomarqueur-positifs ainsi que dans son complément et, pour améliorer la probabilité de succès de l'essai, la règle de décision sélectionne la population globale ou les deux populations pour l'évaluation finale la plupart du temps. La probabilité de ne choisir que la population des patients biomarqueur-positifs est assez faible dans ce scénario (environ 20%). Avec un effet de traitement plus faible, la règle de décision est plus susceptible d'identifier la population des patients biomarqueur-positifs (avec ou sans la population globale) comme la meilleure population à examiner lors de l'analyse finale. Par exemple, dans le Scénario 3, l'effet bénéfique du traitement expérimental est limité au sous-ensemble de patients biomarqueur-positifs et cette population est choisie plus de 50% du temps et, en outre, les deux populations sont sélectionnées environ 30% du temps. Dans ce cas, il sera contre-intuitif d'évaluer l'effet du traitement uniquement dans l'ensemble de la population et, comme prévu, la règle de sélection de la population recommande rarement cette population (seulement environ 13% du temps).

Références

Dmitrienko, A., Paux, G. (2017). Subgroup analysis in clinical trials. *Clinical Trial Optimization Using R*. Dmitrienko, A., Pulkstenis, E. (editors). Chapman and Hall/CRC Press, New York.

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

Millen, B., Dmitrienko, A., Ruberg, S., Shen, L. (2012). A statistical framework for decision making in confirmatory multi-population tailoring clinical trials. *Drug Information Journal*. 46, 647-656.

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

Table 2. 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	12.8%
	Design traditionnel : Puissance	79.8%
	Design Adaptif : Puissance	79.2%
Scénario 2	Probabilité d'arrêt pour futilité lors de la première analyse intermédiaire	20.7%
	Design traditionnel : Puissance	66.8%
	Design Adaptif : Puissance	68.7%
Scénario 3	Probabilité d'arrêt pour futilité lors de la première analyse intermédiaire	31.1%
	Design traditionnel : Puissance	49.1%
	Design Adaptif : Puissance	58.4%

Table 3. Sélection de la population dans le design adaptatif

Scénario pour l'effet du traitement	Paramètre	Valeur
Scénario 1	Probabilité de sélectionner uniquement la population globale pour l'analyse finale	39.9%
	Probabilité de sélectionner uniquement la population biomarqueur-positif pour l'analyse finale	19.9%
	Probabilité de sélectionner les deux populations pour l'analyse finale	40.2%
Scénario 2	Probabilité de sélectionner uniquement la population globale pour l'analyse finale	25.6%
	Probabilité de sélectionner uniquement la population biomarqueur-positif pour l'analyse finale	34.1%
	Probabilité de sélectionner les deux populations pour l'analyse finale	40.3%
Scénario 3	Probabilité de sélectionner uniquement la population globale pour l'analyse finale	12.8%
	Probabilité de sélectionner uniquement la population biomarqueur-positif pour l'analyse finale	57.3%
	Probabilité de sélectionner les deux populations pour l'analyse finale	29.9%