UCLouvain

FACULTE DES SCIENCES

Faculté des sciences

Aspects statistiques des essais cliniques dans les maladies rares

Auteur·es : Yousra Bahbah
Promoteur·rices : Catherine Legrand
Lecteur·rices : Céline Bugli et Sandy Tubeuf
Année académique 2022-2023

Table des matières

Chapitre 1

Introduction

Les maladies rares sont, selon la définition donnée par l'OMS, des maladies touchant moins d'une personne sur deux mille. Ces maladies sont souvent chroniques, non curables et impactent fortement la qualité de vie des malades. Le nombre de traitements est souvent limité et ceux-ci agissent essentiellement sur les symptômes, sans traiter la cause de ces maladies.

Le développement de nouveaux traitements pour les maladies rares est nécessaire mais comporte de nombreuses difficultés. Cela est dû à plusieurs facteurs.

Tout d'abord, le manque de connaissances au sujet de ces maladies limite le développement de nouvelles thérapies. Souvent, leurs mécanismes physiopathologiques et leurs causes sont mal compris, ce qui empêche l'identification de potentiels nouveaux traitements.

De plus, lorsque de nouveaux traitements sont identifiés, la mise en place d'un essai clinique permettant d'étudier leur impact sur des êtres humains est difficile. La principale difficulté est liée à la taille de l'échantillon limité, qui est essentiellement dû à la faible prévalence des maladies rares. La taille d'échantillon nécessaire aux designs les plus utilisés dans les essais cliniques est généralement impossible à obtenir dans ce contexte, ne permettant pas à ces essais de tirer des conclusions robustes. Face à ces difficultés, des nouvelles stratégies sont nécessaires pour faire avancer la recherche sur ces maladies rares.

Des designs d'essais cliniques alternatifs existent et s'avèrent plus appropriés dans le cadre de ces maladies. De nombreux scientifiques ont étudié et écrit des articles sur ce sujet. C'est notamment le cas des articles "Alternative designs for clinical trials in rare diseases" de Abrahamyan et al. [1] et "Methodology of clinical trials for rare diseases" de Smith et al. [48].

Objectifs du mémoire

L'objectif de ce mémoire sera d'identifier les designs d'essais cliniques les plus adaptés aux maladies rares. Dans un second temps, un cas clinique sera étudié et un design d'essai clinique sera proposé. Le choix de ce design sera basé sur les spécificités de la maladie, sur l'objectif de l'étude ainsi que sur les informations disponibles.

Le chapitre 2 décrira les étapes du développement d'un traitement, de la phase préclinique à la phase de pharmacovigilance. Les essais contrôlés randomisés, qui sont des designs fortement utilisés dans les essais cliniques, seront présentés, ainsi que leurs limites dans le cadre des maladies rares.

Le chapitre 3 aura pour objectif de présenter différents designs alternatifs adaptés aux maladies rares. Ceux-ci répondent aux principales limitations causées par ce type de pathologie. Les raisons pour lesquelles ils sont adaptés à ces maladies et les aspects statistiques de ces designs y seront développés.

Le chapitre 4 présentera une proposition de design d'essai clinique pour une étude visant à comparer l'efficacité de quatre traitements dans le cadre des malformations artérioveineuses. Il s'agit d'une maladie touchant 1 personne sur 20 000 et où les options thérapeutiques sont fortement limitées.

Et enfin, le chapitre 5 présentera les conclusions générales de ce mémoire ainsi qu'une discussion sur les aspects qui auraient pu être explorés différemment ou approfondis.

Chapitre 2

Etapes du développement clinique d'un traitement

L'élaboration d'un traitement est un processus complexe qui s'étend sur plusieurs phases. Il peut s'écouler entre dix et quinze années de sa découverte à sa mise sur le marché.

Le développement clinique s'articule autour de plusieurs étapes cruciales, incluant la recherche préclinique, les essais cliniques, et également l'évaluation économique des nouveaux traitements

2.1 Phase préclinique

Le développement d'un nouveau traitement commence d'abord avec une phase préclinique. Cette section se base essentiellement sur l'article "Preclinical drug development" de Brodniewicz et al. [11].

Durant cette phase, différentes propriétés de la molécule sont étudiées. Premièrement, une étude des propriétés pharmacologiques de la molécule est effectuée. Celle-ci est composée de deux parties : une partie portant sur les propriétés pharmacodynamiques, et l'autre, sur les propriétés pharmacocinétiques.

L'étude des propriétés pharmacodynamiques d'une molécule passe par l'étude de son mécanisme d'action, de sa spécificité ainsi que de son affinité avec la cible, qui est l'entité biologique spécifique avec laquelle celle-ci interagit. L'efficacité et la puissance de la molécule sont aussi définies. L'efficacité d'un médicament est l'effet maximal qu'une molécule peut produire. Il s'agit de l'impact maximal causé par une dose déterminée après un temps donné. On considère généralement que l'effet maximal est atteint lorsque la dose administrée provoque une amélioration maximale de l'état de santé, sans entraîner des effets secondaires de grande ampleur. La puissance, quant à elle, est définie par la concentration efficace médiane (EC50), qui est la concentration nécessaire pour obtenir 50 % de l'effet maximal de la molécule étudiée.

Les propriétés pharmacocinétiques, quant à elles, englobent quatre aspects : l'absorption, la distribution, la métabolisation et l'excrétion de la molécule étudiée dans l'organisme. L'étude de l'absorption se focalise sur différents paramètres, tels que la solubilité aqueuse, mais aussi

la biodisponibilité orale et l'absorption intestinale dans le cas où la molécule serait prise par voie orale, ... La distribution est étudiée en évaluant la liaison de cette molécule aux protéines plasmatiques, sa capacité à traverser la barrière hémato-encéphalique, qui est la structure protectrice du système nerveux central, ... Le temps de demi-vie de la molécule et les voies enzymatiques par lesquelles elle est métabolisée sont des points caractérisant la métabolisation. Et enfin, l'excrétion étudie les voies d'élimination de la molécule telle que les voies urinaires et fécales. Tous ces aspects sont étudiés sur des modèles in vitro, en culture cellulaire ou sur des tissus, mais surtout sur des modèles in vivo, et donc, des animaux de laboratoire.

Ensuite, la sécurité de la molécule est évaluée. Pour ce faire, des tests de toxicité sont effectués. Ils permettent d'évaluer la toxicologie génétique, la cancérogénicité, les impacts sur la reproduction, ... que peut avoir la molécule étudiée.

La dose toxique minimale (MTD) et la dose sans effet observable (NOEL) seront aussi évaluées, tout comme l'impact de la molécule sur les différents systèmes, tels que le système nerveux, les voies respiratoires, le système cardiovasculaire, ... Ces tests sont effectués sur des animaux de laboratoire.

Tout au long de la phase préclinique, un dossier doit être rédigé. Celui-ci doit reprendre les points développés ci-dessus ainsi que d'autres aspects du développement de la molécule, tels que la description du processus de fabrication de celle-ci et les procédés analytiques utilisés pour contrôler la qualité du médicament. Ce dossier sera ensuite soumis aux autorités réglementaires afin d'obtenir l'autorisation de passer à la phase d'essai clinique.

2.2 Déroulement d'un essai clinique

Si les résultats obtenus s'avèrent concluants, le traitement pourra être étudié au cours d'un essai clinique. Un essai clinique est défini comme étant une étude menée pour tester l'efficacité et l'innocuité d'un médicament, d'un dispositif médical, d'un traitement ou d'une intervention de santé chez l'homme. Celui-ci se déroule sur plusieurs phases. Cette section se base essentiellement sur le livre "Fundamentals of clinical trials" de Friedman et al. [22].

La phase I évalue la tolérance des participants au traitement étudié. Elle a surtout pour objectif d'estimer la dose à administrer aux patients dans les prochaines phases de l'étude. Cela passe par la définition de la Dose Maximale Tolérée (DMT). Bien qu'il y ait différentes méthodes pour déterminer la DMT, l'approche la plus couramment utilisée consiste à augmenter progressivement les doses administrées jusqu'à atteindre un seuil où un nombre inacceptable de participants présente des signes de toxicité. Les propriétés pharmacodynamiques et pharmacocinétiques de la molécule continuent à être évaluées durant cette phase. Tous ces tests sont généralement effectués sur des volontaires sains, mais cela peut ne pas être le cas, lors d'essais cliniques sur des traitements pour le cancer par exemple. Le nombre participant à cette phase est limité (entre 20 et 80).

La phase II est la phase durant laquelle le traitement est évalué sur des patients malades. La taille de l'échantillon est plus importante qu'en phase I, elle est généralement comprise entre 50 et 200. Seules les doses définies durant la phase précédente seront administrées aux

patients. La posologie et le dosage utilisés à la phase III seront définis sur base des résultats obtenus à cette étape. L'activité biologique de la molécule est aussi évaluée, en étudiant la relation existant entre le taux sanguin et l'activité de la molécule ainsi que la proportion de patients ayant répondu aux traitements par exemple.

La phase III vise à évaluer l'efficacité et les éventuels effets indésirables du traitement étudié pour déterminer sa valeur potentielle dans un contexte clinique. Cette phase inclut un grand nombre de participants, généralement compris entre 300 et 3000. Le traitement est alors comparé à un placebo ou à une molécule déjà existante, ayant la même indication que ce dernier. Cette comparaison sert à mesurer la valeur clinique potentielle du nouveau traitement. Afin de garantir la comparabilité des groupes de l'étude, la phase III introduit un processus de randomisation, répartissant aléatoirement les participants entre ces groupes. Si aucun traitement contrôle n'existe et qu'il n'est pas possible d'inclure un placebo dans l'étude, la comparaison peut se faire avec un "Best Support Care (BSC)" dont le but est de soulager le patient au mieux. Le BSC diffère en fonction de la pathologie. Il peut s'agir de soins agissant uniquement sur les symptômes de la maladie, de soins psychologiques, palliatifs, ... [58]. Le groupe contrôle peut aussi ne rien prendre lorsque qu'aucune autre alternative n'est possible.

A la fin de cette phase, un rapport est rédigé. Il reprend les données de cette phase, ainsi que les résultats sur l'efficacité et la sécurité du traitement étudié. Ce rapport sera ensuite soumis aux autorités réglementaires. Si les résultats sont positifs et que les avantages l'emportent sur les risques, le traitement sera approuvé et obtiendra une autorisation de mise sur le marché.

Il est important de souligner que l'échantillon de patients des trois premières phases de l'essai clinique est limité. Il sera donc important de suivre l'effet du traitement une fois qu'il aura été mis sur le marché. L'objectif de la phase IV, aussi appelée phase de pharmacovigilance, est de continuer à évaluer l'effet du traitement lorsque celui-ci sera pris par un plus grand nombre de personnes, sur une durée plus longue, après sa commercialisation.

2.3 Evaluation du rapport coût-efficacité d'un nouveau traitement

Durant le développement d'un traitement, d'autres processus auront lieu en parallèle des phases cliniques. Ceux-ci permettent d'évaluer l'impact économique du traitement en cours de développement.

Le "Health Economic Assessment (HEA)" évalue les coûts et les avantages qu'une intervention de santé entraîne. Il compare ces facteurs à ceux d'autres traitements alternatifs déjà disponibles. Ce processus est réalisé dès les premières étapes d'un essai clinique et évalue donc l'impact économique que pourrait entraîner la mise sur le marché du traitement en cours de développement. Cette évaluation est souvent réalisée à partir de modèles économiques qui prennent en compte les coûts et les bénéfices sur le long terme. Le HEA permettra ensuite de prendre des décisions sur le financement de ces interventions [47].

Le "Health Technology Assessment (HTA)" est un processus évaluant l'efficacité clinique,

économique, la sécurité et l'efficacité des interventions de santé sur les patients, la société et le système de santé. Ce processus peut aussi inclure une analyse d'impact budgétaire, qui examine l'impact financier d'une intervention de santé sur le système de santé. Le but sera ensuite de prendre des décisions sur la réglementation, le remboursement et l'usage de cette dernière. Elle est menée par des organismes gouvernementaux, des agences d'évaluation des soins de santé, ... L'intervention de santé évaluée peut être un médicament, un dispositif médical ou une intervention chirurgicale par exemple. Le HTA a lieu une fois que les données évaluant l'efficacité et la sécurité d'un traitement sont récoltées. Elle a donc lieu dans les phases plus tardives du développement clinique [8].

2.4 Essais contrôlés randomisés

Les essais cliniques sont donc des processus complexes, constitués de plusieurs phases. Chaque phase joue un rôle clé, que ce soit pour établir la tolérance et la posologie d'un traitement, évaluer son efficacité ou surveiller ses effets une fois que celui-ci est mis sur le marché. De plus, l'évaluation de l'impact économique du traitement à travers le HEA et le HTA est également importante dans la prise de décision concernant son utilisation.

Cependant, la réussite du processus ne repose pas uniquement sur ces phases et ces évaluations, mais également sur le choix du design de l'étude. Parmi les différentes possibilités, l'Essai Contrôlé Randomisé (ECR) est couramment utilisé en phase III et présente de nombreux avantages. Nous détaillerons ce design dans les lignes qui suivent.

Un essai contrôlé randomisé, aussi appelé ECR, est un type d'essai clinique qui se déroule de la façon suivante. Dans le cas où deux traitements sont étudiés, les patients sont assignés aléatoirement à un groupe de traitement expérimental ou à un groupe de traitement contrôle. L'étude est contrôlée avec un traitement contrôle. Celui-ci peut être une molécule connue ayant la même indication que le traitement expérimental, un placebo, un BSC, ou même une absence de traitement comme expliqué à la section 2.2. La figure 2.4.1 illustre le déroulement d'un ECR parallèle simple. Les patients sont assignés au groupe contrôle et au groupe expérimental par randomisation. Après une période définie, une mesure de l'effet du traitement (appelé "output") est effectuée. L'output peut prendre diverses formes, en fonction de la nature de l'étude. Il peut s'agir d'indicateurs cliniques telles que la tension artérielle, le poids, ou d'autres variables spécifiques à la maladie étudiée [22].

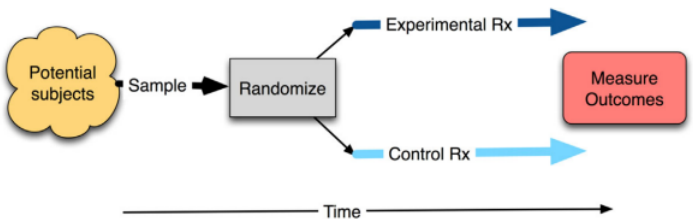

FIGURE 2.4.1 – Schéma représentant le déroulement d'un ECR parallèle [1].

Le caractère randomisé de l'ECR permet à tous les patients d'avoir la même probabilité

d'être assigné aux différents groupes de traitement, dans les cas les plus basiques de randomisation [22]. L'application de la randomisation n'est recommandée que dans une situation d'équipoise. Cela signifie que les scientifiques de l'étude ne peuvent pas affirmer qu'un des traitements étudiés est supérieur à l'autre [23].

On considère généralement que les ECR permettent d'apporter le plus haut niveau de preuve sur les bénéfices d'un traitement. La randomisation contribue principalement aux avantages amenés par ces essais [1].

Tout d'abord, elle permet d'éliminer le biais qui pourrait apparaître durant l'allocation des patients au sein des groupes de traitement. Pour cela, il faut que la randomisation soit correctement effectuée et qu'elle ne puisse pas être prédite [22].

Ensuite, la randomisation permet de contrôler les facteurs confondants. Les facteurs confondants sont des variables qui peuvent influencer à la fois l'output et le traitement, créant ainsi une fausse association entre les deux. Grâce à la randomisation, ces facteurs sont répartis équitablement entre les groupes, rendant ces derniers comparables. Toute différence observée dans l'output peut alors être attribuée au traitement [48].

Un autre atout des ECR est qu'ils permettent de contrôler les risques d'erreurs de type I (α) et de type II (β) du test d'hypothèse effectué sur les résultats de l'essai. L'erreur de type I est la probabilité de rejeter à tort l'hypothèse nulle, et donc, de conclure à tort qu'un traitement expérimental a un effet significativement supérieur à celui d'un traitement contrôle. L'erreur de type II, par contre, est la probabilité de ne pas rejeter l'hypothèse nulle alors qu'elle est fausse, et de ne pas identifier l'efficacité d'un traitement quand elle est réellement présente. Pour ce faire, une certaine taille d'échantillon est nécessaire. Celle-ci est également influencée par la différence d'effet Δ à mettre en évidence entre les deux groupes de traitement [31].

La puissance d'une étude est proportionnelle à la taille de l'échantillon et à l'effet du traitement dans la population considérée. Plus ceux-ci sont importants, plus la puissance de l'essai l'est aussi. La probabilité de l'étude de mettre en évidence l'effet d'un traitement expérimental, s'il est présent, sera alors plus importante [31].

Lorsque l'output étudié est continu et suit une distribution approximativement normale, il est possible d'utiliser les formules 2.4.1 et 2.4.2 pour calculer respectivement la puissance de l'étude et la taille de l'échantillon nécessaire pour un ECR parallèle [35, 31].

$$1 - \beta = \Phi\left(\frac{Z_{1-\frac{\alpha}{2}} - \sqrt{n} \cdot \Delta/\sigma}{\sqrt{2}}\right) + \Phi\left(\frac{-Z_{1-\frac{\alpha}{2}} - \sqrt{n} \cdot \Delta/\sigma}{\sqrt{2}}\right) \tag{2.4.1}$$

où
β est l'erreur de type II,
$\Phi(\cdot)$ est la fonction de répartition de la loi normale centrée réduite,
$Z_{1-\frac{\alpha}{2}}$ est le percentile d'ordre $1 - \alpha/2$ de la loi normale centrée réduite,
n est la taille de l'échantillon,

Δ est la différence à mettre en évidence entre deux groupes pour que l'effet du traitement soit considéré comme cliniquement important,

σ est l'écart-type résiduel.

La taille d'échantillon, n, est alors donnée par

$$\frac{2(Z_{1-\beta} + Z_{1-\frac{\alpha}{2}})^2\sigma^2}{\Delta^2} \tag{2.4.2}$$

où $Z_{1-\beta}$ est le percentile d'ordre $1-\beta$ de la loi normale centrée réduite,

$Z_{1-\frac{\alpha}{2}}$ est le percentile d'ordre $1-\frac{\alpha}{2}$ de la loi normale centrée réduite,

σ^2 est la variance résiduelle,

Δ est la différence à mettre en évidence entre deux groupes pour que l'effet du traitement soit considéré comme cliniquement important.

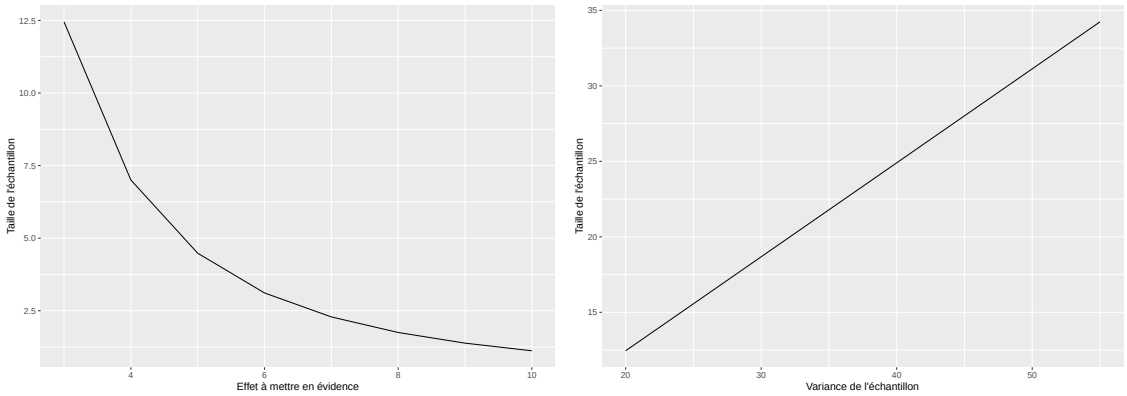

FIGURE 2.4.2 – Graphique représentant l'évolution de la taille de l'échantillon en fonction de la taille de l'effet à mettre en évidence Δ (gauche) et la variance σ^2 (droite).

La formule 2.4.2 montre que la taille de l'échantillon est inversement proportionnelle à la différence d'effet à mettre en évidence et proportionnelle à la variance. La figure 2.4.2 ci-dessus illustre l'évolution de la taille de l'échantillon en fonction de ces deux paramètres.

A gauche, l'influence de Δ est illustrée. Plus celui-ci est grand, plus la taille de l'échantillon nécessaire à l'ECR parallèle est petite. Mettre en évidence une grande différence d'effet requiert une taille d'échantillon moins importante.

A droite, le graphique représentant l'évolution de la taille de l'échantillon en fonction de la variance de la population σ^2. Il montre que plus la variance sera importante, plus la taille de l'échantillon nécessaire à l'ECR parallèle sera importante. En effet, une variance importante signifie que les données sont plus dispersées. Elles sont donc moins fiables pour estimer la taille d'échantillon requise. Ce manque de fiabilité doit être compensé par un plus grand échantillon.

Cependant, il peut être compliqué, voire impossible, de recruter le nombre de patients nécessaire à un ECR. C'est le cas dans les essais cliniques pour les maladies rares, défini comme

étant des maladies touchant moins d'une personne sur deux mille par l'OMS [1]. Le nombre de personnes touchés par ces maladies n'est donc pas élevé, ce qui limite le nombre de patients pouvant être recruté. Il n'est parfois pas possible de recruter le nombre de patients calculé par la formule 2.4.2. Une taille d'échantillon insuffisante peut entraîner une étude dont la puissance est en dessous de la valeur fixé $1 - \beta$. Cette étude risquerait de ne pas mettre en évidence la supériorité d'un traitement expérimental, alors que ce dernier est réellement supérieur [48].

Un autre problème des ECR dans le contexte des maladies rare est qu'il n'est parfois pas possible de contrôler l'étude avec un traitement actif car les traitements disponibles pour ces maladies sont très limités [48]. De plus, lorsqu'une nouvelle option thérapeutique est disponible, les participants de l'étude sont plus enclins à vouloir la prendre, surtout lorsque ce traitement est une des seules possibilités thérapeutiques possibles. Certains sujets s'opposent même aux essais s'ils ont une probabilité d'être assignés à un bras potentiellement inférieur, ce qui est le cas dans les ECR. Cela affecte le recrutement des patients au sein de ces études et limite davantage la taille de l'échantillon qui pourrait être recrutée durant un essai [25].

Pour ces raisons, un ECR classique est parfois difficile à mettre en place dans le contexte des maladies rares. Il faudra alors se tourner vers des designs d'essais cliniques alternatifs. Ceux-ci seront plus adaptés aux maladies rares, lorsque la réalisation d'un ECR n'est pas possible. Ils présentent même certains avantages par rapport aux ECR parallèles dans ce contexte. En effet, ils peuvent nécessiter une taille d'échantillon inférieure à celle d'un ECR parallèle. Cette diminution est rendue possible grâce à l'utilisation de données historiques, ou par la diminution de la variabilité interindividuelle. Comme le montre l'équation 2.4.2, la variance est proportionnelle à la taille de l'échantillon calculée. Diminuer cette variance permet donc de diminuer la taille de l'échantillon nécessaire à l'essai clinique. D'autres designs permettent à tous les participants de prendre le traitement expérimental lors de l'étude, ce qui favorise le recrutement de l'essai clinique [1].

Ces designs alternatifs d'essais cliniques adaptés aux maladies rares seront présentés dans le chapitre suivant.

Chapitre 3

Designs d'essais cliniques alternatifs adaptés aux maladies rares

3.1 Design d'essais cliniques permettant la diminution de la taille de l'échantillon

Réduire la taille de l'échantillon nécessaire à une étude clinique est une stratégie intéressante pour les maladies rares. Une façon d'aboutir à ce résultat est de donner, à chaque patient, le traitement contrôle et le traitement expérimental. Chaque patient sera alors son propre contrôle. Cette stratégie permet de diminuer la variance observée dans les données, augmentant la puissance de l'essai. A puissance égale, la taille de l'échantillon nécessaire sera moindre [48].

Le design "cross-over" et le design "n-of-1" utilisent cette stratégie. Leurs conceptions visent à minimiser la variabilité au sein des données. Cela permet une diminution de la taille de l'échantillon nécessaire à l'étude par rapport à la taille d'échantillon d'un ECR parallèle classique, à puissance, niveau de significativité et différence d'effet à mettre à évidence égale [48].

3.1.1 Design "cross-over"

3.1.1.1 Généralité

Dans le cas où deux traitements sont étudiés, un traitement contrôle A et un traitement expérimental B, le design "cross-over" est un essai qui se déroule en deux périodes. Dans un premier temps, les patients sont randomisés, à probabilité égale, au sein du groupe contrôle et du groupe expérimental. Après un temps donné, les patients changent de groupe de traitement. Le groupe qui prenait le traitement A finira l'étude en prenant le traitement B. L'autre groupe, qui prenait initialement le traitement B, finira l'étude avec le traitement A [1].

Il y a donc deux séquences de traitements possibles dans cet exemple. Soit la séquence 1 AB, soit la séquence 2 BA. Durant la période 1 de la séquence 1, l'effet du traitement contrôle est évalué et durant la période 2 de cette même séquence, c'est l'effet du traitement expérimental qui est évalué. Tandis que pendant la période 1 de la séquence 2, l'effet du traitement

expérimental est évalué et pendant la période 2 de cette même séquence, c'est l'effet du traitement contrôle qui est évalué [1].

