

# Designs traditionnels avec des objectifs multiples

MedianaDesigner package

## 1. Introduction

Ce document décrit la méthodologie statistique utilisée dans le module qui prend en charge les calculs de puissance basés sur des simulations dans les essais cliniques à objectifs multiples (fonction MultAdj).

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. Essais cliniques avec des objectifs multiples

Il est très courant d'utiliser plusieurs objectifs, par exemple plusieurs comparaisons dose-contrôle, dans les essais cliniques de stade avancé pour mieux décrire les profils d'efficacité des traitements expérimentaux. L'évaluation d'objectifs multiples donne lieu à des problèmes de multiplicité et entraîne une probabilité accrue de décisions incorrectes (risque global d'erreur de première espèce). Diverses méthodes statistiques connues sous le nom d'ajustements de la multiplicité ou, plus formellement, de procédures de tests multiples ont été proposées dans la littérature pour traiter l'inflation du risque d'erreur de première espèce. Pour un résumé des ajustements de multiplicité couramment utilisés dans les essais cliniques, voir Dmitrienko et D'Agostino (2013) et Alosh, Bretz et Huque (2014)

### 2.1. Essais cliniques avec une source unique de multiplicité

Les problèmes de multiplicité sont souvent classés en utilisant le concept de source de multiplicité. Une source de multiplicité est définie comme un facteur qui contribue à l'inflation du risque d'erreur de première espèce. Voici des exemples d'essais cliniques de phase III avec une source unique de multiplicité :

- Essais avec plusieurs critères d'évaluation cliniques.
- Essais avec plusieurs doses ou schémas posologiques par rapport à un groupe contrôle commun.

- Essais portant sur plusieurs populations de patients prédéfinies, par exemple la population globale de patients présentant une maladie d'intérêt et un sous-ensemble de patients présentant une certaine caractéristique de base.

Dans tous ces cas, une stratégie analytique unique telle que l'évaluation de plusieurs critères cliniques induit une multiplicité qui doit être traitée à l'aide de méthodes statistiques appropriées.

Pour protéger le risque d'erreur de première espèce dans les problèmes de multiplicité avec une seule source de multiplicité, les procédures de tests multiples suivantes peuvent être appliquées :

- Procédure de Bonferroni.
- Procédure de Holm.
- Procédure à séquence fixe.
- Procédures en chaîne ou graphiques.
- Procédure de Hochberg.
- Procédure de Hommel.

Ces procédures de tests multiples sont prises en charge par le paquet MedianaDesigner. La fonction MultAdj peut être utilisée pour effectuer des calculs de puissance basés sur des simulations dans les essais cliniques avec une seule source de multiplicité en utilisant l'un de ces ajustements de la multiplicité.

## 2.2. Essais cliniques avec plusieurs sources de multiplicité

De plus, il est de plus en plus courant d'utiliser des ensembles complexes d'objectifs dans les essais cliniques de phase III, par exemple plusieurs critères d'évaluation cliniques évalués à plusieurs niveaux de dose ou dans plusieurs populations de patients. Étant donné qu'une combinaison de plusieurs stratégies analytiques est utilisée dans ce contexte, les problèmes de multiplicité qui en résultent sont appelés problèmes avec plusieurs sources de multiplicité.

Pour protéger le risque d'erreur de première espèce dans ces problèmes complexes de multiplicité, une classe de procédures de tests multiples connues sous le nom de procédures gatekeeping a été introduite. Les procédures gatekeeping sont conçues pour les cas avec deux sources de multiplicité ou plus et tiennent compte des informations sur les relations cliniquement pertinentes entre les objectifs individuels. La fonction MultAdj permet des calculs de puissance basés sur des simulations dans des essais cliniques avec plusieurs sources de multiplicité en utilisant des procédures de contrôle efficaces dérivées des procédures de Holm, Hochberg et Hommel.

## 2.3. Essais cliniques avec de multiples objectifs connexes

Enfin, plusieurs objectifs tels que plusieurs critères d'évaluation peuvent être évalués dans un essai clinique dans le but de quantifier un effet global du traitement sur ces objectifs. Le plus souvent, des objectifs de ce type sont poursuivis dans des essais cliniques avec des critères

d'évaluation biologiquement liés. Dans ce cas, les critères d'évaluation individuels sont traités comme des composants d'une seule variable clinique sous-jacente et l'effet du traitement dans l'ensemble des critères d'évaluation est évalué à l'aide de procédures de test globales, par exemple, la procédure globale O'Brien basée sur les moindres carrés ordinaires. Les procédures de test globales sont le plus souvent utilisées dans les essais cliniques de phase II. La fonction MultAdj prend en charge les calculs de puissance basés sur des simulations dans les essais cliniques avec plusieurs critères d'évaluation connexes à l'aide de la procédure O'Brien.