Tout le déroulement de ce design est illustré à la figure 3.1.1, ci-dessous. Cet exemple montre le cas du design cross-over le plus simple : une étude qui se déroule en deux temps où seul deux traitements sont évalués et où seul deux séquences sont possibles. Dans le cadre de ce chapitre, seul ce cas de figure sera abordé. Mais il existe également des designs cross-over étudiant plus de deux traitements, où plus de deux séquences sont possibles, et avec un plus grand nombre de périodes [1].

Dans un design cross-over, une période de "washout" a lieu entre la prise de deux traitements. Ce washout est une pause durant laquelle aucun traitement n'est administré. Elle permet à l'effet du premier traitement de s'estomper et d'éviter la présence d'un effet carryover à la seconde période de l'étude [1].

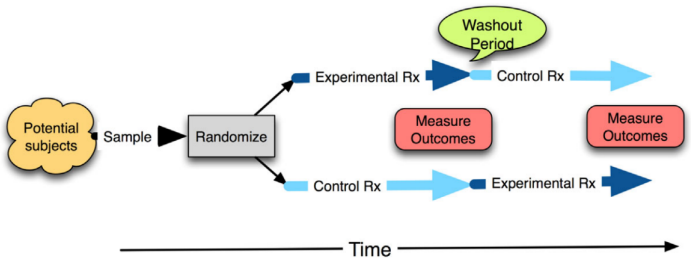

FIGURE 3.1.1 – Schéma représentant le déroulement d'un design cross-over [1].

Un effet carryover se produit lorsque le premier traitement continue d'agir alors que le deuxième traitement a été initié. Il influence alors l'effet du second traitement. Cela pose un problème lors de l'analyse des résultats. La présence d'une période washout est donc souvent nécessaire. Cette étape augmente le temps de l'étude. Il faut donc veiller à ce que le temps de demi-vie des molécules étudiées soit raisonnable. De plus, l'effet des molécules doit être rapidement observé afin de pouvoir évaluer l'effet des traitements rapidement [36].

L'effet des traitements doit aussi être réversible afin que la seconde molécule soit évaluée dans les mêmes conditions que la première. Il faudra aussi veiller à ce que la maladie étudiée soit stable, l'état du patient ne doit pas évoluer de façon trop importante. Les traitements étudiés ne doivent agir que sur les symptômes sans avoir un effet curatif. Ce type de design ne s'applique donc que pour des maladies chroniques et stables, ce qui est souvent le cas dans les maladies rares [1].

Ce design comporte de nombreux avantages dans le contexte des maladies rares. En plus de permettre à tous les patients d'avoir accès au traitement expérimental, sa conception diminue la variabilité dans les données. Dans un essai randomisé contrôlé, une forme de variabilité peut être introduite par des différences entre les patients. Or, dans un design cross-over, chaque patient joue à la fois le rôle de l'individu traité et de l'individu contrôle, ce qui réduit la variabilité interindividuelle de la réponse au traitement. Cette réduction est due à l'élimination des variations causées par les différences individuelles entre les patients. La variance totale étant

la somme des variances interindividuelles et intraindividuelles, cela entraîne une diminution de la variance totale observée. La variance étant proportionnelle à la taille d'un échantillon, cette diminution entraînera donc une diminution de la taille d'échantillon nécessaire, dans le cadre d'un design cross-over [48].

3.1.2 Méthodologie

3.1.2.1 Analyse statistique

Il existe différentes méthodes d'analyse pour comparer l'efficacité des différents traitements étudiés au cours d'un essai clinique utilisant un design cross-over. Cela dépendra du type d'output mesuré. Cet output peut être continu, binaire ou encore discret. Le principe de ces méthodes restera assez similaire.

Un type d'output peut néanmoins poser problème pour les designs cross-over, il s'agit des données du type "time to event", qui sont des données propres aux analyses de données de survie. En fonction de l'évènement étudié, le retour à la condition initiale peut être impossible. Ce sera le cas quand l'évènement attendu est la mort d'un individu par exemple. Or, le retour à la condition initiale du patient est une condition nécessaire à l'utilisation d'un design cross-over. Il faudra donc s'assurer qu'une fois que l'évènement étudié aura lieu, l'état initial du patient pourra être retrouvé [30].

Dans cette section, seules les méthodes évaluant l'effet de deux traitements et utilisant un output continu seront décrites.

La comparaison des résultats de deux groupes se fait généralement avec un test t pairé lorsque la différence entre les données appariées suit une distribution approximativement normale. Lorsque ce n'est pas le cas, il est possible d'utiliser un "Wilcoxon signed-rank test". Cependant, ces deux tests ne permettent pas de tenir compte de la structure des données d'un design cross-over. En effet, ils ne tiennent pas compte du fait que l'étude est structurée en séquences et en périodes, et des effets que ceux-ci pourraient induire, tels que la présence potentielle d'effets carryover [56].

En présence d'effets carryover, une analyse classique peut amener un biais dans l'estimation de l'effet des traitements, faussant la comparaison des bras de l'étude. Nous avons réalisé une petite étude de simulation afin d'illustrer ce propos. Pour évaluer l'effet de deux traitements (A et B) ayant le même effet, une simulation a été réalisée en utilisant une distribution normale $N(10, 4)$. Cela suppose que les deux traitements ont un effet moyen de 10, et que la variance présente dans les résultats est de 4. Dans cette simulation, 50 patients ont été assignés à la séquence AB et 50 patients ont été assignés à la séquence BA.

Supposons que le traitement B ait un temps de demi-vie très long et que son effet persiste lorsque les patients de la séquence BA prennent le traitement A. L'effet résiduel du traitement B chez les patients de la séquence BA a été généré de manière aléatoire en utilisant une distribution normale $N(5, 4)$. Cela suppose que l'effet carryover du traitement B a une moyenne de 5, et que la variance présente dans les résultats est de 4.

Une analyse des résultats a été réalisée, dans le but de comparer l'effet du traitement A à l'effet du traitement B à l'aide d'un test t pairé. Ce test a mis en évidence la supériorité du traitement A par rapport au traitement B avec une p-valeur de 5×10^{-9} alors que les deux traitements ont le même effet. Cette différence dans les résultats est attribuée à l'effet carryover induit par le traitement B chez les patients de la séquence BA. Cela montre bien l'importance de prendre en compte les effets carryover dans l'analyse des résultats d'un design cross-over, afin d'éviter des conclusions erronées sur l'effet des traitements étudiés.

D'autres méthodes ont donc été développées afin de tester et comparer l'effet des traitements étudiés.

Wang et al. [53] ont proposé une méthode d'analyse permettant de tenir compte de ces effets carryover. Elle permet de tester la présence d'effets carryover. En fonction des conclusions de ce test, ils proposent d'utiliser la statistique de test adéquate afin d'analyser les résultats d'un design cross-over. Cette méthode suppose que les données suivent une distribution normale et l'homoscédasticité de la variance, c'est-à-dire une variance des erreurs constante.

Voici un exemple afin d'illustrer cette méthode. Lorsque deux traitements sont étudiés, n_1 patients se voient attribuer la séquence AB. Cela veut dire qu'il commence l'étude avec le traitement A pour ensuite prendre le traitement B. n_2 patients sont assignés à la séquence de traitements BA. Ils commencent donc l'étude avec le traitement B et finissent celle-ci avec le traitement A. Dans ce cas, le traitement A est le traitement contrôle tandis que le traitement B est le traitement expérimental. L'étude compte $n = n_1 + n_2$ patients.

L'output continu évalué est noté Y. Il représente la réponse continue aux traitements et μ est le paramètre d'intérêt. μ_{ik} est l'espérance de Y pour la séquence i (à savoir $i = 1$ lorsque la séquence est AB et $i = 2$ lorsque la séquence est BA) à la période k ($k = 1, 2$). $\bar{Y}_{ik.}$ est la moyenne des observations y_{ikj} (j = 1,...,n_i), où n_i est le nombre de patients assigné à la séquence i. Si n_i est suffisamment grand, $\bar{Y}_{ik.}$ suivra approximativement une distribution normale, de moyenne μ_{ik} et de variance $\frac{\sigma^2}{n_i}$, où σ^2 est la variance résiduelle.

Le paramètre de l'effet du traitement est défini par $\theta = \mu_B - \mu_A$. θ quantifie la différence entre l'effet moyen du traitement expérimental (B) et l'effet moyen du traitement contrôle (A). L'hypothèse nulle stipule que $H_0 : \theta \leq 0$ et l'hypothèse alternative est $H_a : \theta > 0$. Ce test d'hypothèse sera évalué avec un taux d'erreur de type I $\frac{\alpha}{2}$. Un test unilatéral est utilisé car l'hypothèse alternative affirme que l'effet du traitement expérimental est supérieur à celui du traitement contrôle. L'objectif de formuler une telle hypothèse est de détecter une éventuelle supériorité du traitement expérimental par rapport au traitement de contrôle. Bien que les tests bilatéraux soient généralement recommandés dans le cadre d'un essai de supériorité, les auteurs ont opté pour ce type de test, d'où la décision de fixer l'erreur de type I à $\frac{\alpha}{2}$ plutôt qu'à α.

Le paramètre de l'effet carryover est égal à $\lambda = (\mu_{21} - \mu_{11}) - (\mu_{12} - \mu_{22})$. En absence d'effet carryover, la différence de l'effet moyen des traitements devrait être la même, et donc, la soustraction de ces deux différences entraînerait le fait que $\lambda = 0$.

Dans un premier temps, la présence d'effets carryover est évaluée par le test d'hypothèse suivant, $H_0 : \lambda = 0$ et $H_a : \lambda \neq 0$. Elle est testée avec un niveau d'erreur de type I de niveau α avec la statistique de test T_λ, qui suit une distribution asymptotiquement normale. Elle est calculée à l'aide de la formule suivante :

$$T_\lambda = \frac{\hat{\lambda}}{\hat{\sigma}_\lambda} = \frac{\bar{Y}_{21.} - \bar{Y}_{11.} - \bar{Y}_{12.} + \bar{Y}_{22.}}{\sqrt{2\tilde{\sigma}^2(1+\hat{\rho})n/(n_1 n_2)}}$$

où
$\bar{Y}_{ik.}$ est la moyenne des observations des sujets j dans la séquence i à la période k,
$\hat{\rho} = 1 - \frac{\tilde{\sigma}_w^2}{\tilde{\sigma}^2}$ est l'estimation de la corrélation entre les données d'un même individu,
$\tilde{\sigma}^2 = \sum_{i=1}^{2}\sum_{k=1}^{2}\sum_{j=1}^{n_i} \frac{(y_{ikj} - \bar{Y}_{ik.})^2}{2(n-2)}$ est l'estimation de la variance totale des observations,
$\tilde{\sigma}_w^2 = \frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{n_i} \frac{[y_{i2j} - y_{i1j} - (\bar{Y}_{i2.} - \bar{Y}_{i1.})]^2}{n-2}$ est l'estimation de la variance intraindividuelle,
n_i est le nombre de patients dans la séquence i,
$n = n_1 + n_2$ est la taille totale de l'échantillon.

Le niveau α est choisi en fonction de l'erreur de type I souhaité, qui est le risque de conclure à la présence d'effets carryover alors que ceux-ci ne sont pas présents. Dans l'article, les auteurs ont choisi de faire varier la valeur de α entre 0.1 et 0.2.

Plusieurs statistiques de test ont été proposé par Wang et al. [53] en fonction du résultat de ce premier test.

Si l'hypothèse évaluant la présence d'effets carryover n'est pas rejetée, on considère que ceux-ci ne sont pas présents. L'ensemble des données obtenues au cours de l'étude sera alors exploité. Les hypothèses suivantes seront alors évaluées, $H_0 : \mu_{21} + \mu_{12} \leq \mu_{11} + \mu_{22}$ et $H_a : \mu_{21} + \mu_{12} > \mu_{11} + \mu_{22}$. Ces hypothèses seront testées avec un taux d'erreur de type I $\frac{\alpha}{2}$. La statistique de test T_p sera alors utilisée. T_p suit une distribution normale asymptotique et est calculée de la façon suivante :

$$T_p = \hat{\theta}_p / \hat{\sigma}_p$$

où
$\hat{\theta}_p = \frac{1}{2}(\bar{Y}_{21.} - \bar{Y}_{11.} + \bar{Y}_{12.} - \bar{Y}_{22.})$
$\hat{\sigma}_p^2 = \frac{1}{2}\tilde{\sigma}^2(1-\hat{\rho})\frac{n}{n_1 n_2}$ est l'estimation de la variance de la différence entre les traitements,
$\hat{\rho} = 1 - \frac{\tilde{\sigma}_w^2}{\tilde{\sigma}^2}$ est l'estimation de la corrélation entre les données d'un même individu,
$\tilde{\sigma}_w^2 = \frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{n_i} \frac{[y_{i2j} - y_{i1j} - (\bar{Y}_{i2.} - \bar{Y}_{i1.})]^2}{n-2}$ est l'estimation de la variance intraindividuelle,
$\tilde{\sigma}^2 = \sum_{i=1}^{2}\sum_{k=1}^{2}\sum_{j=1}^{n_i} \frac{(y_{ikj} - \bar{Y}_{ik.})^2}{2(n-2)}$ est l'estimation de la variance totale des observations,
n_i est le nombre de patients dans la séquence i, $i = 1, 2$,
$n = n_1 + n_2$ est la taille totale de l'échantillon.

En présence d'effets carryover, $\lambda \neq 0$. Cela induit un biais dans l'estimation des effets de la seconde période. Seuls les résultats de la première période seront donc pris en compte lors de l'analyse des résultats. Les résultats obtenus à la seconde période ne seront pas utilisés. La statistique de test utilisée sera T_1. Elle suit une distribution normale asymptotique. Cette statistique ne teste que l'effet des traitements qui ont été étudié à la première période. Dans le cas de la première séquence AB, l'effet du traitement contrôle A sera étudié et dans le cas de la seconde séquence BA, celui du traitement B expérimental le sera aussi. Les hypothèses suivantes seront alors testées, $H_0 : \mu_{21} \leq \mu_{11}$ et $H_a : \mu_{21} > \mu_{11}$. Ce test d'hypothèse sera évalué avec un taux d'erreur de type I $\frac{\alpha}{2}$. La statistique de test T_1 se calculera alors de la façon suivante :

$$T_1 = \hat{\theta}_1 / \hat{\sigma}_1$$

où
$\hat{\theta}_1 = \bar{Y}_{21.} - \bar{Y}_{11.}$,
$\hat{\sigma}_1^2 = \frac{n}{(n_1 \times n_2)} \hat{\sigma}^2$ est une estimation de la variance de l'effet moyen des traitements,
$\hat{\sigma}^2 = \sum_{i=1}^{2} \sum_{j=1}^{n_i} (y_{i1j} - Y_{i1.})^2 / (n-2)$ est l'estimation de la variance totale des observations,
n_i est le nombre de patients dans la séquence i,
$n = n_1 + n_2$.

L'utilisation de la statistique de test T_1 réduit le design cross-over à un design ECR parallèle où seule la moitié des données est exploitée. Comme la puissance d'un essai dépend de la taille de l'échantillon, la présence d'effets carryover impacte grandement celle-ci et la précision de l'analyse des résultats. Cette inflation est due à l'absence d'indépendance entre les tests évaluant la présence d'effets carryover et l'effet des traitements. En effet, une erreur commise lors de la première étape du test peut entraîner une augmentation de l'erreur de type I $\frac{\alpha}{2}$ lors de la seconde étape, qui évalue l'efficacité des traitements étudiés. Pour contrecarrer ce problème, l'utilisation d'une correction séquentielle, comme celle proposée par Holm [28], est possible. Elle permet d'ajuster le niveau de signification du deuxième test en fonction du résultat du premier.

Dans notre contexte, la procédure d'Holm serait appliquée de la façon suivante. D'abord, le premier test statistique qui évalue la présence d'effets carryover serait réalisé. Si ce test s'avère significatif, alors le second test comparant l'efficacité des traitements serait effectué. Dans ce cadre, l'hypothèse nulle du deuxième test serait rejetée si la p-valeur obtenue est inférieure ou égale à $\frac{\alpha}{N}$, où $N = 2$ représente le nombre total de tests effectués et α le taux d'erreur de type I (ici, $\frac{\alpha}{2}$ dans le cas d'un test unilatéral).

Il existe encore de nombreuses façons analogues à celle de Wang et al. [53] d'analyser les résultats d'un design cross-over. Ces méthodes évaluent aussi, dans un premier temps, la présence d'effets carryover. En fonction de ce premier résultat, une méthode est proposée avec le test d'hypothèse et la statistique de tests adéquats. Le test d'hypothèse a pour but de prouver la supériorité de l'effet du traitement expérimental à celui du traitement contrôle. Des articles

écrits par Willan et Pater [57], Senn [45], ... traitent de ce sujet.

Il est aussi possible d'analyser les résultats d'un design cross-over à l'aide d'un modèle linéaire mixte. Généralement, ces modèles reposent sur l'hypothèse de normalité des effets aléatoires et des résidus. Ces effets aléatoires ainsi que les erreurs résiduelles sont supposés indépendants [44]. Une fois que ces hypothèses ont été validées, il sera possible de procéder à l'estimation des paramètres du modèle, incluant celle de l'effet du traitement. Dans un design cross-over à i séquences et k périodes, la réponse aux traitements peut être modélisée de la façon suivante [30] :

$$Y_{ikj} = \omega + \pi_k + \tau_{d[k,i]} + s_{ij} + \epsilon_{ikj}$$

où
ω est l'intercept du modèle,
π_k est un l'effet associé à la période k, $k = 1, 2$,
$\tau_{d[i,k]}$ est l'effet du traitement pris à la période k de la séquence i, $i = 1, 2$, $d[i, k] = 1, ..., 4$,
s_{ij} est l'effet aléatoire associé au sujet j de la séquence i. $s_{ij} \sim N(0, \sigma_s^2)$, $j : 1, ..., n_i$,
n_i est le nombre de patients dans la séquence i,
σ_s^2 est la variance interindividuelle et la covariance entre les réponses Y_{ikj} et $Y_{ikj'}$,
ϵ_{ikj} est le terme de l'erreur aléatoire, qui suit une distribution N(0, σ^2).

Dans le cas le simple illustré jusqu'à présent, le design étudie l'effet de deux traitements, les séquences sont au nombre de deux (i = 1,2) et chacune d'entre elles sont composées de deux périodes (k = 1,2). Dans les designs croisés 2 × 2, l'effet carryover est supposé être inclus dans l'effet $\tau_{d[i,k]}$ [30].

Dans un modèle où les patients sont traités comme des effets aléatoires, la variance d'une observation Y_{ikj} est égale à la somme de la variance résiduelle et de la variance interindividuelle et donc, $V(Y_{ikj}) = \sigma^2 + \sigma_s^2$. σ_s^2 correspond également à la covariance entre les réponses Y_{ikj} et $Y_{ik'j}$, où $k \neq k'$. Le modèle induit une matrice de covariance 2×2 pour un patient j d'une étude utilisant un design cross-over où deux traitements sont étudiés. Cette matrice peut s'écrire sous la forme suivante [30] :

$$\mathrm{V}\begin{bmatrix} Y_{i1j} \\ Y_{i2j} \end{bmatrix} = \begin{bmatrix} \sigma^2 + \sigma_s^2 & \sigma_s^2 \\ \sigma_s^2 & \sigma^2 + \sigma_s^2 \end{bmatrix}$$

Les observations peuvent être représentées par une matrice $\mathbf{Y}$, dont la matrice de covariance est $\mathbf{\Sigma}$. L'espérance de la matrice $\mathbf{Y}$ est $E(\mathbf{Y} = \mathbf{X}\beta)$, où $\mathbf{X}$ est la matrice de design et β est le vecteur contenant la valeur des effets fixes de la population [30].

Pour évaluer l'efficacité des traitements étudiés, il sera nécessaire d'estimer leur effet. L'estimation des paramètres des effets fixes se fait par "Restricted Maximum Likelihood (REML)" [33]. Une REML se déroule en deux étapes. Dans un premier temps, les composantes de la

variance σ^2 et σ_s^2 sont estimées à partir d'une fonction de vraisemblance marginale, indépendamment des effets fixes. Ensuite, les paramètres des effets fixes sont estimés en utilisant la méthode des moindres carrés généralisés, qui utilise la matrice de covariance reconstituée avec les estimations faites à la première étape. L'estimation des coefficients est obtenue à l'aide de l'équation suivante [30] :

$$\tilde{\beta} = (\mathbf{X}^T\hat{\mathbf{\Sigma}}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\hat{\mathbf{\Sigma}}^{-1}\mathbf{Y}$$

où
$\tilde{\beta}$ est l'estimateur des effets fixes,
$\hat{\mathbf{\Sigma}} = \mathbf{\Sigma}(\hat{\sigma}^2, \hat{\sigma}_s^2)$ est l'estimation de la matrice de covariance de $\mathbf{Y}$,
$\hat{\sigma}^2$ et $\hat{\sigma}_s^2$ sont les estimateurs de σ^2 et σ_s^2.

Il est possible d'estimer la matrice de covariance $\mathbf{\Phi}$ de $\tilde{\beta}$ avec la formule $(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}$, en utilisant l'estimateur $\hat{\mathbf{\Sigma}}^{-1}$ de $\mathbf{\Sigma}^{-1}$. Cependant, l'introduction d'un effet aléatoire induit une mauvaise estimation de la matrice de covariance $\mathbf{\Sigma}$. Et lorsque la taille de l'échantillon est limitée, ce qui est le cas dans le cadre des maladies rares, $\hat{\mathbf{\Phi}}$ est biaisée pour $\mathrm{V}(\hat{\beta})$ [30].

L'inférence pose également un problème lorsque le nombre d'observations est limité, car cela affecte l'estimation des degrés de liberté. En règle générale, le test de Wald est employé pour évaluer l'hypothèse nulle qui affirme que les valeurs des paramètres à estimer sont nulles. Le test d'hypothèse est évalué avec la statistique de Wald, qui est donné par la formule suivante [30] :

$$W = c^{-1}(\tilde{\beta} - \beta_0)^T\mathbf{C}^T(\mathbf{C}\hat{\mathbf{\Phi}}\mathbf{C}^T)^{-1}\mathbf{C}(\tilde{\beta} - \beta_0)$$

où
$\mathbf{C}$ est un "contrast vector" de dimension $1 \times c$,
$\tilde{\beta}$ est l'estimateur des effets fixes,
β_0 est la valeur des effets fixes sous l'hypothèse nulle,
$\hat{\mathbf{\Phi}}$ est l'estimateur de la matrice de covariance $\mathbf{\Phi}$

Dans le contexte d'un design cross-over 2×2, le vecteur β des paramètres de population à estimer pourrait être $(\mu, \tau_A, \tau_B, \pi_1, \pi_2)$, où μ est l'intercept du modèle, τ_i est l'effet du traitement i (A ou B), et π_k est l'effet de la période k (1 ou 2). Le vecteur de contraste $\mathbf{C}$ permettant de tester la différence d'effets des traitements A et B pourrait être (0, 1, -1, 0, 0).

Il est possible d'utiliser un test F pour tester cette hypothèse car la statistique de Wald W suit une distribution $F_{c,\nu}$, où c représente le nombre de degrés de liberté du numérateur et ν, le nombre de degrés de liberté du dénominateur.

Cependant, l'estimation des degrés de liberté peut se révéler complexe dans les modèles mixtes, surtout lorsque le nombre d'observations est limité. En effet, cela réduit la précision de l'estimation des paramètres, dont celle de la variance des effets aléatoires. Le nombre de degrés de liberté ν sera alors mal estimé, étant donné que son estimation dépend de celle de la variance de ces effets aléatoires, faussant ainsi les résultats du test de Wald. Kenward et al. [32] ont développé une méthode permettant de corriger l'estimateur $\hat{\mathbf{\Phi}}$ et de calculer le

nombre de degré de liberté ν afin d'utiliser la bonne valeur de ν dans le test de Wald. Cette procédure a été implémentée dans le logiciel SAS et est disponible dans la procédure PROC MIXED en utilisant l'option **ddfm = kenwardroger**.

3.1.2.2 Calcul de taille d'échantillon

Avant de déterminer la taille d'un échantillon nécessaire à un design cross-over, plusieurs décisions doivent être prises et certains paramètres doivent être fixés. Dans un premier temps, l'objectif de l'étude doit être défini. Dans le cadre de ce travail, le but sera de mettre en évidence la supériorité d'un traitement expérimental par rapport à un traitement contrôle. Ensuite, il faudra définir l'effet à mettre en évidence ainsi que sa taille. L'erreur de type I α et la puissance $1 - \beta$ de l'étude doivent aussi être fixés [30].

Lorsque l'output mesuré est continu, et qu'il suit une distribution approximativement normale, il est possible d'utiliser la formule de calcul de taille d'échantillon total 2.4.2 d'un ECR parallèle. Toutefois, un élément de cette équation se modifie, il s'agit du terme représentant la variance dans la formule. La variance totale au numérateur sera remplacée par la variance intraindividuelle, qui est plus faible que la variance totale. La diminution de la variance mène à une augmentation la puissance de l'essai pour une taille d'échantillon donnée [15]. Comme la taille de l'échantillon est proportionnelle à la variance, cela entraînera une diminution de la taille d'échantillon nécessaire. Pour une même puissance, l'échantillon nécessaire sera donc plus petit que pour un ECR parallèle. La formule utilisée sera donc la suivante [31] :

$$n = \frac{2(Z_{1-\beta} + Z_{1-\frac{\alpha}{2}})\sigma_w^2}{\Delta^2} \tag{3.1.1}$$

où
$Z_{1-\beta}$ est le percentile d'ordre $1 - \beta$ de la loi normale centrée réduite,
$Z_{1-\frac{\alpha}{2}}$ est le percentile d'ordre $1 - \frac{\alpha}{2}$ de la loi normale centrée réduite,
σ_w^2 est le paramètre de la variance intraindividuelle,
Δ est la différence à mettre en évidence entre deux groupes pour que l'effet soit cliniquement important.

Il est supposé qu'un nombre égale de patients $\frac{n}{2}$ sera assigné à chaque séquence [30].

Afin d'illustrer la différence entre les tailles d'échantillon nécessaires pour un ECR parallèle et un design cross-over, prenons en compte les résultats d'une étude menée par Riddlesworth et al. [42], où la glycémie était l'output évalué. Dans cette étude, la variance totale des observations était de 16^2, tandis que la variance intraindividuelle était de 10^2.

Le quotient des formules de calcul de la taille d'échantillon pour un ECR parallèle (équation 2.4.2) et un design cross-over (équation 3.1.1) est égal à la formule suivante :

$$\frac{\sigma^2}{\sigma_w^2} = \frac{n_{ECR}}{n_{CO}}$$

.

où
σ^2 est la variance résiduelle,
n_{ECR} est la taille d'échantillon nécessaire pour un ECR,
σ_w^2 est la variance intraindividuelle,
n_{CO} est la taille d'échantillon nécessaire pour un design cross-over.

Dans cet exemple spécifique, la taille d'échantillon requise pour un ECR parallèle est donc $\frac{16^2}{10^2} = 2.56$ fois plus grande que celle nécessaire pour un design cross-over. Ainsi, dans cet exemple spécifique, la taille d'échantillon requise pour un ECR parallèle serait 2.56 fois plus grande que celle nécessaire pour un design cross-over.

3.2 Design "n-of-1"

3.2.1 Généralité

Un design "n-of-1" est un design où le principe du design cross-over est appliqué plusieurs fois sur un même patient. Son déroulement est illustré à la figure 3.2.1 ci-dessous [1].

Un design n-of-1 est divisé en plusieurs cycles. Le nombre de cycles est choisi par les professionnels de la santé et le patient. Le but est de maximiser le nombre de périodes tout en respectant ce que le patient peut accepter et supporter. Ce nombre varie entre deux et sept en général. Une séquence de traitements est d'abord assignée par randomisation à chaque période, pour chaque patient. Dans le cas simple que nous envisageons, deux traitements, A et B, sont étudiés. Chaque cycle est composé de deux périodes. Le patient commence soit un cycle avec le traitement A, en première période, pour ensuite prendre le traitement B à la seconde période (séquence AB). Ce cycle peut aussi commencer par le traitement B pour ensuite finir avec le traitement A (séquence BA). Chaque traitement est pris durant un même intervalle de temps. Une période de washout est souvent nécessaire entre la prise de deux traitements différents afin d'éviter la présence d'effets carryover [1].

Ce design comporte de nombreux avantages dans le contexte des maladies rares. Premièrement, il permet un meilleur recrutement car chaque patient sait qu'il recevra le traitement expérimental. Le design n-of-1 permet aussi une réduction de la taille de l'échantillon, pour les mêmes raisons que le design cross-over. Augmenter le nombre de cycles contribue également à cette réduction, car la taille de l'échantillon est inversement proportionnelle au nombre de cycles de l'étude [1].

Cependant, l'application des designs n-of-1 est limitée aux maladies incurables, chroniques et stables. Les traitements étudiés ne doivent pas avoir d'effet irréversible et la réponse aux traitements doit être rapidement observée. Ces caractéristiques correspondent à celles des maladies rares, ce qui rend la réalisation de ces essais possibles dans ce contexte [1].

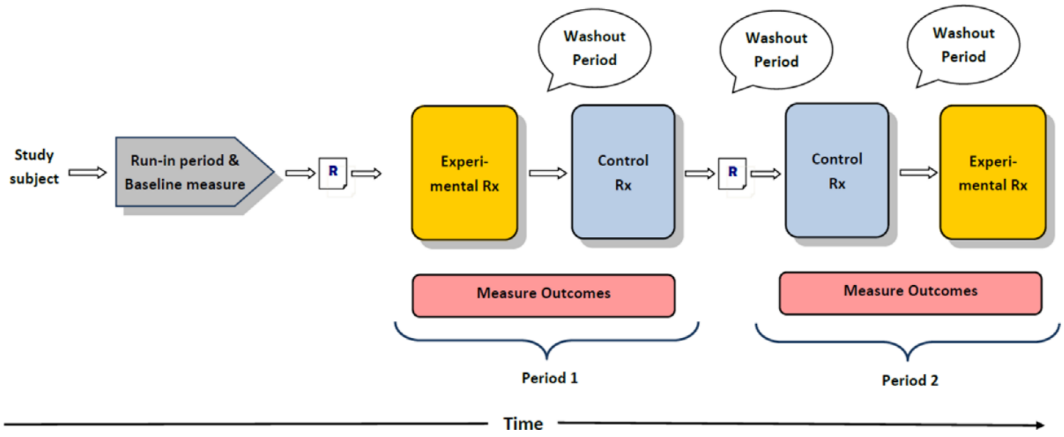

FIGURE 3.2.1 – Schéma représentant le déroulement d'un design n-of-1 [1].

Ensuite, les designs n-of-1 peuvent être plus ou moins longs, en fonction du nombre de périodes de l'étude. La longue durée d'une étude peut mener à des taux d'abandon plus élevés. C'est la raison pour laquelle ce point doit bien être discuté avec les patients [1].

De plus, la réalisation d'une méta-analyse, qui vise à rassembler et à analyser les résultats de plusieurs études indépendantes, n'est pas simple. Rassembler suffisamment d'essais n-of-1 pour une méta-analyse peut s'avérer complexe, car chaque essai implique généralement un seul patient, ce qui réduit la taille totale de l'échantillon. Il peut donc être nécessaire d'avoir un grand nombre d'études pour réaliser une méta-analyse valide. De plus, les méta-analyses classiques supposent aussi une homogénéité des effets entre les études incluses dans l'analyse, ce qui n'est pas forcément le cas entre plusieurs études n-of-1. Les résultats de ce type de design peuvent être très différents, en fonction de leur conception et des protocoles utilisés dans le design n-of-1, qui peuvent fortement différer d'une étude à l'autre. Cela peut parfois rendre les résultats de ces études incomparables, lorsque celles-ci diffèrent trop les unes des autres [55].

3.2.2 Méthodologie

3.2.2.1 Analyse statistique

Il existe différentes méthodes permettant d'analyser les résultats d'un design n-of-1 et de comparer l'effet des traitements étudiés. Dans cette section, seule la comparaison d'un traitement expérimental à un traitement contrôle sera abordée, mais il serait aussi possible d'étudier l'effet de plusieurs traitements. L'output mesuré peut être de différents types : continu, binaire, de comptage, ... [14] Seul le cas des outputs continus sera abordé dans le cadre de ce mémoire.

Une première méthode pouvant être utilisée pour comparer l'efficacité de deux traitements étudiés dans un design n-of-1 est le test t pairé. Cette méthode est applicable lorsque les données sont appariées et que la différence entre ces données appariées suit une distribution normale. La comparaison est faite entre les résultats obtenus avec le traitement contrôle et le traitement expérimental. Imaginons qu'un design n-of-1 est composé de 3 cycles et que 10

patients ont intégré l'étude, 30 paires de données sont obtenus à la fin de l'étude. Le test t pairé est ensuite utilisé pour analyser les 30 paires de données dans leur ensemble, sans prendre en compte les différences qui peuvent exister entre les individus [14].

Lorsque les données sont normalement distribuées, le test t pairé est généralement recommandé. Une étude menée par Chen et al. [14] a comparé la puissance et l'erreur de type I de différentes méthodes telles que le test t pairé et différents modèles mixtes. Ces comparaisons ont été effectuées avec des simulations de Monte-Carlo. Cette étude de simulation a montré que la puissance et l'erreur de type I obtenus avec le test t pairé étaient proches du α et de la puissance $1 - \beta$ fixés initialement. Cependant, tout comme pour les designs cross-over, ce type de test ne tient pas compte de la structure de l'étude, qui est constituée de séquences, de périodes et de cycles. Il ignore donc les effets que ceux-ci pourraient induire.

Une seconde méthode consiste à modéliser la réponse au traitement d'une étude à l'aide d'un modèle mixte, permettant d'intégrer des effets fixes ainsi que des effets aléatoires au modèle. Supposons une nouvelle fois que le design n-of-1 est composé de 3 cycles et que 10 patients ont intégré l'étude. Chaque patient peut être affecté à la séquence 1 ou à la séquence 2. Le groupe assigné à la séquence 1 commence l'étude avec le traitement expérimental et prend ensuite le traitement contrôle. Le sujet assigné à la séquence 2 prend d'abord le traitement contrôle et ensuite, le traitement expérimental. Au début de chaque cycle, chaque patient est assigné à une séquence de traitement par randomisation, avec une probabilité égale d'être assigné à l'une des séquences [14].

La réponse au traitement pour un cycle donné dans ce design n-of-1 peut être exprimée comme suit [14] :

$$y_{kj} = \omega + \alpha * x_{kj} + \tau_k + \lambda_1 Z_1 + \lambda_2 Z_2 + \gamma_j + \epsilon_{kj}$$

où

y_{kj} est la réponse au traitement du patient k à la période j,

ω est l'intercept du modèle et représente l'effet moyen du traitement de référence (A ou B),

α est la moyenne de la différence des réponses aux traitements, par rapport au traitement de référence,

x_{kj} est une variable indicatrice pour l'effet du traitement. $x_{kj} = 1$ si le sujet j reçoit le traitement de la séquence 1 lors de la période k, et $x_{kj} = 0$ si le sujet j reçoit le traitement de la séquence 2 lors de la période k,

τ_k est l'effet induit par la période k,

λ_1 et λ_2 sont les effets carryover observés respectivement à la séquence 1 et à la séquence 2,

Z_1 et Z_2 sont des variables indicatrices pour les effets de carryover. $Z_1 = 1$ et $Z_2 = 0$ si le sujet j est assigné à la séquence 1 lors de la période k, tandis que $Z_1 = 0$ et $Z_2 = 1$ si le sujet j est assigné à la séquence 2 lors de la période k,

γ_j est l'effet aléatoire du patient j et $\gamma \sim iid\ N(0, \sigma_j^2)$,

ϵ_{kj} est l'erreur aléatoire du patient j à la période k. Elle est supposée indépendante de γ.

L'utilisation d'un tel modèle requiert de valider les hypothèses permettant l'utilisation d'un modèle mixte, déjà citées ci-dessus, dans la section traitant de l'analyse d'un design cross-over, à la section 3.1.2.1.

Lorsqu'un effet carryover est présent, l'utilisation d'un modèle mixte est plus adaptée que le test t pairé. En effet, ce modèle permet de distinguer l'effet du traitement de l'effet carryover, évitant que celui-ci ne biaise l'estimation de la réponse au traitement. De plus, ce type de modèle permet également de tenir compte de l'effet que pourrait induire la période à laquelle le traitement est pris [14].

Une autre méthode d'analyse des données d'un design n-of-1 consiste à combiner différentes études n-of-1 et d'estimer l'effet d'un traitement. Cela est possible en utilisant une "Individual Participant Data (IPD)" méta-analyse. C'est une méthode adaptée aux études où la variabilité entre les études est importante, ce qui est le cas des designs n-of-1, comme expliqué ci-dessus [59].

Une IPD méta-analyse utilise des modèles linéaires mixtes pour combiner les résultats de différentes études et estimer la valeur des paramètres du modèle. Les lignes qui suivent décrivent un modèle établi par Zucker et al. [59].

Dans un premier temps, un modèle ainsi que les paramètres inclus dans ce dernier sont choisis. Il faudra ensuite mentionner si les effets inclus dans le modèle sont des effets fixes ou des effets aléatoires. Les hypothèses concernant la distribution des composantes de la variance doivent aussi être définies. Les notations utilisées pour cette IDP méta-analyse sont les suivantes.

$\boldsymbol{Y}_j$ est le vecteur reprenant les résultats de la variable réponse d'un même patient et Y_{kj} est l'observation du patient j pour le traitement pris à la période k. On suppose que ces données suivent une distribution normale multivariée de moyenne $\boldsymbol{\mu}_j$ et de matrice covariance $\boldsymbol{\Sigma}_j$.

Le but de Zucker et al. [59] est de comparer l'efficacité de deux traitements. Les traitements étudiés sont les suivants : un traitement contrôle, que nous appellerons traitement A et un traitement expérimental, que nous appellerons traitement B. Dans ce contexte, il est possible de considérer la moyenne de la réponse pour chaque participant j et chaque période k comme étant ω_j pour le traitement A. αX_{kj} est la différence entre l'effet du traitement A et l'effet du traitement B pour ce même patient.

La moyenne μ_{kj} peut être exprimée selon l'équation suivante : $\mu_{kj} = \omega_j + \alpha_j X_{kj}$, où ω_j représente la réponse moyenne au traitement A pour le patient j et α_j la différence entre l'effet du traitement A et du traitement B pour le participant j. La variable X_{kj} est utilisée pour représenter la condition testée. $X_{kj} = 0$ lorsque le traitement pris par le patient j à la période k est le traitement A alors que $X_{kj} = 1$ lorsqu'il s'agit du traitement B.

Afin de tenir compte d'autres covariables pouvant influencer la réponse au traitement, $\boldsymbol{\mu}_j$ peut être écrit sous la forme $\boldsymbol{\mu}_j = \mathbf{X}_j \boldsymbol{\theta}_j$.

Supposons que $\boldsymbol{\theta}_j = (\omega_j, \alpha_j)$. $\boldsymbol{\theta}_j$ peut être le même pour chaque patient. Les effets du modèle seront alors considérés comme des effets fixes. C'est ce qui est recommandé quand les protocoles des différentes études n-of-1 sont communs.

Il est aussi possible de voir le vecteur de paramètres comme étant $\boldsymbol{\theta}_j = \boldsymbol{\theta} + \boldsymbol{\delta}_j$ où $\boldsymbol{\theta}$ est un effet fixe, $\boldsymbol{\delta}_j \sim N(0, \mathbf{D})$ et $\boldsymbol{D}$ est la matrice de covariance de $\boldsymbol{\delta}_j$. Le modèle prendra alors la forme suivante, $\mathbf{Y}_j = X_j\boldsymbol{\theta}_j + \boldsymbol{\epsilon}_j$ où $\boldsymbol{\epsilon}_j \sim N(\mathbf{0}, \Sigma_j)$ et $\boldsymbol{\theta}_j \sim N(\boldsymbol{\theta}, \mathbf{D})$.

Dans une situation simple, où la réponse d'un patient j est évaluée, la réponse au traitement sera modélisée de la façon suivante.

$$Y_{kj} = \omega + \alpha X_{kj} + \epsilon_{kj}$$

où
Y_{kj} est la mesure du patient j faite à la période k,
ω est l'effet du traitement A,
α est la différence entre l'effet du traitement A et du traitement B,
X_{kj} est une variable indicatrice. $X_{kj} = 0$ lorsque le patient j prend le traitement A à la période k et $X_{kj} = 1$ lorsqu'il prend le traitement B à cette même période,
ϵ_{kj}, l'erreur aléatoire à la période k pour le patient j.

La structure de la matrice de covariance $\boldsymbol{\Sigma}_j$ peut prendre différentes formes. Si le nombre d'observations est limité et que le nombre total de données par patient est faible, ou que peu d'études ont été menées, il est envisageable que tous les individus aient la même matrice $\boldsymbol{\Sigma}$. $\boldsymbol{\Sigma}$ peut être autorégressive de premier ordre. Cela veut dire que chaque mesure a la même variance et que les mesures séparées de k visites ont une corrélation de ρ^k. Une structure où les variances sont égales pour tous les patients mais où les observations de ces patients sont non corrélées peut-être envisagée ($\boldsymbol{\Sigma}_j = \sigma_j^2\mathbf{I}$).

La composante de la variance devra donc être étudiée afin de sélectionner celle qui donnera le modèle le plus adapté. Il est possible de faire ce choix en se basant sur l'Critère d'Information d'Akaike (AIC) ou le Critère d'Information Bayésien (BIC) [52]. Ceux-ci peuvent être utilisés afin de comparer des modèles avec différentes structures de matrice de covariance.

3.2.2.2 Calcul de taille d'échantillon

Afin de déterminer la taille d'échantillon nécessaire à un essai clinique utilisant un design n-of-1, il est nécessaire de définir certains critères. Le but de l'étude sera d'inclure un certain nombre de patients suivant un même protocole. Tout d'abord, l'objectif de l'étude ainsi que l'effet à mettre en évidence doivent être clairement déterminés. Ensuite, il est important de choisir le taux d'erreur de type I et la puissance de l'étude, en fonction des objectifs spécifiques de celle-ci et du niveau de risque acceptable. Nous aborderons uniquement le cas des études à deux bras de traitement où l'output mesuré est continu et suit approximativement une distribution normale.

La formule classique de calcul de taille de l'échantillon, telle que celle décrite pour un ECR parallèle, ne peut pas être utilisée pour un design n-of-1. En effet, celle-ci surestime la taille de l'échantillon nécessaire. En voici les raisons [46].

En statistique, le nombre de degrés de liberté reflète le nombre de valeurs dans l'échantillon qui sont libres de varier. En général, l'estimateur de la variance utilise un nombre de degrés de liberté de $n-1$, où n est le nombre d'observations. Cependant, dans le cas des designs n-of-1, celui-ci est égal à $n(k-1)$. Le nombre de degrés de liberté est déterminé par le nombre d'observations n, mais aussi par le nombre de cycles k. Cette particularité conduit souvent à une sous-estimation du nombre de degrés de liberté. Or, l'estimateur de la variance est inversement proportionnel au nombre de degrés de liberté. Une sous-estimation de ce nombre mène donc à une surestimation de la variance. Cette surestimation de la variance impacte la taille de l'échantillon nécessaire pour atteindre la puissance statistique définie, car la taille de l'échantillon est proportionnelle à la variance. Dans le contexte des maladies rares où le recrutement limité, cette surestimation est problématique [46].

Une stratégie permettant de calculer la taille de l'échantillon minimum pour un design n-of-1 a été proposée par Senn [46]. Celle-ci se déroule en deux temps. Cela permettra d'obtenir une estimation des paramètres dans un premier, pour ensuite les utiliser dans une étude de simulation permettant de calculer la taille de l'échantillon nécessaire.

Premièrement, une taille d'échantillon approximative est calculée en utilisant l'approximation normale. Cette approche repose sur l'hypothèse que la réponse aux traitements est une variable continue qui suit une distribution normale. La formule suivante est alors utilisée :

$$n = \frac{2(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta})^2\sigma^2}{k\Delta^2} \tag{3.2.1}$$

où
$Z_{1-\beta}$ est le percentile d'ordre β de la loi normale réduite centrée,
$Z_{1-\frac{\alpha}{2}}$ est le percentile d'ordre $1-\frac{\alpha}{2}$ de la loi normale réduite centrée,
σ^2 est le paramètre de la variance résiduelle,
k est le nombre de périodes,
Δ est la différence à mettre en évidence entre deux groupes pour que l'effet soit cliniquement important.

La formule 3.2.1 montre que plus le nombre de périodes est élevé, plus la taille de l'échantillon sera faible. Cependant, il faudra tenir compte des limites des patients. En effet, augmenter le nombre de périodes rallonge le temps de l'étude, ce qui peut s'avérer contraignant pour eux. Il sera donc important de parvenir à un équilibre optimal entre un nombre de périodes suffisamment élevé pour minimiser la taille de l'échantillon, et la prise en compte de ce qui est acceptable pour les participants.

En utilisant cette formule, il est possible que la puissance désirée ne soit pas atteinte. Cela peut être dû à une mauvaise estimation des degrés de liberté, comme mentionnée ci-dessus. Une mauvaise estimation d'un des paramètres intervenant dans la formule 3.2.1 peut aussi

mener à ce problème. Il est donc nécessaire d'évaluer la puissance du design à travers une étude de simulation. En spécifiant l'effet à démontrer et la variance associée à cet effet, une série d'essais peut être simulée, chaque simulation utilisant une taille d'échantillon différente, proche de celle estimée initialement. La puissance est ensuite calculée en déterminant le taux d'essais qui ont mis en évidence une différence entre les deux traitements étudiés. La puissance calculée est évaluée par rapport à la valeur prédéfinie. Si le niveau fixé n'est pas atteint, un ajustement de la taille de l'échantillon est nécessaire. Cela peut impliquer une modification du nombre de patients à inclure dans l'essai, ou du nombre de périodes k du design. Cela permettra d'optimiser la taille de l'échantillon en fonction des objectifs de l'étude.

3.3 Designs d'essais cliniques exploitant des données existantes et celles obtenues au cours d'un essai clinique

Lors d'un essai clinique évaluant un traitement pour une maladie rare, l'accumulation d'informations suffisantes pour répondre à la question de recherche peut s'avérer complexe. Cela est principalement dû à la faible prévalence de la maladie, qui limite le nombre d'observations qui pourrait être recueilli. De plus, comme mentionné dans l'introduction, le recrutement dans ces essais cliniques peut s'avérer plus difficile [1].

Une stratégie permettant de contrecarrer ce problème consiste à utiliser les données disponibles avant la fin d'un essai clinique en cours. Ces données peuvent provenir d'études antérieures, disponibles avant même que l'étude en cours n'ait commencé. Cela permettra de réduire la taille de l'échantillon nécessaire. Elles peuvent également provenir des étapes précédentes de l'essai clinique en cours, permettant d'introduire des analyses intermédiaires des données accumulées. Cela pourrait augmenter la probabilité d'assigner aux patients le traitement le plus efficace, facilitant ainsi le recrutement. Cela peut aussi permettre de diminuer le temps de l'étude et la taille de l'échantillon si des conclusions assez robustes peuvent être tirées de ces analyses intermédiaires [1].

Trois exemples de designs exploitant des données disponibles avant et tout au long de l'étude seront présentés dans les sections ci-dessous. Il s'agit des designs adaptatifs, des essais cliniques utilisant des données historiques et des essais cliniques utilisant des concepts issus de la statistique bayésienne.

3.3.1 Designs adaptatifs

3.3.1.1 Généralité

Un essai clinique adaptatif peut être défini comme un essai clinique durant lequel il est possible de procéder à des adaptations du design de l'étude sans compromettre la validité de celle-ci [16].

Il existe trois grands types d'adaptation : l'adaptation prospective, la "concurrent adaptation" et l'adaptation rétrospective [16].

L'adaptation prospective permet de modifier la conception du design. Cela peut concerner le ratio d'allocation aux traitements, le calcul de la taille de l'échantillon, mais aussi la durée de l'étude. En effet, l'étude peut prendre fin prématurément si les traitements mettent en danger la santé des patients de l'essai ou provoquent l'apparition d'effets indésirables. L'essai peut aussi prendre fin pour des raisons de futilité, ce qui signifie que les traitements étudiés n'ont pas su montrer leur efficacité. Cela permet d'éviter la poursuite d'un essai dont les premiers résultats démontrent une absence de bénéfice pour le patient. Il est aussi possible d'écourter un essai pour des raisons d'efficacité, lorsqu'un traitement montre sa supériorité significative de façon précoce, par exemple [16].

Les "concurrent adaptations" sont des adaptations faites au cours de l'essai clinique. Ces adaptations peuvent impliquer les critères d'inclusion et d'exclusion, qui déterminent l'éligibilité des patients à intégrer l'étude. De plus, elles peuvent concerner d'autres aspects de l'essai clinique tels que la posologie du traitement, la durée du traitement, la fréquence de suivi, ... Ces ajustements sont généralement effectués afin d'améliorer l'efficacité de l'essai, d'assurer la sécurité des patients ou de répondre à des problèmes imprévus qui peuvent survenir durant l'essai [16].

Les adaptations rétrospectives sont les modifications apportées au plan d'analyse statistique avant la collecte et l'analyse complètes des données de l'essai. Ces adaptations sont généralement mises en place en réponse à des observations intermédiaires, à des progrès dans la compréhension statistique ou clinique, ... Cependant, il est important de s'assurer que ces modifications n'impactent pas la validité de l'essai [16].

L'avantage principal des designs adaptatifs est de pouvoir ajuster certains paramètres en fonction du déroulement de l'essai. Cela permet à l'essai d'être plus éthique. En effet, le patient aura plus de chances d'être dans le groupe du traitement le plus efficace lorsque le ratio d'allocation est adapté en fonction des résultats de l'étude, par exemple. Dans le contexte des maladies rares, ces adaptations pourraient contribuer à une amélioration du recrutement et à rendre disponible plus rapidement de nouvelles solutions thérapeutiques, si l'essai démontre la supériorité d'un traitement avant son terme [1].

Il faudra néanmoins surveiller la puissance statistique et l'erreur de type I, qui pourrait être affectées par les adaptations effectuées au cours de l'essai. Par exemple, une réestimation de la taille de l'échantillon pourrait influencer la puissance de l'essai. De même, le nombre de tests effectués pourrait impacter l'erreur de type I [16].

Compte tenu du nombre de possibilités de designs d'essais adaptatifs, seuls certains ont été sélectionnés pour être présentés dans ce travail. Les designs choisis sont particulièrement adaptés aux contextes des maladies rares et sont largement reconnus et repris dans la littérature scientifique.

Response Adaptative Randomization Design (RARD)

Un RARD permet d'ajuster le ratio d'allocation entre les groupes de l'étude, en fonction des résultats obtenus au cours de celle-ci. Cette adaptation est une adaptation du type prospective [1].

Dans le cadre des maladies rares, l'ajustement du ratio l'allocation aux traitements est intéressant. En effet, le nombre de patients attribués au traitement le plus efficace sera plus élevé, permettant à l'étude d'être plus éthique tout en améliorant le recrutement au sein de l'étude [1].