### 3. Étude de cas

Le processus de calculs de taille d'échantillon basés sur des simulations dans les essais cliniques à objectifs multiples sera illustré à l'aide d'un essai de phase III avec une source unique de multiplicité. Cette étude de cas est basée sur un essai de confirmation pour le traitement du diabète de type 2 (Rosenstock et al., 2009). Un critère d'évaluation principal unique est supposé dans cette étude de cas et la multiplicité est induite par la comparaison de trois doses d'un traitement expérimental à un placebo (les doses seront marquées L, M et H). Pour contrôler le risque d'erreur de première espèce au niveau nominal (niveau unilatéral de 0.025) dans cet essai, un ajustement de multiplicité doit être défini prospectivement.

Les informations de l'essai utilisées dans les calculs de la taille de l'échantillon sont résumées ci-dessous :

- Un design équilibré est envisagé dans l'essai et 30% des patients devraient être perdus de vue avant la fin de la période de traitement. La taille de l'échantillon commune pour chaque groupe (nombre de patients recrutés) varie entre 120 et 150 patients.
- Un certain nombre de scénarios d'effet du traitement (ensembles de tailles d'effet pour les comparaisons dose-placebo individuelles) seront pris en compte pour soutenir un calcul robuste de la taille de l'échantillon :
  - Scénario 1 (scénario réaliste) : Les tailles d'effet aux doses L, M et H sont de 0.25, 0.3, 0.35, respectivement.
  - Scénario 2 (scénario optimiste) : Les tailles d'effet aux doses L, M et H sont respectivement de 0.3, 0.35, 0.4.
  - Scénario 3 (scénario pessimiste) : La taille de l'effet commune entre les doses est de 0.3.

En général, étant donné qu'un effet bénéfique plus fort est attendu à la dose la plus élevée (dose H), des pondérations seront prédéfinies pour les trois comparaisons dose-placebo afin de quantifier leur importance relative :

- Dose L par rapport au placebo : Le poids est de 0.25.
- Dose M par rapport au placebo : Le poids est de 0.35.

- Dose H par rapport au placebo : Le poids est de 0.4.

Un ajustement de la multiplicité basé sur la procédure de Hommel sera utilisé dans cette étude de cas. Cette procédure de tests multiples garantit un contrôle global du risque d'erreur de première espèce dans ce contexte car les statistiques de test associées aux trois comparaisons suivent une loi normale trivariée et que les trois coefficients de corrélation sont positifs (Dmitrienko et D'Agostino, 2013). Cet ajustement de multiplicité sera mis en place pour tenir compte du schéma de pondération défini ci-dessus.

La probabilité globale de succès sera définie de manière standard, c'est-à-dire comme la probabilité d'établir un effet statistiquement significatif pour au moins une comparaison dose-placebo tout en contrôlant le risque d'erreur de première espèce global à un niveau unilatéral de 0.025 (cette définition de la puissance est connue sous le nom de puissance disjonctive). La taille de l'échantillon sera choisie pour garantir une puissance de 80% dans les scénarios décrits ci-dessus.

Des calculs de puissance ont été effectués dans le cadre de 12 scénarios au total afin d'évaluer la probabilité de succès dans les scénarios d'effet de traitement sélectionnés pour une gamme de tailles d'échantillon différentes. Les résultats des simulations sont présentés dans la Table 1. Comme prévu, cette table montre que la puissance disjonctive augmente avec la taille de l'échantillon dans chaque scénario. Si le Scénario 1 était le seul scénario d'effet de traitement à considérer pour le calcul de la taille de l'échantillon, le nombre commun de patients recrutés par bras serait fixé à 130 puisque cette taille d'échantillon donne une puissance de 80%. Pour permettre des calculs plus robustes, le Scénario 3 a été inclus et il est clair qu'un échantillon de plus grande taille sera nécessaire pour atteindre l'objectif de puissance de 80% pour cet essai. La taille finale de l'échantillon (en termes de patients recrutés) sera fixée à 140 patients par bras.

## Références

Alosh, M., Bretz, F., Huque, M. (2014). Advanced multiplicity adjustment methods in clinical trials. *Statistics in Medicine*. 33, 693-713.

Dmitrienko, A., D'Agostino, R. Sr. (2013). Traditional multiplicity adjustment methods in clinical trials. *Statistics in Medicine*. 32, 5172-5218.

Rosenstock et al. (2009). Effect of saxagliptin monotherapy in treatment-naive patients with type 2 diabetes. *Current Medical Research and Opinion*. 25, 2401-2411.

**Table 1. Puissance disjonctive en fonction de la taille de l'échantillon dans les trois scénarios d'effet de traitement**

<b>Taille d'échantillon commune par bras</b>	<b>Scénario 1</b>	<b>Scénario 2</b>	<b>Scénario 3</b>
120 patients	0.765	0.878	0.736
130 patients	0.805	0.904	0.776
140 patients	0.834	0.926	0.809
150 patients	0.865	0.941	0.824

La taille de l'échantillon est définie comme le nombre de patients recrutés.