L'utilisation d'un RARD nécessite que les réponses au traitement des patients soient disponibles avant la randomisation du patient suivant. Cela est souvent le cas dans les essais cliniques des maladies rares. Le recrutement est lent et les réponses aux traitements sont fréquemment disponibles avant que le patient suivant ne soit assigné à un bras de l'étude. Cela s'explique par le fait que les outputs sélectionnés sont souvent des critères à court terme, dont la réponse peut être évaluée rapidement [1].

La figure 3.3.1 illustre le déroulement d'un RARD, où le ratio d'allocation est modifié suivant le principe du "Randomized Play-the-Winner" (RPW). Ce type de design est particulièrement adapté pour les essais cliniques où deux traitements sont comparés et où l'output étudié est binaire . C'est ce cas précis qui sera étudié dans le cadre de ce travail. Dans d'essai clinique, certains paramètres doivent être fixés avant le début de l'essai [13].

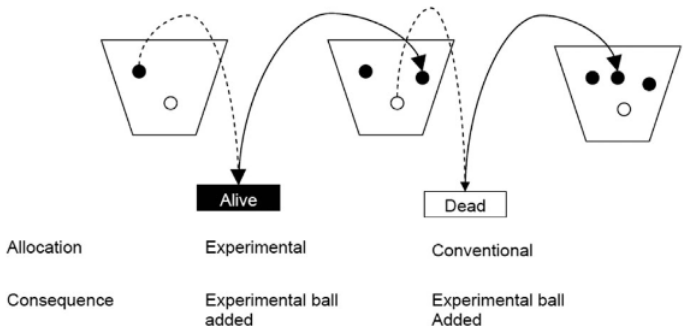

FIGURE 3.3.1 – RARD utilisant le principe du RPW. Dans ce schéma, les boules noires et blanches illustrent deux traitements distincts. La proportion de boules de chaque couleur présente dans chaque urne indique la probabilité d'assignation de ces traitements aux futurs patients. Les évènements observés au cours de l'essai ("Alive" et "Dead") influenceront le ratio d'allocation des traitements au cours de l'étude. Ces ajustements sont illustrés par les modifications dans la composition des urnes au cours du temps [1].

Tout d'abord, il est nécessaire de définir le nombre d'analyses intermédiaires après lesquelles la probabilité d'assignation des traitements étudiés sera ajustée.

Ensuite, chaque urne de la figure 3.3.1 indique la probabilité d'assignation de chaque traitement. Le nombre initial de boules représentant chaque traitement doit être déterminé. Le nombre de boules initiales est souvent noté a_0 et b_0 lorsque deux traitements sont étudiés [13]. Dans la figure 3.3.1, ces paramètres sont fixés à $a_0 = 1$, qui est le nombre de Boule Noire (BN) dans la première urne, et à $b_0 = 1$, qui est le nombre de Boule Blanche (BB) dans cette même urne. Cela signifie que le traitement BN a une probabilité de 0.5 d'être assigné aux patients

au début de l'étude, tout comme le traitement BB. Lorsqu'un patient entre dans l'étude, une boule est tirée dans l'urne contenant une BN et une BB. Si une BN est tirée, le patient sera traité avec le traitement BN. Si une BB est tirée, alors, ce dernier sera traité avec le traitement BB [1].

Chaque urne de la figure 3.3.1 représente un intervalle de temps de l'étude, avant qu'une nouvelle adaptation ne soit faite. Les critères qui influenceront le ratio d'allocation des traitements au cours de l'étude doivent être définis. Ceux-ci comprennent les effets "bénéfiques" et "négatifs" des traitements, ainsi que la fréquence à laquelle ils sont observés. En fonction des évènements, les ratios d'allocation sont adaptés en modifiant la composition des urnes. Lorsque le traitement BN montre un effet bénéfique ou que le traitement BB montre un effet négatif, un certain nombre a_1 de BN seront ajoutées dans l'urne, et inversement pour un certain nombre b_1 de BB [13].

Dans l'exemple illustré à la figure 3.3.1, les paramètres a_1 et b_1 ont été fixés à 1. Lorsque le traitement BN montre des effets bénéfiques chez les patients qui le prennent, une BN est ajoutée dans l'urne, ce qui signifie que la probabilité qu'il soit assigné au prochain patient intégrant l'étude augmente à $\frac{2}{3}$. La probabilité que le traitement BB soit assigné diminue alors à $\frac{1}{3}$. Par la suite, le traitement BB entraîne des effets négatifs sur un patient. La probabilité qu'il soit assigné au prochain patient diminue à $\frac{1}{4}$ alors que la probabilité que le traitement BN soit assigné augmente à $\frac{3}{4}$ [1].

Un RARD utilisant le principe du RPW permet d'optimiser la proportion de patients attribués au traitement le plus efficace au moment de leur intégration dans l'étude. Une autre stratégie qui pourrait être utilisée est celle du "drop-the-loser", qui vise à réduire la probabilité d'assignation du traitement le moins efficace aux futurs patients de l'étude [16].

Chang et al. [13] ont proposé une autre façon de déterminer la probabilité d'attribution d'un traitement dans un RARD. Ils utilisent la formule suivante, qui décrit la façon dont la probabilité d'allocation d'un traitement est ajustée en fonction de la valeur des outcomes étudiés durant un essai clinique :

$$Pr(trt = i) = f(\hat{\mathbf{u}})$$

où
$Pr(trt = i)$ est la probabilité qu'un patient soit alloué au groupe recevant le traitement i,
$\hat{\mathbf{u}}$ est le vecteur contenant la valeur des outcomes.

La fonction f détermine la façon dont la valeur des outcomes influence la probabilité d'attribution d'un traitement. Elle peut être représentée de la façon suivante :

$$Pr(trt = i) \propto a_{0i} + b\hat{u}_i^m$$

où
a_{0i} est une constante définie pour chaque traitement i, et représente la probabilité initiale d'attribution du traitement i,

b est une constante ajustant la sensibilité de la probabilité d'attribution aux valeurs de l'outcome,

$\hat{u}_i$ est la moyenne de la valeur des outcomes observées pour le traitement i,

m est un paramètre qui permet d'ajuster l'impact de la valeur des outcome sur la probabilité d'attribution.

Prenons un exemple simple pour illustrer cette méthode. Supposons qu'un essai utilisant un RARD étudie deux traitements, A et B. Au début de l'essai, chaque traitement aura une probabilité de $\frac{1}{2}$ d'être assigné à un patient. $a_{0A} = a_{0B} = 1$ dans ce cas-ci. Supposons que $b = 1$. La probabilité d'attribution des traitements sera directement proportionnelle à la moyenne des outcomes observée pour chaque traitement. m est aussi fixé à $m = 1$, ce qui signifie que l'impact de la valeur des outcomes sur la probabilité d'attribution du traitement est linéaire.

Imaginons qu'au cours de l'essai, le traitement A semble être plus efficace que le traitement B, avec des outcomes ayant respectivement une moyenne de $\hat{u}_A = 0.8$ et $\hat{u}_B = 0.6$. En utilisant la formule proposée par Chang et al. [13], nous obtenons les équations suivantes :

$$Pr(trt = A) \propto 1 + 0.8 = 1.8$$
$$Pr(trt = B) \propto 1 + 0.6 = 1.6$$

Les probabilités d'assignation sont ensuite calculées en normalisant les valeurs de 1.8 et 1.6 par leur somme (3.4). Celle du traitement A ($\frac{9}{17}$) sera légèrement supérieure à celle du traitement B ($\frac{8}{17}$). Ces probabilités continueront à être ajustées tout au long de l'étude, afin de favoriser la probabilité d'assignation du traitement qui semble être le plus efficace.

Lors de la mise en oeuvre d'un RARD, quel que soit le principe utilisé pour adapter les ratios d'allocation, il est essentiel de veiller à ce que le déséquilibre dans l'attribution des traitements ne compromette la qualité de l'étude. Il faudra particulièrement s'assurer que ces adaptations n'entraînent pas une sous-représentation d'un des groupes de traitement. Cela compromettrait la robustesse des conclusions de l'analyse finale.

Considérons une nouvelle fois le scénario illustré à la figure 3.3.1. Le ratio de BB est de $\frac{1}{4}$ tandis que celui de BN est de $\frac{3}{4}$ après que les ratios d'allocation aient été ajustés deux fois. Si les évènements ultérieurs de l'étude devaient encore diminuer la probabilité d'assignation du traitement BB, le nombre d'observations correspondant à ce traitement pourrait se réduire drastiquement. Dans des cas extrêmes, le nombre d'observations pour le groupe de traitement le moins efficace pourrait être réduit à un. Cela compromettrait la validité de l'analyse des résultats, étant donné le peu de données disponibles pour le groupe de traitement le moins efficace [1].

Pour éviter cela, il est possible de mettre en place des seuils d'allocations minimaux, afin d'éviter les probabilités d'assignation extrêmes. Il est aussi possible de mettre en place des périodes de blocage pendant lesquelles les ratios l'allocation sont fixes, empêchant ceux-ci de

devenir trop extrêmes. Ces mesures permettent de garantir que tous les groupes de traitement soient suffisamment représentés dans l'étude.

Ranking and Selection Design (RSD)

Lorsque plusieurs traitements sont étudiés et comparés, un "Ranking and Selection Design (RSD)" peut être envisagé. Les RSD les plus simples se divisent en deux étapes, comme l'illustre la figure 3.3.2 [1].

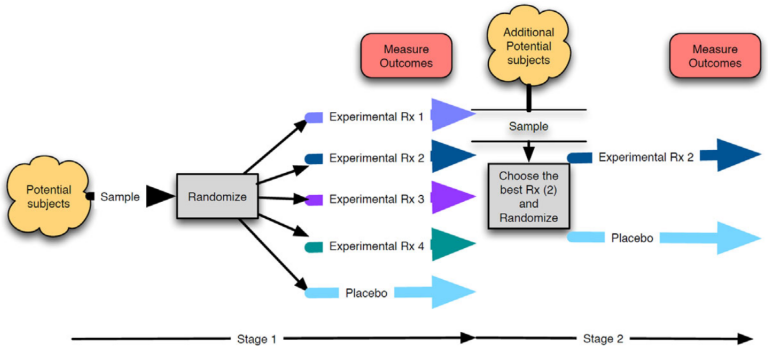

FIGURE 3.3.2 – Schéma représentant le déroulement d'un RSD [1].

Dans un premier temps, les patients sont randomisés au sein des différents bras de traitements étudiés. Chaque traitement possède la même probabilité d'être attribué. Un groupe contrôle peut être inclus dans l'étude, comme le montre la figure 3.3.2, où un traitement placebo permet de contrôler l'essai. A la fin de la première étape, deux options peuvent être envisagées [1].

Soit un plan "drop-the-looser" a été choisi. Les traitements inférieurs seront alors éliminés. Soit les traitements supérieurs seront gardés pour la suite de l'étude, ce qui correspond à un plan "pick-the-winner". Pour chacun de ces plans, des critères sont définis au préalable afin de déterminer les traitements supérieurs ou inférieurs, et nombre de traitement à sélectionner/ éliminer [1]. Ce sont donc des adaptations prospectives qui sont effectuées dans ce design. Il est également nécessaire de définir les critères de sélection des traitements avant le début de l'étude [16]. Le choix des critères d'arrêt sera abordé plus en détail dans la section 3.3.1.2.

Durant la seconde étape d'un RSD, l'efficacité du ou des traitements sélectionnés est étudiée. Le design choisi pour cette partie de l'étude peut être un design adaptatif, un ECR parallèle (comme le montre la figure 3.3.2), ... Cette seconde étape est importante et doit être menée afin d'obtenir des conclusions significatives. En effet, la première étape n'est pas suffisante pour tirer des conclusions définitives. Il est possible que les traitements répondant aux critères de la première étape ne répondent pas aux critères de la seconde étape. Il faudra donc veiller à ce que les traitements répondent aux critères établis tout au long des différentes étapes de l'étude. Cela permettra d'obtenir des conclusions finales valides [1].

L'utilisation d'un RSD dans le contexte des maladies rares présente plusieurs avantages, dont la possibilité de comparer plusieurs options thérapeutiques avec une taille l'échantillon limitée. De plus, en sélectionnant les options thérapeutiques les plus efficaces pour la phase suivante, il est possible d'améliorer le recrutement des patients au sein de l'étude.

Cependant, il faudra veiller à contrôler le taux d'erreur de type I. Ce point-là sera approfondi à la section 3.3.1.2.

Design séquentiel

Le dernier type de design adaptatif qui sera présenté dans cette section est le design séquentiel. Celui-ci est très similaire à un ECR parallèle classique. La principale distinction est qu'il inclut des analyses intermédiaires des résultats, comme l'illustre la figure 3.3.3 ci-dessous. Ces adaptations sont des adaptations prospectives.

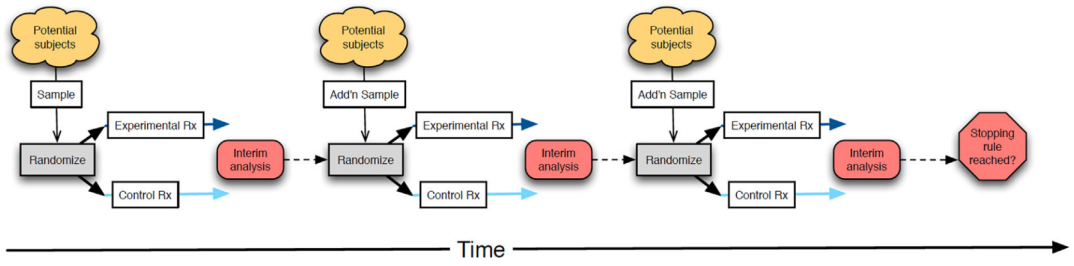

FIGURE 3.3.3 – Schéma représentant le déroulement d'un design séquentiel [1].

Ces analyses peuvent mener à l'arrêt anticipé de l'étude en cours, ce qui peut se produire lorsqu'un traitement se révèle nettement supérieur aux autres bras de traitements avant la fin de l'essai clinique. L'étude peut aussi être arrêtée pour des raisons de futilité, lorsque les traitements étudiés sont jugés inefficace et que les taux de réponse aux traitements ne sont pas suffisants par exemple. D'autres adaptations pourraient être mises en place à la suite d'une analyse intermédiaire, telles que la réestimation de la taille de l'échantillon, l'inclusion d'un nouveau groupe de traitement dans l'étude, ... Les critères d'adaptation doivent être discutés et définis avant le commencement de l'étude. La section 3.3.1.2 ci-dessous aborde plus en détail le choix de ces critères. Le nombre d'analyses intermédiaires ainsi que le moment de leur mise en œuvre doivent également être déterminés avant le début de l'étude [1].

L'introduction d'analyses intermédiaires amène des avantages intéressants dans le cadre de la recherche sur les maladies rares. Tout d'abord, cela peut permettre de diminuer la taille de l'échantillon. En effet, si une analyse intermédiaire mène à des conclusions permettant de justifier l'arrêt de l'étude, il ne sera plus nécessaire de recruter et de suivre les patients qui ne l'ont pas encore intégré. Cependant, il faut souligner que la taille de l'échantillon peut parfois s'avérer plus importante sous l'hypothèse nulle. Cela peut mener à des situations où un échantillon plus grand que celui d'un ECR parallèle est nécessaire. Ensuite, l'intégration d'analyses intermédiaires dans l'étude peut aussi réduire la durée de celle-ci, étant donné que les designs séquentiels peuvent permettre d'obtenir des conclusions plus rapidement.

Cependant un problème majeur des designs séquentiels est l'inflation de l'erreur de type I induit par la multiplicité des tests. Sa prise en compte est primordiale. Des méthodes telles que celles de Benjamini-Hochberg, Sidak et Bonferonni peuvent être utilisées pour remédier à ce problème. [1].

3.3.1.2 Méthodologie

Notions et concepts utilisés dans l'analyse des essais adaptatifs

Un essai clinique adaptatif peut être composé de plusieurs étapes (k étapes). A chaque étape, un test d'hypothèse est effectué. En fonction des résultats obtenus, diverses adaptations peuvent être apportées à la conception de l'étude, telles que l'arrêt prématuré de l'étude, la réestimation de la taille d'échantillon, ... En fonction de l'objectif de l'essai, un test d'hypothèse est formulé à chaque étape. Les hypothèses nulles peuvent s'écrire sous la forme suivante, $H_0 : H_{01} \cap H_{02}$ [13].

Considérons par exemple que l'hypothèse nulle H_{0k} faite à l'étape k représente l'hypothèse nulle selon laquelle un traitement B n'est pas plus efficace qu'un traitement A. Les hypothèses nulle et alternative testées à l'étape k pourraient s'écrire de la façon suivante [13] :

$$H_{0k} : \eta_{Ak} \geq \eta_{Bk}$$
$$H_{ak} : \eta_{Ak} < \eta_{Bk}$$

où
η_{Ak} est la réponse au traitement A à l'étape k,
η_{Bk} est la réponse au traitement B à l'étape k.

Ces hypothèses seront testées à un niveau $\frac{\alpha}{2}$. Le rejet de n'importe quelle H_{0k} (k = 1, ..., K) conduira à la même conclusion clinique. Mais si aucune hypothèse n'est rejetée, l'hypothèse globale ne pourra pas être interprétée [13].

L'évaluation de l'hypothèse nulle H_{0k} sera basée sur les observations des sous-échantillons évalués de la première étape à l'étape k. La statistique de test utilisée pour réaliser ce test d'hypothèse est T_k, et est une combinaison des p-valeurs p_k testant l'hypothèse H_{0k} [13].

La statistique de test T_k peut être construite de différentes façons. Celle que nous utiliserons dans ce travail est la suivante [12] :

$$T_k = \sum_{i=1}^{k} p_i, k = 1, ..., K \tag{3.3.1}$$

où p_i est la p-valeur associée à l'hypothèse H_{0k} à l'étape k.

La fonction de distribution cumulative de la statistique de test T_k est notée $\psi_k(t)$. Elle exprime la probabilité que la statistique de test T_k soit inférieure à une certaine valeur t, en tenant compte des valeurs obtenues aux étapes précédentes et de la valeur t à l'étape actuelle. $\psi_k(t)$ est calculée de la façon suivante [12] :

$$\begin{aligned}\psi_k(t) &= \Pr\left(\alpha_1 < T_1 < \beta_1, \ldots, \alpha_{k-1} < T_{k-1} < \beta_{k-1}, T_k < t\right) \\ &= \int_{\alpha_1}^{\beta_1} \ldots \int_{\alpha_{k-1}}^{\beta_{k-1}} \int_{-\infty}^{t} f_{T_1 \ldots T_k} dt_k dt_{k-1} \ldots dt_1,\end{aligned} \tag{3.3.2}$$

où

$\Pr\left(\alpha_1 < T_1 < \beta_1, \ldots, \alpha_{k-1} < T_{k-1} < \beta_{k-1}, T_k < t\right)$ est la probabilité jointe que la statistique de test T_i soit comprise entre α_i et β_i pour chaque $i < k$, et que $T_k < t$ (la détermination des limites α_k et β_k sera abordée dans le point "Détermination des limites d'arrêt" ci-dessous),

L'intégral $\int_{\alpha_1}^{\beta_1} \ldots \int_{\alpha_{k-1}}^{\beta_{k-1}} \int_{-\infty}^{t} f_{T_1 \ldots T_k} dt_k dt_{k-1} \ldots dt_1$ est une intégrale multiple sur un domaine spécifique de la densité de probabilité jointe $f_{T_1 \ldots T_k}$ des k variables aléatoires $T_1, \ldots, T_k$.

Lorsqu'une étude aboutit à la conclusion de l'efficacité du traitement expérimental à l'étape k, le taux d'erreur de type I global est calculé à l'aide de la formule suivante [12] :

$$\alpha = \sum_{k=1}^{K} \pi_k \tag{3.3.3}$$

où $\pi_k = \psi_k(\alpha_k | H_0)$ est la probabilité de rejeter à tort l'hypothèse nulle H_{0k} à l'étape k alors que celle-ci est vraie.

De manière similaire, la puissance de l'essai est calculée en utilisant la formule suivante [12] :

$$1 - \beta = \sum_{k=1}^{K} \omega_k \tag{3.3.4}$$

où $\omega_k = \psi_k(\alpha_k | H_a)$ est la probabilité de rejeter correctement l'hypothèse nulle H_{0k} à l'étape k.

La structure des designs à étapes multiples introduit un autre concept important, celle de la puissance conditionnelle. La puissance conditionnelle est la probabilité conditionnelle de rejeter correctement l'hypothèse nulle pendant le reste de l'essai sur la base des données observées. Elle est couramment utilisée pour le suivi d'un essai en cours, afin d'adapter la taille de l'échantillon par exemple. La puissance conditionnelle à l'étape k est la somme de la

probabilité de rejeter l'hypothèse nulle de l'étape $k+1$ à K. Pour un essai adaptatif à deux étapes, elle est calculée à l'aide de la formule suivante [12] :

$$cP_1 = Pr(T_2 \leq \alpha_2 | t_1)$$

Ces principes et les notions expliqués ci-dessus sont nécessaires à la détermination des limites d'arrêt et au calcul de la probabilité d'arrêt, qui sont des aspects importants pour la mise en place d'un essai adaptatif. Ces sujets seront traités plus en détail dans cette section.

Détermination des limites d'arrêt

La détermination des limites d'arrêt est indispensable pour les essais cliniques adaptatifs où certains bras de traitement peuvent être éliminés. Cela s'applique notamment aux RSD et aux designs séquentiels. Certaines méthodes utilisées pour établir ces limites seront présentées ci-dessous.

Les règles d'arrêt sont généralement définies de la façon suivante [12] :

$$\begin{cases} \text{Arrêt pour efficacité} & \text{si } T_k \leq \alpha_k \\ \text{Arrêt pour futilité} & \text{si } T_k > \beta_k \\ \text{Continuation de l'étude avec adaptation(s)} & \text{si } \alpha_k < T_k \leq \beta_k \end{cases}$$

où
α_k est la limite d'arrêt pour efficacité à l'étape k,
β_k est la limite d'arrêt pour futilité à l'étape k.

Les limites d'arrêt sont déterminées en se basant sur la statistique de test T_k, calculée à partir de la formule 3.3.1. Dans un premier temps, l'attention est portée sur l'erreur de type I global α, définie par l'équation 3.3.3, ainsi que sur la probabilité de commettre une erreur de type I à l'étape k, π_k. Cette dernière est également appelée "α spend". Elle est ajustée à chaque étape, afin de tenir compte de la multiplicité des tests effectués et est définie de la façon suivante [12] :

$$\pi_k = \psi_k(\alpha_k | H_0) \tag{3.3.5}$$

où $\psi_k(t)$ est la fonction de distribution cumulative de la statistique de test T_k, définie avec la formule 3.3.2.

Les équations 3.3.2, 3.3.5 et 3.3.3 seront utilisées afin de déterminer les limites d'arrêt d'un essai clinique adaptatif. Ces équations permettent de construire les limites d'arrêt et peuvent être appliquées selon différentes approches [12].

Tout d'abord, il est possible d'utiliser une fonction permettant de calculer les limites α_k et β_k. Une méthode utilisant ce principe a été développé par O'Brien et Fleming [40]. Cette méthode permet d'analyser les résultats d'un ECR parallèle avec des critères d'arrêt intermédiaire, avant la fin de l'étude. Cependant, elle n'est pas recommandée dans le contexte des maladies rares. Elle nécessite une étude avec une puissance statistique élevée, ce qui est rarement réalisable avec des petites tailles d'échantillons.

Ensuite, une autre approche consiste à modifier la fonction calculant le taux d'erreur de type I à l'étape k π_k. Généralement, la formule suivante est utilisée $\sum_{k=1}^{K} \pi_k$. Elle est appelée "error-spending function". Les limites peuvent ensuite être déterminées en utilisant cette nouvelle "error-spending function", ainsi que les équations 3.3.2 et 3.3.3 [13].

Et enfin, une méthode développée par Chang et al. [12] permet de déterminer les limites d'arrêt en utilisant une méthode basée sur la somme des p-valeurs. La statistique de test T_k, définie à l'équation 3.3.1, permet de quantifier l'accumulation des erreurs dans un processus séquentiel.

Dans un plan à deux étapes, les taux d'erreur conditionnelle de l'étape 1 π_1 et 2 π_2 sont définis en utilisant les formules 3.3.2 et 3.3.5. Ils sont ensuite définis de la façon suivante :

$$\pi_1 = \int_0^{\alpha_1} dt_1 = \alpha_1 \tag{3.3.6}$$

$$\pi_2 = \begin{cases} \int_{\alpha_1}^{\beta_1} \int_{t_1}^{\alpha_2} dt_2 dt_1 & \text{si } \beta_1 \leq \alpha_2 \\ \int_{\alpha_1}^{\alpha_2} \int_{t_1}^{\alpha_2} dt_2 dt_1 & \text{si } \beta_1 > \alpha_2 \end{cases} \tag{3.3.7}$$

Après avoir effectué les intégrations de l'équation 3.3.7 et substitué les résultats de la formule 3.3.6 dans la formule 3.3.3, les limites d'arrêt peuvent être calculées à l'aide de la formule suivante :

$$\alpha = \begin{cases} \alpha_1 + \alpha_2 (\beta_1 - \alpha_1) - \frac{1}{2} \left(\beta_1^2 - \alpha_1^2\right) & \text{si } \beta_1 < \alpha_2 \\ \alpha_1 + \frac{1}{2} (\alpha_2 - \alpha_1)^2 & \text{si } \beta_1 \geq \alpha_2 \end{cases} \tag{3.3.8}$$

Cette formule 3.3.8 montre la façon dont les limites d'arrêt sont déterminées en fonction de l'erreur de type I global. Ces limites sont utilisées dans les tests d'hypothèse multiples et permettent de contrôler le taux d'erreur global.

Dans le contexte d'une procédure d'essai à deux étapes, α_1 et β_1 sont généralement définis en fonction du contexte de l'étude. Le choix de α_2, le critère d'arrêt pour efficace à l'étape 2, dépend de la première étape et du taux d'erreur global souhaité.

Si β_1 est inférieur à α_2, la limite d'arrêt est déterminée par le premier cas de la formule 3.3.8. Cela signifie que le critère d'arrêt pour efficacité à l'étape 2 a une valeur plus élevée que

le critère d'arrêt pour futilité à l'étape 1. Par contre, si un critère plus strict est souhaité et donc, que β_1 est supérieur ou égal à α_2, la limite d'arrêt est déterminée par le second cas de la formule 3.3.8. La formule utilisée dépendra des seuils α_1 et β_1 choisis lors de la première étape, ainsi que du niveau d'exigence souhaité pour la seconde étape.

En fixant une valeur pour α_1, β_1 et pour le taux d'erreur de type I global, il sera possible de déterminer la valeur de α_2 avec l'équation 3.3.8. Cela permet de garder le contrôle sur le taux d'erreur de type I global tout en tenant compte des résultats obtenus à la première étape.

Chacune des méthodes présentées possède des avantages et des limitations. Le choix de la méthode de détermination des limites d'arrêt dépendra donc des spécificités de chaque essai clinique.

Probabilités d'arrêt et réestimation de la taille de l'échantillon

L'une des caractéristiques essentielles d'un design adaptatif à plusieurs étapes est la probabilité d'arrêt à chaque étape. En effet, celle-ci permet de calculer des éléments importants tels que la durée attendue de l'essai et la taille de l'échantillon attendue. Ces éléments peuvent ensuite être adaptés en fonction des observations faites au cours de l'étude. Si les résultats s'avèrent prometteurs, la durée de l'étude et la taille de l'échantillon pourraient être réduites, ce qui est avantageux dans le contexte des maladies rares.

Deux types de probabilité inconditionnelle d'arrêt peuvent être définies à l'étape k : L'"Efficacy Stopping Probability (ESP)" et la "Futility Stopping Probability (FSP)". L'ESP_k correspond à la probabilité d'arrêt pour raisons d'efficacité à l'étape k, tandis que la seconde, la FSP_k, correspond à la probabilité d'arrêt pour futilité à l'étape k. Ces probabilités sont dérivées de l'équation 3.3.2 et se calculent de la façon suivante :

$$ESP_k = \psi_k(\alpha_k)$$

$$FSP_k = 1 - \psi_k(\beta_k)$$

où $\psi_k(t)$ est la fonction de distribution cumulative de la statistique de test T_k, définie avec la formule 3.3.2.

L'ESP_k et la FSP_k permettent de calculer la durée attendue de l'étude si celle-ci s'arrête pour des raisons d'efficacité, de futilité, ainsi que la durée attendue de l'essai de façon inconditionnelle.

Elles permettent également de réajuster la taille de l'échantillon. Cela peut s'avérer utile dans le cas des maladies rares. En effet, si les résultats sont prometteurs, il pourrait être possible de réduire le nombre de patients à recruter. La formule utilisée pour réestimer la taille de l'échantillon est la suivante :

$$N_{exp} = N_{max} - \sum_{k=1}^{K} n_k(\psi_k(\beta_k) - \psi_k(\alpha_k))$$

où

N_{max} est la limite supérieure du nombre total de patients qui peuvent être inclus dans l'étude,

n_k est le nombre de patients par groupe dans le sous-échantillon de l'étape k,

$\psi_k(t)$ est la fonction de distribution cumulative de la statistique de test T_k, définie avec la formule 3.3.2,

α_k est la limite d'arrêt pour efficacité à l'étape k,

β_k est la limite d'arrêt pour futilité à l'étape k.

3.3.2 Utilisation de contrôle historique dans les essais cliniques

3.3.2.1 Généralité

Dans le cadre d'une étude clinique consacrée aux maladies rares, il est possible d'utiliser des contrôles historiques, qui sont des données issues d'études réalisées dans le passé. Ils sont utilisés dans le bras contrôle de l'étude actuelle. Cela peut se faire de deux façons, soit par une substitution partielle, soit par une substitution totale du bras contrôle par des données historiques. Ces deux approches sont différentes et possèdent leurs propres caractéristiques. Dans le cas d'une substitution totale, le bras contrôle est entièrement remplacé par des données historiques. La substitution partielle, quant à elle, permet de compléter le bras contrôle avec des données historiques [24].

En utilisant des données historiques pour constituer le groupe contrôle, il est possible de réduire le nombre de patients nécessaires à l'étude, venant ainsi répondre à l'une des principales problématiques liées aux maladies rares. De plus, cette approche pourrait améliorer le recrutement des patients. La probabilité accrue d'être assigné au traitement expérimental pourrait atténuer les réticences de certains, facilitant ainsi leur participation à l'essai clinique [25].

Cependant, il faut noter que ces designs ne devraient être utilisés que dans des situations spécifiques, lorsque le recrutement des patients est particulièrement compliqué par exemple. En effet, l'utilisation de données historiques peut mener à des biais. Le choix d'un groupe contrôle dont les résultats sont exceptionnellement bas pourrait provoquer des biais de sélection. De même, l'absence de randomisation pourrait entraîner la présence de facteurs confondant, menant à des conclusions erronées à la fin de l'étude [26].

Le choix des données historiques est une étape importante dans le cadre des essais cliniques utilisant des contrôles historiques. Il existe plusieurs façons de sélectionner les données à inclure dans une étude clinique, afin d'éviter la présence de biais de sélection et de facteurs confondants [24].

Tout d'abord, les données doivent être en accord avec les caractéristiques clés de l'essai actuel, le(s) traitement(s) étudié(s) ainsi que les critères d'inclusion, tels que l'âge, le sexe, ou encore les antécédents cliniques des patients. Ensuite, il est recommandé de choisir des données historiques issues de sources diversifiées. Cela permettra de tenir compte de l'hétérogénéité de l'effet du traitement et de la maladie sur les différents patients [17]. Lors de la publication des résultats, il faudra également justifier l'emploi de contrôles historiques, ainsi que les critères de

sélection utilisés pour les sélectionner. Cela inclut une explicitation de leur utilité dans l'essai et des méthodes de sélection [24].

Une autre façon de sélectionner les données historiques passe par l'utilisation d'une procédure statistique nommée "test-then-pool". Ce test permet d'évaluer la pertinence de l'inclusion des données historiques dans l'essai clinique actuel. Elle est adaptée aux études évaluant un output catégoriel, particulièrement lorsque celui-ci est binaire. Il peut être utilisé lorsque le groupe contrôle est partiellement substitué avec des contrôles historiques [51].

Le "test-then-pool" est un test d'hypothèse dont l'hypothèse nulle affirme que les données des contrôles historiques ont un taux de réponse au traitement similaire à celui du groupe contrôle actuel. Cela peut s'écrire de la façon suivante : $H_0 : \pi_H = \pi_C$ où π_H est la proportion de patients ayant répondu au traitement dans le groupe historique et π_C est la proportion de patients ayant répondu au traitement dans le groupe contrôle de l'étude actuelle. L'hypothèse alternative, quant à elle, affirme que ces proportions sont différentes, $H_a : \pi_H \neq \pi_C$. La statistique de test utilisée par Viele et al. [51] pour évaluer ces hypothèses est la statistique de test z. Cependant, dans le contexte des maladies rares, nous utiliserons plutôt le test de Fisher exact au lieu de la statistique de test z. Le test de Fisher est souvent utilisé lorsque les tailles d'échantillon sont relativement petites, car il ne repose pas sur des approximations normales. Si le résultat du test ne rejette pas l'hypothèse nulle, une analyse combinée peut être envisagée. Si le résultat aboutit au rejet de l'hypothèse nulle, l'analyse devra être effectuée en ignorant les données historiques.

Il faut tout de même noter que ce test se limite à la comparaison des proportions de succès des deux sous-ensembles du groupe contrôle, sans tenir compte des caractéristiques spécifiques des patients. Il est donc essentiel de veiller à ce que les patients dont sont issus les contrôles historiques soient similaires à ceux recrutés dans l'étude actuelle, afin de garantir l'homogénéité de l'échantillon et la comparabilité des données.

3.3.2.2 Méthodologie

3.3.2.3 Analyse statistique

Plusieurs méthodes permettent d'analyser les données provenant d'essais cliniques utilisant des contrôles historiques. Celles qui seront présentées peuvent s'appliquer à différents types d'output, qu'il soit binaire ou continu. Pour illustrer leur utilisation, nous appliquerons à chacune d'elles un type d'output, soit binaire, soit continu.

Le "threshold crossing" est une méthode statistique permettant de déterminer si les contrôles historiques sélectionnés peuvent être inclus dans l'étude. Mais elle permet aussi de fixer les limites à partir desquelles le traitement expérimental est considéré comme futile ou supérieur. Cette méthode n'est utilisable que lorsque le groupe contrôle est exclusivement composé de contrôles historiques. Les étapes d'un "threshold crossing" sont semblables à celles utilisée dans les études de phase II en oncologie, comme dans le design de Simon [19].

Tout d'abord, la population éligible au traitement est définie. Cela est nécessaire au recru-

tement des patients au sein de l'étude, mais aussi à la sélection des données historiques. La sélection des données historiques ainsi qu'une estimation de l'effet du traitement sont ensuite effectuées. Cet effet est calculé en fonction des résultats observés au cours de l'étude et de leur évolution. Les patients du groupe contrôle doivent être inclus avant que les patients du bras expérimental ne soient sélectionnés. De plus, le groupe contrôle doit être identifié à partir d'un protocole précis [19].

Ensuite, des limites d'efficacité et de futilité sont établies par des experts, en fonction du contexte et des objectifs de l'étude. Celles-ci sont calculées à partir des données historiques. Lorsque les résultats du traitement expérimental dépassent le seuil d'efficacité, le traitement sera considéré comme efficace. Si ces résultats atteignent ou sont inférieurs au seuil de futilité, l'étude prendra fin. Il est aussi possible que les résultats se situent entre les deux seuils. Le traitement expérimental sera alors considéré comme prometteur. Un ECR parallèle ou un deuxième essai à un bras pourrait être envisagé afin de confirmer les premières observations. Le "threshold crossing" permet donc d'évaluer les risques et les bénéfices potentiels de la continuation ou de l'arrêt d'un essai en fonction des résultats observés [19].

Suite à cela, les patients recrutés pourront recevoir le traitement expérimental. Il faudra veiller à ce que ceux-ci soient sélectionnés selon les mêmes critères que ceux utilisés pour la sélection des données historiques [19].

Dans la continuité des méthodologies permettant d'analyser des essais cliniques utilisant des contrôles historiques, une autre stratégie permettant d'attribuer des poids différents aux données provenant des contrôles historiques peut être utilisée. Cette stratégie repose sur l'utilisation de la "power prior", une technique issue de l'approche bayésienne [27].

La prior informative utilisée pour le groupe contrôle de l'étude en cours est le résultat du produit entre une prior initiale et une fonction de vraisemblance, qui est calculée sur base des données historiques. Celle-ci est ensuite élevée à une certaine puissance α_0, comprise entre 0 et 1. Il n'existe pas de règle stricte pour choisir la valeur de α_0. Celle-ci repose sur le jugement et l'évaluation d'experts, ainsi que sur la comparaison entre les données actuelles et les données historiques. Lorsque α_0 est proche de 0, cela signifie que les données historiques n'apportent que peu d'informations à l'étude en cours. Mais lorsque celui-ci est proche de 1, les données historiques apportent de l'information à la distribution de la prior informative. Ce produit est ensuite utilisé comme la prior informative de l'étude clinique en cours [51].

Voici un exemple illustrant l'utilisation d'une power prior réduisant l'influence des contrôles historiques lors d'une étude clinique évaluant un output binaire. La prior initiale non-informative est une $Beta(0.001, 0.001)$. La prior beta est une distribution de probabilité couramment utilisée pour modéliser des proportions. Les valeurs choisies indiquent une certaine incertitude quant aux données et à l'état des connaissances a priori. Les contrôles historiques indiquent que 35 cas historiques ont répondu au traitement tandis que 65 n'y ont pas répondu. Cela permet de mettre à jour la prior $Beta(0.001, 0.001)$ qui devient $Beta(0.001 + 35, 0, 001 + 65)$ [51].

Si l'utilisation d'une power prior est faite et que α_0 prend la valeur de 0.4, la prior informative devient $Beta(0.001 + 35 * 0.4, 0, 001 + 65 * 0.4)$, qui est une $Beta(14.001, 26.001)$ et sera

la prior du groupe contrôle de l'étude actuelle. Le fait de fixer la valeur de α_0 à 0.4 reflète un niveau de connaissance plutôt limité [51].

Imaginons qu'une étude est en cours et comporte deux groupes de traitements, un groupe contrôle, prenant le traitement A, et un groupe expérimental, prenant le traitement B. Le groupe contrôle comprend n_A patients, parmi lesquels Y_A répondront au traitement contrôle. Le groupe expérimental, quant à lui, comptabilise n_B individus, avec Y_B répondant au traitement expérimental [51].

Il est ensuite possible de combiner les informations de la prior informative du groupe contrôle $Beta(14.001, 26.001)$ et de la prior non informative du groupe expérimental $Beta(0.001, 0.001)$ aux résultats obtenus dans l'étude actuelle. La distribution a posteriori pour le groupe contrôle sera généralement plus précise par rapport à celle du groupe expérimental, étant donné que la prior utilisée pour le bras de contrôle est informative. La posterior de la probabilité de succès pour les traitements, conditionnée aux données accumulées jusqu'à présent, peut être exprimée de la manière suivante [51].

Pour le traitement contrôle, la posterior est donnée par $p_A|D \sim Beta(14.001+Y_A, 26.001+n_A-Y_A)$, où Y_A représente le nombre de succès observés pour le groupe contrôle et n_A le nombre total d'observations dans ce groupe.

Pour le traitement expérimental, la posterior sera $p_B|D \sim Beta(0.001+Y_B, 0.001+n_B-Y_B)$, où Y_B et n_B représentent respectivement le nombre de succès observés et le nombre total d'observations pour le groupe expérimental.

Ces expressions mathématiques reflètent l'intégration de l'information historique et actuelle dans l'évaluation des probabilités de succès des traitements. Il sera ensuite possible de calculer la probabilité $Pr(p_A < p_B|Data)$ et de comparer le groupe expérimental au groupe contrôle [51].

Lorsque plusieurs études passées peuvent être exploitées pour un essai clinique en cours, il est possible d'utiliser une "Meta-Analytic Predictive (MAP)". Cette méthode est aussi issue de la statistique bayésienne et permet d'effectuer une méta-analyse des contrôles historiques afin d'estimer la posterior du paramètre du bras contrôle de l'étude actuelle. Le groupe contrôle de l'étude en cours doit donc être partiellement substitué [24].

Supposons que H essais historiques soient disponibles. Les données de l'output principal de ces essais sont désignées par $Y_1, ..., Y_H$ et sont associées à des paramètres $\theta_1, ..., \theta_H$. L'output de cette étude peut être un output continu par exemple. Pour un essai en cours, le vecteur de paramètre est noté θ^*. Ce paramètre constitue l'objet d'intérêt dans le cadre d'une MAP, qui diffère des méta-analyses traditionnelles. En effet, une méta-analyse classique se concentre généralement sur l'inférence de μ, qui est le paramètre de la moyenne des effets des traitements observés dans les études historiques. En revanche, dans une MAP, l'attention est portée sur l'estimation de θ^*, qui est le paramètre qui représente l'effet du traitement dans un nouvel essai clinique. Il s'agit d'une quantité inconnue que l'on cherche à estimer à partir des données de l'essai en cours et des informations provenant des essais précédents.

Dans une approche MAP, il y a deux composantes principales. La première est la distribution $Y_h|\theta_h$, qui représente la relation entre les données historiques (Y_h) et leurs paramètres spécifiques (θ_h). La seconde est la distribution $\theta_1, ..., \theta_H, \theta^*|\mu, \tau^2$, qui décrit la relation entre les paramètres des contrôles historiques ($\theta_1, ..., \theta_H$) et le paramètre de l'étude actuelle (θ^*), en fonction des paramètres de la méta-analyse μ et τ^2.

La méthodologie présentée par Neuenschwander et al. [38] pose l'hypothèse selon laquelle les paramètres historiques $\theta_1, ..., \theta_H$ et actuels θ^* sont issus de la même distribution. L'objectif consiste alors à extraire des informations pertinentes sur θ^* à partir des données historiques. Une hypothèse selon laquelle les paramètres et les données sont normalement distribués est également posée. Les distributions de $Y_h|\theta_h$ et de $\theta_1, ..., \theta_H, \theta^*|\mu, \tau^2$ sont formulées de la façon suivante :

$$Y_h|\theta_h \sim N(\theta_h, S_h^2)$$

$$\theta_1, ..., \theta_H, \theta^*|\mu, \tau^2 \sim N(\mu, \tau^2)$$

où S_h^2 est la variance observée dans l'échantillon, $h = 1..., H$.

Lorsque la variance interétude est connue, il est possible d'obtenir la distribution a posteriori $\mu|Y_1, ...Y_H, \tau$. Cette dernière est obtenue à partir de la formule de la distribution normale, en supposant que les données Y_h sont indépendantes et suivent une distribution normale de moyenne μ et de variance τ^2, ainsi qu'en utilisant le fait que la moyenne d'une distribution normale est une combinaison linéaire des observations, pondérées par leurs précisions respectives. Cela permet d'aboutir à la formule suivante :

$$\mu|Y_1, ...Y_H, \tau \sim N(\frac{\sum w_h Y_h}{\sum w_h}, \frac{1}{\sum w_h})$$

où $w_h = \frac{1}{\sigma_h^2 + \tau}$.

Dans le contexte d'une MAP, l'objectif principal est d'estimer le paramètre du groupe contrôle d'un nouvel essai clinique en cours θ^*. La posterior de θ^* peut être obtenue en utilisant les données des essais cliniques historiques et l'information sur la variabilité interessai τ^2. Elle peut être dérivée de la façon suivante.

Tout d'abord, nous savons que $Y_h|\theta_h \sim N(\theta_h, S_h^2)$. Nous supposons aussi que les θ_h suivent une distribution normale de moyenne μ et de variance τ^2 et donc, $\theta_1, ..., \theta_H, \theta^*|\mu, \tau^2 \sim N(\mu, \tau^2)$.

À partir de ce modèle, nous pouvons obtenir la posterior de μ sachant $Y_1, ..., Y_H$ et la variabilité interessai τ, qui est donnée par $\mu|Y_1, ...Y_H, \tau \sim N(\frac{\sum w_h Y_h}{\sum w_h}, \frac{1}{\sum w_h})$.

Par ailleurs, nous savons que la distribution conditionnelle de θ^* sachant μ et τ^2 est donnée par $\theta^*|\mu, \tau^2 \sim N(\mu, \tau^2)$. Cela est dû au fait que, selon notre hypothèse, chaque θ_h, dont θ^*, suit une distribution normale de moyenne μ et de variance τ^2.

Enfin, en utilisant ces deux distributions, nous pouvons obtenir la distribution a posteriori de θ^* sachant les résultats des essais historiques $Y_1, ..., Y_H$ et le paramètre τ, qui est donnée par :

$$\theta^*|Y_1, ..., Y_H, \tau \sim N(\frac{\sum w_h Y_h}{\sum w_h}, \frac{1}{\sum w_h} + \tau^2)$$

Cette formule est obtenue grâce à l'application du théorème de Bayes, qui permet de combiner la posterior de μ et la distribution conditionnelle de θ^* pour obtenir la posterior de θ^*.

Il est important de souligner que pour tester des hypothèses avec les posteriors $\mu|Y_1, ...Y_H, \tau$ et $\theta^*|Y_1, ..., Y_H, \tau$, la variance interétudes doit être connue. Cependant, la variabilité est un paramètre rarement connu. Et en présence d'un nombre limité d'essais historiques, ce qui peut être le cas dans le contexte des maladies rares, une méta-analyse entièrement bayésienne pourrait s'avérer plus appropriée qu'une MAP, car elle prend en compte l'incertitude concernant le paramètre τ.

3.3.2.4 Calcul de taille d'échantillon

La taille d'échantillon nécessaire pour le groupe expérimental dans un essai clinique utilisant des contrôles historiques est déterminée par plusieurs facteurs. Elle dépend de la taille de l'échantillon historique, du niveau d'erreur de type I accepté, de la puissance de l'étude ainsi que de l'effet traitement qui doit être mis en évidence Δ. Plusieurs méthodes permettent de calculer cette taille d'échantillon. Seule la situation où deux traitements sont comparés et où l'outcome évalué est binaire sera abordée dans le cadre de ce travail.

Une approche couramment utilisée pour déterminer le nombre de patients nécessaires dans le groupe expérimental lorsque le bras contrôle est entièrement substitué par des contrôles historiques est celle proposée par Makuch et Simon. La formule de Makuch-Simon (MS) intègre les informations historiques en prenant en compte l'estimation de la proportion de succès dans le bras de contrôle, basée sur les données historiques. Cette formule est ensuite utilisée pour calculer la puissance de l'essai [34].

Supposons que π_A et π_B désignent les taux de succès réels du traitement A (le groupe contrôle), et du traitement B (le groupe expérimental). Ces probabilités sont inconnues.

L'échantillon des contrôles historiques est constitué de n_A patients. Ceux-ci reçoivent le traitement A. Le nombre de patients répondant à ce traitement est noté Y_A. Le taux de réponse pour le groupe contrôle est de $p_A = \frac{Y_A}{n_A}$. Des notations analogues peuvent être utilisées pour les patients de l'étude actuelle qui reçoivent le traitement expérimental, donnant un taux de réponse de $p_B = \frac{Y_B}{n_B}$, où Y_B est le nombre de patients ayant répondu au traitement expérimental et n_B est le nombre de patients traités [34].

Lorsque l'étude vise à démontrer la supériorité du traitement expérimental, le test suivant sera posé, $H_0 : \pi_B \leq \pi_A$ et $H_a : \pi_B > \pi_A$. La statistique de test qui sera utilisée est la suivante [34] :

$$z = \frac{x_B - x_A}{0,5\sqrt{1/(n_B + 0,5) + 1/(n_A + 0,5)}}$$

$x_i = arcsin\sqrt{\frac{Y_i+0.375}{n_i+0,75}}$ où $i = A, B$ et est le résultat d'une transformation arcsinus.

La statistique z permet d'incorporer les données historiques dans le calcul de x_A. Elle suit asymptotiquement une distribution normale standardisée. Le test aboutira donc au rejet de l'hypothèse nulle lorsque $z > Z_{1-\frac{\alpha}{2}}$ où $Z_{1-\frac{\alpha}{2}}$ est le percentile 1 - $\frac{\alpha}{2}$ de la loi normale centrée réduite [34].

Pour obtenir une puissance $1 - \beta$ afin de détecter une différence $\Delta > 0$ avec un taux d'erreur de type I de α, la méthode MS stipule que [34]

$$z_{1-\alpha}\sqrt{\frac{1}{n_A+0,5}+\frac{1}{n_B+0,5}}+z_{1-\beta}\sqrt{\frac{1}{n_B+0,5}} = 2arcsin\sqrt{\frac{Y_B+0,375}{n_B+0,75}}arcsin\sqrt{\frac{Y_A+0,375}{n_A+0,75}}$$

pour un $\alpha, \beta, \Delta, Y_A, n_A$ et un π_B définis. Cette égalité est résolue de façon itérative, en faisant varier n_B et en utilisant $Y_B = \pi_B n_B$. Le n_B sélectionné correspondra à l'arrondi à l'entier supérieur de la valeur de n_B qui satisfait cette équation [34].

L'une des contraintes associées à la méthode MS est l'utilisation du test z, qui induit une dépendance à la théorie asymptotique. Celle-ci requiert une taille d'échantillon importante, ce qui pose problème dans le contexte des maladies rares. Il faudra s'assurer que les hypothèses sous-jacentes à l'utilisation de ce type de test soient remplies. Ces hypothèses stipulent que les données suivent une distribution normale, et qu'une taille d'échantillon suffisamment grande est atteinte, afin que l'approximation normale soit valide. Dans le cas où ces conditions ne sont pas satisfaites, l'utilisation de la méthode MS peut entraîner des résultats faussés.

De plus, un autre défaut de la méthode MS est qu'elle ne tient pas compte des différences entre le groupe historique et le groupe expérimental. Elle se base essentiellement sur les données historiques, sans prendre en compte les données actuelles. Cela peut poser problème lorsque les conditions de l'essai actuel diffèrent grandement de celles des contrôles historiques en termes de protocoles de traitement ou de critères d'inclusion, par exemple. Pour tenir compte de ces différences, des méthodes plus sophistiquées sont nécessaires, telles des approches de modélisation hiérarchique ou bayésienne, qui permettent de modéliser les différences entre les groupes tout en intégrant les informations historiques [38, 51].

3.3.3 Utilisation de concepts bayésiens dans l'analyse de design d'essai clinique classique

3.3.3.1 Généralité

La statistique fréquentiste et bayésienne sont deux approches utilisées dans l'analyse des essais cliniques. Elles offrent des perspectives différentes sur l'inférence statistique.

Dans l'approche fréquentiste, les paramètres sont traités comme des constantes inconnues. Ils n'ont qu'une seule valeur fixe mais inconnue. Les tests d'hypothèse et les intervalles de

confiance sont des outils clés de l'inférence fréquentiste.

En revanche, l'approche bayésienne traite les paramètres comme des variables aléatoires ayant une distribution de probabilité. Cela permet d'intégrer des informations antérieures, à l'aide d'une distribution a priori, aussi appelée prior. Les données collectées lors de l'essai sont alors utilisées pour mettre à jour la prior en une distribution a posteriori, appelée posterior. L'inférence est ensuite basée sur cette dernière. L'approche bayésienne permet donc d'intégrer des informations externes et de fournir des estimations plus précises, ce qui est utile lorsque la taille d'échantillon est limitée.

Dans l'approche bayésienne, la première étape consiste à définir une prior $P(\theta)$ pour les paramètres θ du modèle. Une prior est une distribution de probabilité définie sur les valeurs que peut prendre un paramètre avant que des données ne soit recueilli. Le choix de cette dernière est basé sur les connaissances disponibles concernant ces paramètres. Ensuite, une fonction de vraisemblance $P(D|\theta)$ est définie en fonction des données disponibles et de la nature de ces données. Elle permet de quantifier la probabilité que de telles données soient observées, étant donné la valeur des paramètres du modèle. Selon le théorème de Bayes, le produit de la prior et de la fonction de vraisemblance est un produit proportionnel à la posterior $P(\theta|D)$. Dans le cadre d'un essai clinique évaluant l'effet de traitements, la postérior permet de quantifier la probabilité qu'un traitement ait un certain effet. La posterior est mise à jour en fonction des nouvelles données obtenues tout au long de l'étude [48].

Dans le contexte des maladies rares, l'utilisation de données a priori permet de réduire la taille de l'échantillon nécessaire à un essai clinique, ce qui est particulièrement bénéfique étant donné le nombre limité de patients disponibles pour ces essais [50]. De plus, compte tenu de la taille limitée de l'échantillon dans ces essais, l'utilisation de données a priori peut aider à compenser la quantité d'informations limitée générée par l'échantillon [1].

Néanmoins, le choix de la prior dans l'approche bayésienne peut s'avérer complexe. De plus, il influence significativement les résultats obtenus ainsi que les conclusions de l'étude [48]. Un autre inconvénient est que les méthodes bayésiennes sont souvent moins familières aux chercheurs, aux comités d'éthique, ... ce qui peut rendre leur mise en place plus compliquée [1].

L'objectif de cette section sera de présenter des applications de méthodes bayésiennes dans l'analyse des essais cliniques pour les maladies rares, en mettant en évidence leurs avantages dans ce contexte.

3.3.3.2 Définition de critères d'arrêt et optimisation de l'allocation des traitements

Comme cela a été expliqué dans la section 3.3.1.1, il est possible de modifier l'allocation de traitements en fonction des résultats obtenus au cours d'un essai clinique. Cette adaptation peut être réalisée grâce à la mise en place d'analyses intermédiaires tout au long de l'essai. Suite à ces analyses intermédiaires, le principe du RARD peut être appliqué suivant les principes bayésiens. Cela permet d'augmenter le nombre de patients assignés au meilleur groupe de traitement. Plusieurs méthodes ont été proposées pour calculer les probabilités d'allocations aux différents bras de l'essai dans le cadre d'un RARD.

Des critères d'arrêt bayésiens, pour futilité et/ou pour efficacité, peuvent aussi être utilisés dans le cadre des designs adaptatifs. Ces critères peuvent être établis à partir d'une posterior et permettent de mettre fin à un essai de manière précoce si les résultats sont insuffisamment prometteurs ou, au contraire, suffisamment concluants.

Ryan et al. [43] ont établi un design permettant d'ajuster les probabilités d'assignations de traitement et d'établir des critères d'arrêt en se basant sur la probabilité qu'un traitement i soit le traitement le plus efficace de l'étude. L'objectif principal de cette étude était de comparer quatre traitements et d'identifier le bras de traitement le plus prometteur tout en minimisant l'allocation des patients aux traitements les moins efficaces.

Au commencement de l'étude, les patients sont randomisés au sein des différents groupes de traitement avec la même probabilité. Par la suite, lors des analyses intermédiaires, les proportions d'attribution sont réévaluées, et l'étude pourrait être interrompue en raison de son efficacité ou de sa futilité. Ces décisions sont basées sur la formule proposée par Wathen et al. [54] :

$$Pr(\pi_t = \max(\pi_1, ..., \pi_M)|data)^\gamma \tag{3.3.9}$$

où

M est le nombre total de traitements étudié,

π_t est la probabilité que le traitement t, celui ayant la plus grande probabilité d'être le meilleur, soit effectivement le meilleur,

π_m sont les probabilités respectives que chaque groupe de traitement soit le meilleur,

$0 < \gamma < 1$ est le facteur prévenant les probabilités de randomisation extrêmes, qui est déterminé par l'équipe de recherche.

Le calcul de la posterior $Pr(\pi_t = \max(\pi_1, ..., \pi_M)|data)$ utilisé dans la formule 3.3.9 implique plusieurs étapes. L'output évalué au cours de l'étude de Ryan et al. [43] était le score FAOS QoL (Foot and Ankle Outcome Score Quality of Life), qui est un indicateur de la qualité de vie liée à la santé du pied et de la cheville. Ce score Y a été modélisé à l'aide d'une distribution normale en suivant le modèle suivant :

$$Y \sim N(\theta_t + \beta Z, \sigma^2)$$

A partir de cette formule, une prior a été défini pour chacun des paramètres, en fonction de l'état des connaissances actuelles de ces paramètres. La fonction de vraisemblance a aussi été établie sur base des données disponibles, permettant d'obtenir la posterior $Pr(\pi_t = \max(\pi_1, ..., \pi_M)|data)$. Tout cela a été établit par les scientifiques et les cliniciens travaillant sur l'étude clinique, tout comme les valeurs données aux paramètres, qui ne seront pas détaillés ici.

Lors de chaque analyse intermédiaire, les probabilités de randomisation des M groupes de traitement sont ajustées de façon à être proportionnelles à la probabilité a posteriori que ces groupes soient le groupe de traitement le plus efficace. Les nouveaux ratios d'allocation sont calculés en divisant chaque probabilité par la somme de toutes les probabilités. Ainsi, la

somme des probabilités d'attribution normalisées est égale à 1.

Sur la base de la formule 3.3.9, des critères d'exclusion peuvent être établis. On peut envisager d'abandonner un bras de traitement lorsque sa probabilité d'attribution est inférieure à un certain seuil. Cela permet d'éviter d'exposer inutilement des patients à un traitement inefficace. Les patients précédemment inclus dans ce groupe de traitement seraient alors randomisé au sein des groupes de traitement restants.

Des critères d'arrêt peuvent aussi être établis à partir de cette même formule. Si la probabilité qu'un traitement soit le meilleur dépasse un seuil prédéfini, l'étude pourrait être arrêtée pour des raisons d'efficacité. Cela permet de mettre fin à l'étude de manière précoce et de réduire la taille de l'échantillon et le nombre de patients exposés à des traitements moins efficaces. Une étude peut également être interrompue pour des raisons de futilité. Cela peut se produire lorsque la probabilité $Pr(\pi_t = \max \pi_1, ..., \pi_M | data)$ du traitement le plus prometteur devient très faible. Cela a une importance éthique, car cela minimise l'exposition des patients à des traitements potentiellement moins efficaces.

3.3.3.3 Calcul de taille d'échantillon

La taille de l'échantillon d'un essai clinique calculée dans le cadre fréquentiste peut être trop élevée dans le contexte des maladies rares. Étant donnée la faible prévalence de ces maladies, obtenir un échantillon suffisamment large pour détecter l'effet prédéfini, à la puissance désirée et avec le taux d'erreur de type I requis, peut s'avérer complexe, voire impossible. L'approche bayésienne offre une alternative pour déterminer la taille de l'échantillon et permet souvent d'obtenir une taille d'échantillon moins élevée que celle obtenue dans le cadre fréquentiste [50].

Une façon de calculer la taille de l'échantillon dans le cadre bayésien se fait en tenant compte de la taille de la population étudiée et des bénéfices pour la santé que pourrait apporter le traitement expérimental à cette population. Le traitement qui présente le plus grand bénéfice à posteriori sera recommandé. Les bénéfices doivent être définis au préalable [50].

Ces bénéfices sont modélisés par une fonction d'utilité $h_i(\xi_i)$, qui représente l'efficacité du traitement i chez les patients qui le reçoivent. Il est supposé que les données suivent une distribution de la famille exponentielle à un paramètre, dont la moyenne est ξ_i. Dans la situation où deux traitements sont comparés, à savoir, les traitements i (i = A, B), le gain attendu pour un patient qui recevra le traitement i à l'avenir sera défini par une fonction $g_i(\xi_i)$ [49].

Si les fonctions h_i et g_i respectent certaines conditions, il est possible d'utiliser des formules assez simples pour calculer la taille de l'échantillon nécessaire à un essai clinique. Pour commencer, g_i doit être strictement croissante et différentiable, c'est-à-dire que sa dérivée doit exister et être définie pour toutes les valeurs de son domaine. De plus, $E_0(h_i(\xi_i)) \leq E_0(max g_A(\xi_A), g_B(\xi_B))$, où $i = A, B$, garantissant que le bénéfice du traitement d'un patient au sein de l'essai ne dépassera pas celui d'un patient prenant ce même traitement en dehors de cet essai. Ces fonctions dépendent uniquement de ξ, supposant que les bénéfices observés ne proviennent que du traitement administré. Il est alors possible de calculer l'espérance de l'utilité nette pour différentes tailles d'échantillon, et de choisir la taille d'échantillon qui maximise cette utilité [49].

La taille optimale de l'échantillon, lorsque la population est composée de N personnes, est alors de $O(N^{1/2})$ où $N \to \infty$. La notation $O()$ est une notation asymptotique décrivant l'évolution de la taille de l'échantillon en fonction de la taille de la population. Cette formule, qui s'applique lorsque les données collectées durant l'essai clinique suivent une distribution exponentielle à un paramètre, signifie que lorsque la population tend vers l'infini, la taille optimale de l'échantillon varie proportionnellement à la racine carrée de N.

D'autres méthodes permettent de calculer la taille d'un échantillon dans le cadre bayésien. L'une d'entre elles a été proposée par Brakenfford et al. [10]. Elle permet de déterminer la taille de l'échantillon pour un ECR. Cette méthode s'applique au cas où les données de l'essai clinique suivent une distribution normale. Elle utilise ensuite des distributions a priori pour les paramètres. L'objectif est de définir la taille de l'échantillon qui permet de conclure avec une probabilité prédéfinie, que le traitement expérimental est plus efficace que le traitement contrôle, ou que cette probabilité n'excède pas un certain seuil, démontrant la futilité du traitement expérimental.

3.3.3.4 Utilisation des concepts bayésiens pour l'analyse de designs cités précédemment

Comme décrit dans les sections précédentes, les concepts bayésiens peuvent aussi être utilisés pour l'analyse d'autres designs, tels que les designs adaptatifs (section 3.3.4.2), les essais cliniques utilisant données historiques (section 3.3.3.3), ... Les méthodes permettant de mettre ces concepts en application ont été présentées dans les sections précédentes.

3.4 Autres designs adaptés aux maladies rares

D'autres designs qui répondent aux problématiques liées aux maladies rares existent, en plus de ceux présentés dans les sections précédentes [37].

Par exemple, un "Enriched Enrollment Randomized Withdrawal Design (EERWD)" peut être utilisé dans ce contexte. Ce design sélectionne uniquement les patients répondant au traitement expérimental pour participer à l'essai clinique [37].

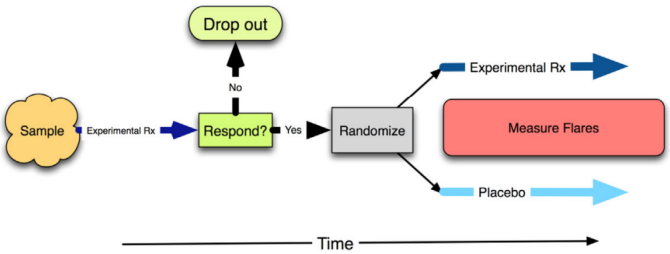

FIGURE 3.4.1 – Schéma représentant le déroulement d'un EERWD [1].

Son déroulement est illustré à la figure 3.4.1. Premièrement, tous les patients de l'étude sont traités avec le traitement expérimental. Seuls ceux qui répondent à ce traitement passent à la deuxième étape. Les patients sélectionnés sont ensuite randomisés dans deux groupes : le groupe contrôle et le groupe expérimental. Seuls les résultats obtenus lors de cette seconde étape seront utilisés pour l'analyse finale des données [37].

Les patients du groupe de contrôle pourront reprendre le traitement expérimental à la fin de l'étude, ce qui répond aux préoccupations éthiques et limite les réticences des patients concernant leur éventuelle randomisation dans le groupe de contrôle lors de la deuxième phase de l'étude [37].

L'avantage principal de ce design est qu'il permet à tous les patients de l'étude d'accéder au traitement expérimental. Le EERWD permet aussi de réduire le temps durant lequel les patients sont traités avec le traitement contrôle [48]. Tout cela permet d'améliorer le recrutement des patients au sein de l'étude. Cependant, son application est limitée aux maladies chroniques qui évoluent de façon prévisible. De plus, la sélection conduit à un échantillon non représentatif de la population spécifique, ce qui limite la généralisation des résultats [1].

Un autre design adapté au contexte des maladies rares est le "Randomized Placebo-Phase Design (RPPD)". Comme dans un ECR parallèle à deux bras de traitements, les patients sont d'abord randomisés dans deux groupes, le groupe contrôle et le groupe expérimental. Après un certain temps, le groupe contrôle reçoit également le traitement expérimental. Ainsi, tous les patients terminent l'étude avec le traitement expérimental [21]. Son déroulement est illustré à la figure 3.4.2.

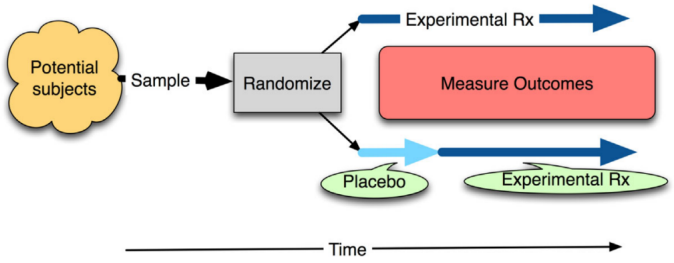

FIGURE 3.4.2 – Schéma représentant le déroulement d'un RPPD [1].

Le raisonnement de ce design est le suivant. Un traitement produit un effet sur un patient malade après un temps donné. Si le traitement expérimental est efficace, le groupe ayant commencé l'essai avec ce traitement devrait montrer une amélioration plus rapide que le groupe ayant commencé plus tard ce même traitement [21].

Ce design présente les mêmes avantages que l'EERWD. Il permet à tous les patients de l'étude de recevoir le traitement expérimental et il minimise le temps passé sous le traitement contrôle [1]. Cependant, certains aspects sont complexes à définir, comme le fait de déterminer l'intervalle de temps durant lequel le groupe contrôle devra recevoir le traitement contrôle [2].

Chapitre 4

Essai clinique évaluant l'efficacité de quatre traitements pour les malformations artérioveineuses

4.1 Physiopathologie des malformations artérioveineuses

Les Anomalies Vasculaires (AVs) sont des lésions du système vasculaire, qui apparaissent généralement entre la naissance et l'enfance. Ces lésions peuvent causer des douleurs chroniques, affecter le fonctionnement des organes touchés et impacter la qualité de vie ainsi que l'état psychologique des patients [20].

Les AVs sont souvent associées à des mutations perturbant la régulation de la prolifération cellulaire. Les protéines K-RAS et B-RAF, impliquées dans le contrôle de la croissance cellulaire, l'apoptose et la différenciation cellulaire, sont fréquemment touchées par ces mutations. La dérégulation de ces voies peut mener à une vascularisation non contrôlée exacerbée, provoquant les symptômes caractéristiques des AVs [7].

Il existe différents sous-types d'AVs, qui dépendent principalement des mutations et des zones touchées. Les lésions peuvent affecter divers types de vaisseaux, y compris les artères, les capillaires veineux et les vaisseaux du système lymphatique [7].

Les Malformations Artérioveineuses (MAVs) sont un type spécifique d'anomalie vasculaire et sont considérées comme des maladies rares. En effet, ces pathologies affectent environ 1 personne sur 20 000 [41]. Ces maladies sont souvent associées à des mutations menant à une dérégulation de l'activité des protéines régulant la prolifération cellulaire, telles que la protéine RAS [7].

Les options thérapeutiques actuelles pour les MAVs sont limitées. Souvent, seule une prise charge du patient par des professionnels de la santé afin de leur fournir un BSC est possible. Il est donc nécessaire d'étudier et de développer de nouvelles options thérapeutiques dans le but de traiter les patients atteints de MAVs.

Certaines molécules utilisées en oncologie pourraient offrir des perspectives intéressantes

pour le traitement des MAVs. Ces molécules ciblent souvent des protéines impliquées dans la régulation des voies de signalisation cellulaire, dont la dérégulation contribue également au développement des MAVs [29]. C'est notamment le cas de quatre molécules : la thalidomide, le trametinib, le regorafenib ainsi que le cabozantinib.

La thalidomide est une molécule dont le mécanisme d'action est encore mal connu mais qui est utilisée dans le traitement de certains cancers pour ses propriétés anti-inflammatoires et immunomodulatrices [6]. Cette molécule est également un inhibiteur de l'angiogenèse et inhibe par conséquent la croissance des vaisseaux sanguins [18]. Cette propriété rend donc son utilisation intéressante dans le traitement des MAVs. Dans une étude menée par Boon et al. [9], les effets de la thalidomide ont été évalués chez 18 patients atteints de MAVs sévères résistantes aux thérapies conventionnelles. Les résultats ont montré une amélioration significative des symptômes chez ces patients.

Le trametinib est un inhibiteur de la protéine MEK, qui régule l'activité de la protéine ERK impliquée dans les voies de régulation cellulaire. Cette voie est perturbée par les mutations de B-RAF, conduisant à une suractivation de MEK, qui à son tour, suractive la protéine ERK. Le trametinib inhibe l'activité de MEK en empêchant son activation par B-RAF [4]. MEK est aussi une protéine dont la dérégulation est impliquée dans le développement des MAVs. L'effet du trametinib a également été évalué chez des patients atteints de MAVs. Deux cas de patients atteints d'anomalies lymphatiques complexes ont montré une amélioration de leur état de santé suite à la prise du trametinib [39].

Le régorafénib et le cabozantinib sont des inhibiteurs de tyrosine kinases, qui sont des protéines impliquées dans la croissance tumorale et la formation de nouveaux vaisseaux sanguins [3, 5]. Leurs mécanismes d'action pourraient justifier leur utilisation dans le traitement des MAVs. Étant donné leur rôle dans la régulation de l'angiogenèse, qui est une voie impactée dans les MAVs, le régorafénib et le cabozantinib pourraient offrir de nouvelles voies thérapeutiques pour ces patients.

Dans ce chapitre, nous proposons un design d'essai clinique visant à évaluer l'efficacité de ces quatre traitements chez des patients atteints de MAVs. Ce design devra tenir compte des contraintes imposées par les maladies rares, comme décrit dans le chapitre 2, étant donné que les MAVs sont considérées comme telles.

4.2 Méthodologie

Recrutement des patients

Le recrutement des patients est basé sur des critères d'inclusion et d'exclusion, établis par un conseil composé de scientifiques spécialisés dans les MAVs : des dermatologues, des internistes, des radiologues, des chirurgiens, des hémato-oncologues ainsi que des pharmacologues ayant une expertise sur les médicaments étudiés dans cet essai.

Les patients pouvant intégrer l'étude clinique sont les adultes souffrant de MAVs extracrâniennes ("Schobinger" de stade III ou de stade IV ou "Yakes" de stade II, III ou IV) ainsi

que ceux dont le pronostic vital est affecté par les MAVs. Les patients présentant des complications évolutives ou résistant aux soins standard peuvent aussi être inclus. Une fois recrutés, ils ne doivent pas subir d'intervention chirurgicale pendant l'essai. Si cela devait se produire, on considéra que le traitement a été un échec pour le patient concerné.

Les patients asymptomatiques atteints de MAVs, portant un certain type de mutation (mutation germinale ou somatique touchant la protéine PIK3CA) ou présentant une tumeur maligne (à l'exception du carcinome basocellulaire) ne pourront pas intégrer cette étude. Les patients pour lesquels les traitements étudiés sont contre-indiqués ne pourront pas non plus faire partie de l'étude, tout comme ceux qui ne pourront pas suivre le protocole imposé ou qui n'acceptent pas de signer le consentement éclairé. Les femmes enceintes, allaitantes, ou ne prenant pas de mesures contraceptives ne pourront pas non plus intégrer l'étude.

Traitements étudiés

Les médicaments décrits dans la section précédente, à savoir la thalidomide, le trametinib, le regorafenib ainsi que le cabozantinib, sont les traitements qui seront évalués lors de cette étude. Ces médicaments ont déjà obtenu leur autorisation de mise sur le marché pour d'autres indications. L'utilisation de médicaments déjà approuvés permet de réduire la durée de l'étude. En effet, étant donné que la production de ces médicaments a déjà été établie et que la sécurité de ces molécules a déjà été étudiée, certaines étapes peuvent être omises, permettant d'accélérer le développement de nouvelles thérapies pour les MAVs.

Tous ces traitements sont administrés par voie orale. Ils seront étudiés en monothérapie. Les doses les plus faibles seront administrées aux patients afin de minimiser le risque d'apparition d'effets secondaires.

Objectifs de l'étude

L'objectif de cette étude sera donc d'évaluer l'efficacité de la thalidomide, du trametinib, du regorafenib et du cabozantinib dans le traitement des MAVs.

L'impact de ces traitements sur l'amélioration de la qualité de vie des patients sera l'endpoint principal. L'analyse statistique du design présenté ci-dessous sera basée sur l'analyse de cet endpoint, qui est binaire. Pour chaque patient, un comité d'experts évaluera l'impact du traitement sur le patient. Deux résultats seront donc possibles. Soit le traitement entraîne une amélioration de la qualité de vie du patient, auquel cas le traitement sera considéré comme un succès. Si aucune amélioration n'est observée, le traitement sera considéré comme un échec.

L'amélioration de la qualité de vie sera évaluée en fonction du soulagement de la douleur procuré par les traitements, de la progression des lésions, des complications causées par les malformations liées à la maladie, du volume de ces malformations, et de l'état des lésions observées par imagerie. L'arrêt des saignements et des suintements, la cicatrisation des ulcérations, et le bon fonctionnement des organes seront également pris en compte pour évaluer l'efficacité des traitements.

4.2.1 Design d'essai clinique proposé

Le design d'essai clinique choisi pour évaluer l'efficacité de quatre traitements dans le cadre les MAVs s'est porté sur un design adaptatif et est représenté à la figure 4.2.1.

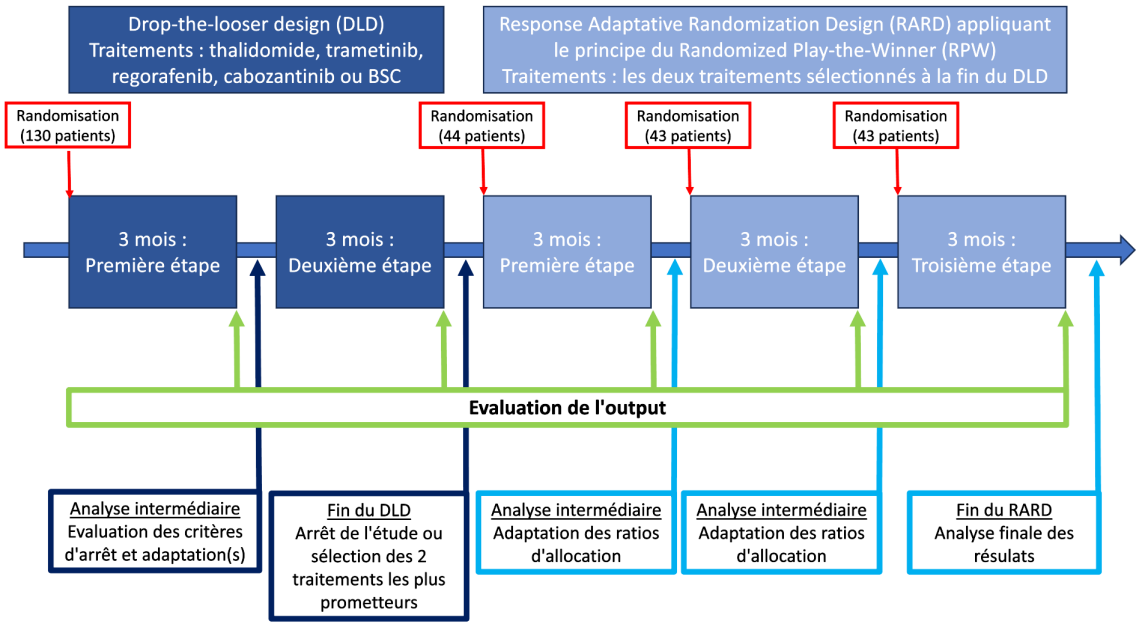

FIGURE 4.2.1 – Schéma représentant le déroulement du design d'essai clinique proposé pour évaluer l'efficacité de quatre traitements pour les malformations artérioveineuses.

Dans un premier temps, pendant une période de huit mois, un RSD en deux étapes, incorporant le principe du "drop-the-looser" aura lieu. À chaque étape, des analyses intermédiaires seront réalisées, et les traitements les moins efficaces seront éliminés en fonction de critères bien définis. Des critères d'efficacité et de futilité seront établis pour chaque traitement envisagé. Dans l'éventualité où la première phase de l'étude ne permet pas d'identifier le traitement le plus efficace, et que l'étude ne s'arrête pas de façon précoce pour des raisons de futilité, une seconde phase d'étude sera mise en place. Les deux traitements les plus efficaces à la dernière étape du RSD précédente seront alors sélectionnés. L'efficacité de ces traitements sera ensuite évaluée à l'aide d'un RARD utilisant le principe du RPW, sur une période d'un an. Le déroulement de chaque étape de ce design sera détaillé dans les sections qui suivent.

L'élaboration de cette étude s'est basé sur le livre "Adaptive design theory and implementation using SAS and R" de Chang et al. [13].

L'utilisation de designs adaptatifs pour cet essai clinique est justifiée par le nombre de traitements à évaluer et le manque de connaissances disponibles sur l'efficacité de ces traitements dans le cadre de MAVs. Un tel design permettra d'évaluer l'efficacité des quatre traitements sélectionnés et de le faire évoluer en fonction des résultats obtenus au cours des différentes étapes. Cela offrira la possibilité de prendre des décisions assurant que les patients de l'étude soient randomisés dans les groupes de traitement les plus efficaces. Ce niveau de flexibilité

ne serait pas possible avec un design cross-over ou un design n-of-1, qui ne permettent pas d'éliminer un groupe de traitement au cours de l'essai clinique.

L'utilisation de critères bayésiens, que ce soit dans un ECR parallèle à quatre bras ou dans l'élaboration des critères d'arrêt du design adaptatif proposé, aurait pu être intéressant. Toutefois, comme décrit dans la section 3.3.3, une telle analyse nécessite l'utilisation d'une prior informative. À l'heure actuelle, il n'existe pas d'études similaires dans la littérature traitant des MAVs qui pourraient fournir une telle prior. L'élaboration et la définition d'une prior, sa distribution et ses paramètres, auraient nécessité une consultation avec des experts ayant d'excellentes connaissances en statistiques bayésiennes et sur les MAVs, afin d'obtenir une posterior utilisable. Cette démarche n'a pas été réalisable dans le cadre de ce mémoire.

Etant donné qu'une prior non informative n'aurait pas apporté de bénéfices supplémentaires par rapport à une analyse fréquentiste, cette option n'a pas été retenue. En effet, en l'absence d'information a priori, les deux approches tendent à donner des résultats similaires.

4.2.1.1 Première partie : Drop-Loser design

Déroulement

Tout d'abord, la qualité de vie des patients intégrant l'étude sera évaluée afin de pouvoir évaluer efficacité des traitements. L'output utilisé est donc un output binaire. La réponse au traitement sera considérée comme un succès si le patient observe une amélioration par rapport à sa qualité de vie initiale, et sera considérée comme un échec si aucune amélioration n'est observée.

Ensuite, les patients seront randomisés au sein des groupes de traitements, à savoir la thalidomide, le trametinib, le regorafenib et le cabozantinib. Un groupe de contrôle sera inclus. Les patients intégrant ce groupe de traitement seront traités avec le BSC. Chaque bras de l'étude a une probabilité d'affectation de $\frac{1}{5}$.

Le Drop-Loser Design (DLD) se déroulera en deux étapes, chacune comportant une analyse : une analyse intermédiaire et une analyse finale. À la fin de chaque étape, il sera possible soit d'éliminer les bras qui ne répondent pas aux critères de l'étude, soit d'arrêter l'essai clinique pour des raisons de futilité ou d'efficacité. Le groupe de contrôle, traité avec le BSC, sera présent tout au long du RSD, afin de permettre une comparaison avec le taux de succès des traitements étudiés.

Pour la constitution de l'échantillon, la première moitié des participants doit être recrutée avant le début de l'étude, et la seconde moitié doit être recrutée au plus tard quatre mois après le lancement de l'essai clinique.

L'output sera évalué après une période de trois mois. Une première analyse suivra cette évaluation, avec un délai d'un mois pour permettre une analyse des données et la mise en place des adaptations nécessaires. La seconde mesure de l'output de l'essai clinique débutera dès que toutes les données seront disponibles, soit trois mois après le commencement de cette étape. Les résultats seront une nouvelle fois évalués après une période de trois mois. L'analyse

finale des résultats du DLD aura lieu durant un mois, offrant un intervalle de temps suffisant pour l'étude détaillée des données.

La puissance de l'étude sera fixée à 80% et le taux d'erreur de type I à 0.025.

A la fin de chaque étape, le design pourra être adapté en fonction des règles de décision présentées ci-dessous. Ces décisions seront prises sur base des hypothèses suivantes et de la p-valeur associée à chacune d'entre elles.

Le test d'hypothèse étudié est le suivant. L'hypothèse nulle à l'étape k stipule que le taux d'amélioration de la qualité de vie induit par le traitement i n'est pas supérieur à celui du BSC. L'hypothèse alternative, quant à elle, affirme que le traitement i entraîne une amélioration de la qualité de vie chez un pourcentage de patients supérieur de celui du BSC à l'étape k. Les hypothèses s'écrivent de la façon suivante :

$$H_{0ik} : \pi_{BSCk} \geq \pi_{ik}$$

$$H_{aik} : \pi_{BSCk} < \pi_{ik}$$

où
π_{BSCk} est la taux de succès du BSC à l'étape k, où $k = 1, 2$,
π_{ik} est la taux de succès du traitement i à l'étape k, où $i = A, B, C, D$; $k = 1, 2$.

Nous utiliserons la statistique de test Z pour évaluer ces hypothèses, où $Z \sim N(0, 1)$. Elle est calculée avec la formule suivante :

$$Z = \frac{p_{ik} - p_{BSCk}}{\sqrt{\frac{\bar{p}_k \bar{q}_k}{n_{BSCk}} + \frac{\bar{p}_k \bar{q}_k}{n_{ik}}}}$$

où
$p_{BSCk} = \frac{Y_{BSCk}}{n_{BSCk}}$ est l'estimateur de la probabilité de succès du BSC à l'étape k,
Y_{BSCk} est le nombre de succès chez les patients ayant été traité avec le BSC à l'étape k,
n_{BSCk} étant le nombre de patients ayant été traité avec le BSC à l'étape k,
$p_{ik} = \frac{Y_{ik}}{n_{ik}}$ est l'estimateur de la probabilité de succès du traitement i à l'étape k,
Y_{ik} est le nombre de succès chez les patients traités avec le traitement i à l'étape k,
n_{ik} est le nombre de patients traités avec le traitement i à l'étape k,
$\bar{p}_k$ est l'estimation des probabilités de succès combinées du BSC et du traitement i à l'étape k,
$\bar{q}_k$ est l'estimation des probabilités d'échec combinées du BSC et du traitement i à l'étape k.

Lors de la première analyse, nous évaluerons quatre tests d'hypothèses H_{0i1} pour comparer l'efficacité des quatre traitements à celle du BSC. Chacun de ces tests est associés à une p-valeur $p_{i1}, i = 1, ..., 4$, qui est la p-valeur associée au test d'hypothèse H_{0i1}.

Une correction de la p-valeur doit être effectuée afin de contrôler le taux d'erreur de type I. La méthode suggérée par Chang et al. [13] est la méthode de Bonferroni. Cette correction sera appliquée séparément à chaque étape.

Règles de décision

Les règles de décision définissent les critères guidant les adaptations à la fin de chaque analyse. Une fois que les résultats des analyses sont disponibles, les p-valeurs obtenues sont ajustées puis comparées à des valeurs seuils.

La statistique de test $T_{ik} = \sum_i^k p_{ik}$ est calculée pour chaque traitement i à chaque étape k. Elle est ensuite comparée aux valeurs seuils de la façon suivante.

Lors de la première analyse intermédiaire, les p-valeurs associées aux différents tests d'hypothèse H_{0i1} seront d'abord ajustées selon la méthode de Bonferonni en utilisant la formule suivante : $\tilde{p}_{i1} = 4 \times p_{i1}$, étant donné que le nombre de tests effectués est de quatre.

Si $\tilde{p}_{i1} \leq \alpha_1$, l'hypothèse nulle sera rejetée. Le traitement i - soit la thalidomide, le trametinib, le regorafenib ou le cabozantinib - répondant à ce critère sera considéré comme plus efficace que le BSC, de façon significative. Si $\tilde{p}_{i1} > \beta_1$, l'hypothèse nulle sera acceptée. Cela signifiera que le traitement i n'a pas une meilleure efficacité que celle du BSC. Le traitement sera alors abandonné et sera exclu de l'étude. Si $\alpha_1 < \tilde{p}_{1min} \leq \beta_1$, où $\tilde{p}_{1min} = \min\{\tilde{p}_{11}, ..., \tilde{p}_{41}\}$, l'étude continuera et une seconde étape aura lieu. Cependant, si $\tilde{p}_{1min} > \beta_1$, l'étude prendra fin pour des raisons de futilité.

Lors de la seconde étape, en fonction du nombre de traitements restants, m_2 tests seront effectués. Un ajustement des p-valeurs sera à nouveau fait par la méthode de Bonferonni en utilisant la formule $\tilde{p}_{i2} = m_2 p_{i2}$. Les critères d'acceptation seront défini de la façon suivante pour cette seconde étape.

Si $\tilde{p}_{i1} + \tilde{p}_{i2} \leq \alpha_2$, l'hypothèse nulle sera rejetée, concluant à la supériorité du traitement i par rapport au BSC. Si ce n'était pas le cas, l'hypothèse nulle ne pourra pas être rejetée. L'hypothèse nulle pourra être rejetée globalement seulement si $\tilde{p}_{1min} + \tilde{p}_{2min} \leq \alpha_2$, signifiant qu'au moins un traitement i est plus efficace que le BSC.

Les valeurs critiques ont été fixées à $\alpha = 0.025$, qui est l'erreur de type I global de l'étude. Les valeurs de α_1, le seuil d'efficacité à l'étape 1, et β_1, le seuil de futilité à l'étape 1, ont été déterminées à partir du livre de Chang et al. [13]. Elles ont été fixé à $\alpha_1 = 0.01$ et $\beta_1 = 0.15$.

Le choix de $\alpha_1 = 0.01$ implique la chose suivante. Si la p-valeur ajustée associée à H_{0i1} est inférieure à 0,01, cela signifie qu'il y a moins de 1% de chances d'observer une différence de 25% entre la proportion de succès du traitement expérimental et du traitement contrôle, en faveur du traitement expérimental, sous l'hypothèse nulle. Cela signifie donc qu'il est très peu probable que le traitement expérimental ne soit pas supérieur au traitement contrôle, en termes d'efficacité. L'hypothèse nulle pourra donc être rejetée.

Par contre, en fixant $\beta_1 = 0.15$, on suppose que lorsque la p-valeur ajustée associée à H_{0i1} est supérieure à 0,15, il y a une probabilité supérieur à 15% d'observer cette même différence sous l'hypothèse nulle. Lorsque ce seuil est franchi, le bras de traitement sera exclu de l'étude.

La deuxième limite d'arrêt d'efficacité, α_2, est calculée en utilisant la formule 3.3.8 :

$$\alpha = \alpha_1 + \alpha_2 (\beta_1 - \alpha_1) - \frac{1}{2} \left(\beta_1^2 - \alpha_1^2\right)$$

Avec cette formule, une valeur de 0.1871 est obtenue pour α_2 lorsque $\alpha = 0.025$, $\alpha_1 = 0.01$ et $\beta_1 = 0.15$, où $\beta_1 < \alpha_2$.

A la fin du DLD, l'étude pourrait prendre fin. Cela sera le cas si seul un traitement est sélectionné à la fin de cette première phase ou si aucun des traitements n'est considérés comme plus efficace que le BSC. Si au moins deux traitements sont considérés comme plus efficaces que le BSC, un RARD appliquant le principe du RPW sera mis en place afin de comparer les traitements les plus prometteurs. Seuls les deux traitements les plus efficaces seront comparés avec le RARD. Les deux meilleurs traitements seront ceux pour lesquels la valeur de $\tilde{p}_{i1} + \tilde{p}_{i2}$ est la moins élevée et n'ayant pas répondu au critère de futilité à la première étape.

4.2.2 Détermination de la taille de l'échantillon

La taille de l'échantillon requise pour la première phase de cette étude est de 130 patients. Le calcul de taille d'échantillon s'est effectué avec la macro SAS %DrpLsrNRst, développée par Chang et al. [13].

Cette macro permet d'effectuer une simulation de Monte-Carlo du design de notre étude, paramétrée selon certaines valeurs, définies ci-dessous. Cette approche a permis d'évaluer la puissance attendue de l'étude et de calculer la taille de l'échantillon nécessaire à celle-ci.

Pour chaque scénario envisagé, 100 000 simulations ont été effectuées. L'étude inclut cinq bras de traitement. Les paramètres définis pour la macro %DrpLsrNRst sont les suivants. La taille de l'échantillon choisie permet d'obtenir une puissance statistique de 80% et un taux d'erreur de type I de 0.025. Les limites d'arrêt ont été fixées à $\alpha_1 = 0.01$, $\beta_1 = 0.15$ et $\alpha_2 = 0.1871$. Les simulations ont été effectuées pour un design qui étudie un output binaire, qui est le taux de succès du traitement i, avec un écart-type de 0.5. La valeur de l'écart-type peut être définie en utilisant la formule pour l'écart type d'une distribution binomiale $\sigma = \sqrt{p(1-p)}$, où p est la proportion de succès attendue. Comme il n'y a pas d'estimation cette probabilité de succès, la valeur de p a été fixée à 0.5, valeur à laquelle l'écart-type d'une proportion est maximisé. Cela donne alors une valeur d'écart-type de $\sigma = 0.5$

La taille de l'échantillon a été déterminée en exécutant la macro%DrpLsrNRst sous l'hypothèse nulle et l'hypothèse alternative. Sous l'hypothèse alternative, nous avons supposé qu'une différence de 25% devait être mise en évidence entre le taux de succès du traitement standard (BSC) et celui de chaque traitement i. Cette valeur représente donc le seuil de différence d'efficacité que nous considérons comme significatif entre les traitements. Sous l'hypothèse nulle,

nous avons supposé que le taux de succès était identique dans tous les groupes de traitement.

Sous l'hypothèse alternative, la taille moyenne de l'échantillon requise était de 129.7, tandis que sous l'hypothèse nulle, elle était de 128.4. Par conséquent, une taille d'échantillon de 130 patients est nécessaire pour la réalisation de cette étude.

Une première analyse intermédiaire sera effectuée une fois que les résultats de 50% des participants auront été générés. Les hypothèses H_{0i1} seront alors évalués et en fonction des valeurs obtenues, les adaptations nécessaires seront effectuées. Nous avons fixé une puissance conditionnelle cible de 80%, qui représente la probabilité conditionnelle de rejeter l'hypothèse nulle si celle-ci est fausse après la première étape, sur la base des données intermédiaires.

4.2.2.1 Deuxième partie : RARD utilisant le principe du RPW

Dans le cadre de notre étude, nous utiliserons un RARD utilisant le principe du RPW. Cette stratégie permet d'ajuster le ratio d'allocation des traitements en fonction des résultats intermédiaires, maximisant les chances qu'un patient reçoive le traitement le plus efficace à partir de la seconde randomisation.

Si le DLD ne parvient pas à déterminer le traitement le plus efficace et que l'essai n'est pas arrêté pour des raisons de futilité, ce RARD aura lieu. Nous évaluerons et comparerons les deux traitements les plus prometteurs, en nous basant sur le même critère binaire que celui utilisé précédemment : la proportion de patients ayant vu leur qualité de vie s'améliorer.

Cette deuxième phase de l'étude visera à comparer les deux traitements qui ont montré la plus grande efficacité lors de la première phase de l'étude (DLD). Cette approche permet non seulement de résoudre les problèmes éthiques potentiels, mais aussi d'améliorer le recrutement, car les patients qui étaient dans le groupe témoin lors de la première phase seront désormais randomisés dans l'un des deux groupes de traitement les plus prometteurs.

Déroulement du design RPW

Dans le cadre du RARD utilisant le principe du RPW, les patients de l'étude sont randomisés entre deux groupes de traitement. L'attribution des traitements est modélisée à l'aide d'une urne contenant deux types de balles. Le terme "urne" est utilisé pour illustrer le mécanisme de randomisation. Chaque type de balles représente un traitement. La proportion de chaque type de balles dans l'urne reflète la probabilité qu'un patient reçoive un certain traitement. Au début de l'essai, l'urne contient a_0 balles pour le traitement A et b_0 balles pour le traitement B, ces deux traitements étant ceux qui se sont montrés les plus prometteurs lors du DLD.

Après chaque analyse intermédiaire, la composition de l'urne est mise à jour. Si le traitement A est plus efficace que le traitement B, a_1 balles supplémentaires représentant le traitement A sont ajoutées à l'urne, et inversement. Si le traitement B est plus efficace, alors b_1 balles supplémentaires pour le traitement B sont ajoutées.

Pour notre essai clinique, nous avons fixé les valeurs initiales a_0 et b_0 à 1. Les deux traitements ont donc la même probabilité d'être attribués au début de cette phase de l'étude. De même, les valeurs a_1 et b_1 sont fixées à 1, indiquant qu'une seule balle, représentant le traitement le plus efficace de l'analyse intermédiaire k, est ajoutée à l'urne à chaque analyse intermédiaire. Cela permet de faire évoluer le ratio d'allocation sans atteindre des valeurs où les ratios de randomisation sont extrêmes. Si aucun traitement ne démontre une efficacité supérieure, aucune boule ne sera ajoutée dans l'urne.

L'évaluation de la qualité de vie des patients se fera tous les trois mois. La comparaison des taux de succès des traitements s'effectuera ensuite à l'aide d'un test chi-carré. Les résultats devront être disponibles un mois après la mesure de l'output. Sur la base de ces résultats, les ratios d'allocation seront ajustés, permettant ainsi de randomiser les nouveaux patients de l'étude en utilisant ces ratios révisés. Le traitement qui présente le meilleur taux de succès à chaque évaluation voit sa probabilité d'allocation augmenter. Ces étapes sont répétées deux fois, tous les quatre mois. Une dernière analyse finale aura lieu, trois mois après la dernière adaptation des ratios d'allocation. Son déroulement est détaillé dans la section qui suit.

Analyse finale des résultats

Une fois que tous les résultats auront été assemblés, ceux-ci seront analysés afin d'évaluer si la différence d'efficacité observée entre les traitements est statistiquement significative. Un test est effectué afin de d'évaluer le test d'hypothèse suivant :

$$H_0 : \pi_A \leq \pi_B$$
$$H_a : \pi_A > \pi_B$$

où
π_A est le taux de succès du traitement A,
π_B est le taux de succès du traitement B.

Nous utiliserons la statistique de test Z pour évaluer ces hypothèses, où $Z \sim N(0,1)$. Elle est calculée avec la formule suivante :

$$Z = \frac{p_A - p_B}{\sqrt{\frac{\bar{p}\bar{q}}{n_B} + \frac{\bar{p}\bar{q}}{n_A}}}$$

où $p_A = \frac{Y_A}{n_A}$ est l'estimateur de la probabilité de succès du traitement A,
Y_A est le nombre de succès chez les patients traités avec le traitement A,
n_A est le nombre de patients traités avec le traitement A,
$p_B = \frac{Y_B}{n_B}$ est l'estimateur de la probabilité de succès du traitement B,
Y_B est le nombre de succès chez les patients traités avec le traitement B,
n_B est le nombre de patients traités avec le traitement B,
$\bar{p} = \frac{Y_A+Y_B}{n_A+n_B}$ est l'estimation de la probabilité de succès global, du traitement A et du traitement B,

$\bar{q} = 1 - \bar{p}$ est l'estimation de la probabilité d'échec global, du traitement A et du traitement B.

La puissance ainsi que l'erreur de type I seront fixées respectivement à 0.8 et 0.025. Pour garantir que l'erreur de type I ne dépasse pas la valeur prévue de 0.025, la valeur limite de Z sera fixée à 2.05. Si nous avions choisi une valeur limite de Z égale à 1.96, le taux d'erreur de type I pourrait potentiellement dépasser la valeur cible de 0.025. La valeur limite $Z = 2.05$ a été définie à l'aide de la macro %MacroRPW, développée par Chang et al. [13] et conçue pour le logiciel statistique SAS. Pour ce faire, nous avons simulé un grand nombre de scénarios (200), dans lesquels le taux de réussite était identique pour les deux traitements étudiés. L'erreur de type I est alors étudiée en calculant la proportion de tests qui conduisent au rejet de l'hypothèse nulle alors que cette dernière est vraie.

Calcul de taille d'échantillon

La taille de l'échantillon requise pour la deuxième phase de cette étude est de 130 patients. Celle-ci a été déterminée à l'aide de la macro SAS %MacroRPW, développée par Chang et al. [13].

Cette macro permet d'effectuer des simulations de Monte-Carlo du design de notre étude, en se basant sur des paramètres prédéfinis, tels que la puissance souhaitée, la valeur critique Z, le nombre de patients inclus dans l'étude et l'écart d'efficacité entre les traitements qui doit être mis en évidence pour qu'un traitement soit considéré comme significativement plus efficace. %MacroRPW utilise un algorithme de simulation pour générer des distributions de réponse aux traitements basées sur les probabilités de succès de chaque traitement. L'objectif principal de cette macro est d'estimer la puissance du design étudié.

La taille de l'échantillon a été déterminée en exécutant la macro %MacroRPW sous l'hypothèse nulle et l'hypothèse alternative. Sous l'hypothèse alternative, nous avons postulé qu'une différence de 25% entre les traitements étudiés devait être démontrée pour que l'un d'eux soit considéré comme significativement plus efficace. Sous l'hypothèse nulle, nous avons supposé qu'il n'y avait pas de différence entre les taux de succès des deux traitements.

Sous les deux hypothèses, la taille moyenne de l'échantillon nécessaire était de 130. Cette taille d'échantillon permet de mettre en évidence une différence de 25% entre les proportions de succès des deux traitements examinés dans l'étude, avec une puissance minimale de 80% et une erreur de type I de 0.025, la valeur critique de Z étant fixée à 2.05.

La deuxième partie de l'étude clinique commencera avec 44 patients. Le design inclut deux analyses intermédiaires au cours desquelles les ratios d'allocation sont ajustés. Après chaque ajustement, 43 nouveaux patients sont ajoutés à l'étude. Ceux-ci sont répartis entre les groupes de traitement en respectant les nouveaux ratios d'allocation. Une dernière analyse est effectuée à la fin de l'étude pour déterminer si un des deux traitements est plus efficace, de façon significative.

Chapitre 5

Conclusions et discussion

La recherche et l'innovation dans les essais cliniques pour les maladies rares sont primordiales. Développer ces études est essentiel pour explorer de nouvelles options thérapeutiques et améliorer la qualité de vie des patients affectés par ces pathologies.

La mise en place de ces essais cliniques soulève des défis spécifiques, tels que le recrutement limité de patients et la petite taille des populations cibles. L'intégration d'un groupe de contrôle dans le design de ces études peut également poser des problèmes éthiques, en raison du nombre restreint de traitements disponibles pour ces maladies.

Ces contraintes rendent parfois l'application des ECR parallèles difficile voire impossible. Ce type de design est considéré comme la référence en matière d'essais cliniques. Par conséquent, l'étude et l'utilisation de designs d'essai clinique alternatifs sont pertinentes. Nous avons exploré trois stratégies dans le cadre de ce mémoire : la réduction de la variabilité, l'utilisation de données historiques et obtenues au cours de l'étude, ainsi que l'assurance pour tous les patients de bénéficier du traitement expérimental.

Concernant le design de l'essai clinique évaluant l'efficacité de quatre traitements dans le cadre des MAVs, notre approche s'est basée sur un design adaptatif en deux parties. Cette stratégie permet d'identifier les traitements les plus prometteurs parmi les quatre sélectionnés pour l'étude, tout en adaptant le design en fonction des résultats obtenus et des critères pré-spécifiés. Si aucun traitement ne se révèle supérieur et que l'étude n'est pas arrêtée précocement pour des raisons de futilité, les deux traitements présentant la meilleure efficacité seront retenus pour une deuxième étape.

Cette deuxième étape est un RARD appliquant le principe du RPW. Ce type de design permet d'augmenter progressivement le nombre de patients alloué au traitement démontrant une efficacité supérieure, en fonction des données recueillies au cours de l'essai.

Etant donné le manque d'informations sur l'effet des traitements étudiés sur les MAVs, l'utilisation des données générées pendant l'essai est un grand avantage. De plus, cela pourrait entraîner une réduction de la taille de l'échantillon nécessaire, si l'efficacité ou la futilité des traitements sur les MAVs était démontrée de façon prématurée.

Néanmoins, si aucun traitement ne se révèle prometteur, l'essai pourrait être interrompu précocement, évitant une perte de temps, de ressources financières, et protégeant les patients d'un traitement potentiellement inefficace, voire toxique.

Discussion

Le modèle que nous avons proposé offre de nombreux avantages. Il serait à présent judicieux de le valider par des études de simulations pour confirmer sa robustesse. Ceci représente une étape essentielle pour garantir la qualité du design proposé.

La présence d'un groupe de contrôle dans notre modèle est un atout majeur, car elle renforce la validité de nos conclusions et aide à établir le rapport bénéfice-risque du traitement. Cependant, en cas de difficultés, il serait envisageable d'adapter le modèle et d'enlever le groupe contrôle. Ceci nécessiterait une modification de la macro %DrpLsrNRst pour calculer correctement la taille de l'échantillon et la puissance de l'étude. Cela pourrait se produire si le recrutement ne permet pas d'obtenir la taille d'échantillon nécessaire (n = 130) ou si la présence du groupe contrôle entraînait des réticences chez les patients, impactant le recrutement de l'étude.

Dans le cadre de l'essai clinique mis en place dans ce travail, l'analyse des données a été effectuée en utilisant un test Z. Ce test statistique est largement utilisé pour comparer des taux ou des proportions de succès entre deux groupes de traitement. Cependant, son application est basée sur l'approximation de la loi binomiale par la loi normale lorsque les tailles des échantillons sont suffisamment grandes. Cette approche peut être appropriée lorsque np et $n(1-p)$ sont supérieurs à 5, où n est la taille de l'échantillon et p la probabilité de succès. Ce scénario est plausible dans le contexte de notre étude. Néanmoins, il est important de souligner que l'usage du test Z n'est pas optimal dans le cas des petits échantillons, comme cela est souvent le cas pour les études sur les maladies rares. Dans ces situations, l'utilisation du test exact de Fisher pourrait être plus appropriée. Ce dernier, qui ne repose pas sur l'approximation par une distribution spécifique, s'avère plus adapté aux petites tailles d'échantillon.

Ensuite, la mise en place d'un essai clinique utilisant des concepts bayésiens aurait pu être une alternative intéressante. Néanmoins, elle nécessite une certaine expertise pour définir les priors et leurs paramètres, nécessaire au calcul de la postérior. Cette approche pourrait faire l'objet de travaux futurs, car elle offre une flexibilité et une intégration des connaissances précédentes qui pourraient être bénéfiques dans le contexte des maladies rares.

Pour conclure, bien que chaque maladie rare n'affecte qu'un nombre limité de patients, leur impact cumulé sur la santé publique est important. Les cliniciens, les chercheurs et les organismes de réglementation doivent continuer à rechercher de nouvelles méthodes pour évaluer les risques et les bénéfices de nouveaux traitements pour ces maladies. Ce travail a proposé et analysé des conceptions d'études alternatives adaptées au contexte des maladies rares.

Ceci n'est que le début et il reste beaucoup à faire. L'adoption de modèles plus sophistiqués, tels que des modèles bayésiens, et le développement de nouvelles macros pour faciliter l'analyse statistique de ces essais cliniques, sont des sujets de recherche prometteur pour l'avenir. Il est

essentiel la recherche dans le domaine des maladies rares se poursuive afin d'améliorer la vie des patients qui en sont atteints.

Glossaire

AIC Critère d'Information d'Akaike. 24
AVs Anomalies Vasculaires. 51

BB Boule Blanche. 28–30
BIC Critère d'Information Bayésien. 24
BN Boule Noire. 28–30
BSC Best Support Care. 5, 6, 51, 55–58

DLD Drop-Loser Design. 55, 58, 59
DMT Dose Maximale Tolérée. 4

ECR Essai Contrôlé Randomisé. 6–9, 11, 16, 19, 25, 31, 32, 36, 40, 48, 49, 55, 63
EERWD Enriched Enrollment Randomized Withdrawal Design. 48, 49
ESP Efficacy Stopping Probability. 37

FSP Futility Stopping Probability. 37

HEA Health Economic Assessment. 5, 6
HTA Health Technology Assessment. 5, 6

IPD Individual Participant Data. 23

MAP Meta-Analytic Predictive. 41–43
MAVs Malformations Artérioveineuses. 51–55, 63
MS Makuch-Simon. 43, 44

RARD Response Adaptative Randomization Design. ii, 27–30, 45, 54, 58, 59, 63
REML Restricted Maximum Likelihood. 17
RPPD Randomized Placebo-Phase Design. 49
RPW Randomized Play-the-Winner. ii, 28, 29, 54, 58, 59, 63
RSD Ranking and Selection Design. 31, 32, 35, 54, 55

Bibliographie

[1] Lusine Abrahamyan, Brian M Feldman, George Tomlinson, Marie E Faughnan, Sindhu R Johnson, Ivan R Diamond, and Samir Gupta. Alternative designs for clinical trials in rare diseases. In *American Journal of Medical Genetics Part C : Seminars in Medical Genetics*, volume 172, pages 313–331. Wiley Online Library, 2016.

[2] Lusine Abrahamyan, Chuan Silvia Li, Joseph Beyene, Andrew R Willan, and Brian M Feldman. Survival distributions impact the power of randomized placebo-phase design and parallel groups randomized clinical trials. *Journal of clinical epidemiology*, 64(3) :286–292, 2011.

[3] European Medicines Agency. Cabometyx : Epar - product information, 2023. Accessed : 2023-05-30.

[4] European Medicines Agency. Mekinist : Epar - product information, 2023. Accessed : 2023-05-30.

[5] European Medicines Agency. Stivarga : Epar - product information, 2023. Accessed : 2023-05-30.

[6] European Medicines Agency. Thalidomide bms previously thalidomide celgene : Epar - product information, 2023. Accessed : 2023-05-30.

[7] Lara Al-Olabi, Satyamaanasa Polubothu, Katherine Dowsett, Katrina A Andrews, Paulina Stadnik, Agnel P Joseph, Rachel Knox, Alan Pittman, Graeme Clark, William Baird, et al. Mosaic ras/mapk variants cause sporadic vascular malformations which respond to targeted therapy. *The Journal of clinical investigation*, 128(4) :1496–1508, 2018.

[8] David Banta. The development of health technology assessment. *Health policy*, 63(2) :121–132, 2003.

[9] Laurence M Boon, Valérie Dekeuleneer, Julien Coulie, Liliane Marot, Anne-Christine Bataille, Frank Hammer, Philippe Clapuyt, Anne Jeanjean, Anne Dompmartin, and Miikka Vikkula. Case report study of thalidomide therapy in 18 patients with severe arteriovenous malformations. *Nature Cardiovascular Research*, 1(6) :562–567, 2022.

[10] Timo B Brakenhoff, Kit CB Roes, and Stavros Nikolakopoulos. Bayesian sample size re-estimation using power priors. *Statistical methods in medical research*, 28(6) :1664–1675, 2019.

[11] Teresa Brodniewicz and Grzegorz Grynkiewicz. Preclinical drug development. *Acta Pol Pharm*, 67(6) :578–85, 2010.

[12] M Chang. Recursive two-stage adaptive design, 2006.

[13] Mark Chang. *Adaptive design theory and implementation using SAS and R*. CRC Press, 2014.

[14] Xinlin Chen and Pingyan Chen. A comparison of four methods for the analysis of n-of-1 trials. *PloS one*, 9(2) :e87752, 2014.

[15] Vernon M Chinchilli and James D Esinhart. Design and analysis of intra-subject variability in cross-over experiments. *Statistics in Medicine*, 15(15) :1619–1634, 1996.

[16] Shein-Chung Chow and Mark Chang. Adaptive design methods in clinical trials–a review. *Orphanet journal of rare diseases*, 3(1) :1–13, 2008.

[17] Olivier Collignon, Anna Schritz, Riccardo Spezia, and Stephen J Senn. Implementing historical controls in oncology trials. *The Oncologist*, 26(5) :e859–e862, 2021.

[18] Kristina M Cook and William D Figg. Angiogenesis inhibitors : current strategies and future prospects. *CA : a cancer journal for clinicians*, 60(4) :222–243, 2010.

[19] H-G Eichler, Brigitte Bloechl-Daum, Peter Bauer, Frank Bretz, Jeffrey Brown, Lisa Victoria Hampson, Peter Honig, Michael Krams, Hubert Leufkens, Robyn Lim, et al. "threshold-crossing" : a useful way to establish the counterfactual in clinical trials ? *Clinical Pharmacology & Therapeutics*, 100(6) :699–712, 2016.

[20] Ulrike Ernemann, Ulrich Kramer, Stephan Miller, Sotirios Bisdas, Hans Rebmann, Helmut Breuninger, Christine Zwick, and Jürgen Hoffmann. Current concepts in the classification, diagnosis and treatment of vascular anomalies. *European journal of radiology*, 75(1) :2–11, 2010.

[21] Brian Feldman, Elaine Wang, Andrew Willan, and John Paul Szalai. The randomized placebo-phase design for clinical trials. *Journal of clinical epidemiology*, 54(6) :550–557, 2001.

[22] Lawrence M Friedman, Curt D Furberg, David L DeMets, David M Reboussin, and Christopher B Granger. *Fundamentals of clinical trials*. Springer, 2015.

[23] James F Fries and Eswar Krishnan. Equipoise, design bias, and randomized controlled trials : the elusive ethics of new drug development. *Arthritis Res Ther*, 6(3) :1–6, 2004.

[24] Mercedeh Ghadessi, Rui Tang, Joey Zhou, Rong Liu, Chenkun Wang, Kiichiro Toyoizumi, Chaoqun Mei, Lixia Zhang, CQ Deng, and Robert A Beckman. A roadmap to using historical controls in clinical trials–by drug information association adaptive design scientific working group (dia-adswg). *Orphanet Journal of Rare Diseases*, 15(1) :1–19, 2020.

[25] Robert C Griggs, Mark Batshaw, Mary Dunkle, Rashmi Gopal-Srivastava, Edward Kaye, Jeffrey Krischer, Tan Nguyen, Kathleen Paulus, Peter A Merkel, et al. Clinical research for rare disease : opportunities, challenges, and solutions. *Molecular genetics and metabolism*, 96(1) :20–26, 2009.

[26] Marlene E Haffner. Designing clinical trials to study rare disease treatment. *Drug information journal : DIJ/Drug Information Association*, 32 :957–960, 1998.

[27] Kathryn T Hall, Lene Vase, Deirdre K Tobias, Hesam T Dashti, Jan Vollert, Ted J Kaptchuk, and Nancy R Cook. Historical controls in randomized clinical trials : opportunities and challenges. *Clinical Pharmacology & Therapeutics*, 109(2) :343–351, 2021.

[28] Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, pages 65–70, 1979.

[29] Soonsil Hyun and Dongyun Shin. Small-molecule inhibitors and degraders targeting kras-driven cancers. *International Journal of Molecular Sciences*, 22(22) :12142, 2021.

[30] Byron Jones and Michael G Kenward. *Design and analysis of cross-over trials*. CRC press, 2014.

[31] Steven A Julious. Sample sizes for clinical trials with normal data. *Statistics in medicine*, 23(12) :1921–1986, 2004.

[32] Michael G Kenward and James H Roger. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, pages 983–997, 1997.

[33] Michael G Kenward and James H Roger. An improved approximation to the precision of fixed effects from restricted maximum likelihood. *Computational Statistics & Data Analysis*, 53(7) :2583–2595, 2009.

[34] James Kepner and Dennis Wackerly. Some observations on the makuch/simon approach to sample size determination in clinical trials with historical controls. *Communications in Statistics-Simulation and Computation*, 30(3) :611–621, 2001.

[35] Edward L Korn and Boris Freidlin. Conditional power calculations for clinical trials with historical controls. *Statistics in medicine*, 25(17) :2922–2931, 2006.

[36] Chi-Yeon Lim and Junyong In. Considerations for crossover design in clinical study. *Korean Journal of Anesthesiology*, 74(4) :293–299, 2021.

[37] R Andrew Moore, Philip J Wiffen, Christopher Eccleston, Sheena Derry, Ralf Baron, Rae F Bell, Andrea D Furlan, Ian Gilron, Simon Haroutounian, Nathaniel P Katz, et al. Systematic review of enriched enrolment, randomised withdrawal trial designs in chronic pain : a new framework for design and reporting. *Pain*, 156(8) :1382–1395, 2015.

[38] Beat Neuenschwander, Gorana Capkun-Niggli, Michael Branson, and David J Spiegelhalter. Summarizing historical information on controls in clinical trials. *Clinical Trials*, 7(1) :5–18, 2010.

[39] Cynthia L Nicholson, Siobhan Flanagan, Michael Murati, Christina Boull, Erin McGough, Rebecca Ameduri, Brenda Weigel, and Sheilagh Maguiness. Successful management of an arteriovenous malformation with trametinib in a patient with capillary-malformation arteriovenous malformation syndrome and cardiac compromise. *Pediatric Dermatology*, 39(2) :316–319, 2022.

[40] Peter C O'Brien and Thomas R Fleming. A multiple testing procedure for clinical trials. *Biometrics*, pages 549–556, 1979.

[41] Stephan Quintin, John W Figg, Yusuf Mehkri, Chadwin O Hanna, Maxwell G Woolridge, and Brandon Lucke-Wold. Arteriovenous malformations : An update on models and therapeutic targets. *Journal of neuroscience and neurological surgery*, 13(1), 2023.

[42] Tonya D Riddlesworth, Roy W Beck, Robin L Gal, Crystal G Connor, Richard M Bergenstal, Sooji Lee, and Steven M Willi. Optimal sampling duration for continuous glucose monitoring to determine long-term glycemic control. *Diabetes technology & therapeutics*, 20(4) :314–316, 2018.

[43] Elizabeth G Ryan, Sarah E Lamb, Esther Williamson, and Simon Gates. Bayesian adaptive designs for multi-arm trials : an orthopaedic case study. *Trials*, 21(1) :1–16, 2020.

[44] Henry Scheffe. A" mixed model" for the analysis of variance. *The Annals of Mathematical Statistics*, pages 23–36, 1956.

[45] SJ Senn. Cross-over trials, carry-over effects and the art of self-delusion. *Statistics in Medicine*, 7(10) :1099–1101, 1988.

[46] Stephen Senn. Sample size considerations for n-of-1 trials. *Statistical methods in medical research*, 28(2) :372–383, 2019.

[47] Steven Simoens. Health economic assessment : a methodological primer. *International journal of environmental research and public health*, 6(12) :2950–2966, 2009.

[48] Catrin Tudur Smith, Paula R Williamson, and Michael W Beresford. Methodology of clinical trials for rare diseases. *Best practice & research Clinical rheumatology*, 28(2) :247–262, 2014.

[49] Nigel Stallard, Frank Miller, Simon Day, Siew Wan Hee, Jason Madan, Sarah Zohar, and Martin Posch. Determination of the optimal sample size for a clinical trial accounting for the population size. *Biometrical Journal*, 59(4) :609–625, 2017.

[50] Moreno Ursino and Nigel Stallard. Bayesian approaches for confirmatory trials in rare diseases : opportunities and challenges. *International Journal of Environmental Research and Public Health*, 18(3) :1022, 2021.

[51] Kert Viele, Scott Berry, Beat Neuenschwander, Billy Amzal, Fang Chen, Nathan Enas, Brian Hobbs, Joseph G Ibrahim, Nelson Kinnersley, Stacy Lindborg, et al. Use of historical control data for assessing treatment effects in clinical trials. *Pharmaceutical statistics*, 13(1) :41–54, 2014.

[52] Scott I Vrieze. Model selection and psychological theory : a discussion of the differences between the akaike information criterion (aic) and the bayesian information criterion (bic). *Psychological methods*, 17(2) :228, 2012.

[53] Sue-Jane Wang and HM James Hung. Use of two-stage test statistic in the two-period crossover trials. *Biometrics*, pages 1081–1091, 1997.

[54] J Kyle Wathen and Peter F Thall. A simulation study of outcome adaptive randomization in multi-arm clinical trials. *Clinical Trials*, 14(5) :432–440, 2017.

[55] Anke Wegman, Daniëlle AWM van der Windt, Wim AB Stalman, and Theo PGM de Vries. Conducting research in individual patients : lessons learnt from two series of n-of-1 trials. *BMC Family Practice*, 7(1) :1–8, 2006.

[56] Stefan Wellek and Maria Blettner. On the proper use of the crossover design in clinical trials : part 18 of a series on evaluation of scientific publications. *Deutsches Ärzteblatt International*, 109(15) :276, 2012.

[57] Andrew R Willan and Joseph L Pater. Carryover and the two-period crossover clinical trial. *Biometrics*, pages 593–599, 1986.

[58] S Yousuf Zafar, David Currow, and Amy P Abernethy. Defining best supportive care. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology*, 26(31) :5139–5140, 2008.

[59] Deborah R Zucker, Robin Ruthazer, and Christopher H Schmid. Individual (n-of-1) trials can be combined to give population comparative treatment effect estimates : methodologic considerations. *Journal of clinical epidemiology*, 63(12) :1312–1323, 2010.

Annexe

Code permettant de réaliser les graphiques de la figure 2.4.2

```
require(ggplot2)

# Erreur de type I(alpha) et II (beta)
beta = rep(0.2, 8)
alpha = rep(0.05, 8)

# Variance variable, pour voir son influence sur la taille de l'échantillon
var = seq(20,55,5)
# Variance constante, pour voir l'influence de l'effet à mettre en évidence sur la taille de l'échantillon
varD = rep(20, 8)

# Effet à mettre en évidence variable, pour voir son influence sur la taille de l'échantillon
delta = seq(3,10,1)
# Effet à mettre en évidence constante, pour voir l'influence de la Variance sur la taille de l'échantillon
deltaV = rep(3, 8)

# Calcul de la taille de l'échantillon et influence de la taille de l'effet à mettre en évidence
nD = (2*(qnorm(1 - beta,0,1) + qnorm(1 - alpha/2,0,1))*varD)/delta^2
nD
# Calcul de la taille de l'échantillon et influcence de la variance
nV = (2*(qnorm(1 - beta,0,1) + qnorm(1 - alpha/2,0,1))*var)/(deltaV^2)
nV

# Création des dataframes pour construire les graphiques
datD = data.frame(nD, varD, delta)
datV = data.frame(nV, var, deltaV)

# Evolution de la taille de l'échantillon en fonction de la taille de l'effet à mettre en évidence
ggplot(datD, aes(delta,nD)) + geom_line() + xlab("Effet à mettre en évidence") + ylab("Taille de l'échantillon")

# Evolution de la taille de l'échantillon en fonction de la variance
ggplot(datV, aes(var, nV)) + geom_line() + xlab("Variance de l'échantillon") + ylab("Taille de l'échantillon")
```

Code démontrant que l'utilisation d'une analyse classique peut amener un biais dans l'analyse des résultats dans un design cross-over

```
rm(list = ls())
```

```
# Nombre de sujets
n <- 100

# Pour la reproductibilité des résultats
set.seed(123)

# Ajout des effets carry-over aux données de traitement suivant
P11 = rnorm(n/2, mean = 10, sd = 2) # Effet du premier traitement de la séquence 1 donc, le traitement A
P21 = rnorm(n/2, mean = 10, sd = 2) # Effet du premier traitement de la séquence 2 donc, le traitement B

P12 = rnorm(n/2, mean = 10, sd = 2) # Effet du deuxième traitement de la séquence 1 donc, le traitement B
ECO <- c(rnorm(n/2, mean = 5, sd = 2)) # Génération des effets carry-over produit par le traitement B
P22 = rnorm(n/2, mean = 10, sd = 2) + ECO # Effet du deuxième traitement de la séquence 2 donc, le traitement A additionné à l'effet carryover du traitement B

A = c(P11, P22) # vecteur rassemblant l'effet du traitement A
B = c(P21, P12) # vecteur rassemblant l'effet du traitement B
t.test(A, B, alternative = "greater", paired = TRUE)
```

Macro %DrpLsrNRst et utilisation de celle-ci pour calculer la taille de l'échantillon et la puissance du DLD

```
%Macro DrpLsrNRst(nSims=100000, CntlType="strong", nArms=5,
alpha=0.025, beta=0.2, NId=0, cPower=0.9, nInterim=50,
Nmax=150, nAdj="Y", alpha1=0.01, beta1=0.15,
alpha2=0.1871, EP="normal", sigma=1, tStd=24, tAcr=10);
Data DrpLsrNRst; Set dInput;
Keep FSP ESP AveN Power cPower Nmax;
Array us{&nArms};
Array u1{&nArms};
Array u2{&nArms};
Array cs{&nArms};
seedx=1736; seedy=6214; alpha=&alpha; NId=&NId;
Nmax=&Nmax; nArms=&nArms; n1=&nInterim; sigma=&sigma;
cPower=&cPower; FSP=0; ESP=0; AveN=0; Power=0;
If &EP="mean" Then sigma=&sigma;
If &EP="binary" Then sigma=(us1*(1-us{1}))**0.5;
If &EP="survival" Then
sigma=us{1}*(1+exp(-us{1}*&tStd)*(1-exp(us{1}*&tAcr))
/(&tAcr*us{1}))**(-0.5);
Do isim=1 to &nSims;
```

```
TotalN=nArms*n1 ; uMax=us{1} ; Cntrst=0 ; SumSqc=0 ;
Do i=1 To nArms ;
u1{i}=Rannor(seedx)*sigma/Sqrt(n1)+us{i} ;
If u1{i}>uMax Then Do uMax=u1{i} ; iMax=i ;
End ;
Cntrst=Cntrst+cs{i}*u1{i} ;
SumSqc=SumSqc+cs{i}*cs{i} ;
End ;
Z1 = Cntrst*Sqrt(n1)/Sqrt(SumSqc)/sigma ;
p1=1-ProbNorm(Z1) ;
* Bonferroni Adjustment ;
If &CntlType="strong" Then
p1=(nArms-1)*(1-ProbNorm((uMax-us1)/sigma*Sqrt(n1/2))) ;
If p1>&beta1 Then FSP=FSP+1/&nSims ;
If p1<=&alpha1 Then do ;
Power=Power+1/&nSims ; ESP=ESP+1/&nSims ;
End ;
If iMax=1 Then Goto myJump ;
If p1>&alpha1 and p1<=&beta1 Then do ;
BF=Probit(1-max(0,&alpha2-p1))-Probit(1-cPower) ;
n2=2*(sigma/(u1iMax-u11)*BF)**2 ;
nFinal=min(n1+n2, Nmax) ;
If &nAdj="N" Then nFinal=Nmax ;
If nFinal>n1 Then do ;
TotalN=2*(nFinal-n1)+nArms*n1 ;
u2{1}=Rannor(seedx)*sigma/Sqrt(nFinal-n1)+us{1} ;
u2{iMax}=Rannor(seedy)*sigma/Sqrt(nFinal-n1)+us{iMax} ;
T2=(u2{iMax}-u2{1}+NId)*Sqrt(nFinal-n1)/2**0.5/sigma ;
p2=1-ProbNorm(T2) ; TS2=p1+p2 ;
If .<TS2<=&alpha2 Then Power=Power+1/&nSims ;
End ;
End ;
myJump :
AveN=AveN+TotalN/&nSims ;
End ;
Output ;
Run ;
Proc Print Data=DrpLsrNRst (obs=1) ;
Run ;
%Mend DrpLsrNRst ;

Title "Simulations of Drop-loser Design under Ha" ;
Data dInput ;
Array us{5} (.1, .35, .35, .35, .35) ;
%DrpLsrNRst(CntlType="strong", nArms=5, alpha=0.025, beta=0.2,
NId=0, cPower=0.8, nInterim=25, Nmax=100, nAdj="N",
alpha1=0.01, beta1=0.15, alpha2=0.1871, EP="binary", sigma=0.2) ;
```

```
Run ;

Title "Simulations of Drop-loser Design under Ho" ; Data dInput ;
Array us{5} (.4, .4, .4, .4, .4) ;
%DrpLsrNRst(CntlType="strong", nArms=5, alpha=0.025, beta=0.2,
NId=0, cPower=0.8, nInterim=25, Nmax=100, nAdj="N",
alpha1=0.01, beta1=0.15, alpha2=0.1871, EP="binary", sigma=0.2) ;
Run ;
```

Macro %RPW et utilisation de celle-ci pour calculer la taille de l'échantillon et la puissance du RARD utilisant le principe du RPW

```
%Macro RPW(nSims=100000, Zc=1.96, nSbjs=200, nAnlys=3,
RR1=0.2, RR2=0.3, a0=1, b0=1, a1=1, b1=1, nMin=1) ;
Data RPW ; Keep nSbjs aveP1 aveP2 Zc Power ;
seed1=364 ; seed2=894 ; Power=0 ; aveP1=0 ; aveP2=0 ;
Do isim=1 to &nSims ;
nResp1=0 ; nResp2=0 ; a0=&a0 ; b0=&b0 ; Zc=&Zc ; N1=0 ; N2=0 ;
nSbjs=&nSbjs ; nAnlys=&nAnlys ; nMax=nSbjs-&nMin ;
a=a0 ; b=b0 ; r0=a/(a+b) ;
Do iSbj=1 To nSbjs ;
If Mod(iSbj,Round(nSbjs/nAnlys))=0 Then r0=a/(a+b) ;
rnAss=Ranuni(seed1) ;
If (rnAss < r0 And N1<nMax) Or N2>=nMax Then
Do ;
N1=N1+1 ; rnRep=Ranuni(seed2) ;
if rnRep <=&RR1 Then Do ;
nResp1=nResp1+1 ; a=a+&a1 ;
End ;
End ;
Else
Do ;
N2=N2+1 ; rnRep=Ranuni(seed2) ;
If rnRep <=&RR2 Then Do ;
nResp2=nResp2+1 ; b=b+&b1 ;
End ;
End ;
End ;
p1=nResp1/N1 ; p2=nResp2/N2 ;
aveP1=aveP1+p1/&nSims ; aveP2=aveP2+p2/&nSims ;
sigma1=sqrt(p1*(1-p1)) ; sigma2=sqrt(p2*(1-p2)) ;
Sumscf=sigma1**2/(N1/(N1+N2))+sigma2**2/(N2/(N1+N2)) ;
TS = (p2-p1)*Sqrt((N1+N2)/sumscf) ;
If TS>Zc Then Power=Power+1/&nSims ;
End ;
```

```
Output ;
Run ;
Proc Print data=RPW ; Run ;
%Mend RPW ;

%RPW(Zc=2.05, nSbjs=130, nAnlys=3, RR1=0.4, RR2=0.4,
a0=1, b0=1,a1=1, b1=1, nMin=1) ;

%RPW(Zc=2.05, nSbjs=130, nAnlys=3, RR1=0.15, RR2=0.4,
a0=1, b0=1,a1=1, b1=1, nMin=1) ;
```

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
Faculté des sciences
Place des sciences, 2 bte L6.06.01, 1348 Louvain-la-Neuve, Belgique | **www.uclouvain.be/sc**